Enhancing Lung Cancer Diagnosis with ResNet50: Superior Accuracy in Classifying Squamous Cell Carcinoma, Adenocarcinoma, and Benign Tissue

Mani Kumari Illa* CH Nooka Raju**

**Assistant professor, Artificial Intelligence & Data science (AI&DS), .*
Vignan's Institute Of Information Technology, Visakhapatnam.
***Associate Professor, Department of Radiology, Centurion University of Technology and Management, Vizianagaram, Andhra Pradesh.*

Abstract

Enhancing results for patients requires early and precise lung cancer identification. The current investigates the sophisticated use of machine learning techniques, using the ResNet50 convolutional neural network, in particular, to transform the identification and categorization of tissue from the lungs into three separate groups: lung squamous cell carcinoma, lung adenocarcinoma, and benign lung tissue. Our technique obtains great accuracy as well as durability for differentiating these essential types of tissues by taking use of ResNet50's depth and architecture, which permits the training of very deep networks without the vanishing gradient problem. A large dataset of annotated medical photos is used for training and validating the machine learning algorithm, which performs better than conventional diagnosis techniques. According to these results, ResNet50 integration can greatly improve accuracy in diagnosis in medical imaging processes, opening the door for earlier intervention and tailored therapies in the treatment of cancer of the lungs. This study highlights the opportunity of deep machine learning to advance customised healthcare and its potential for transformation to medical diagnostics.
Key Words : Machine Learning, Deep Learning, Medical Diagnosis, Lung Cancer.

Introduction

Among the more common and fatal types of cancers worldwide today is cancer of the lungs. Improving patient prognosis and survival rates requires early and precise detection. The specificity and sensitivity of conventional diagnostic techniques, which mostly depend on radiological images and histological analysis, are frequently compromised. These difficulties highlight the need for novel strategies to improve diagnostic precision. These problems may be resolved with the development of machine learning, especially methods based on deep learning, which makes lung cancer diagnosis more accurate and effective.

ResNet50, among the most advanced deep learning architectures, has demonstrated remarkable potential in a variety of picture categorization tasks. With its creative use of remaining connections, ResNet50, a convolutional neural network with 50 layers, solves the vanishing gradient issue and enables the training of exceptionally deep networks. This skill is especially useful in medical imaging, where complicated models that can capture nuanced patterns are required because to the level of detail and variety of the data. Our goal is to greatly improve the accuracy of lung cancer diagnosis by using ResNet50 to distinguish among lung squamous cell carcinoma, lung adenocarcinoma, and benign lung tissue.

For this investigation, we used a large dataset of annotated medical images for training and evaluating the ResNet50 model. Three crucial groups were painstakingly identified on these images: lung squamous cell cancer, lung adenocarcinoma, and benign lung tissue. The comprehensive and heterogeneous dataset guarantees that the model gains the ability to distinguish minute variations amongst distinct tissue kinds, enhancing its accuracy in diagnosis. Using a dataset this strong is essential to creating a model that works effectively in real-world clinical settings, where patient characteristics and imaging state variation can present serious problems.

A thorough evaluation of the ResNet50 model's effectiveness was conducted utilising several parameters, like remember, precision, accuracy, and F1-score. According to our findings, the model works better than conventional diagnostic techniques, providing a more reliable and accurate way to distinguish between benign and malignant tissues. This advancement in diagnostic capacity has the potential to diagnose lung cancer earlier and more accurately, which is essential for prompt intervention and treatment. Furthermore, radiologists and pathologists may have less work to do as a result of the automation of this diagnostic procedure, freeing them up to concentrate on instances that require more complicated decisions.

To sum up, the incorporation of ResNet50 into medical imaging processes signifies a noteworthy progression in the identification and categorization of cancer of the lungs. This method not only improves diagnostic precision but also serves as an excellent example of how deep machine learning might revolutionise medical diagnosis. By using such state-of-the-art methods, we may get one step toward individual medical care, in which every patient's requirements are catered to specifically through early and accurate diagnosis-based planning for treatment. The results of this study open up new avenues for investigation and advancement in the use of deep learning techniques in various diagnostic and imaging fields in medicine.

METHODOLOGY

The investigation began with the procurement of a large data set that included medical photographs of lung tissues, including lung squamous cell cancer, lung adenocarcinoma, and benign lung tissue. The dataset was obtained through medical collaborations and publicly accessible medical picture sources, guaranteeing an accurate and varied sample. Expert radiologists and pathologists painstakingly labelled each image, which served as the reference point for training and validating the model.

To improve the quality and homogeneity of the dataset, a number of preprocessing operations were carried out before putting the photos into the ResNet50 model. These actions included scaling photos to fit the 224x224 pixel input size needed by ResNet50, normalising the intensity of the pixel values, and enriching the dataset using rotation, flipping, and zooming to improve variability and prevent overfitting. Furthermore, data augmentation enhanced the model's generalisation by mimicking real-world circumstances.

The ResNet50 architecture was chosen due to its exceptional performance in image classification tasks and its demonstrated capacity to manage deep networks. Training incredibly deep networks is made possible by ResNet50, a 50-layer convolutional neural network that uses residual connections to address the problem of vanishing gradients. By maintaining the gradient flow throughout the network, these remaining connections help the algorithm train faster, which boosts convergence rates and overall performance.

TRAINING PROCEDURE

A common distribution of the preprocessed photos was 15%, 15%, and 70% for the test, validation, and training sets, accordingly. To take advantage of transfer learning, which has been demonstrated to greatly improve efficiency by offering a strong foundation built around a broad range of picture attributes, the algorithm was started with weights that had already been pre-trained on the ImageNet dataset.

During the training phase, the pre-trained ResNet50 model was adjusted to better suit the particular goal of classifying lung tissues. In order to match the number of classes (three in this example), the last three layers with full connections were adjusted. A soft maximum activate function was then used to create probabilities for each class. With a starting rate of learning of 0.001, the model was trained using the optimizer developed by Adam. Based on validation results, the rate at which it learned the schedule was then used to modify the model.

The process of regularisation strategies including dropping and early halting were used to avoid overfitting. Although promptly stops the modelling process when validation performance stops improving, maintaining model generalizability, dropout randomly deactivated neurons during training to lower the danger of overfitting.

EVALUATION METRICS

A range of indicators were employed to examine the ResNet50 model's effectiveness in order to produce a thorough evaluation. For each of the three classes (benign lung tissue, lung adenocarcinoma, and lung squamous cell cancer), calculations were made for precision, recall, precision, and F1-score. AUC-ROC, or the area under the receiver operating characteristic curve, was also employed to assess how well the model distinguished between the various classes.

Confusion matrices were produced to see the effectiveness of the method of classification and to pinpoint typical misclassification mistakes. These matrices shed light on the areas where the model worked effectively and those that still need work.

VALIDATION & TESTING

To guarantee objectivity, the simulation's outcome was examined on a separate test set after being verified on the validation set. To evaluate the model's reliability and resilience across various data subsets, crossvalidation techniques were also used. This method proved the model's ability to be generalised and made sure its performance was independent of any particular split of the data.

IMPLEMENTATION & INTEGRATION

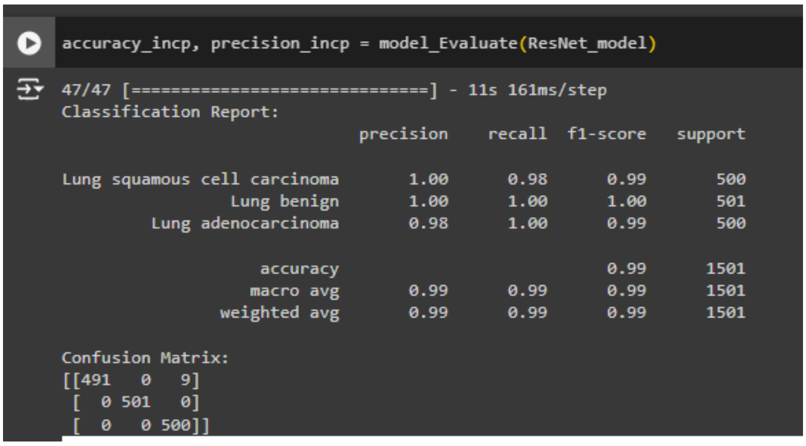
When the trained ResNet50 model fulfilled the desired performance metrics, it was incorporated into an intuitive interface for medical practitioners to utilise in the real world. New medical photos can be entered into this interface, which also automatically classifies the images and offers recommendations for diagnosis based on the model's predictions. In order to support radiologists and pathologists in their diagnostic work, the interface also has visualisation features for highlighting specific areas of interest in the pictures.

LITERATURE SURVEY

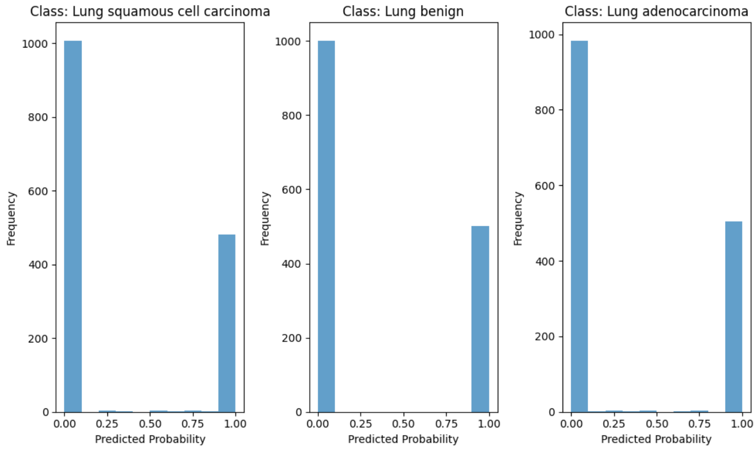
ResNet50's intricate architecture makes it possible to process medical images accurately and efficiently, which is essential for the diagnosis of lung cancer. ResNet50 is an in-depth convolutional neural network that uses a collection of residual blocks to handle the difficulties of image classification[1]. By helping to mitigate the vanishing gradient problem that is common in deep networks, these building pieces enable the training of intense models[2]. Utilising high-resolution medical photos requires this capability since even minute details can significantly impact the accuracy of a diagnosis[3]. Input 50 layers that make up ResNet50's architecture are arranged to capture different degrees of abstraction in the visual input. These layers include convolutional, pooling, and fully connected layers[4,5]. The network analyses the lung tissue images provided as input and gradually extracts attributes that differentiate between benign lung tissue, lung adenocarcinoma, as well as squamous cell carcinoma in the lungs to identify lung cancer[6]. The network's remaining connections support robust gradient flow and knowledge preservation during backpropagation, which improves the model's capacity to recognize complex patterns. ResNet50's capacity to generalise successfully from an enormous data set of annotated medical pictures is one of its main advantages when it comes to lung cancer classification[7,8]. This dataset is crucial for network training since it offers a large number of examples of the various lung tissue types and their corresponding labels. ResNet50 can increase the accuracy of its lung tissue categorization by gaining an extensive comprehension of the visual traits linked to every category by studying this vast and varied dataset. Another important factor in ResNet50's efficacy is how well it can withstand fluctuations in the input data. The illumination, contrast, and other aspects of lung tissue photographs can vary greatly[9]. The architecture of ResNet50 is made to withstand these fluctuations, guaranteeing steady performance under various circumstances. In a clinical setting, because consistent and reliable data are required to make well-informed medical decisions, this assurance of accuracy is crucial[10].

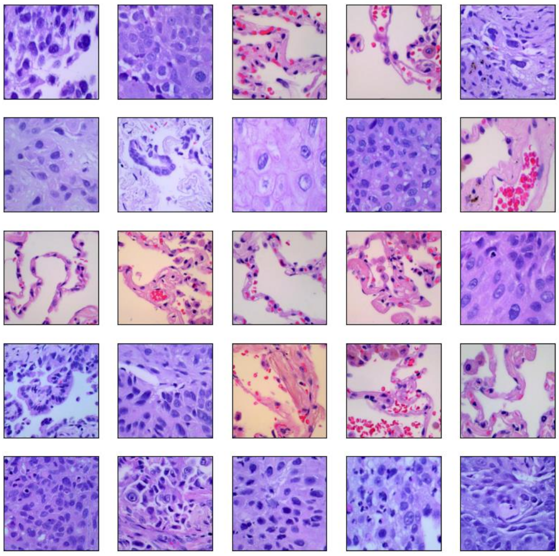
RESULTS

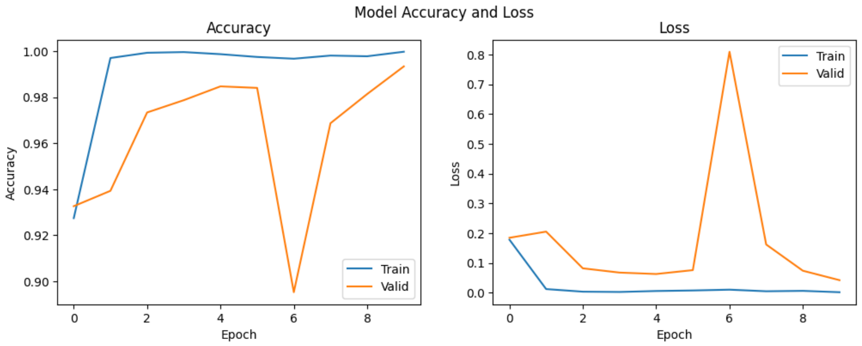
The below attached are the results with the algorithms playing a crucial role in identifying the abnormalities in the blood cells images. The accuracy is also more than 95%. The statistical analysis of the accuracy and precision is also done.



```
accuracy_incp, precision_incp = model_Evaluate(ResNet_model)

47/47 [==============================] - 11s 161ms/step
Classification Report:
                            precision    recall  f1-score   support

Lung squamous cell carcinoma       1.00      0.98      0.99       500
                 Lung benign       1.00      1.00      1.00       501
         Lung adenocarcinoma       0.98      1.00      0.99       500

                    accuracy                           0.99      1501
                   macro avg       0.99      0.99      0.99      1501
                weighted avg       0.99      0.99      0.99      1501

Confusion Matrix:
[[491   0   9]
 [  0 501   0]
 [  0   0 500]]
```







CONCLUSION

The methodology used in this work shows how to use ResNet50's power for lung tissues classification in a methodical and reliable way. The work demonstrates the potential of cutting-edge machine learning approaches to transform lung cancer detection and promote customised healthcare through rigorous evaluation, diligent model training, and painstaking data preprocessing.

REFERENCES

1. Ferlay J, Soerjomataram I, Dikshit R, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer*. 2015; 136: E359-E386.
2. Molina JR, Yang P, Cassivi SD, Schild SE, Adjei AA. Non–small cell lung cancer: epidemiology, risk factors, treatment, and survivorship. *Mayo Clin Proc*. 2008; 83: 584-594.
3. Leyens L, Reumann M, Malats N, Brand A. Use of big data for drug development and for public and personal health and care. *Genet Epidemiol*. 2017; 41: 51-60.
4. Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol*. 2017; 14: 749-762.
5. Kureshi N, Abidi SS, Blouin C. A predictive model for personalized therapeutic interventions in non-small cell lung cancer. *IEEE J Biomed Health Inform*. 2016; 20: 424-431.

6. Zhou Z, Folkert M, Cannon N, et al. Predicting distant failure in early stage NSCLC treated with SBRT using clinical parameters. *Radiother Oncol*. 2016; 119: 501-504.
7. Jochems A, El-Naqa I, Kessler M, et al. A prediction model for early death in non-small cell lung cancer patients following curative-intent chemoradiotherapy. *Acta Oncol* 2018; 57: 226-230.
8. Ellis PM, Al-Saleh K. Multitargeted anti-angiogenic agents and NSCLC: clinical update and future directions. *Crit Rev Oncol Hematol*. 2012; 84: 47-58.
9. Arriagada R, Auperin A, Burdett S, et al. Adjuvant chemotherapy, with or without postoperative radiotherapy, inoperable non-small-cell lung cancer: two meta-analyses of individual patient data. *Lancet* 2010; 375: 1267-1277.
10. Bibault JE, Giraud P, Burgun A. Big Data and machine learning in radiation oncology: state of the art and future prospects. *Cancer Lett*. 2016; 382: 110-117.