Journal of Basrah Researches (Sciences) **51**(1), 1 (2025)
DOI: https://doi.org/10.56714/bjrs.51.1.1

Classification of Lung and Colon Cancer Using Stacked Ensemble Learning of Multiple CNN Architectures

Amal O. Hasan, Zakariya A. Oraibi*

Department of Computer Science College of Education for Pure Sciences University of Basrah, Basrah, Iraq.

ARTICLE INFO

Received 25 Decermber 2024
Revised 31 Decermber 2024
Accepted 7 January 2025
Published 30 June 2025

Keywords:

Lung cancer, Colon cancer, Classification, Xception, MobileNet.

Citation: A. O. Hasan, Z. A. Oraibi , J. Basrah Res. (Sci.) **51**(1), 1 (2025). DOI:https://doi.org/10.56714/bjrs.51.1.1

ABSTRACT

The two most prevalent and deadly types of cancer globally are lung and colon. Early and accurate detection of these cancers is necessary for improving patient outcomes. It was found that machine learning techniques, especially deep learning models, have the ability to generate accurate detection and classification results when trained on big databases. This paper introduces a novel framework to classify five classes of lung and colon cancer using ensemble learning of multiple Convolutional Neural Network (CNN) architectures. First, these models are trained and the predictions resulting from applying them on the test subset of images are preserved, then an ensemble voting is used to select the best result from each model. Our approach was applied on a challenging dataset of 25000 images. The classification accuracy achieved was 98.47% which is higher than the accuracy achieved using the standalone models applied to the same dataset. The potential of our ensemble approach can be further improved in the future.

1. Introduction

One of leading causes of cancer are Lung and colon cancer worldwide [1]. In the meantime, one of the primary strategies to avoid deaths of lung and colon cancer is early detection. [2]. Recently, machine learning techniques, specifically deep learning, have shown huge potential in accurately detecting these cancers by analyzing medical images efficiently. Hence, they can offer an automatic detection and classification of various diseases [3,4].

The vast revolution of Convolutional Neural Networks (CNNs) especially deep learning, allowed for efficient detection and classification of cancer diseases using CT scans or X-ray images. Models and architectures developed for this purpose usually require large amounts of images to allow them to be trained efficiently to produce superb classification accuracy [5]. For research purposes, in 2019, a large dataset of lung and colon images with 25000 images was introduced. This dataset is divided into five classes with 5000 samples per class. The challenge hence lies in the ability to handle such amounts of data during the training stage of a deep learning model since the resources required impose the use of supercomputers to perform this task [6]. In several field including bioinformatics

*Corresponding author email : zakaria au@uobasrah.edu.iq

N: 1817-2695 (Print); 2411-524X (Online)
line at: https://jou.jobrs.edu.iq

field, ensemble methods are increasingly often used to perform prediction tasks such as regression and classification [7]. One of the ensemble learning techniques is Vote-based learning, where each classifier is situated on a variety of weighted training dataset categories [8]. Voting can be used to increase the performance of the model. It is combined the predictions from various models. Also, voting may be utilized for data prediction or classification [9].

In this paper, a framework that is based on stacked ensemble learning of multiple CNN architectures is introduced. These models are: Xception, Inception, and MobileNet. A meta-model is produced using the assembly of these models and voting is applied to pick the best outcome. Hence, the ensemble model with voting was found to produce better classification results that the standalone models. Experiments were carried out on a challenging dataset of 25000 images with four-fold cross validation. In addition, our results were compared with state-of-the-art results applied to the same dataset.

The rest of this paper is organized as follows: Section II presents related work that were applied on the LC25000 datase. Section III describes in details the methodology of the proposed ensemble model. Section IV presents the results of the proposed approach with a comparison with state-of-the-art methods. Finally, we present conclusions and future work in Section V.

2. Literature review

In literature, many techniques were proposed for classification images of LC25000 dataset using deep learning. Some of these methods were applied on the full LC25000 datasets others were applied on a subset of the dataset. Authors in [10] used the BICLCD-TSADL technique. This technique contained many approaches of preprocessing, feature extraction and classification such as GhostNet method for feature extraction, the echo state network (ESN) and Tuna Swarm Algorithm classifiers were utilized for lung and colon cancer images detecting. They applied these models on five class of LC25000 dataset and achieved 99.33% accuracy. The work proposed in [11] employed a prediction framework based on InceptionV3, Daisy features, and Histogram of Gradients (HoG) to classify lung tissues into two classes: benign and malignant using 15000 images of LC25000 and scoring a high accuracy of 99.6%. On the other hand, a hybrid CNN method of VGG-16 architecture CLAHE technique were proposed in [12] achieved an accuracy of 98.96. Multiple deep learning architectures were employed in [13] and achieved an accuracy of 99.30% in classifying LC25000 dataset samples. However, they reported results on two classes which means they applied their framework on 10000 samples only. In [14] a CNN created in this work by using a liner stack of layers. Three hidden layers, one input layer, and one fully connected layer were utilized. This model applied on 15000 lung cancer images and achieved 97.20 accuracy.

Table 1. Summary of literature review

No.	References	Methodology	Database	Accuracy
1	Obayya et al. [10] (2023)	BICLCD-TSADL	LC25000	99.33
2	Chen et al. [11] (2021)	Inception V3, HOG and daisy feature extractions	15000 lung cancer images	99.60
3	Hadiyoso et al. (2023) [12]	CNN-CLAHE-VGG16	LC-25000	98.96
4	Talukder et al. [13] (2022)	Hybrid Model	10000 images of colon cancer	99.30
5	Hatuwal et al. [14]	CNN	15000 lung cancer images	97.20

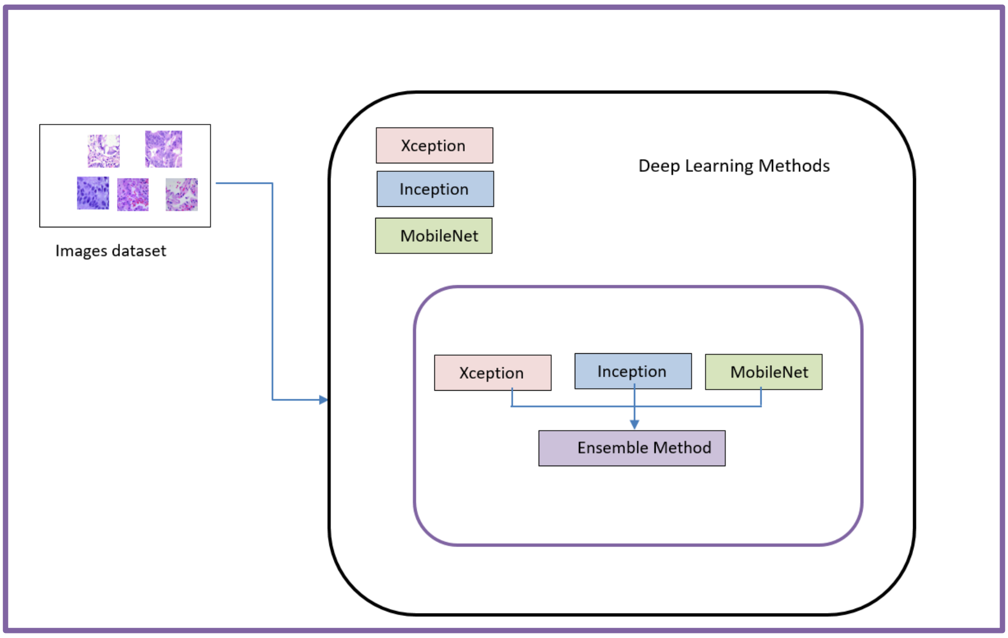

Fig. 1. The proposed model.

3. Methodology

The proposed methodology relies on three powerful CNN models and the ability to combine them using stacked ensemble learning with voting. In this section, details of the methodology will be presented thoroughly.

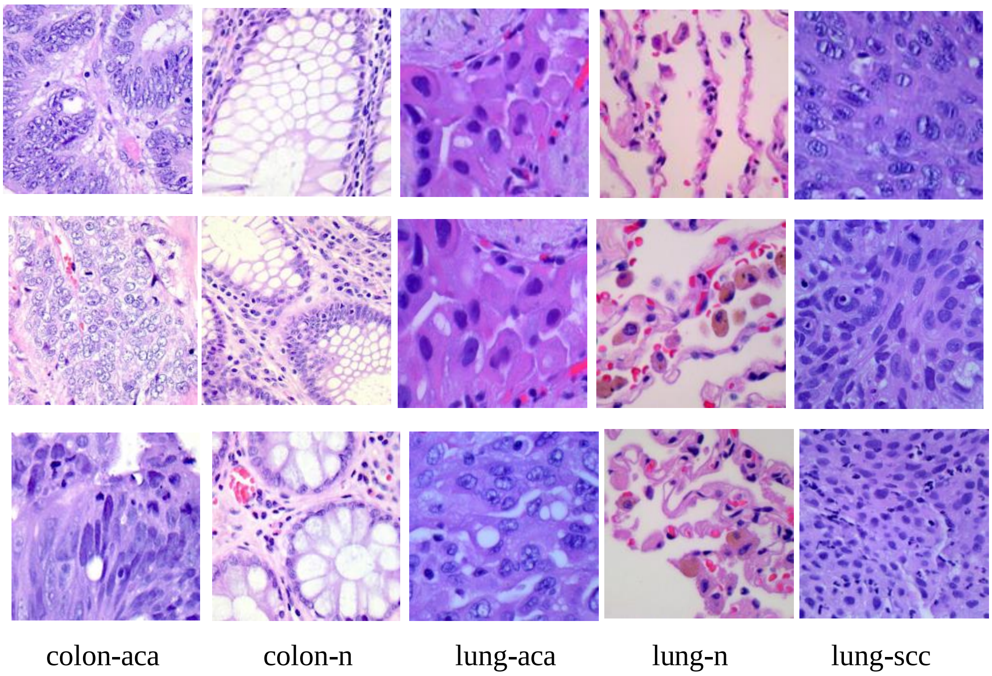

Fig. 2. Sample images from the Lung and Colon (LC25000) dataset.

3.1. Inception Architecture

This model employs a factorization process that separates convolutions into distinct branches to operate on space and channels in succession [15]. A wide range of strategies are utilized for network optimization. In order for Inception to learn multi-scale representations of the input images, tiny kernels are swapped for bigger ones. In this case, the amount of restrictions and complexities are also downsized with a resolution of 768 × 768. The unique criteria of this dataset is its diversity and volume which make it suitable for research purposes.

3.2. Xception Architecture

It is considered a development of Inception architecture. According to [16], Xception is a linear stack of separable convolutional layers with connections. The purpose of these layers is to reduce the need for memory and the expense of computing. The number of these layers is 36 divided into 14 modules. Space-wise along with channel-wise features are learned when the separable convolutions are divided in the Xception model.

3.3.MobileNet Architecture

This architecture is based on depthwise separable convolutions, this convolutions factorize a conventional convolution into a pointwise convolution, which is a 1x1 convolution and a depthwise convolution. A single filter applies to each input channel by using the depthwise convolution. Then, to combine the outputs of the depthwise convolution, a 1×1 convolution then applied by the pointwise convolution [17].

3.4.Stacked Ensemble Model

The proposed methodology depends on the ability to extract the best predictions from the multiple architectures applied on the dataset. Hence, an ensemble strategy is applied to find the best prediction based on hard voting. This ensemble technique ensures less overfitting with high classification performance to create the meta-classifier, Inception, Xception, and MobileNet models are utilized. In the experiments, we demonstrate the reasons to choose this ensemble strategy by showing the amount of accuracy improvement resulting from this combination in comparison with the accuracy resulting from each standalone model.

In order to create the stacked ensemble model. A set of individual models are trained which have been previously described. These individual models (Inception, Xception, and MobileNet) are trained on the LC25000 dataset. After that, the predictions from each model are stored. The meta-learner uses the predictions from the individual models as features. Finally, the meta-learner uses voting to choose the label that has the best probability among the three produced probabilities. Figure (1) shows the proposed model.

3.5. Dataset

The LC25000 dataset utilized in this paper can be considered as a comprehensive set of Lung and Colon cancer images [18]. This dataset consists of five classes, with 5000 images per class. In total, we are provided with 5000 images which makes the dataset a challenging one. Image samples of these classes are shown in Figure (2). The original dataset is divided into two folders, the first one is the lung folder. The lung folder has three classes: Lung Adenocarcinoma (Lung aca), Lung Benign (Lung n), and Lung squamous cell carcinoma (Lung scc). The second folder is colon folder which has two classes, Colon Adenocarcinoma (Colon aca) and Colon Benign (Colon n).

In the experiments, 75% of images from each class were used during the training stage and the remaining 25% images were used for the testing stage. Although some papers in the literature reported the accuracy of using only a subset sample of images from the original dataset, in our paper, we used all images from the five classes during training and testing and reported accuracy using four-fold cross-validation. LC25000 dataset is available to the public and can be used for research purposes.

Table 2. Hyperparameters were used in the experiments

Hyperparameter	**Value**
Epochs	15
Optimizer	Adam
Loss Function	Categorical Crossentropy
Learning Rate	0.001
Batch-Size	32

4. EXPERIMENTAL RESULTS

4.1 Results of Classification

The results of experiments conducted on the LC25000 are elaborated in this section. First, we introduce the list of hyperparameters used during training. After that, we demonstrate the classification results based on the four metrics: accuracy, precision, recall, and F1-score. Finally, a comparison between our proposed approach with state-of-the-art results is given to show the robustness of our methodology. Table (1) lists the hyperparameters used during the training stage. As we can see, we used 15 epochs along with a learning rate of 0.001 since the majority of classification tasks proved to work well using this setting. A batch size of 32 with Adam optimizer has also been utilized.

In the experiments, four-fold cross-validation was used and the accuracy was reported based on the average of four resultant accuracies. All models were trained on 18750 images while testing was performed on 6250 images. The performance metrics used in the paper are as shown in equations 1,2,3, and 4.

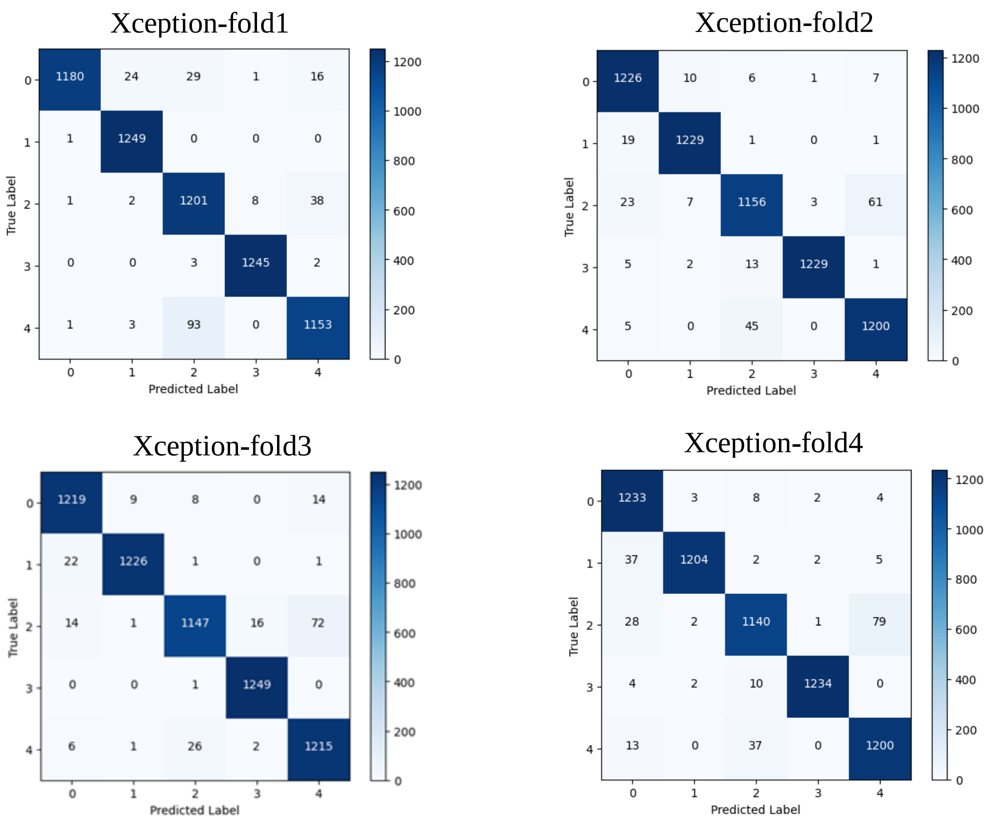

Fig. 3. Confusion Matrix for Xception model in four folders.

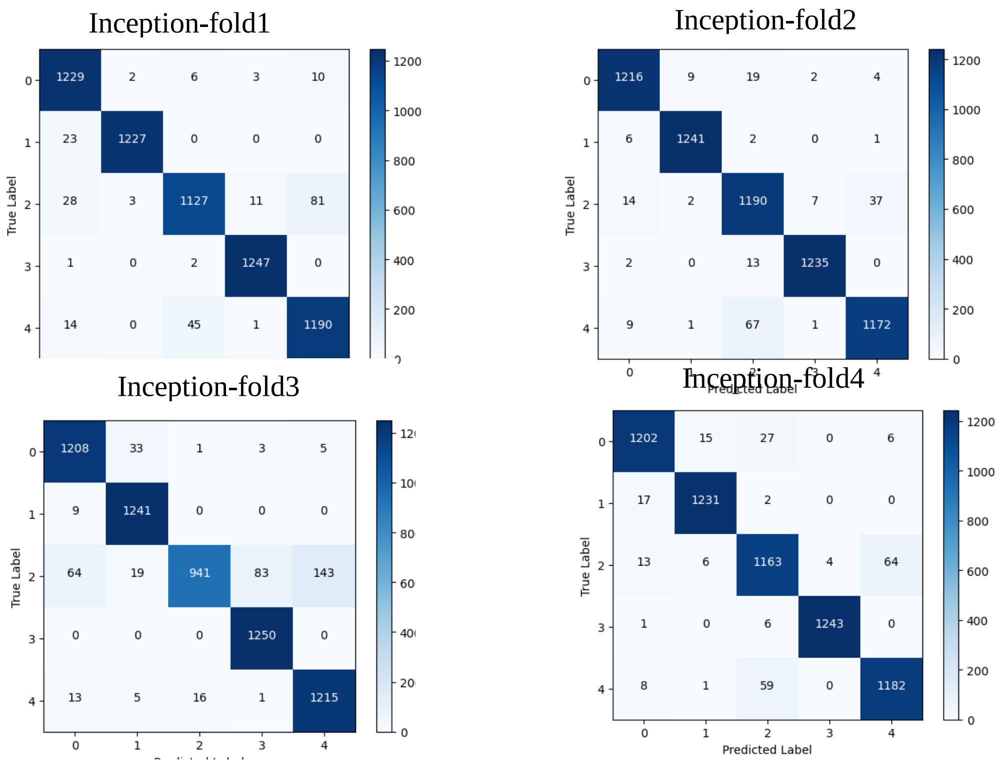

Fig. 4. Confusion Matrix for Inception model in four folders.

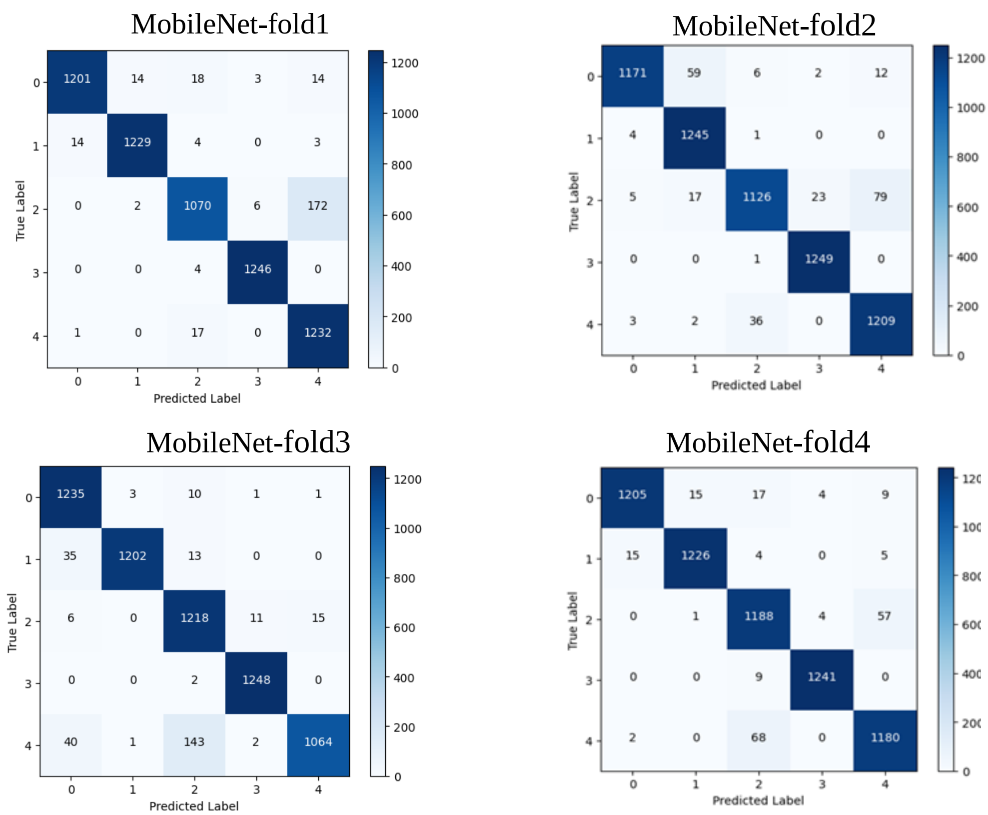

Fig. 5. Confusion Matrix for MobileNet model in four folders.

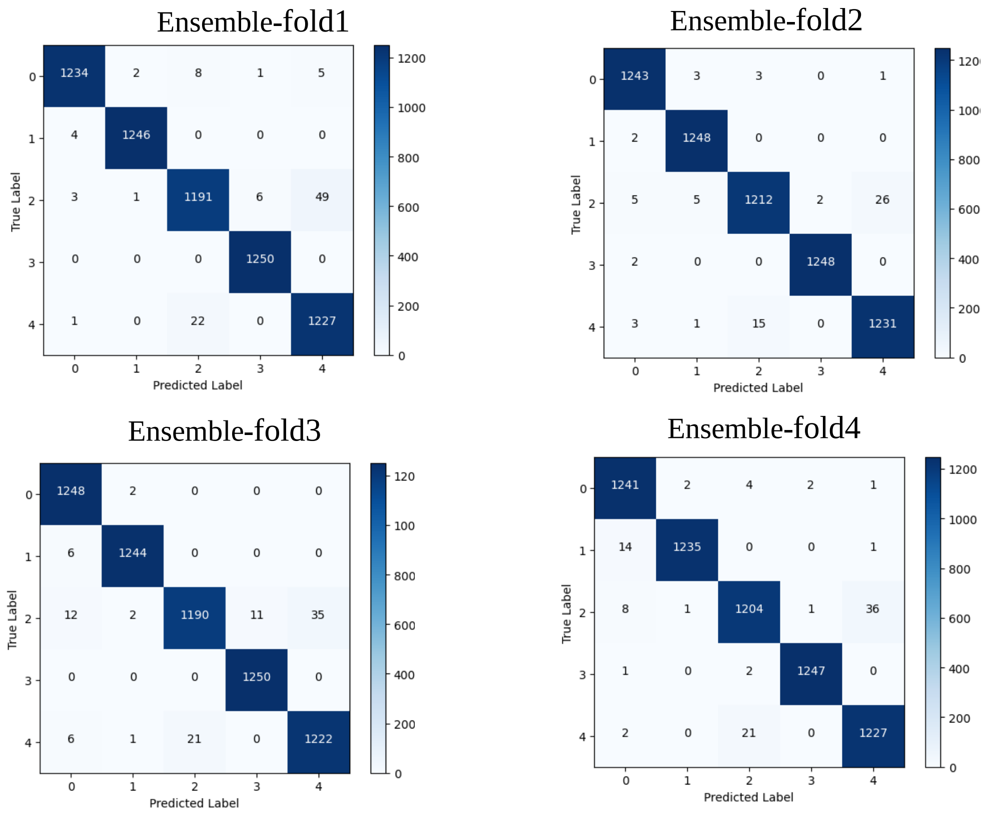

Fig. 6. Confusion Matrix for Ensemble method in four folders.

$$\text{Accuracy (for each class)} = \frac{TP+TN}{TP+FP+TN+FN} \quad \text{......(1)}$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad \text{......(2)}$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad \text{......(3)}$$

$$\text{F1} - \text{Score} = \frac{2*\text{precision}*\text{recall}}{\text{precision}+\text{recall}} \quad \text{......(4)}$$

Figures (3, 4, 5, 6) show the confusion matrices of the four-fold cross-validation of the proposed ensemble methodology and the three deep learning architectures.

Table (2) introduces the results of applying our stacked ensemble approach to the LC25000 images. The results of each standalone model were improved when the ensemble learning approach was applied.

Table 3. The Results of Our Models

Method	precision	Recall	F1-Score	Accuracy
Xception	96.60	96.50	96.50	96.53
Inception	95.95	95.70	95.75	95.79
MobileNet	96.15	95.95	95.80	95.93
Ensemble Method	98.60	98.55	98.65	98.47

4.2. Comparison with State-of-the-Art Classification Techniques

In this section, we compare the performance of the proposed stacked ensemble model of multiple deep learning architectures with the results of other techniques applied on the same LC25000 dataset to show the robustness and efficiency of our model. In Table (3), we elaborate on the results of recent methodologies. Authors in [19] proposed utilizing a multiresolution EfficientNet architecture achieving a high prediction accuracy of 97.24% on the five classes LC25000 database using the EffcientNet-B0 model. In [20] authors, Masud et al. proposed to use a CNN architecture to classify five classes of lung and colon dataset. The accuracy score was just 96.33%. Ijaz et al. [21] proposed a fusion mechanism of both deep learning and Gray Wolf optimization algorithms. The accuracy achieved was 87%. In [22], authors introduced a deep learning framework and applied it on a subset of LC25000 images and achieved a classification score of 97.73%. Wadekar et al. [23] proposed to use a modified version of the VGG-19 model to predict only three classes of Lung cancer images. The accuracy scored using their method was 97.73%. The accuracy achieved by our proposed methodology surpasses the proposed techniques in the literature. Moreover, the framework can be enhanced by using alternative powerful CNN models.

Table 4. Comparison between our approach and state-of-the-art methods

References	Methodology	Dataset	Year	Accuracy
Anjum et al. [19]	EfficientNet-B2	LC-25000	2023	97.24
Masud et al. [20]	CNN	LC-25000	2021	96.33
Ijaz et al. [21]	ResNet50,EfficientNetB0,KNN	LC25000	2022	98.37
Provath et al. [22]	CNN	LC25000	2023	97
Wadekar et al. [23]	Modified VGG-19	15000 Lung	2023	97.73
Our work	Stacked Ensemble Model	LC25000	2024	98.47

5. Conclusion

In this paper, an ensemble deep learning method was applied to classify lung and colon cancer diseases. Three state-of-the-art architectures named: Inception, Xception, and MobileNet were combined using a hard voting strategy to produce the best prediction accuracy. In the experiment, we applied the proposed method on a challenging dataset of 25000 images with five classes. Each class has 5000 images and 75% samples of the original data were used for training and the remaining samples were used for testing. A high classification accuracy of 98.3% was achieved which outperforms the results of each model. This approach has the potential for further enhancement to improve the classification accuracy.

In the future, another ensemble approach will be tested on different deep learning architectures. The purpose is to reach to an optimal accuracy to make the prediction results from the machine as accurate as the human. Moreover, testing on many classes of lung and colon cancer will be carried out to determine how sensitive the model will be in handling such diversity.

References

[1] N. Shahadat , R. Lama and A. Nguyen,” Lung and Colon Cancer Detection Using a Deep AI Model,” Cancers, 16, 3879, November 2024,DOI: 10.3390/cancers16223879.

[2] J. D. Akinyemi, A. A. Akinola, O. O. Adekunle, T. O. Adetiloye, E. J. Dansu,” Lung and Colon Cancer Detection from CT Images Using Deep Learning,” Machine Graphics & Vision 32(1):85–97, 2023, Available: https://bibliotekanauki.pl/articles/3147620.pdf.

[3] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," nature, vol. 521, no. 7553, pp. 436-444, 2015.DOI: https://doi.org/10.1038/nature14539.

[4] A. Esteva et al., "Dermatologist-level classification of skin cancer with deep neural networks," nature, vol. 542, no. 7639, pp. 115-118, 2017. DOI: https://doi.org/10.1038/nature21056.

[5] 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 Proceedings of the 2020 3rd Artificial Intelligence and Cloud Computing Conference, 2020, pp. 38–45, DOI:https://doi.org/10.48550/arXiv.2103.12155.

[6] 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, pp. 2020–08, 2020, DOI:https://doi.org/10.1101/2020.08.15.20175760.

[7] M. Hosni, I. Abnane, A.Idri, J. M. C.D. Gea, J.L.F. Aleman, ” Reviewing ensemble classification methods in breast cancer,” Computer Methods and Programs in Biomedicine, Volume 177, pp. 89-112, August 2019, DOI: 10.1016/j.cmpb.2019.05.019.

[8] A. A. Abro,” Vote-Based: Ensemble Approach,” Journal of Science, 25(3), pp. 871-879, 2021,DOI: https://doi.org/10.16984/saufenbilder.901960.

[9] T. Akhtar, S. O. Gilani, Z. Mushtaq, S. Arif, M. Jamil, Y. Ayaz, S. I. Butt and A. Waris,” Effective Voting Ensemble of Homogenous Ensembling with Multiple Attribute-Selection Approaches for Improved Identification of Thyroid Disorder,” Electronics 10, 3026, 2021, DOI: 10.3390/electronics10233026.

[10] M. Obayya, M.A.Arasi, N.Alruwais, R.Alsini, A.Mohamed, and I. Yaseen,” Biomedical Image Analysis for Colon and Lung Cancer Detection Using Tuna Swarm Algorithm With Deep Learning Model,” IEEE Access, 10.1109, 3309711, 2023, DOI: 10.1109/ACCESS.2023.3309711

[11] M. Chen, S. Huang, Z. Huang, and Z. Zhang, “Detection of lung cancer from pathological images using cnn model,” in 2021 IEEE International Conference on Computer Science, Electronic Information Engineering and Intelligent Control Technology (CEI), IEEE, 2021. pp. 352–358, DOI: 10.1109/CEI52496.2021.9574590.

[12] S. Hadiyoso, S. Aulia, I. D. Irawati, “Diagnosis of lung and colon cancer based on clinical pathology images using convolutional neural network and CLAHI framework,” International Journal of Applied Science and Engineering, vol. 20, no. 1, pp. 1–7, 2023, Available:https://gigvvy.com/journals/ijase/articles/ijase-202303-20-1-006.pdf.

[13] M. A. Talukder, M. M. Islam, M. 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, vol. 205, p. 117695, 2022, DOI: 10.1016/j.eswa.2022.117695.

[14] B.K., Hatuwal, H.C. Tapa, “Lung Cancer Detection Using Convolutional Neural Network on Histopathological Images,” International Journal of Computer Trends and Technology, vol. 68, Issue 10, 21–24, 2020, DOI: 10.14445/22312803/IJCTT-V68I10P104 .

[15] G. K. Raju, P. Padmanabham, and A. Govardhan, “Enhanced content-based image retrieval with trio-deep feature extractors with multi-similarity function,” International Journal of Intelligent Engineering & Systems, vol. 15, no. 6, 2022, DOI: 10.22266/ijies2022.1231.46.
[16] A. Jimenez, J. M. Alvarez, and X. Giro-i Nieto, “Class weighted convolutional features for visual instance search,” arXiv:1707.02581, v. 1, 2017, DOI:https://doi.org/10.48550/arXiv.1707.02581.
[17] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, “Mobilenets: Efficient convolutional neural networks for mobile vision applications,” arXiv:1704.04861, v. 1, 2017, DOI:https://doi.org/10.48550/arXiv.1704.04861.
[18] 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,” arXiv:1912.12142,v. 1, 2019,DOI: https://doi.org/10.48550/arXiv.1912.12142.
[19] S. Anjum, I. Ahmed, M. Asif, H. Aljuaid, F. Alturise, Y. Y. Ghadi, R. Elhabob, et al., “Lung cancer classification in histopathology images using multiresolution efficient nets,” Computational Intelligence and Neuroscience, (1):1-12, 2023, DOI:https://doi.org/10.1155/2023/7282944
[20] M.Masud, N.Sikder, A. A. Nahid, A. K. Bairagi, and M. A. AlZain, “A machine learning approach to diagnosing lung and colon cancer using a deep learning-based classification framework,” Sensors, vol. 21, no. 3, p. 748, 2021, DOI: https://doi.org/10.3390/s21030748
[21] M.Ijaz, I.Ashraf ,U.Zahid, A.Yasin, S.Ali, M.A.Khan, S.A.Alqahtani,” : A Decision Support System for Lung Colon Cancer Classification using Fusion of Deep Neural Networks and Normal Distribution based Gray Wolf,” ACM Transactions on Asian and Low-Resource Language Information Processing, 2375-4699, 2023, DOI:https://doi.org/10.1145/3625096.
[22] A. Provath, K. Deb, and K. Hyun,” Classification of Lung and Colon Cancer Using Deep Learning Method,” in “Frontiers of Computer Vision”, G.Irie, C.Shin,T.Shibata, K. Nakamura, IW-FCV 2023. Communications in Computer and Information Science, vol 1857, pp 56–70, 2023. DOI: 10.1007/978-981-99-4914-4_5.
[23] S. Wadekar and D. K. Singh, “A modified convolutional neural network framework for categorizing lung cell histopathological image based on residual network, healthcare analytics, volume 4, 2023, 100224, 2023, DOI: 10.1016/j.health.2023.100224.

Journal of Basrah Researches (Sciences) **51**(1), 1 (2025)
DOI: https://doi.org/10.56714/bjrs.51.1.1

تصنيف سرطان الرئة والقولون باستخدام التعلم الجمعي المكدس لمجموعة من الشبكات العصبية

آمال عودة حسن ، زكريا أحمد عريبي*

قسم علوم الحاسبات,كلية التربية للعلوم الصرفة,جامعة البصرة,البصرة,العراق

معلومات البحث

الاستلام 25 كانون الأول 2024
المراجعة 31 كانون الأول 2024
القبول 7 كانون الثاني 2025
النشر 30 حزيران 2025

الكلمات المفتاحية

سرطان الرئة، سرطان القولون،تصنيف, Xception , MobileNet

Citation: A. O. Hasan, Z. A. Oraibi , J. Basrah Res. (Sci.) **51**(1), 1 (2025). DOI:https://doi.org/10.56714/bjrs.51.1.1

الملخص

أكثر نوعين من السرطانات انتشارًا وفتكًا على مستوى العالم هما سرطان الرئة والقولون. يعد الكشف المبكر والدقيق عن هذه السرطانات ضروريًا لتحسين النتائج المقدمة للمرضى. وقد وجد أن تقنيات التعلم الآلي، وخاصة نماذج التعلم العميق، لديها القدرة على توليد نتائج كشف وتصنيف دقيقة عند تدريبها على قواعد البيانات الكبيرة. تقدم هذه الورقة البحثية إطارًا جديدًا لتصنيف خمس فئات من سرطان الرئة والقولون باستخدام التعلم الجمعي لمجموعة من الشبكات العصبية التلافيفية المتعددة (CNN). حيث يتم أولاً، القيام بتدريب هذه النماذج ويتم الاحتفاظ بالتنبؤات الناتجة عن تطبيقها على مجموعة الصور الفرعية للاختبار، ثم يتم استخدام التصويت الجماعي لاختيار أفضل نتيجة من كل نموذج. تم تطبيق الطريقة المقترحة في هذه الورقة البحثية على مجموعة بيانات صورية كبيرة مكونة من ٢٥٠٠٠ صورة. بلغت دقة التصنيف التي تم تحقيقها ٩٨,٤٧٪ وهي أعلى من الدقة التي تم تحقيقها باستخدام النماذج المستقلة المطبقة على نفس مجموعة البيانات المستخدمة. يمكن تحسين إمكانات الطريقة المقترحة بشكل أكبر في المستقبل.

***Corresponding author email :** zakaria au@uobasrah.edu.iq

N: 1817-2695 (Print); 2411-524X (Online)
line at: https://jou.jobrs.edu.iq