Classification of lung cancer from histopathology Images using a Deep Ensemble Classifier

Onkar Singh
Machine Vision and Intelligence Lab
National Institute of Technology,
Jamshedpur, India
onkar122@gmail.com

Koushlendra Kumar Singh
Machine Vision and Intelligence Lab
National Institute of Technology,
Jamshedpur, India
koushlendra.cse@nitjsr.ac.in

Saikat Das
Department of Radiotherapy
All India Institute of Medical Sciences
Bhopal, India
saikat.radiotherapy@aiimsbhopal.edu.in

Akbar Sheikh Akbari
Leeds Beckett University, UK
a.sheikh-akbari@leedsbeckett.ac.uk

Nurulfajar Abd Manap
Universiti Teknikal Malaysia Melaka,
Malesia
nurulfajar@utem.edu.my

Abstract— Lung cancer continues to be the leading disease of patient death and disability all over the world. Many metabolic abnormalities and genetic illnesses, including cancer, can be fatal. Histological diagnosis one of the important part to determine form of malignancy. Thus, one of the most significant research challenges is explore the classification of lung cancer based on histopathology images. The proposed method encompasses the ensemble learning for classification of lung cancer and its subtype which employing pre-train deep learning models (EfficientNetB3, InceptionNetV2, ResNet50, and VGG16). The ensemble model has been created utilizing VotingClassifier in soft voting mode. The ensemble model is fit using the extracted features (features_train) and training labels (y_train). The LC25000 database's images of lung tissues are utilized to train and evaluate the ensemble classifiers. Our proposed method has an average F_I score of 99.33%, recall of 99.33%, precision of 99.33%, and accuracy of 99.00% for lung cancer detection. The findings of the analysis demonstrate that our proposed approach performs noticeably better compared to existing models. This technology is more suited to handle a wide range of classification challenges than using a single classifier alone and could improve the accuracy of predictions.

Keywords— lung cancer, deep learning, Histopathology Images, VGG16, Resnet50, EfficientNetB3, InceptionNetV2

I. INTRODUCTION

Lung cancer is a terrible condition that seriously threatens global health because it is responsible for a sizable portion of all deaths caused by cancer. It is characterized by the unchecked proliferation of aberrant cells in the lungs, which affects their function and can have potentially fatal results. Creating successful preventive, early diagnosis, and treatment plans require a thorough understanding of the complexity of lung cancer. The World Health Organisation (WHO) estimates that 2.2 million new instances of lung cancer diagnosed globally in 2020, making it among the deadliest cancers [1, 2, 3]. It accounts for roughly one in five cancer fatalities globally, making it the main cause of cancer-related mortality. These ominous figures demonstrate the urgent need for better lung cancer prevention and management knowledge. Tobacco usage is commonly acknowledged as the main factor contributing to lung cancer. The report by the International Agency for Research on Cancer (IARC), smoking is directly responsible for 80–90% of occurrences of lung cancer. Tobacco smoke's toxic substances and carcinogens considerably increase the chance of developing lung cancer. It is crucial to remember that exposure to secondhand smoke, workplace dangers, environmental toxins, and genetic factors can all cause the disease to harm non-smokers as well [4]. According to the disease's phase and category, indicators of lung cancer may alter. Common symptoms include a chronic cough, chest discomfort, breathlessness, hoarseness, exhaustion, and recurrent respiratory infections [5]. Lung cancer, sadly, frequently goes undiagnosed until it has spread to an advanced stage, at which point treatment choices are few and the outlook is dismal. It mostly falls under the categories of small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC) [6, 7]. Statistics show that between 80 and 85 percent of cases of lung cancer are caused by NSCLC. Cells make up the inner layer and surfaces of the cavities in the body, which account for 30% of these tutors. Adenocarcinomas are a subclass that persists on the outside of the lungs [7]. Lung cancer comprehension and treatment have advanced significantly as a result of developments in medical science and technology. Early identification rates have increased as a result of several diagnostic approaches, including genetic testing and imaging procedures. Lung cancer treatments include a multidisciplinary strategy that is individualized for each patient and includes surgery, radiotherapy, chemotherapy, targeted therapy, and immunotherapy. Researchers in a variety of fields, including online crime, artificial intelligence, computational biology, automation, and management, with a focus on medical diagnostics, are becoming more interested in deep learning-based approaches [8]. According to recent polls, one of the more time-consuming duties in the healthcare industry is the identification and classification of lung cancer illnesses. It is challenging to compile a repository containing healthcare images that could be categorized and quickly used for illness identification.

In this study, four deep learning models EfficientNetB3, InceptionNetV2, ResNet50, and VGG16 models are employed to classify lung cancer using an ensemble technique. Following is a breakdown of the remaining sections of the manuscript: The literature review is covered in Section 2. The suggested methodology is described in more detail in Section 3. Section 4 analyzed the investigation's findings, and Section 5 presents a conclusion.

II. LITERATURE SURVEY

The categorization of lung cancer is essential for maximizing treatment results, diagnosis, and therapy strategy. The accuracy of lung cancer classification based on imaging data has been a topic under investigation for many years, especially for huge datasets like the LC25000. The objective of this review of the literature is to provide a succinct description of the many classification schemes used to identify lung cancer. Researchers used DL and non-DL-

oriented approaches across almost all cancer treatment subcategories. We discuss the techniques provided for identifying lung cancer in this review. There are variations amongst these processes in regard to the type of images applied, the methods used to analyze those images, the characteristics that can be retrieved, and the machine learning model's methodology for detecting malignancy [9]. Considering their capability to autonomously acquire and retrieve pertinent information from medical pictures, convolutional neural network models have attracted a lot of curiosity in the categorization of lung cancer. In order to classify pulmonary nodules, Zhang et al. [10] built ensemble learners by combining many deep CNN learners. The LIDC-IDRI database is used to extract CT imaging patches of 743 nodules. To begin, 10-fold cross-validation was employed to train and assess eight deep CNN learners with various architectures. Yinghuan Shi et al. [11] classified four features using various varieties of Support Vector Machines (SVMs) from a collection of histopathological colon images. To identify various cancer types across various picture regions, researchers adopted a multi-class classification method as opposed to a single-class classification. Sirinukunwattana [12] suggested Spatially Constrained CNN (SC-CNN) technique categorizes four different nucleus types in colon tumors using histopathological images. A supervised machine-learning approach using infrared spectral pictures was proposed by Kuepper [13] as an automatic class-free diagnosis tool for colon cancer. Teramoto et al. [14] demonstrated deep convolutional neural network (DCNN) architecture's ability to identify lung cancer images and their propensity for avoiding overfitting were both enhanced by the use of various techniques for data augmentation. Shen et al. [15] presented multi-crop CNN for nodule malignancy categorization. Their strategy is distinct from other approaches because they did not apply segmentation or extracted-feature procedures to the computed tomography (CT) images they examined. The deep learning approach has been presented by Yuan et al. [16] for automatically detecting tumors in colonoscopy videos. They used AlexNet, an established Convolutional neural network and achieved a classification accuracy of 91.47%, to perform the classification. In order to differentiate between lung nodules and non-nodules, Moradi [17] explored different techniques. To reduce the likelihood of erroneous findings being assumed, they introduced the three-dimensional CNN Methods. They employed a logistic regression categorization, which produces a categorization outcome using information from four CNN approaches. They successfully created logistic regression by combining gradient improvement and decision tree classification. Bohdan Chapliuk [18] identified lung cancer using computed tomography (CT) and three-dimensional DenseNet. Two different neural network models have been constructed for identification and segmentation, and the methods have been tested employing three-dimensional images of the entire lungs. According to a study published by Toraman et al. [19] colon cancer may have been detected using Fourier Transform Infrared (FTIR) spectroscopy waves. The accuracy percentage for the classification of the waveforms' many statistical parameters, which the authors extracted and utilized SVM and ANN to classify, was 95.71%. A technique for tumor prediction determined by glowworm swarm optimization (GSO), which gathers data from multiple sources, was presented by Selvanambi et al. [20]. They were successful in achieving the greatest accuracy of 98% using the Recurrent Neural Network (RNN) learning method. Trishna Saikia et al. [21] have proposed a revolutionary approach that integrates the K-Nearest Neighbour (KNN) technique with four distinct advanced deep convolutional neural networks: VGG-16, VGG-19, Inception-V3, and Mobilenet-V2. The proposed innovative technique had a 98.3% accuracy rate. Mehta et al. [22] proposed an approach for ear recognition utilizing the ensemble method. They use three classification methods based on deep CNNs, and they average each classifier's estimations to predict with an accuracy of 96.83%.

Phankokkruad [23] employed three transfer learning techniques to detect lung cancer. The outcomes demonstrated that the suggested approaches for VGG16, ResNet50V2, and DenseNet201, respectively, received assessment scores of 62%, 90%, and 89%. The ensemble technique, which combines the three suggested CNN algorithms, performs better than the others with a 91% validation success rate. Goswami et al.[24] presented Pulmonary Lung Cancer Detection Using Deep Neural Networks, leading to the use of chest computed tomography (CT) and underwent training using transfer learning employing five different deep neural network techniques. The method was utilized for the classification of three distinct verities of lung cancer. The VGG frameworks connected using random forest and support vector machines proposed by Trishna Saikia et al. [25] for the classification of lung cancer also minimize the computing overhead for categorization. The hybrid approaches' extraordinary accuracy of 98.70% in classifying lung nodules is impressive. Hatuwal et al. [26] obtained training and validation accuracy in the categorization of lung cancer of 96.11% and 97.2%, respectively, employing a cross-entropy error function. Hatuwal used 15,000 samples from the LC25000 dataset in this investigation. Bukhari et al.'s [27] description of the recognition of histopathological pictures using 10000 images from the LC25000 imaging dataset and 193 images from Colorectal Adenocarcinoma Gland (CARD). ResNet50 (96.77%), ResNet30 (95.74%), and ResNet18 (94.79%) were found to have the highest levels of sensitivity, according to him. MIN LI et al. [28] used histopathological pictures to categorize the subgroups of the cancer. The multi-dimensional characteristics of 121 LC25k histopathology pictures were then retrieved, and the suitable features (Relief) methodology was utilized to select features. Support vector machines (SVMs) are used in the method for categorizing lung cancer variants, and it shows classification accuracy up to 83.91%. The histopathological pictures of colorectal and lung cancers were categorized by Naresh et al. [29]. The conventional hand-crafted feature extraction process involving transfer learning was carried out using CNN models that had been trained as feature extraction. The categorization process uses the dataset LC25000. With DenseNet121, the RF classification approach has an F1-score of 98.5%, 98.60% accuracy and recall, 98.63% precision, and the ability to recognize aberrant cells in the colon and lungs. Onkar et al. [36] proposed the mesh-free technique for the enhancement of lung CT images for the

identification of lung nodules. Masud et al.[31] provided a machine learning model for diagnosing lung and colon cancer based on deep learning approaches that looked at histological images of lung and colon tumors to differentiate within five different types of tissues. The histological images were categorized as carcinomas of squamous cells, adenocarcinomas, and lung-healthy by using a shallow neural network architecture that had been recommended by Mangal et al. [19]. The classification of adenocarcinomas and normal colons was performed on the 2500 instances of the LC25000 database using the same approach. Shandilya et al. [34] created a CAD approach to categorize histological pictures of pulmonary cells. They utilized a freely accessible library of histological pictures of lung cells for the creation and evaluation of CAD. Multi-scale analysis was applied to the extraction of image features. A comparative examination of lung tumor diagnosis using seven pre-trained CNN-based systems revealed that ResNet101 performed the best overall, with a score of 98.67%. This discovery will aid researchers in developing improved CNN-based lung cancer identification systems. To distinguish between lung benign, adenocarcinoma (ADC), and cancer of squamous cells (SCC) from histopathological pictures, Halder et al. [32] established the Morphology-based Attention Network (MorphAttnNet). The framework's development depends on morphological and convolutional techniques. To pick significant characteristics from pictures of histology, an attention-based approach is employed. Analysis of the presented framework's effectiveness utilizing the open-access LC25000 dataset revealed that it classified lung cancer subtypes with accuracy (98.96%).

.

In this study, deep learning pre-trained base models (EfficientNetB3, InceptionNetV2, ResNet50, and VGG16) were used in ensemble learning to classify lung cancer. The ensemble model is defined as a VotingClassifier with soft voting mode, whereas the deep learning models were first initialized as KerasClassifier objects. The extracted features (features_train) and training labels (y_train) are used to fit the ensemble model. The classifiers are trained and tested using images of lung tissues from the LC25000 database. The ensemble classifier methodology is used to integrate the testing dataset assessments from each classifier.

III. Methodology

This study proposed an ensemble classifier using four deep-learning models EfficientNetB3, InceptionNetV2, ResNet50, and VGG16. Pre-processing the data, extracting the features, training the ensemble models, and assessing the predictions are all steps in the ensemble learning process for classifying lung cancer into three groups Lung Adenocarcinoma, Lung Benign Tissue and Lung Squamous Cell Carcinoma. The data is first produced by gathering file locations and labels from the sets of images of lungs. Next, features from the training data are extracted using pre-trained models EfficientNetB3, InceptionNetV2, ResNet50, and VGG16. A feature array is produced by concatenating these features. The foundation models are then initialized, and an ensemble model is built using a voting classifier in soft voting mode. Utilizing the feature array and associated labels, the ensemble model is trained.

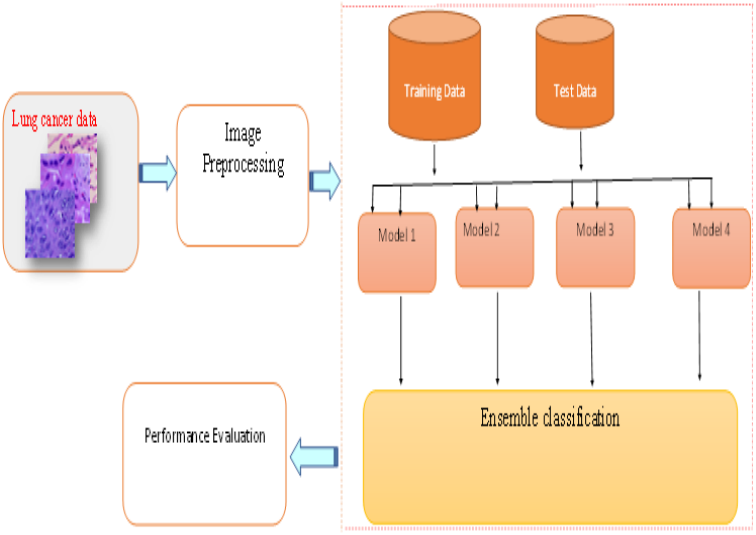


Figure 1 Proposed ensemble classifier model

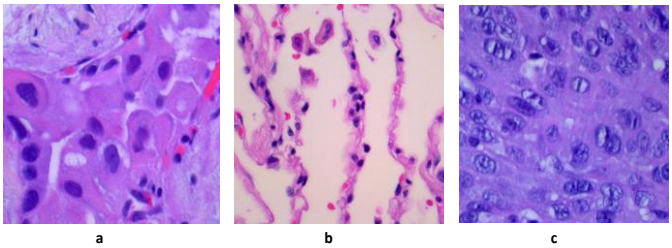
The pre-trained classifier is employed to extract characteristics from the test data for evaluation, and the ensemble model predicts the labels. Figure 1 demonstrated the proposed ensemble model for the classification of lung cancer.

A. Dataset

In this investigation, lung cancer histology pictures from the LC25000 database source were employed [35]. The LC25000 archive includes 25,000 color images of lung and colon cells together with five distinct kinds of tissue. The lung cancer data collection, which has 5000 pictures in three subclasses, is described in detail in Table 1. These classes are Lung Adenocarcinoma (Lung_aca), Benign Lung Tissue (Lung_n), and Lung Squamous Cell Carcinoma (Lung_scc) with class id 0, 1, and 2 respectively. The lung dataset, which consists of 15,000 pictures, has been separated into training data (80%) and validation and test data (20%). Using the LC25000 dataset Figure 2 displays the lung sample images for all three classes.

Table 1. Lung Cancer subcategory with Class ID from the LC 25000 Database

Cancer type	Class name	Class id	Number of samples
Lung Adenocarcinoma	Lung_aca	0	5000
Lung Benign Tissue	Lung_n	1	5000
Lung Squamous Cell Carcinoma	Lung_scc	2	5000



a b c

Figure 2. Lung sample images from the LC25000 dataset (a) lung adenocarcinoma (b) lung benign tissue and (c) lung squamous cell carcinoma are examples of cancerous tissues

B. Image Pre-processing

The pre-processing stage begins with scaling down the image to 224 by 224 pixels. In order to maintain consistency with the pixel intensity of the scaled photos, the original conversion of the images into bgr2rgb and subsequent conversion into a NumPy array. The next step is feature scaling, which involves applying the generalization technique to the image and dividing the simple image data in the array by 255 to reflect the image's greater intensity level. Finally, we complete labelling, in which we allocate the numbers 0, 1, and 2 to every type of lung image subclass.

C. Feature Extraction and Ensemble Classification

Ensemble learning (EL), often referred to as multiple-classifier learning, is a strategy for organizing and bringing together a sizable number of participants to successfully address a task [30]. In this stage process of feature extraction, utilizing pre-trained deep-learning classification architecture EfficientNetB3, InceptionNetV2, ResNet50, and VGG16 for extracting the deep feature and using it for the classification. Deep Learning pre-trained classification architecture is used in the feature extraction step by initializing them as KerasClassifier entities and combining these with the training data for learning the fundamental Keras models. The prediction method is called on every model using the training data to retrieve the features, which are then combined to form the features_train array. Regarding the ensemble learning scenario, the models that underlie the learning (EfficientNet, InceptionNet, ResNet50, and VGG16) are initialized as KerasClassifier objects, while the ensemble model is specified as a VotingClassifier using a soft voting method. The training labels (y_train) and the features that were obtained (features_train) are used to fit the ensemble model. The pre-trained models are utilized in the forecasting and assessment phase, and the test set characteristics are obtained utilizing a similar feature collection method. The features_test array is made by concatenating the retrieved features. Using an ensemble model along with extracted features' prediction approach, the trained ensemble model is utilized to forecast the labels for the test set. The efficiency of the ensemble model is evaluated using classification metrics such as accuracy, precision, recall, and F1-score. The classification outcome and confusion matrix have been displayed for the purpose of quantifying the model's efficacy. Table 2 depict the employed variable in pre-trained CNN models.

Table 2. Applied Variable

Variable of Models	**Value**
Dimension of image	224 x 224
Channels	3
Epochs	10
Batch size	40
Convolution layer activation	Relu
Dropout	45%
Dense layer activation	Softmax
Compiler optimizer	Adam
Compiler loss	Categorical Crossentropy
Learning rate	0.001

IV. EXPERIMENT RESULT AND DISCUSSION

The experiment was carried out utilizing the publicly available 15000 lung images from the LC25000 data collection. An ensemble classification model, based on four deep learning models, has been established for classifying lung cancer. The 15000 lung dataset is divided into 80 % (12000) for the training dataset and 20% (3000) for validation and testing using the train_test_split () method. To assess ensemble classifier performance, precision, recall, F1-score, accuracy, and ROC curve metrics are utilized. The experiment was performed on Microsoft Windows 10, Intel(R) Xeon(R) Silver 4208 CPU @ 2.10GHz 2.10 GHz with 32 GB of RAM. Anaconda Navigator is examined in a Jupyter Notebook. Python and a variety of libraries, including pandas, NumPy, Matplotlib, TensorFlow, Keras, Scikit-learn, and many more, were used to construct the proposed approach. A technique for assessing the potency of deep learning classifications is the confusion matrix. It is structured solely as a Table 3 and possesses four (TP, TN, FP, FN) permutations of expected and actual numbers. In Table 2, TP (True Positive) denotes a favorably anticipated value that was accurately projected, whereas TN (True Negative) denotes a negatively predicted value. False Positive (FN) indicates a positively anticipated value that was not intended, while False Negative (FN) displays a negatively anticipated value that was not intended. It is quite useful for figuring out the f1-score, ROC curve, accuracy, recall, and precision.

Table 3 Confusion Matrix

	Positive (1)	Negative (0)
Positive (1)	TP	FP
Negative (0)	FN	TN

The metrics of precision, recall, F1-score, and accuracy are employed to assess the classification outcomes which have been described as

$$\text{Pr}ecision = \frac{TP}{TP+FP} \tag{1}$$

$$\text{Re}call = \frac{TP}{TP+FN} \tag{2}$$

$$F1\_score = 2 \times \frac{(\text{Pr}ecision + recall)}{(\text{Pr}ecision + recall)} \tag{3}$$

$$\text{Re}call = \frac{TP+TN}{TP+FP+TN+FN} \tag{4}$$

The ratio of correctly anticipated positive values to all favorably predicted values serves as a measure of precision. The recall is defined as the ratio of positively anticipated values that were successfully predicted to all actual values.

The precision and recall scores for a classification problem are harmonically averaged to produce the f1-score. Accuracy, which favorably is inversely correlated with the total sample's number of correctly predicted occurrences, is the most important efficacy measure.

Table 4 displays the performance metrics for recall, accuracy, FI score, and precision. It demonstrated that the proposed approach obtained an average accuracy of 99.00%, FI score of 99.33%, recall of 99.33%, and precision of 99.33% for detecting lung cancer. It is demonstrated that the classification of the lung benign tissue (Lung_n) achieved 100% precision, recall, and FI_score. Lung adenocarcinoma (Lung_aca) and lung squamous cell carcinoma (Lung_scc) classifications achieved 99% recall, precision, and FI_score.

Table 4 Performance Matrices

Classes	Precision	Recall	FI_score	Accuracy
Lung_aca	0.99	0.99	0.99	99.00
Lung_n	1	1	1	
Lung_scc	0.99	0.99	0.99	

Figure 3 shows a confusion matrix that illustrates how well the ensemble classifier performed on the test dataset. The Receiver Operating Characteristic (ROC) curve and area under the curve (AUC) for the ensemble classification model are illustrated in Figure 4. To evaluate the model's effectiveness in terms of the true positive rate and false positive rate, the ROC curve is calculated and demonstrated. The ensemble classifier's AUC value of 1 demonstrates that the suggested approach has full discriminatory strength, or the capacity to precisely distinguish between positive and negative data. This is a really impressive outcome that demonstrates how well the ensemble classifier is doing. An AUC of 1 indicates the classifier's perfect classification performance. The genuine positive rate and the false positive rate are traded off, as displayed by the ROC curve.

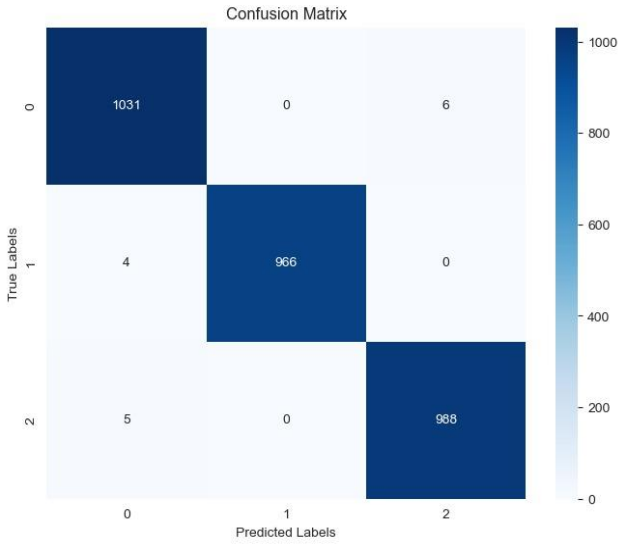


Figure 3 Ensemble classifier's confusion matrix for lung cancer

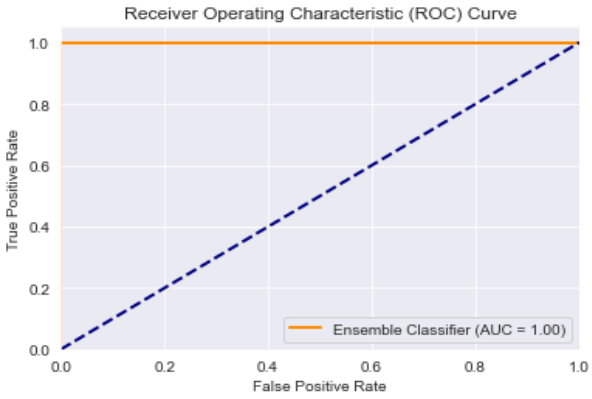


Figure 4 Ensemble classifier's ROC curve for lung cancer

Table 5 Analysis of comparisons with the proposed work and prior approaches

Authors	Cancer_type	Image type	Classifier	Accuracy	Precision	Recall	FI_Score
Masud[31]	Lung, Colon	Histopathological	CNN	96.33	96.39	96.37	96.38
Hatuwal [26]	Lung	Histopathological	CNN	97.2	97.33	97.33	97.33
Halder [32]	Lung	Histopathological	MorphAttnNet	98.96	98.33	99.12	98.72
Mangal [33]	Lung	Histopathological	CNN	97.89	–	–	–
Shandilya [34]	Lung	Histopathological	CAD, CNN	98.67	–	–	–
Naresh [29]	Lung, Colon	Histopathological	CNN	98.60	98.60	98.63	98.50
Proposed	Lung	Histopathological	Ensemble	99.00	99.33	99.33	99.33

In Table 5, the accuracy, precision, recall, and FI Scores of the proposed work are displayed alongside those of the authors' recommended strategy. It illustrates circumstances in which the most beneficial result from applying several classes has been reported. The - indicates that the relevant and significant measure is not covered in the article that relates to it. Compared to the existing approach shown in Table 5, it has been concluded that the proposed work performs more effectively.

V. Conclusion

The effectiveness of the investigation's center for the quick and precise detection of cancer will rise with the development of techniques that can identify certain qualities of a given malignant tumor by scrutinizing digital images. This would lower the possibility of human mistakes. In this study, three categories of lung cancer are classified using ensemble learning utilizing the deep learning models, EfficientNetB2, InceptionNetV2, ResNet50, and VGG16. The VotingClassifier with the soft voting manner is used for the development of the ensemble classifier. Our test results on the LC25000 dataset demonstrated that the proposed approach outperformed other convolutional neural network models in terms of accuracy (99.00%), precision (99.33%), recall (99.33%), and FI score (99.33%). In future studies, we will assess our approach using robust computational resources and use other contrasted learning-based techniques for obtaining deeper information using the unlabeled database. Furthermore, histopathology pictures

might be created utilizing generative techniques to view and investigate abnormalities across numerous ontologies.

Competing Interests: Authors declared that no funding has been received for this research work.

Conflict of interest: Authors have no conflict of interest.

REFERENCES

1. Ferlay, J., Ervik, M., Lam, F., Colombet, M., Mery, L., Piñeros, M., & Bray, F. (2021). Global cancer observatory: cancer today. Lyon: International Agency for Research on Cancer; 2020.
2. WHO Cancer https://www.who.int/news-room/fact sheets/detail/cancer (Last accessed June 2021).
3. World Health Organization (WHO). Cancer. Retrieved from https://www.who.int/news-room/fact-sheets/detail/cancer
4. American Cancer Society, Lung Cancer Causes (2019) [Online]. Available: https://www.cancer.org/cancer/lung-cancer/causes-risks-prevention/what-causes.html
5. Siegel, R. L., Miller, K. D., & Jemal, A. (2019). Cancer statistics, 2019. CA: a cancer journal for clinicians, 69(1), 7–34 https://doi.org/10.3322/caac.21451
6. Molina, J. R., Yang, P., Cassivi, S. D., Schild, S. E., & Adjei, A. A. (2008, May). Non-small cell lung cancer: epidemiology, risk factors, treatment, and survivorship. In *Mayo clinic proceedings* (Vol. 83, No. 5, pp. 584-594). Elsevier. https://doi.org/10.4065/83.5.584.
7. El-Regaily, S. A., Salem, M. A., Abdel Aziz, M. H., & Roushdy, M. I. (2018). Survey of computer aided detection systems for lung cancer in computed tomography. *Current Medical Imaging*, *14*(1), 3-18.
8. Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., ... & Farhan, L. (2021). Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *Journal of big Data*, *8*, 1-74.
9. Thawani, R.; McLane, M.; Beig, N.; Ghose, S.; Prasanna, P.; Velcheti, V.; Madabhushi, A. Radiomics and radiogenomics in lung cancer: A review for the clinician. Lung Cancer 2018, 115, 34–41.
10. Zhang, B., Qi, S., Monkam, P., Li, C., Yang, F., Yao, Y. D., & Qian, W. (2019). Ensemble learners of multiple deep CNNs for pulmonary nodules classification using CT images. *IEEE Access*, *7*, 110358-110371.
11. Shi, Y., Gao, Y., Yang, Y., Zhang, Y., & Wang, D. (2013). Multimodal sparse representation-based classification for lung needle biopsy images. *IEEE Transactions on Biomedical Engineering*, *60*(10), 2675-2685.
12. 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.
13. 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.
14. Teramoto, A., Tsukamoto, T., Kiriyama, Y. and Fujita, H., (2017) Automated classification of lung cancer types from cytological images using deep convolutional neural networks. BioMed research international, 2017, p.4067832.
15. Shen, W., Zhou, M., Yang, F., Yu, D., Dong, D., Yang, C., ... & Tian, J. (2017). Multi-crop convolutional neural networks for lung nodule malignancy suspiciousness classification. *Pattern Recognition*, *61*, 663-673.
16. Yuan, Z., Izadyyazdanabadi, M., Mokkapati, D., Panvalkar, R., Shin, J. Y., Tajbakhsh, N., ... & Liang, J. (2017, February). Automatic polyp detection in colonoscopy videos. In *Medical Imaging 2017: Image Processing* (Vol. 10133, pp. 718-727). SPIE
17. Moradi, P., & Jamzad, M. (2019, March). Detecting lung cancer lesions in CT images using 3D convolutional neural networks. In *2019 4th International Conference on Pattern Recognition and Image Analysis (IPRIA)* (pp. 114-118). IEEE.
18. Chapaliuk, B., & Zaychenko, Y. (2018, October). Deep learning approach in computer-aided detection system for lung cancer. In *2018 IEEE First International Conference on System Analysis & Intelligent Computing (SAIC)* (pp. 1-4). IEEE.
19. Toraman, S., Girgin, M., Üstündağ, B., & Türkoğlu, İ. (2019). Classification of the likelihood of colon cancer with machine learning techniques using FTIR signals obtained from plasma. *Turkish Journal of Electrical Engineering and Computer Sciences*, *27*(3), 1765-1779.
20. Selvanambi, R., Natarajan, J., Karuppiah, M., Islam, S. H., Hassan, M. M., & Fortino, G. (2020). Lung cancer prediction using higher-order recurrent neural network based on glowworm swarm optimization. *Neural Computing and Applications*, *32*, 4373-4386.
21. Saikia, T., Hansdah, M., Singh, K. K., & Bajpai, M. K. (2022, June). Classification of Lung Nodules based on Transfer Learning with K-Nearest Neighbor (KNN). In *2022 IEEE International Conference on Imaging Systems and Techniques (IST)* (pp. 1-6). IEEE.
22. Mehta, R., & Singh, K. K. (2022, December). Ear recognition system using averaging ensemble technique. In *International Conference on Machine Learning, Image Processing, Network Security and Data Sciences* (pp. 220-229). Cham: Springer Nature Switzerland.
23. Phankokkruad, M. (2021). Ensemble transfer learning for lung cancer detection. In 2021 4th international conference on data science and information technology (pp. 438–442).
24. Goswami, J., & Singh, K. K. (2023). Pulmonary Lung Cancer Classification Using Deep Neural Networks. In *Machine Vision and Augmented Intelligence: Select Proceedings of MAI 2022* (pp. 395-407). Singapore: Springer Nature Singapore.
25. Saikia, T., Kumar, R., Kumar, D., & Singh, K. K. (2022). An automatic lung nodule classification system based on hybrid transfer learning approach. *SN Computer Science*, *3*(4), 272.
26. Hatuwal BK, Thapa HC (2020) Lung cancer detection using convolutional neural network on histopathological images. Int J Comput Trends Technol 68:21–24
27. Bukhari, S. U. K., Syed, A., Bokhari, S. K. A., Hussain, S. S., Armaghan, S. U., & Shah, S. S. H. (2020). The histological diagnosis of colonic adenocarcinoma by applying partial self supervised learning. *MedRxiv*, 2020-08.
28. Li M, Ma X, Chen C, Yuan Y, Zhang S, Yan Z, Chen C, Chen F, Bai Y, Zhou P, et al. (2021) Research on the auxiliary classification and diagnosis of lung cancer subtypes based on histopathological images. IEEE Access 9:53687–53707.
29. Kumar N, Sharma M, Singh VP, Madan C, 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 75:103596
30. Zhou, Z. (2016). Machine learning. China Merchants, 3, 30–35 (in Chinese).
31. Masud, M., Sikder, N., Nahid, A. -A., Bairagi, A. K., & AlZain, M. A. (2021). A machine learning approach to diagnosing lung and colon cancer using a deep learning-based classification framework. Sensors, 21, 748.
32. Halder, A., & Dey, D. (2023). MorphAttnNet: An Attention-based morphology framework for lung cancer subtype classification. *Biomedical Signal Processing and Control*, *86*, 105149.
33. Mangal S, Chaurasia A, Khajanchi A (2020) Convolution neural networks for diagnosing colon and lung cancer histopathological images. arXiv preprint arXiv:200903878
34. Shandilya, S., & Nayak, S. R. (2022). Analysis of lung cancer by using deep neural network. In Innovation in electrical power engineering, communication, and computing technology (pp. 427–436). Springer
35. Borkowski, A. A., Bui, M. M., Thomas, L. B., Wilson, C. P., DeLand, L. A., & Mastorides, S. M. (2019). Lung and colon cancer histopathological image dataset (lc25000). *arXiv preprint arXiv:1912.12142*.
36. Singh, O., Kashyap, K. L., & Singh, K. K. (2023). Mesh-free technique for enhancement of the lung CT image. Biomedical Signal Processing and Control, 81, 104452.