A comparative study for lung, colon and breast cancer diagnosis using different convolutional neural networks

Hussein Akar[1], Jacqueline Abou Diab[1], Fatima Sbeity[1,2], Mohamad Abou Ali[1,3], Abdallah Kassem[4], Lara Hamawy[1]

[1]Dept. of Biomedical Engineering, Lebanese International University, Lebanon
[2]Dept. of Physics & Electronics, Lebanese University, Lebanon
[3]Dept.of Computer Science & Artificial Intelligence, University of the Basque Country, Spain
[4] Dept. of Electrical &Computer Engineering, Notre Dame University, Lebanon

2023 Seventh International Conference on Advances in Biomedical Engineering (ICABME) | 979-8-3503-2585-0/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICABME59496.2023.10293120

ABSTRACT— **In modern times, the prevalence of diseases, particularly cancer, is rising rapidly. Diagnostic errors pose a critical safety concern in healthcare due to the complexity of medical decision-making, disease variability, and human cognitive limitations. This paper aims to compare different convolutional neural networks (CNNs) to find the most accurate model that can classify lung, colon, and breast cancer using histopathological images. A convolutional neural network is a subset of machine learning. It is one of the various types of artificial neural networks which are used for different applications and data types. A CNN is a kind of network architecture for deep learning algorithms and is specifically used for image recognition and tasks that involve the processing of pixel data. We created a baseline model that consists of a basic CNN architecture. It serves as a starting point for comparison with more complicated models. The baseline model's performance is then compared to the performance of various pre-trained models (ResNet-50, ResNet-101, VGG-16, VGG-19, Inception V3, and DenseNet-121). The dataset used consists of 25,000 histopathological images of both the lung and colon and 277,000 histopathological of breast images. Results show that the ResNet-50 model is the most efficient among all the other models, scoring a validation accuracy of 98.3%. Additionally, ResNet-101, VGG-16, VGG-19, and DenseNet-121 also demonstrated strong performance, scoring 95.8%, 96.6%, 95.1%, and 92.4%, respectively. Meanwhile, Inception-V3 displayed a fair result, scoring a validation accuracy of 73.2%.**

Keywords—CNN, histopathological images, deep learning, baseline model, lung cancer, colon cancer, breast cancer.

I. INTRODUCTION

Nowadays, diseases are spreading more and more, and there is no denying the fact that cancer is a disease that commands a lot of shock, fear, and stress in popular perception. Lung, colon, and breast cancer are the most frequent types of cancer observed in the whole world, with a lifetime risk. This makes most of the world's organizations and countries put effort into the early discovery of these types of cancers. The early discovery of cancer can improve the rate at which treatment is successful and help the patient live normally [1]. Deep learning is vital in cancer diagnosis as it utilizes advanced machine learning techniques to analyse complex medical imaging data. Convolutional neural networks (CNNs) excel at automatically extracting relevant features from histopathological images, enabling accurate cancer cell and tumor detection and classification. Cancer is a rapid reaction of aberrant cells that expand beyond their acceptable boundaries, allowing them to infect and travel to neighbouring tissues, which is known as metastasis [2]. Metastasis is one of the prime factors of cancer-related death because it spreads from the place where it first formed to another part of the body. There are two types of tumors: benign and malignant tumor. A benign tumor is an abnormal but noncancerous collection of cells. It grows more slowly, has even borders, and doesn't spread to other parts of your body. Malignant tumors are cancerous; they spread to various parts via the bloodstream or the lymphatic system [3]. Colorectal cancer, often referred to as colon cancer, is a leading cause of death characterized by uncontrolled cell growth in the colon or rectum, often originating from benign polyps. Lung cancer, the most widely recognized malignancy, forms in lung tissues and can affect one or both lungs, either as primary or secondary cancer. It ranks as the top cause of cancer-related deaths for both genders. Breast cancer on the other hand is another disease marked by uncontrolled breast cell growth, with various types depending on the affected cells. It can spread through blood vessels and lymph nodes and is a significant cause of mortality among women worldwide.

The rest of the paper is arranged as follows: related work is presented in section II, some background about deep learning and CNN is provided in section III, proposed methodology is described in section IV, results and discussion are presented in section V.

II. RELATED WORK

M. Naji et al. in (2021) performed a study about machine learning algorithms for breast cancer prediction and diagnosis. In their study, they applied five machine learning algorithms to the breast cancer Wisconsin diagnostic dataset to predict and diagnose breast cancer, finding that Support Vector Machine achieved the highest accuracy of 97.2% and precision of 97.5% and outperformed all other algorithms. The methodology included data acquisition, pre-processing, and model evaluation through splitting labelled data into training and test sets, and

performance metrics such as confusion matrix, accuracy, precision, sensitivity, F1 score, and AUC were used to evaluate and compare the models. The dataset has 569 instances and 11 attributes, and the features are computed from a digitized image of a breast cancer sample obtained from a fine-needle aspirate [5]. Moreover, D. Nayagam et al. in (2022) accomplished a study in colon cancer classification on histopathological images using deep learning techniques; a convolutional neural network (CNN) was used to classify colon adenocarcinoma and benign colonic tissue histopathological images. Supervised learning and deep learning techniques were employed for efficient and high-performance classification. The image dataset was split into 20% for testing and 80% for training, with a total of 10,000 images. The trained CNN model achieved 99.7% accuracy in classifying the images. This approach demonstrated high accuracy, reduced computational time, and minimal resource requirements compared to alternative methods, indicating superior performance and lower error rates [6]. Md. Alamin Taukder et al. in (2022) introduced a hybrid ensemble feature extraction model for efficient identification of lung and colon cancer. The model combines deep feature extraction, ensemble learning, and high-performance filtering. Evaluation was conducted on histopathological lung and colon datasets (LC25000) comprising 25,000 images. Pre-processing involved resizing to 128x128, bgr2rgb conversion, feature scaling, and labeling. Five transfer learning models were used for feature extraction, followed by six ML algorithms. High-Performance Filtering selected the top three algorithms for Ensemble Learning. The proposed model achieved impressive accuracy rates: 99.05% for lung cancer, 100% for colon cancer, and 99.30% for lung and colon cancer. Performance metrics, including accuracy, recall, precision, f1-score, MAE, MSE, RMSE, confusion matrix, AUC score, and ROC Curve, confirmed the superiority of the selected model [7].

III. DEEP LEARNING

Deep learning is a subset of machine learning that utilizes artificial neural networks with multiple layers to model and solve complex problems. It involves training a model on a large dataset, allowing it to learn from experience, and then using the trained model to make predictions or decisions about new data. Deep learning models are typically composed of multiple layers of interconnected nodes, each of which performs a simple mathematical operation on its inputs. By chaining together many such layers, deep learning models can learn increasingly complex representations of the data they are trained on. This allows them to discover patterns and relationships that may not be immediately apparent to humans [10].

A. Convolutional neural networks

Convolutional Neural Networks (CNNs) are a type of deep learning model that is commonly used for image recognition and analysis tasks. CNNs are designed to automatically and adaptively learn spatial hierarchies of features from input images, making them well-suited for tasks such as object detection, segmentation, and classification. CNNs are composed of multiple layers, including convolutional layers, pooling layers, and fully connected layers. In a convolutional layer, the model applies a set of filters to the input image, which allows it to detect patterns and features at different scales and orientations. The pooling layer then downsamples the output of the convolutional layer to reduce the size of the feature maps. Finally, the fully connected layers process the features to make predictions or classifications.

IV. METHODOLOGY

The following block diagram shows our proposed model (see figure 1)

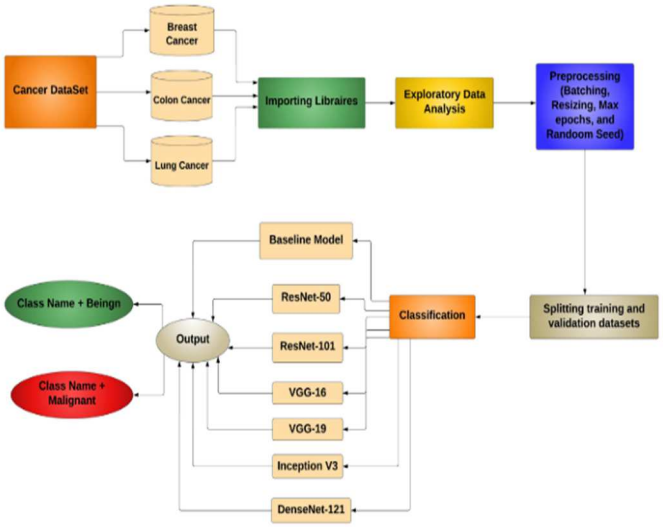

Figure 1 The proposed hybrid model for the classification of lung, colon and breast cancer

First, two datasets were downloaded from Kaggle and loaded into the Google Collaboratory environment, which was stored and accessed from Google Drive. The first dataset was for lung and colon. It contains 25000 histopathological images of 5 classes including lung and colon cancer and healthy samples. There are five classes in the dataset, each with 5,000 images:

- Lung benign tissue
- Lung adenocarcinoma
- Lung squamous cell carcinoma
- Colon adenocarcinoma
- Colon benign tissue [8]

The second dataset was for breast. It was also downloaded from Kaggle; it provides a large, high-quality dataset of breast cancer images. The collection comprises around 277,000 color histopathological images with a total dimension of 50 by 50 [9]. We imported the necessary libraries for building and training the different models, including modules for data manipulation, model definition, performance evaluation, and image processing, like: including os for file system operations, matplotlib for visualization, NumPy for numerical computations, pandas for data manipulation, and TensorFlow for building and training machine learning models.

A. Exploratory data analysis

An early inspection of the dataset was conducted. Figure 2 shows the number of observations or data-points for each class. It can be seen that lung cancer (adenocarcinoma) and lung cancer (squamous cell carcinoma) had the highest number of observations.

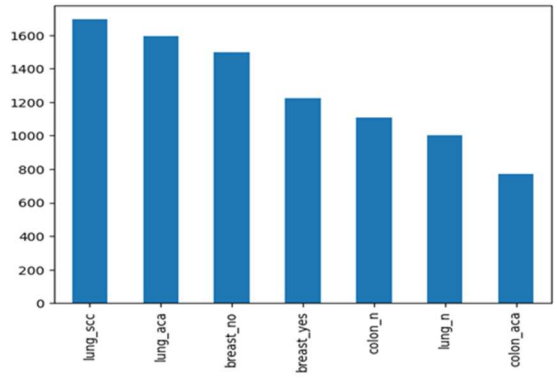

Figure 2 Distribution of Images by Class

In the next step, the samples were also visualized randomly for each class, as shown in figure 3. It can be clearly observed that there are quite a few differentiations in each of the images, so we can assume the model will learn the image features very well.

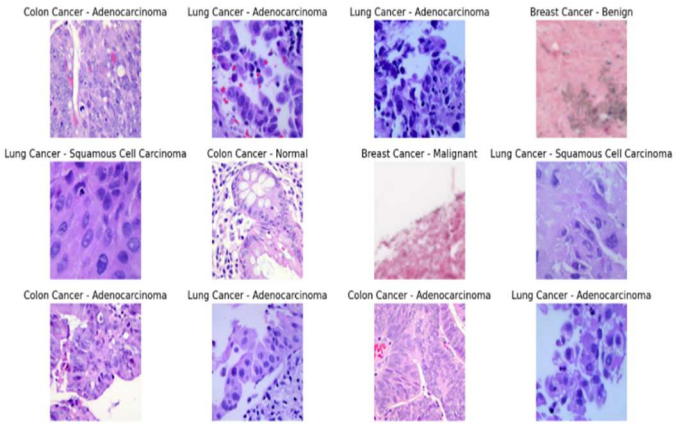

Figure 3 Visualizing Random Images for Each Classes.

B. Preprocessing

The labelled dataset contained different types of histopathological cancer images which includes Breast Cancer - Benign, Breast Cancer - Malignant, Colon Cancer - Adenocarcinoma, Colon Cancer - Normal, Lung Cancer - Adenocarcinoma, Lung Cancer - Benign, and Lung Cancer - Squamous Cell Carcinoma. The dataset is read using the image dataset from the directory function in Keras and is split into an 80% training set and a 20% validation set. The variables batch size, img size, and max epoch are defined to set the size of the batches used for training, the size of the images, and the maximum number of epochs used to train the model. Then set the batch size to 100, img size to 240, and the max epoch to 20, respectively. A smaller batch size, such as 100. It allows for more frequent updates of the model's parameters during training, which can help the model converge faster and potentially reach a better solution. On the other hand, a larger batch size, such as 512, is often used for pre-trained models because these models are already well-optimized and can benefit from larger batches due to their increased computational efficiency. An image size of 240 pixels provides enough detail for the model to learn relevant features while also being computationally efficient. In addition, a maximum epoch value of 20 is often sufficient to allow the model to converge to a good solution while minimizing the risk of overfitting. Then a random seed is set to ensure the reproducibility of the results.

C. Classification

We create a baseline model that consists of a basic CNN structure of 16 layers (6 Conv2D layers including MaxPooling layers, 6 Batch normalization layers, 1 flatten layer and 3 Dense layers). The baseline model and six pretrained models (ResNet-50, ResNet-101, VGG16, VGG19, Inception V3 and DenseNet 121) are trained on the histopathological images from the labelled dataset. ResNet-50 and ResNet-101 each boasts an impressive stack of 50 and 101 consecutive layers, respectively, comprising a mix of convolutional, pooling, and fully connected layers. Inception-v3 relies on a foundation of multiple stacked Inception modules, seamlessly transitioning into fully connected layers and concluding with a softmax layer for classification purposes. In contrast, DenseNet-121 embraces a comprehensive set of architectural elements, encompassing convolutional layers, batch normalization, ReLU activations, pooling layers, fully connected layers, and a final softmax layer to facilitate classification. The VGG architectures, on the other hand, exhibits a strikingly uniform and sequential structure, marked by the repetitive application of identical convolutional and pooling layers throughout their design.

V. RESULTS AND DISCUSSION

Table 1 represents the results from different models using various evaluation metrics.

Table 1: comparison between different model's performance using various evaluation metrics.

	Training Accuracy	Validation Accuracy	Precision	Recall	F1 Score	Specificity	Matthew Correlation Coefficient	Cohen Kappa	Training speed (Sec/Step)	Predication speed (Sec/Step)
Baseline CNN Model	0.754	0.746	0.831	0.746	0.709	0.690	0.719	0.699	18.428	0.521
ResNet50	**1.000**	**0.983**	**0.983**	**0.983**	**0.983**	0.962	**0.979**	**0.979**	**37.115**	**0.233**
ResNet101	0.981	0.958	0.958	0.958	0.958	0.955	0.951	0.950	51.09	1.25
VGG16	**0.998**	0.966	0.966	0.966	0.966	0.949	0.960	0.960	78.67	0.89
VGG19	**0.999**	0.951	0.952	0.951	0.951	0.933	0.942	0.942	21.49	1.14
Inception V3	0.738	0.732	0.763	0.732	0.723	0.898	0.693	0.684	37.36	0.81
DenseNet-121	0.949	0.924	0.929	0.924	0.924	**0.978**	0.912	0.911	36.011	0.6

Excellent V. Good Good Fair

After conducting a thorough evaluation of various deep learning models, using various evaluation metrics like Precision measures the accuracy of positive predictions, while recall evaluates the model's ability to identify positive instances correctly. The F1 score combines precision and recall to provide a balanced accuracy measure [11]. Specificity assesses the model's ability to identify negative instances correctly. The Matthew Correlation Coefficient (MCC) provides a balanced evaluation of the model's performance by considering true positives, true negatives, false positives, and false negatives. Cohen's Kappa coefficient is a statistical measure that assesses agreement among raters, accounting for chance agreement [12]. we can conclude that ResNet50 outperforms the other models (as shown in table 1) with the highest testing accuracy (98.3%), precision (98.3%), recall (98.3%), F1 score (98.4%), specificity (96.2%), and MCC (97.9%). VGG16, VGG19, and ResNet101 also demonstrated successful performance with validation accuracies of 96.6%, 95.1%, and 95.8%, respectively, but their scores were lower than those of ResNet50.

As the graph below (figure 4) shows, the ResNet-50 Model is the perfect classifier for all classes because most of the curves passed through the top left corner that indicate that the model achieves a high True Positive Rate TPR (1) while maintaining a low False Positive Rate FPR (0) or very close to it, for instance, the class of lung cancer-benign achieves the highest AUC rate (100 %) indicating strong discrimination between the positive and negative instances for this class.

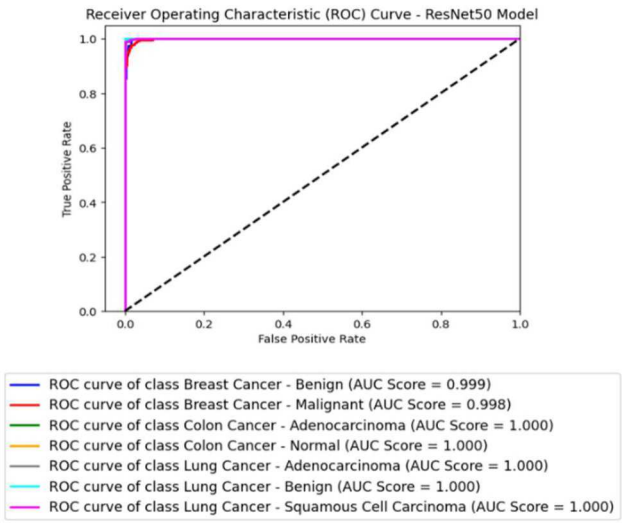

Figure 4 Multiclass Receiver Operating Characteristic (ROC) -- ResNet-50.

CONCLUSION

This paper presents a deep learning approach that utilizes histopathological images for the classification of lung, colon, and breast cancer. The goal is to use deep learning techniques to increase the precision and effectiveness of cancer diagnosis. This study establishes a baseline model which serves as a reference point for evaluating more complex models. The baseline model's result is then compared with ResNet-50, ResNet-101, VGG-16, VGG-19, Inception V3, and Dense Net 121 results. The evaluation metrics used in this study include precision, recall, F1 score, and Matthew's correlation coefficient (MCC). The results show that ResNet-50 outperforms the other models in terms of these metrics, while ResNet-101, VGG-16, and VGG-19 also demonstrate strong performance. This study highlights the significance of architectural changes in generating higher performance and contributes to the growing body of research in the field of deep learning-based systems. Overall, this research has significant implications for the field of cancer diagnosis and represents a valuable contribution to the advancement of medical science.

ACKNOWLEDGMENT

I would like to extend my sincere appreciation to our advisor, Dr. Lara Hamawy, whose guidance, expertise, and unwavering commitment have been instrumental in shaping my research skills and fostering my passion for this field.

REFERENCES

[1] V patil et. al, "Breast cancer detection using MATLAB functions," Advancement in Image Processing and Pattern Recognition, vol. 03, no. 02, p. 06, 2020.

[2] "Cleveland clinic," 20 December 2021. [Online]. Available: https://my.clevelandclinic.org/health/diseases/22213-metastasis-metastatic-cancer.

[3] "Cleveland Clinic," 2022. [Online]. Available: https://my.clevelandclinic.org/health/diseases/22121-benign-tumor#:~:text=A%20benign%20tumor%20has%20distinct,other%20parts%20of%20your%20body.

[4] "American Cancer Society," 18 January 2023. [Online]. Available: https://www.cancer.org/cancer/types/about/what-is.html.

[5] M. Naji et al. "Machine Learning Algorithms For Breast Cancer Prediction and Diagnosis," ScienceDirect, p. 06, 2021.

[6] D. Nayagam, k. Aarthi, S. Mirra "Colon Cancer Classification on Histopathological Images using Deep Learning Techniques," *International Journal of Engineering* Research & Technology (IJERT), vol. 10, no. 08, p. 05, 2022.

[7] Md. Alamin Talukder et al. "Machine Learning-based Lung and Colon Cancer Detection using Deep Feature," *Expert Systems With Applications,* vol. 205, 2022.

[8] Larxel, "Kaggle," 2020. [Online]. Available: https://www.kaggle.com/datasets/andrewmvd/lung-and-colon-cancer-histopathological-images.

[9] P.mooney, "Kaggle," 2018. [Online]. Available: https://www.kaggle.com/datasets/paultimothymooney/breast-histopathology-images.

[10] PK. Suzuki, "Overview of deep learning on medical imaging," Japanese Society of Radiological Technology and Japan Society of Medical Physics, p. 17, 2017.

[11] D. Martin Ward Powers, "Evaluation From Precision, Recall and F-Factor," School of Informatics and Engineering, South Australia, 2007.

[12] S. Maitra, "Importance of Mathews Correlation Coefficient & Cohen's Kappa for Imbalanced Classes," 27 June 2021. [Online]. Available: https://sarit-maitra.medium.com/mathews-correlation-coefficient-for-imbalancedclasses705d93184aed#:~:text=Here%2C%20we%20will%20talk%20about,coefficient%20to%20a%20confusion%20matrix.