Classification of Lung Cancer using Pre-Trained Deep Learning Models

P.Deepa
Research Scholar
Department of CSE
Annamalai University
Chidambaram
deepu.prithiv@gmail.com

M.Arulselvi
Associate Professor
Department of CSE
Annamalai University
Chidambaram
marulcse.au@gmail.com

S.MeenakshiSundaram
Principal Cum Director
VPMM Engg. College
Krishnan Koil
Srivilliputhur
bosemeena@gmail.com

ABSTRACT

Many Computer Assisted Diagnosis (CAD) systems have been designed and used in recent past for diagnosingdifferenttypes of cancer. Identification of carcinoma at an earlier stage is more important, and it is made possible due to the use modern image processing and deep learning methods.The occurrence of Lung cancer is seen to be increased and Computed Tomography (CT) scan images were utilized in investigation to locate and classify lung cancer, also for determining the severity of those cancer. This study is aimed at employing pre-trained deep neural networks for classification of lung cancer images. A gaussian-based approach is used to segment CT scan images. In this research exploits a transfer learning-based classification method for the chest CT images acquired from Cancer Image Archive and available in the Kaggle platform. Pre-trained models VGG and RESNET were trained using segmented chest CT images, and their performance was evaluated using different optimization algorithms were analyzed.

Keywords: Computer Aided Diagnosis, lung cancer, deep learning, CT image, gaussian, transfer learning, pre-trained models, optimization algorithms.

1.INTRODUCTION

The association among both tobacco and lung cancer was originally proposed in 1912, and it continues to be a public health concern. In India, it is estimated that 10-15% of the population has been exposed to tobacco.Lung carcinoma is a disease that originates in the lung region and may spread to the other human organs. When the body's cells start to grow out of control, cancer develops. In human chest, two sponge-like organs termed as lungs were present. The right lung has three lobes and the lobes of the left lung are two. As the cardiac system occupies considerable more space, the left lung appears to be smaller. Air passing through the mouth or nose enter in to the lung region through trachea. The bronchi tubes continue the trachea and enter the lung region, which further divides as smaller bronchi. Smaller branches called bronchioles branches ends at the little air sacs termed as alveoli.

During inhaling phase of respiration the tiny air sacs supply oxygen in to the bloodflow and during exhale phase they remove carbon dioxide from the blood. The important activity performed by lungs is to ingest oxygen and expel carbon dioxide. The lung carcinoma starts from the cells that are present in the bronchi and other region of the lungs (bronchioles and alveoli). The pleura is a thin layer that protects the lungs during inhale and exhale phase by allowing the lungs to glide backward and forward against the ribcage. During breathing the diaphragm of the lungs swings up and down thus propelling the air in and out of the lungs.

2.LITERATURE SURVEY

Accurate assessment or detection of disorders or anomalies is a difficult task in medicine. As a result, machine learning approaches based on automated or semi-automatic computer-aided diagnostic (CAD) systems can help medical experts forecast or diagnose disease or anomalies, allowing them to make informed decisions or plan accurate treatment schedules.

With the advancement of technology, deep learning-based CAD systems work admirably, and the system's output assists specialists in making appropriate selections. Deep learning can be used to a wide range of research fields in medical health monitoring, according to a thorough review of the literature.Deep learning is used in this study to classify lung cancer because it is a popular and powerful machine intelligence and classification technology [9, 10]. If merely a massive quantity of data is inferred, deep learning, notably CNN, is considered to have high success rate. Although numerous deep learning-based lung cancer algorithms have been proposed [18 - 23], their accuracy can still be improved.In [24], they introduced a weakly supervised strategy for classifying whole-slide lung cancer images quickly and effectively. For discriminative block retrieval, their solution uses a patch-based fully convolutional network. Furthermore, to build a globally holistic WSI descriptor, context-aware extraction of features and aggregation algorithms were developed.The suggested method in [25] intends to investigate the efficiency of pre-trained deep models for malignancy classification of lung nodules. VGG16, VGG19, MobileNet, Xception, InceptionV3, ResNet50, Inception-ResNet-V2, DenseNet169, DenseNet201, NASNetMobile, and NASNetLarge were utilized as feature extraction methods to process the Lung Image Database Consortium and Image Database Resource Initiative (LIDC/IDRI). Following

- **RESNET50**

Deep neural networks requires a longer training time and likelihood of overfitting is high. To overcome such issues a residual learning approach was proposed to reduce the training timeof the model that are significantly deeper when compared to the conventional neural models. During the neural training the accuracy reaches a saturation point and tends to decrease

that, Naive Bayes, MultiLayer Perceptron (MLP), Support Vector Machine (SVM), Near Neighbors (KNN), and Random Forest (RF) classifiers were used for classifying the features extracted by the deep neural models.

3. PROPOSED METHOD

3.1 TRANSFER LEARNING

By transfer learning process a model which is trained to solve one problem with respect to the samples collected from that respective problem domain is utilized to solve a problem in another distinct domain using the knowledge gained from the earlier domain. It involves the trained model to solve the problem in a newer domain without having to retrain the model with samples from the newer domain. It is believed that the parameters learned from training the model in a domain can be reused to solve a problem in a newer or fresh domain. Currently numerous pre- trained models are available to solve image classification and other computer vision applications.VGG, GoogLeNet, and Residual Network are three of the more prominent models.As these pre-trained models are considered to be more efficient and consistent, they are used extensively in transfer learning tasks. VGG-16/ 19 which is known as the first exclusive deep neural network with numerous layers, residual network based models (Resnet- 50) – a model with symmetric- asymmetric building blocks are used in this study.

if model training is continued further. The "degradation problem" is the term for this. This demonstrates that not all neural network topologies are created equal.To overcome this problem, ResNetfollows residual mapping. Instead of assuming that stacking of layers will suffice to map a input to output, the residual network allows explicit mapping of residuals. The essential components of a Residual network is shown below in Fig. 3.
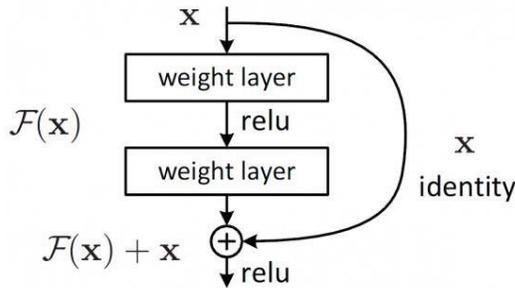
Fig.3 Residual Network: building block [27]

Feedforward neural networks having shortcut connections can be used to realize mathematical expression F(x)+x. ResNets can be used to solve a variety of problems. They're simple to optimize and achieve increasing accuracy as the network's depth grows, resulting in better results than previous networks.

- ▪ **VGG16**

VGG16 is a variant of CNN architecture (presented in Fig. 4)considered to be one of the efficientdeep neural models. An important and unique feature of VGG16 is the usage of convolutional layers with 3x3 size filters, and a maximum pooling using 2x2 size filter. The convolution and max pooling layers are arranged uniformly in the architecture. The classification layer consists of two fully connected layer configured with softmax activation. The network has in total 138 million trainable parameters.
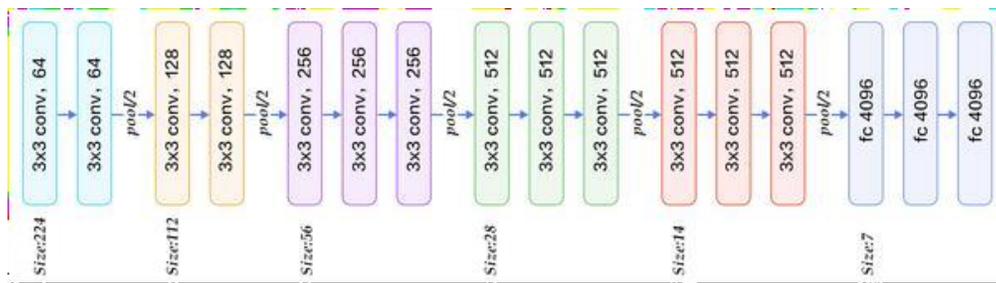
Fig. 4 Architecture of VGG-16 [28]

Gaussian Based Segmentation of CT Lung Image

Consider a pixel from a color image represented as $x_i = (x_i^R, x_i^G, x_i^B)$ and the segmentation of the color image with N number of pixels can be summarized as follows;

a) Choose the number of segments in the image represented as M and Initialize $\theta = (\pi_1, \pi_2, \ldots, \pi_M, \mu_1, \mu_2, \ldots\ldots\ldots, \mu_M, C_1, C_2, \ldots, C_M)$ where μ and C represents the mean and center of the cluster.

b) Let us assumed that color image is formed by a mixture of m single

Gaussian. The initial value of it can be fixed by $\pi^0 = \pi_1^0 = \cdots_2 = \pi^0 = \dfrac{1}{M}$

$$\overline{M} \qquad \text{Eq. 1}$$

$$\mu_l^{(t+1)} = \frac{1}{N\,\pi_l^{(t+1)}} \sum_{i=1}^{N} x_i \gamma(Z_{il})$$

$$C_l^{(t+1)} \qquad \text{Eq. 2}$$

$$= \frac{1}{N\,\pi_l^{(t+1)}} \sum_{i=1}^{N} (x_i - \mu_l^{(t+1)})^2 \gamma(Z_{il})$$

$$where\ l = 1, 2, \ldots, M$$

c) Using the iterative Expectation and Maximization algorithm the optimal value of θ was calculated and if the difference between the current and the previous iteration results are less than

10^{-6} then the iterative procedure is stopped.

d) Based on the optimal value of the θ obtained by the iterative process, each pixel x_i in the color image is clustered based on the following mathematical expression;

$I_l =$ Eq. 3

$$max\{E[\ln p(X, Z \mid \pi, \mu, C)] = \sum_{i=1}^{N}\sum_{l=1}^{M} \gamma(Z_{il})\{\ln \pi_l + \ln N(x_i|\mu_l, C_l)\}, l = 1,2, \ldots, M\}$$
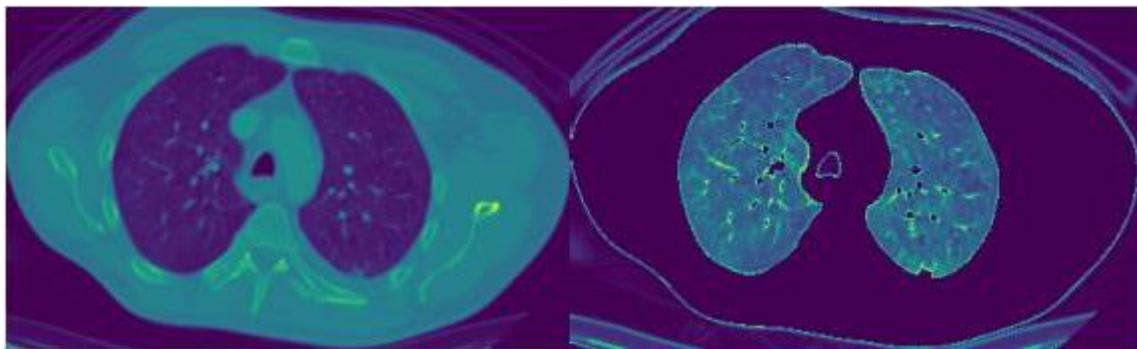

Fig. 5 Result of Gaussian Segmentation (left: CT image | right: segmented lung region)

As can be seen from the above results of segmentation shown in Fig. 5, the Gaussian mixture model responds to the distribution of color information available in the image to a degree. The Gaussian mixture model is a half- parametric density estimation method that combines the benefits of parameter estimation and non-parametric estimation while ignoring the probability density function's specific form. Furthermore, the model's complexity is primarily related to solution issues and has nothing to do with the volume of the training samples. To perform image segmentation, the Gaussian mixture model was used to describe a color image and the EM technique to estimate the various parameters of the Gaussian model

3.2 OPTIMIZER ALGORITHM

Gradient descent is a method of minimizing an objective function $J(\theta)$ that is parameterized by the parameters $\theta \in \mathbb{R}^d$ of a model by updating the parameters in the opposite direction as the gradient of the objective function $\nabla_\theta J(\theta)$ with respect to the parameters. The number of the steps taken to reach a (local) minimum is determined by the learning rate. The goal is to follow the slope of the surface formed by the objective function downhill until it reaches a global minimum or maximum. A low learning rate causes painfully slow convergence, whereas a high learning rate prevents convergence and causes the loss function to oscillate around the minimum or even diverge.Stochastic Gradient

Descent (SGD) has problems navigating ravines, which are widespread around local optima and are regions where the slope curves significantly more sharply in one direction than in another [29]. SGD oscillates throughout the ravine's slopes while only making halting progress along the bottom towards the local optimum in these cases. Momentum [30] is an approach that helps speed SGD in the desired direction while dampening oscillations. This is accomplished by adding a fraction γ of the previous time step's update vector to the current time step's update vector:

$$v_t = \gamma v_{t-1} + \eta \nabla_\theta J(\theta)$$

$$\theta = \theta - v_t$$

When the gradients of two dimensions point in the same direction, the momentum term increases, and when the gradients of two dimensions point in different directions, the momentum term decreases. As a result, there is faster convergence and less oscillation. Another technique for calculating adaptive learning rates for each parameter is Adaptive Moment Estimation (Adam) [31]. Adam, like Adadelta and RMSprop, stores exponentially decaying average of past squared gradients v_t and an exponentially decaying average of past gradients m_t. Adam behaves like a heavy ball with friction, preferring flat minima in the error surface [32], whereas momentum behaves like a ball rolling down a slope. As follows, we compute the decaying averages of past and past squared gradients m_t and v_t:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1)g_t$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2)g^2{}_t$$

The optimization methods Adam and Rmsprop are very similar, although there are a few important distinctions among RMSProp with momentum and Adam: Adam updates are directly estimated using a running average of the

first and second moments of the gradient, whereas RMSProp with momentum creates parameter changes using a momentum on the rescaled gradient. In the case of sparse gradients, RMSProp does not include a bias-correction term, which is necessary because not correcting the bias results to very huge stepsizes and often divergence. The gradient shifting hyper parameter ξ value has a considerable impact on ADAM and RMSProp's performance, and ADAM appears to perform exceptionally well (frequently superseding Nesterov accelerated gradient method) when its momentum parameter β_1 is extremely close to 1. The results presented the plots of Fig.6 and 7 shall present the performance comparison of ADAM and RMSprop algorithm. ADAM yields better performance in case of all pre-trained models considered in the experiments.

4. EXPERIMENTS AND RESULTS

For evaluating the performance of the proposed lung image feature extraction and classification method, a commonly available lung CT images from kaggle platform is used. The dataset comprises of 1200 images for training and testing the deep learning models. Initially the images are normalized and then segmented using Gaussian approach. Segmented images (lung region) from the CT images are used for training the models (VGG-16 and Resnet50). Pre-trained models were used as feature extractors, though they are trained using Imagenet dataset, and then using a fully connected neural layers the images are classified as either large cell, small cell, squamous, or normal lung image. The hyper parameters were tuned using a grid search process.

Table 1. Summary of Performance of Pre-Trained models with RMSprop optimization

Epoch	VGG-16		RESNET-50	
	Train	Test	Train	Test
1	0.331429	0.32197	0.333333	0.225379
10	0.466667	0.475379	0.544762	0.530303
20	0.518095	0.526515	0.548571	0.664773
30	0.535238	0.543561	0.605714	0.698864
40	0.552381	0.560606	0.647619	0.721591
50	0.565714	0.558712	0.674286	0.710227
60	0.571429	0.5625	0.638095	0.746212
70	0.567619	0.581439	0.649524	0.748106
80	0.575238	0.583333	0.660952	0.74053
90	0.586667	0.568182	0.685714	0.731061
100	0.584762	0.613636	0.691429	0.774621

The optimal number of epochs was determined based on average validation accuracy across 05 folds and the accuracy of the model obtained for different epochs are summarized in Table 1 and Table 2. With ADAM optimization algorithm the VGG-16 yield an accuracy of 96% after 100th epoch. Though VGG model has the ability to achieve better accuracy with more number of layers, it cannot converge to global minima when the number of layers is increased beyond

20. This is due to vanishing gradient problem and the learning rate becomes less that there are

no updates to the model weights. Using batch normalization the gradient explosion is controlled. The vanishing and gradient problem can be limited by using the residual learning network (RESNET-50). But from the tabulated results it can be seen that the RESNET yields poor performance since in the original implementation [27] of RESNET the stochastic gradient descent optimization technique was used. ADAM algorithm has issues in converging to the optimal solution and the SGD with a learning rate scheduler may outperform it.

Table 2. Summary of Performance of Pre-Trained models with ADAM optimization

Epoch	VGG-16		RESNET-50	
	Train	Test	Train	Test
1	0.398095	0.513258	0.506667	0.32197
10	0.693333	0.736742	0.649524	0.530303
20	0.796191	0.700758	0.702857	0.748106
30	0.813333	0.859848	0.685714	0.725379
40	0.88381	0.890152	0.72381	0.742424
50	0.899048	0.924242	0.76381	0.780303
60	0.92381	0.916667	0.72381	0.74053
70	0.92381	0.933712	0.752381	0.806818
80	0.912381	0.912879	0.729524	0.785985
90	0.933333	0.916667	0.775238	0.806818
100	0.96	0.960227	0.779048	0.776515

5.CONCLUSION

As a response to the growth of lung cancer, pathologist and the physicians needs an assistive mechanism such as computer aided diagnosis for an efficient diagnosis and treatment. This paper analyzes the scope and limitations of different optimization algorithms and their performance in combination with the pre-trained models. It is observed in case of VGG-16 models the use of ADAM optimization has helped to achieve better accuracy and whereas in case of RESNET model the ADAM or the RMSprop algorithm does not helped the model to converge to an optimal solution. Despite the suggested method's great performance, it could benefit from being tested with a larger number of images from various databases. We plan to add more images for training the model from scratch in the future to validate the model performance more accurately. These models may be stored in the cloud and used to provide quick diagnostics. This should considerably minimize clinician effort. Future works shall focus incollecting local CT scans of lung cancer cases and use them for evaluating the trained models.

REFERENCES

1. The American Cancer Society medical, 2019, https://www.cancer.org/cancer/lung-cancer/about/what-is.html

2. Horn L, Lovly CM (2018). "Chapter 74: Neoplasms of the lung". In Jameson JL, Fauci AS, Kasper DL, Hauser SL, Longo DL,

Loscalzo J (eds.). Harrison's Principles of Internal Medicine (20th ed.). McGraw-Hill.

3. Lu C, Onn A, Vaporciyan AA, et al. (2017). "Chapter 84: Cancer of the Lung". Holland-Frei Cancer Medicine (9th ed.). Wiley Blackwell.

4. Alberg AJ, Brock MV, Samet JM (2016). "Chapter 52: Epidemiology of lung cancer". Murray &Nadel's Textbook of Respiratory Medicine (6th ed.). Saunders Elsevier.

5. Ramada Rodilla JM, CalvoCerrada B, Serra Pujadas C, Delclos GL, Benavides FG (June 2021). "Fiber burden and asbestos-related diseases: an umbrella review". Gaceta Sanitaria. 36 (2): 173–183.

6. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM. et al. Corrigendum: Dermatologist-level classification of skin cancer with deep neural networks. Nature. 2017;546(7660):686.

7. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A. et al. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. JAMA. 2016;316(22):2402–10.

8. Yin X., Han J., Yang J., Yu P.. Efficient classification across multiple database relations: a crossmine approach. Ieee. T. Knowl. Data. E. 2016;18(6):770–783.

9. W. Sun, B. Zheng, and W. Qian, "Computer aided lung cancer diagnosis with deep learning algorithms,"in Medical imaging 2016: computer-aided diagnosis, 2016, vol. 9785: International Society for Optics andPhotonics, p. 97850Z.

10. O. Echaniz and M. Graña, "Ongoing work on deep learning for lung cancer prediction," in InternationalWork-Conference on the Interplay Between Natural and Artificial Computation, 2017: Springer, pp. 42-48.

11. Thun MJ, Hannan LM, Adams-Campbell LL, Boffetta P, Buring JE, Feskanich D, et al. (September 2008). "Lung cancer occurrence in never-smokers: an analysis of 13 cohorts and 22

cancer registry studies". PLOS Medicine. 5 (9): e185.

12. A. El-Baz and J. Suri, Lung Imaging and Computer Aided Diagnosis, Taylor & Francis, 2011.

13. "Tobacco Smoke and Involuntary Smoking". IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. WHO International Agency for Research on Cancer.
83. 2004.

14. "Lung Carcinoma: Tumors of the Lungs". Merck Manual Professional Edition, Online edition. July 2020. Retrieved 21 July 2021.

15. Collins LG, Haines C, Perkel R, Enck RE (January 2007). "Lung cancer: diagnosis and management". American Family Physician. 75 (1): 56–63.

16. Ost D (2015). "Chapter 110: Approach to the patient with pulmonary nodules". In Grippi MA, Elias JA, Fishman JA, Kotloff RM, Pack AI, Senior RM (eds.). Fishman's Pulmonary Diseases and Disorders (5th ed.). McGraw-Hill.
p. 1685.

17. Frank L, Quint LE (March 2012). "Chest CT incidentalomas: thyroid lesions, enlarged mediastinal lymph nodes, and lung nodules". Cancer Imaging. 12 (1): 41–48.

18. D. Riquelme and M. A. Akhloufi, "Deep learning for lung cancer nodules detection and classification in CT scans," AI, vol. 1, no. 1, pp. 28-67, 2020.

19. A. Asuntha and A. Srinivasan, "Deep learning for lung Cancer detection and classification," Multimedia Tools and Applications, vol. 79, no. 11, pp. 7731-7762, 2020.

20. J. H. Lee et al., "Performance of a Deep Learning Algorithm Compared with Radiologic Interpretation for Lung Cancer Detection on Chest Radiographs in a Health Screening Population," Radiology, vol. 297, no. 3, pp. 687-696, 2020. S. Takahashi et al., "Predicting Deep Learning Based Multi-Omics Parallel Integration Survival Subtypes in Lung Cancer Using Reverse Phase Protein Array Data," Biomolecules, vol. 10, no. 10, p. 1460, 2020.

21. A. Bhandary et al., "Deep-learning

framework to detect lung abnormality–A study with chest X-Ray and lung CT scan images," Pattern Recognition Letters, vol. 129, pp. 271- 278, 2020.

22. L. Cong, W. Feng, Z. Yao, X. Zhou, and W. Xiao, "Deep Learning Model as a New Trend in Computeraided Diagnosis of Tumor Pathology for Lung Cancer," Journal of Cancer, vol. 11, no. 12, p. 3615, 2020.

23. X. Wang, H. Chen, C. Gan, H. Lin, Q. Dou, Q. Huang, M. Cai, and P.-A. Heng, ''Weakly supervised learning for whole slide lung cancer image classification,'' in Proc. 1st Conf. Med. Imag. Deep Learn. (MIDL), 2018, pp. 1–10.

24. R. V. M. d. Nóbrega, S. A. Peixoto, S. P. P. da Silva and P. P. R. Filho, "Lung Nodule Classification via Deep Transfer Learning in CT Lung Images," *2018 IEEE 31st International Symposium on Computer-Based Medical Systems (CBMS)*, 2018, pp. 244-249.

25. ViharKurama,
https://blog.paperspace.com/popular-deep-learning-architectures-resnet-inceptionv3-squeezenet/

26. He, Kaiming, et al. "Deep residual learning for image recognition." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.

27. Abhipray Kumar Dash,
https://iq.opengenus.org/vgg16/

28. Sutton, R. S. (1986). Two problems with backpropagation and other steepest-descent learning procedures for networks. Proc. 8th Annual Conf. Cognitive Science Society.

29. Kingma, D. P., & Ba, J. L. (2015). Adam: a Method for Stochastic Optimization. International Conference on Learning Representations, 1–13.

30. Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., &Hochreiter, S. (2017). GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. In Advances in Neural Information Processing Systems 30 (NIPS 2017)