Lung and Colon Cancer Detection from CT Images Using Deep Learning

Joseph D. Akinyemi[1], Akinkunle A. Akinola[2], Olajumoke O. Adekunle[1], Taiwo O. Adetiloye[3], Emmanuel J. Dansu[4,5]

[1]*Department of Computer Science, University of Ibadan, Ibadan, Nigeria*

[2]*Department of Data Science, Bowling Green State University, Bowling Green, Ohio, USA*

[3]*BigCodeGen LLC, Fort Worth, Texas, USA*

[4]*Department of Mathematical Sciences, Federal University of Technology, Akure, Nigeria*

[5]*Ecological Integration Laboratory, Graduate School of Life Sciences, Tohoku University, Sendai, Japan*

jd.akinyemi@ui.edu.ng

Abstract. Cancer is a deadly disease that has gained a reputation as a global health concern. Further, lung cancer has been widely reported as the most deadly cancer type globally, while colon cancer comes second. Meanwhile, early detection is one of the primary ways to prevent lung and colon cancer fatalities. To aid the early detection of lung and colon cancer, we propose a computer-aided diagnostic approach that employs a Deep Learning (DL) architecture to enhance the detection of these cancer types from Computed Tomography (CT) images of suspected body parts. Our experimental dataset (LC25000) contains 25 000 CT images of benign and malignant lung and colon cancer tissues. We used weights from a pre-trained DL architecture for computer vision, EfficientNet, to build and train a lung and colon cancer detection model. EfficientNet is a Convolutional Neural Network architecture that scales all input dimensions such as depth, width, and resolution at the same time. Our research findings showed detection accuracies of 99.63%, 99.50%, and 99.72% for training, validation, and test sets, respectively.

Key words: cancer detection, efficientNet, CT images, healthcare.

1. Introduction

Cancer is one of the leading causes of death globally causing about 10 million deaths or one out of every six deaths in the year 2020. Of the various types of cancer, lung cancer accounts for the highest number of deaths, accounting for about 18% of cancer deaths in the year 2020; the second highest being colon and rectum cancer, which account for 9% of cancer deaths [4, 22]. The presence of cancer in the body is often medically confirmed by medical imaging techniques such as Computed Tomography (CT) scan or Magnetic Resonance Imaging (MRI) which are then analyzed and interpreted by medical experts. To improve detection accuracy and reduce doctors' burdens, Machine Learning (ML) and Deep Learning (DL) techniques have been employed to accelerate the detection of cancer in such medical images. This enables researchers to analyze a large number of patients in much less time and at a lower cost [18]. The accuracy of these models on real-world data has been limited largely because of insufficient datasets available for experimentation, yet, near-perfect accuracy is highly desirable in medical systems.

DL models, especially for computer vision, thrive on large volumes of data. When these are not available, model performance may be unsatisfactory. Until 2019, freely available cancer image datasets have only contained a few thousand images. However, since the release of the LC25000 dataset [3], the research community has seen consistent progress in the accuracy of cancer detection and classification models. However, our major observation is that most published works have performed experiments on subsets of the dataset (either lung cancer or colon cancer). This may be because of the relatively large size of the entire dataset which makes attempts to build a single model to detect both lung and colon cancers not achieve near-perfect accuracy. In this work, a single DL model was developed to detect and classify lung and colon cancer with improved classification accuracy over existing models. We used the pre-trained EfficientNet [19] model to build a lung and colon cancer detection/classification model and obtained 99.72% accuracy on a hold-out (out-of-training) test set of 5000 images in only 44 training epochs. The introspection results of our training and validation also reveal a good training progression and "goodness of fit". Our motivations for developing a single model for lung and colon cancer classification include (i) the fact that both diseases have the highest cancer fatalities, and (ii) the possibility of achieving better accuracy in medical predictions when large datasets are available.

Our research contribution is to develop a novel DL model based on EfficientNet for accurate and robust lung and colon cancer detection. This model is designed to automatically extract the most important features from medical images for accurate and reliable cancer detection. Furthermore, the proposed model is capable of achieving high accuracy and robustness in the detection of lung and colon cancer. Additionally, we have evaluated the proposed model on the LC25000 dataset to demonstrate its effectiveness in detecting both types of cancer.

The rest of this paper is organized as follows. In Section 2, we review literature related to the application of DL in cancer detection. The materials and methods used in the research are given in Section 3. This includes the dataset used, the network architecture, and applicable experiments. Section 4 presents our results and discussions. Finally, conclusions and future work are given in Section 5.

2. Review of related literature

The growing success of ML and DL has made it applicable in several walks of life including cancer detection. In this section, we present a review of recent relevant works in lung and colon cancer detection.

A ML-based lung and colon cancer detection using hybrid model comprising deep feature extraction and ensemble learning was introduced by [18]. Their hybrid model which evaluated the histopathological (LC25000) lung and colon datasets achieved accuracy rates of 99.05%, 100%, and 99.30%, respectively. This feature extraction approach

(a hybrid ensemble feature extraction model that combines deep feature extraction and high-performance ensemble learning for cancer images) is anticipated to be helpful for the diagnosis of lung and colon cancers in medical sectors.

In [8], eight different pre-trained models for lung and colon cancer classification were employed, also using the LC25000 dataset. Five of the eight pre-trained models (Inception V2 & V3, MobileNet, Xception and DenseNet169) achieved 100% accuracy, while the rest (VGG-16, ResNet50 and NASNetMobile) achieved more than 95% accuracy. While this may be a pointer to the effectiveness of these pre-trained models in classifying lung and colon cancer images, it must be stated that the reported results are for binary classifications on subclasses of the dataset with the largest subclass containing 10 000 images. [14] proposed research directions to assist in early-stage detection of cancer while identifying gaps for future development of lung cancer detection in medical IoT devices based on various ML algorithms utilized for detecting a number of diseases.

In [15], a Convolutional Neural Network (CNN) model was proposed that was characterized by its speed of diagnosis and high accuracy with few parameters for diagnosing colon cancer. The model consists of two paths where each path is responsible for creating 256 feature maps to increase the number of features at different levels in order to improve the accuracy and sensitivity of the classification. They compared the performance of their CNN model with VGG-16 model on the 10 000 colon CT images of the LC25000 dataset and reported classification accuracies of 99.6% and 96.2%, respectively.

Having developed a lung cancer detection model using InceptionV3, Histogram of Gradients (HoG) and Daisy feature extraction, [6] obtained 99.6% accuracy in classifying benign and malignant lung tissues using 15000 lung CT scans from the LC25000 dataset. [12] introduced a lung cancer classification using Particle Swarm Optimization (PSO), Genetic Algorithm (GA), and Support Vector Machine (SVM). In terms of classification, PSO-GA-SVM outperformed SVM without parameter optimization. The accuracy, precision, recall, and F1-score values for the PSO-GA-SVM, were discovered to be 97.69%, 98.46%, 98.82%, and 97.66%, respectively.

A DL-based classification framework for lung and colon cancer diagnosis using ML was presented in [11]. A framework was proposed that can help medical professionals identify as well as differentiate among five types of lung and colon tissues. A supervised learning approach with a DL model was used to identify three cancerous and two non-cancerous lung and colon tumors. In that research, pathological images were obtained, relating to these types of cancers from the LC25000 dataset, which was also used to train and validate the approach. After obtaining approximately 96% accuracy and proving the superiority of this method over other similar cancer detection methods, it was concluded that this computer-based identification method would allow less costly pathologists' diagnoses of lung and colon cancer cases with minimal effort and time.

In [20], three CNN models were trained and tested for colon cancer detection using the 10 000 colon CT images of the LC25000 dataset. Two of the three CNN models

were built from scratch while one of them was a pre-trained CNN (MobileNetV2). Of the two models built from scratch, one used average-pooling and reported an accuracy of 95.48%, while the other used max-pooling and reported an accuracy of 97.49%, but the pre-trained MobileNetV2 outperformed them both with a classification accuracy of 99.67%. All accuracies were reported on 20% of the dataset while the remaining 80% were used to train the models.

In [2], a framework was proposed based on multiple lightweight DL models (ShuffleNet, MobileNet and SqueezeNet) for the detection of lung and colon cancer from CT images in the LC25000 dataset. Following the extraction of deep features by these lightweight DL models, features transformations such as Principal Component Analysis (PCA) and Fast Walsh-Hadamard Transform were employed to reduce the dimension of the features and extract a relevant subset of dense features. The resulting feature subsets from the two transforms were then used to train four ML models out of which Support Vector Machine had the best accuracy of 99.6%.

In [9], seven different DL architectures were tested on the colon images subset of the LC25000 dataset. One of the DL architectures was a 9-layer CNN consisting of convolution layers, max-pooling layers, a flatten layer, a dropout layer, and dense layers. The other six models were pre-trained models (VGG-16, EfficientNetB0, ResNet101V2, ResNet50, DenseNet121, and MobileNetV2). All the models were trained with 80% of the data, and validated with 10% of the data while the remaining 10% was held out for testing. The authors reported that their 9-layers CNN had the best accuracy of 99.8% and thus outperformed all the pre-trained models. While this is commendable, it is not justifiable from a computational standpoint, especially because they did not provide details of how the pre-trained models were used for transfer learning (hyperparameter tuning, whether layers were added, retrained etc.). However, the performance of this 9-layer CNN is commendable, the training and validation plots show good generalization, yet we believe that the accuracy of this 9-layer architecture is likely to drop when used for a multi-class classification like ours.

To develop a transfer-learning model for detecting lung and colon cancer using LC25000 histopathological images, [13] tuned and used a pre-trained model, AlexNet, to classify the CT images of lung and colon tissues. Initially, the model achieved an overall accuracy of 89.9%. A so-called class-selective image processing method was employed to identify the underperforming class and selectively preprocess the images in that class. This method improved the model accuracy to 98.4% showing that simple and efficient image processing methods can improve ML model performance.

In [17], an analysis of lung cancer using a deep neural network was presented. 15 000 samples of histopathological photographs of lung adenocarcinoma, lung squamous cell carcinoma, and benign lung tissue from three different types were used. Histopathological photographs of lung tissues were classified using a Computer-Aided Diagnostic (CAD)

method. Using various pre-trained models with hyper-tuning, the best accuracy was achieved from ResNet 101, a CNN network, at 98.67%.

In [23], a method for detecting and classifying colon cancer was proposed, using what was called MA_ColonNET. This MA_ColonNET was built from a 45-layer CNN architecture consisting of 2D convolution layers and max-pooling layers. MA_ColonNET achieved a 99.75% accuracy on 2000 lung CT images which were used for testing the trained model.

From the reviewed works, it has been observed that achieving very high accuracies (over 99%) on "out-of-training" samples has been relatively difficult when building a single model to predict both lung and colon cancer, compared with separate models for lung and colon cancer detection. This is most likely due to the differences in the image structures between the different disease types. However, considering the computational demand of training, validating, testing, and deploying DL models, it is more expensive to train two separate models on one dataset than it is to train a single model. More so, since the LC25000 dataset has come as an answer to a request for larger datasets for medical image analysis, it is sensible for the computer vision research community to take advantage of its large size to develop better predictive models.

3. Materials and methods

In this work a DL approach for cancer detection from CT images is proposed. We have experimented with a relatively large dataset of 25 000 images, details of which can be found in section 3.1. Our DL architecture employs the pre-trained EfficientNet-B7 [19] as the backbone and adds a few dense layers on top of it. The details of the architecture are presented in section 3.2. The experimental setup is presented in section 3.3.

3.1. Dataset

The dataset used in this work is the lung and colon cancer histopathological dataset [3] which is also referred to as LC25000. It originally contains 250 images each of *Lung benign tissue* (`Lung_n`), *Lung adenocarcinoma* (`Lung_aca`), *Lung squamous cell carcinoma* (`Lung_scc`), *Colon adenocarcinoma* (`Colon_aca`), and *Colon benign* (`Colon_n`) tissue totaling 750 images which were then augmented to 25 000 images with a total of 5000 images in each of the five classes. For the purpose of testing the performance of our network, we created a hold-out test set that contained 20% of the total size of the dataset (i.e. 5000 images). The remaining 20 000 images were used for training and validation in order to experiment and improve the architecture before finally testing it on the test set. The validation split is 10% of the remaining 20 000 images (i.e. 2000 images), while the remaining 18 000 images were used for training.

It is noteworthy that the hold-out test set was created with an equal number of

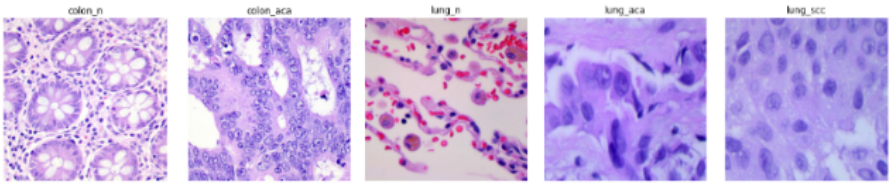

Fig. 1. Samples of CT images from the LC25000 dataset [3].

Layer (type)	Output Shape	Param #	Connected to
global_average_pooling2d_2 (Glo	(None, 2560)	0	top_activation[0][0]
flatten_2 (Flatten)	(None, 2560)	0	global_average_pooling2d_2[0][0]
dense_8 (Dense)	(None, 256)	655616	flatten_2[0][0]
dense_9 (Dense)	(None, 128)	32896	dense_8[0][0]
dropout_2 (Dropout)	(None, 128)	0	dense_9[0][0]
dense_10 (Dense)	(None, 64)	8256	dropout_2[0][0]
dense_11 (Dense)	(None, 5)	325	dense_10[0][0]

Total params: 64,794,780
Trainable params: 697,093
Non-trainable params: 64,097,687

Fig. 2. Architecture of the top layers attached to EfficientNet-B7 for cancer detection from CT images.

images (1000) from each of the five classes. In the same vein, the validation split was done so that equal quantities (400) of images from each class are present in the split. This balance is important to ensure that class imbalance does not bias the classification probabilities of each class at training. Figure 1 shows samples of images in the dataset – one from each of the five classes.

3.2. Network Architecture

The EfficientNet-B7 architecture was developed by the Google Research Brain team. It was first presented in the paper by [19] in which a family of neural network models (B0 to B7) was developed by uniformly scaling all convolution dimensions of depth, width, and resolution using compound scaling with a fixed ratio.

While EfficientNet-B7 showed improved accuracy on the ImageNet dataset [7] at that time (84.4% accuracy), the architecture also had fewer mode parameters compared to the state-of-the-art. This means that it can be trained in less time and with fewer computing resources and these are desirable features for transfer learning.

To create our cancer detection architecture, we reused the pre-trained EfficientNet-B7 without its top layer. Then, we added 4 dense layers on top of the pre-trained architecture as shown in Figure 2. The EfficientNet model architecture is very deep due

Tab. 1. Hyperparameter settings.

Hyperparameter	Value
Input image size	$224 \times 224 \times 3$
Base learning rate	$5 \cdot 10^{-5}$
Batch size	128
Optimizer	Adam
Seed	42
Loss function	Categorical cross entropy
Evaluation metric	Accuracy

Tab. 2. Accuracy and loss values.

	Training	Validation	Testing
Accuracy [%]	99.63	99.50	99.72
Loss	0.0124	0.0145	0.0103

to the compound scaling method adopted. The architecture was trained with images of dimension $224 \times 224 \times 3$, thus, that is the input dimension specified in the input layer of the architecture.

The final dense layer of the network has five units for classifying the input image into one of the five classes. The network was trained with a base learning rate of $5 \cdot 10^{-5}$, using the Adam optimizer and a categorical cross-entropy loss function. We also included a dropout layer to drop 30% of the neurons in the previous layer in order to combat overfitting and obtain a robust model. The model was trained with early stopping for 44 epochs. We set a seed value of 42 to ensure reproducibility in the random selection of the test set. The modified EfficientNet-B7 architecture was trained on 18 000 CT images, validated on 2000 CT images and tested on a hold-out set of 5000 images.

3.3. Experimental setup

Experiments were carried out on Kaggle's GPU $T4 \times 2$ and the Tensorflow library was employed for DL. Table 1 shows the hyperparameter settings employed for the experiments.

4. Results and discussion

The results of training, validation, and testing are presented in Table 2. From the table, it can be observed that the model training and prediction are in step and it is not overfitting.

Tab. 3. Comparison of cancer prediction models on LC25000 dataset.

No.	Ref.	Cancer type	ML/DL model	Validation protocol	Acc. [%]
1	[8]	Lung and colon	Inception, Xception, VGG-16, ResNet, MobileNet, DenseNet, NASNetMobile	Train (80%); val. (20%)	96-100
2	[15]	Colon	VGG-16	Train (70%); val. (30%)	99.6
3	[20]	Colon	MobileNetV2	Train (80%); val. (20%)	99.67
4	[17]	Lung	ResNet 101	Train (80%); val. (20%)	98.67
5	[23]	Colon	MA_Colon NET	Train (80%); val. (20%)	99.75
6	[21]	Lung and colon	DarkNet+ SVM	Train (70%); val. (30%)	99.69
7	[10]	Lung	CNN	Train (90%); val. (10%)	97.2
8	[16]	Colon	DenseNet, ResNet, SVM, RF, KNN, XGB	Train (75%); val. (25%)	98.53
9	[5]	Lung	Ensemble	Train (80%); val. (20%)	99.6
10	[1]	Lung and colon	DHS-CapsNet	Train (70%); val. (15%); test (15%)	99.23
11	Ours	Lung and colon	EfficientNetB7	Train (∼70%); val. (∼10%); test (20%)	99.72

The accuracy and loss plots in Figures 3 and 4 are a confirmation of the fact that the model is not overfitting. They show a smooth progression of training towards convergence and this is an indication of the robustness of the EfficientNet architecture for transfer learning on a task such as cancer detection from CT images.

Figure 5 shows the precision, recall, and F1-score values on the test set, and Figures 6 and 7 show the confusion matrix with and without normalization of the predictions on the test set. These two figures again show the details of the predictive power of the developed cancer detection model as there were very negligible misclassifications at test time. It is noteworthy that this model performance was achieved without further data augmentation on the 18000 training images and in 44 training epochs. This is an indication that the compound scaling method of EfficientNet is effective in detecting cancerous cell nodules in tissue CT images.

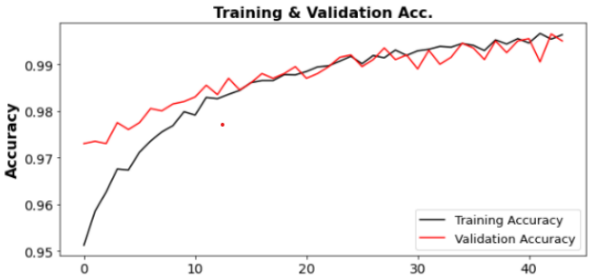

Fig. 3. Training and validation accuracy.

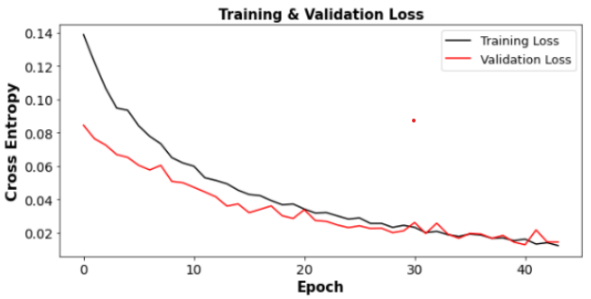

Fig. 4. Training and validation loss.

	precision	recall	f1-score	support
0	0.9990	0.9970	0.9980	1000
1	1.0000	1.0000	1.0000	1000
2	0.9930	0.9950	0.9940	1000
3	0.9990	0.9980	0.9985	1000
4	0.9950	0.9960	0.9955	1000
accuracy			0.9972	5000
macro avg	0.9972	0.9972	0.9972	5000
weighted avg	0.9972	0.9972	0.9972	5000

Fig. 5. Precision, recall and F1-score values.

We compared the performance of our developed model with other existing models on the LC25000 dataset. Table 3 shows our comparative analysis of works published between the years 2020 and 2022 on cancer detection from CT images using the LC25000 dataset. Most of the works have employed various pre-trained DL models, but none had employed the EfficientNet model for this task. Also, it can be observed that our developed model

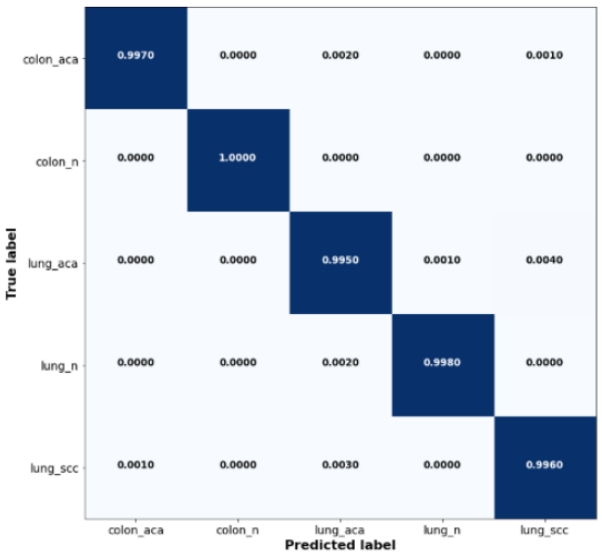

Fig. 6. Normalized confusion matrix.

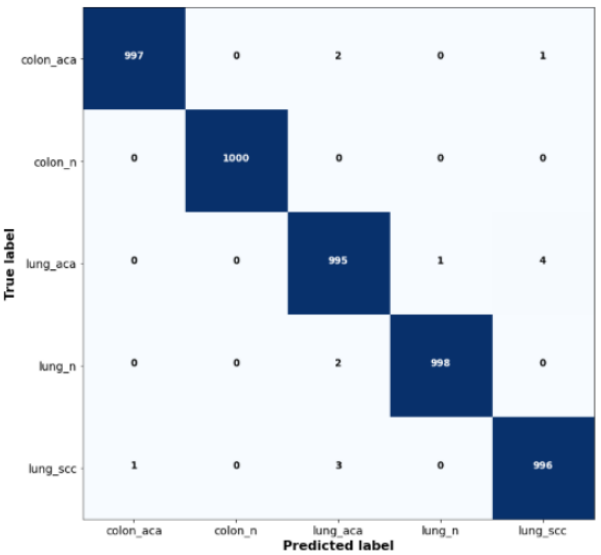

Fig. 7. Confusion matrix without normalization.

significantly outperforms a good number of the works presented. However, a few works, which performed at par with our work, have been closely 'x-rayed' and our analysis and discoveries are presented in the next paragraph.

The paper [8] reported a 100% accuracy on the LC25000 dataset, but upon careful examination, we found that they performed three binary classification tasks on three subsets of the dataset and each of these subsets contains 10000 images. Therefore, their classification model cannot be said to be as robust as ours. [15] used only the colon cancer subset of the LC25000 dataset which contained only 10000 images. [21] used only training and validation split and this method has been shown to result in overfitting to the validation set. More so, the DarkNet part of their model was trained for 3000 epochs and then some feature optimizations were carried out on the features before they were used by SVM for a final classification. This does not compare to our model which was trained for just 44 epochs yet achieved higher accuracy. As for [23], their work was also built on the 10 000 images of colon tissues, which is just a binary classification task and does not come near our multi-class classification of 5 classes ranging from lung to colon cancer.

5. Conclusions and future work

While the research in the detection of lung cancer using ML and DL seems to be getting more attention than colon cancer, both deadly diseases ultimately require early detection and diagnosis by health practitioners in order to mitigate the spread to other parts of the body. The use of ML and DL methods to aid in the diagnosis of these diseases will not only reduce the burdens of relevant stakeholders in cancer diagnosis, but it also holds great prospects for faster and more accurate diagnosis, thus resulting in fewer fatalities.

In this work, a robust DL method for lung and colon cancer has been developed. Trained on only 18 000 images, the model was able to detect cancer in a hold-out test set containing 5000 images to a very high degree of accuracy (99.72%). The fact that existing works also have similar accuracies on the same dataset is a pointer to the fact that we are drawing closer to a breakthrough in AI-assisted diagnosis of cancer disease.

Future work should focus on experimenting on a larger corpus, preferably from heterogeneous sources (e.g., various hospitals, various machines, etc.). More work also needs to be done in terms of interpretation and analysis of the predictions to medical experts and patients in order to improve the trust and confidence in AI-assisted diagnosis of cancer.

Acknowledgments

We would like to express our gratitude to BigCodeGen LLC for its generous support of our research project. We also appreciate the reviewers for their valuable comments.

References

[1] K. Adu, Y. Yu, J. Cai, K. Owusu-Agyemang, B. A. Twumasi, and X. Wang. DHS-CapsNet: Dual horizontal squash capsule networks for lung and colon cancer classification from whole slide histopathological images. *International Journal of Imaging Systems and Technology*, 31(4):2075–2092, 2021. doi:10.1002/ima.22569.

[2] O. Attallah, M. F. Aslan, and K. Sabanci. A framework for lung and colon cancer diagnosis via lightweight deep learning models and transformation methods. *Diagnostics*, 12(12), 2022. doi:10.3390/diagnostics12122926.

[3] A. A. Borkowski, M. M. Bui, L. B. Thomas, C. P. Wilson, L. A. DeLand, and S. M. Mastorides. LC25000 Lung and colon histopathological image dataset. 2019. `https://github.com/tampapath/lung_colon_image_set`.

[4] CDC. An update on cancer deaths in the united states. In *Center for Disease Control (CDC)*. 2022. `https://www.cdc.gov/cancer/dcpc/research/update-on-cancer-deaths/`.

[5] H. Chen, X. Xu, T. Ge, C. Hua, X. Zhu, Q. Wang, Z. Yu, and R. Zhang. A novel tool for the risk assessment and personalized chemo-immunotherapy response prediction of adenocarcinoma and squamous cell carcinoma lung cancer. *International Journal of General Medicine*, 14:5771–5785, 2021. doi:10.2147/IJGM.S327641.

[6] M. Chen, S. Huang, Z. Huang, and Z. Zhang. Detection of lung cancer from pathological images using CNN model. In *Proc. 2021 IEEE Int. Conf. Computer Science, Electronic Information Engineering and Intelligent Control Technology (CEI)*, pages 352–358, Fuzhou, China, 24-26 Sep 2021. doi:10.1109/CEI52496.2021.9574590.

[7] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A large-scale hierarchical image database. In *Proc. 2009 IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, pages 248–255, Miami, FL, USA, 20-25 Jun 2009. doi:10.1109/CVPRW.2009.5206848.

[8] S. Garg and S. Garg. Prediction of lung and colon cancer through analysis of histopathological images by utilizing pre-trained CNN models with visualization of class activation and saliency maps. In *Proc. 3rd Artificial Intelligence and Cloud Computing Conf. (AICCC) 2020*, ACM International Conference Proceeding Series, pages 38–45. Association for Computing Machinery, Kyoto, Japan, Dec 18-20 2020. doi:10.1145/3442536.3442543.

[9] Md I. Hasan, Md S. Ali, Md H. Rahman, and Md K. Islam. Automated detection and characterization of colon cancer with deep convolutional neural networks. *Journal of Healthcare Engineering*, 2022:5269913, 2022. doi:10.1155/2022/5269913.

[10] B. K. Hatuwal and H. C. Thapa. Lung cancer detection using convolutional neural network on histopathological images. *International Journal of Computer Trends and Technology*, 68(10):21–24, 2020. doi:10.14445/22312803/IJCTT-V68I10P104.

[11] 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*, 21(3):748, 2021. doi:10.3390/s21030748.

[12] F. Maulidina, Z. Rustam, and J. Pandelaki. Lung Cancer Classification using Support Vector Machine and Hybrid Particle Swarm Optimization-Genetic Algorithm. In *Proc. 2021 Int. Conf. Decision Aid Sciences and Application (DASA)*, pages 751–755, Sakheer, Bahrain, 07-08 Dec 2021. doi:10.1109/DASA53625.2021.9682259.

[13] S. Mehmood, T. M. Ghazal, M. A. Khan, M. Zubair, M. T. Naseem, T. Faiz, and M. Ahmad. Malignancy detection in lung and colon histopathology images using transfer learning with class selective image processing. *IEEE Access*, 10:25657–25668, 2022. doi:10.1109/ACCESS.2022.3150924.

[14] K. Pradhan and P. Chawla. Medical Internet of things using machine learning algorithms for lung cancer detection. *Journal of Management Analytics*, 7(4):591–623, 2020. doi:10.1080/23270012.2020.1811789.

[15] Y. Qasim, H. Al-Sameai, O. Ali, and A. Hassan. Convolutional neural networks for automatic detection of colon adenocarcinoma based on histopathological images. In F. Saeed, F. Mohammed, and A. Al-Nahari, editors, *Innovative Systems for Intelligent Health Informatics: Proc. Int. Conf. Reliable Information and Communication Technology (IRICT)*, volume 72 of *Lecture Notes on Data Engineering and Communications Technologies*, pages 19–28. Springer, Cham, 22-23 Dec 2021. doi:10.1007/978-3-030-70713-2_3.

[16] D. Sarwinda, A. Bustamam, R. H. Paradisa, T. Argyadiva, and W. Mangunwardoyo. Analysis of deep feature extraction for colorectal cancer detection. In *Proc. 2020 4th Int. Conf. Informatics and Computational Sciences (ICICoS)*, pages 1–5, Semarang, Indonesia, 10-11 Nov 2020. doi:10.1109/ICICoS51170.2020.9298990.

[17] S. Shandilya and S. R. Nayak. Analysis of lung cancer by using deep neural network. In M. Mishra, R. Sharma, Rathore A. K., J. Nayak, and B. Naik, editors, *Proc. 2nd Conf. Innovation in Electrical Power Engineering, Communication, and Computing Technology (IEPCCT 2021)*, volume 814 of *Lecture Notes in Electrical Engineering*, pages 427–436, Bhubaneswar, India, 24–26 Sep 2022. Springer, Singapore. doi:10.1007/978-981-16-7076-3_37.

[18] Md A. Talukder, Md M. Islam, Md A. Uddin, A. Akhter, K. F. Hasan, and M. A. Moni. Machine learning-based lung and colon cancer detection using deep feature extraction and ensemble learning. *Expert Systems with Applications*, 205:117695, 2022. doi:10.1016/j.eswa.2022.117695.

[19] Z. Tan, Y. Yang, J. Wan, G. Guo, and S. Z. Li. Deeply-learned hybrid representations for facial age estimation. In *Proc. 28th Int. Joint Conf. Artificial Intelligence (IJCAI)*, pages 3548–3554, Macao, China, 10-16 Aug 2019. doi:10.24963/ijcai.2019/492.

[20] Z. Tasnim, S. Chakraborty, F. M. J. M. Shamrat, A. N. Chowdhury, H. A. Nuha, A. Karim, S. B. Zahir, and Md M. Billah. Deep learning predictive model for colon cancer patient using CNN-based classification. *International Journal of Advanced Computer Science and Applications*, 12(8):687–696, 2021. doi:10.14569/IJACSA.2021.0120880.

[21] M. Toğaçar. Disease type detection in lung and colon cancer images using the complement approach of inefficient sets. *Computers in Biology and Medicine*, 137:104827, 2021. doi:10.1016/j.compbiomed.2021.104827.

[22] WHO. Cancer: Key facts, 7 Feb 2022. `https://www.who.int/news-room/fact-sheets/detail/cancer`.

[23] M. Yildirim and A. Cinar. Classification with respect to colon adenocarcinoma and colon benign tissue of colon histopathological images with a new CNN model: MA_ColonNET. *International Journal of Imaging Systems and Technology*, 32(1):155–162, 2022. doi:10.1002/ima.22623.