Performance Evaluation and Optimization of Pre-trained Deep Learning Models Using a Weighted Ensemble Approach for Lung Cancer Classification

Neha Raja Panwar[1*], Prashant Kumar Shrivastava[2]

[1]Department of Computer Science and Engineering, Sanjeev Agrawal Global Educational (SAGE) University, Bhopal, Madhya Pradesh, India
[2]School of Computer Technology, Sanjeev Agrawal Global Educational (SAGE) University, Bhopal, Madhya Pradesh, India
neharajap4@gmail.com

Abstract. Lung cancer remains to be among the leading contributor of deaths caused by cancer in the world, and accurate early diagnosis is a key to the better survival of the patient. The recent trends have observed the deep learning models as a promising tool in medical image analysis that can be used to accurately and automatically classify cancer. The paper presents the comparison of experimental study of three pre-trained convolutional neural network (CNN) models, namely VGG16, ResNet50 and DenseNet121, with the use of publicly available LC25000 lung and colon histopathological image dataset. The best architecture of lung cancer detection was evaluated based on the models in terms of accuracy, precision, recall, and loss measures. The experimental findings indicate that VGG16 was the best in this case because it achieved the best validation accuracy of 99.88 percent and the lowest loss rates. The validation accuracy of DenseNet121 and ResNet50 were 99.52 and 97.84 respectively, neither being quite as competitive as VGG16 but still quite competitive by comparison. Besides that, a weighted ensemble model was also constructed by integrating all the three networks using weights of 0.3 (VGG16), 0.2 (ResNet50), and 0.5 (DenseNet121). The accuracy of the ensemble as a whole was 96.88% which, though good, was not able to beat the best individual model. The results show that despite the ability of ensemble technique to increase robustness, individual models, in particular VGG16, may perform better in the classification of histopathological images. The results point to the promise of deep learning for accurate lung cancer detection and underscore the need for comparative model analysis and optimized ensemble approaches to promoting computer-aided diagnostic technology in medical practice.

Keywords: Lung Cancer Detection, VGG16, Histopathological Image, ResNet50, DensNet121, Ensemble Learning.

1 Introduction

Cancer is one of the most common and deadliest diseases to afflict the human body. It is identified by the uncontrolled growth and dissemination of cells into the body. If detected early, cancer can be treated quite successfully [1]. In normal biological processes, old or damaged cells are naturally replaced with new ones. But when this

S. Bhalerao et al. (eds.), *Proceedings of the 2nd International Conference on Recent Advancement and Modernization in Sustainable Intelligent Technologies & Applications (RAMSITA-2026)*, Advances in Intelligent Systems Research 207,
https://doi.org/10.2991/978-94-6239-678-4_6

process does not work properly, damaged cells are not replaced, and abnormal cells are formed that could spread (metastasize) to other organs [2].

Lung cancer is a type of cancer that develops in the lung tissue [5]. It has been recognized as the second most prevalent form of cancer [3-4]. From 2011 to 2015, statistics reported that in the USA, there were about 439.2 individuals per 100,000 diagnosed with cancer per year, while 163.5 per 100,000 lost their lives due to cancer [3]. In 2021 alone, an estimated 235,760 new cases of lung cancer and 131,880 deaths were expected in the USA [6]. Similarly, in the United Kingdom, approximately 44,500 people are diagnosed with lung cancer annually [7]. Globally, lung cancer caused 1.37 million deaths in the year 2008 [8] and 1.6 million in 2012 [9], representing an increase of approximately 17% over this time.

One of the biggest challenges posed by lung cancer is that in the initial stages, it presents no symptoms at all, and signs become noticeable only when the condition has reached a further state [5], [7]. In the USA alone, 17.4% of patients survive five years after diagnosis without any intervention, while in developing nations, survival rates are even smaller [3]. On the contrary, early detection greatly improves chances of recovery, makes treatment less complicated, lower in cost, and above all, raises survival rates. Early diagnosis and treatment can improve the 5-year survival rate from 15% to as much as 65–80% [6].

Different diagnostic approaches are adopted for lung cancer, viz., blood tests, imaging tests, endoscopic tests, and biopsies, each with their own strengths and weaknesses. Alternatively, Computed Tomography (CT) scanning is a rapid, painless, and highly informative technique that gives information regarding the size of the tumor, its arrangement and location [4]. The CT scans create a set of X-rays at various angles of the same anatomy location and make up three-dimensional pictures [10]. They are also imperative in identification of intrathoracic pathologies [11]. To obtain enhanced detection, specialists tend to use contrast-enhanced CT scan, whereby, a contrast agent is injected into the blood to provide a more clear view of the lung structures [7]. Chest imaging by way of such techniques is very useful in detecting lung cancer [12].

CAD systems have been designed to help radiologists identify cancer in medical images. CAD systems use image processing alongside machine learning algorithms to find abnormal regions in CT images that are later analyzed by medical specialists for definitive diagnosis [13].

Researchers have introduced many CAD systems over time to detect and classify lung cancer from CT images [14]. These systems employ various image feature extraction techniques and machine learning algorithms to enhance the accuracy of diagnosis. An important research problem is finding the best combination of image features and algorithms to achieve credible detection. Deep learning, especially convolutional neural networks (CNNs), have shown outstanding success in medical image analysis, with superior performance in disease classification, segmentation, and prediction. Pre-trained CNN models, trained on massive datasets such as ImageNet, can be properly modified for medical imaging applications using transfer learning. VGG16, ResNet50, and DenseNet121 have demonstrated high potential in extracting high-level hierarchical features and are very apt for histopathological image classification. Despite these developments, an indispensable research gap exists in the relative performance of different pre-trained models in lung cancer histopathological images and, in researching the ensemble learning methodologies to enhance the strength of predictions.

It is with this gap that this work fulfills by conducting an experimental comparative analysis of VGG16, ResNet50, and DenseNet121 on the publicly accessible LC25000 Lung and Colon Cancer Histopathological Image Dataset. The paper is an evaluation of model performance based on accuracy, precision, recall and loss and then weighted ensemble model is implemented which combines the strength of the three architectures. The ensemble method is explored using the weighted versions of VGG16, ResNet50, and DenseNet121 with the weights of 0.3, 0.2, and 0.5 correspondingly, in order to achieve higher classification results than single models.

The main contributions of the current study are the following:

1. Comparative investigations of three commonly used pre-trained CNNs (VGG16, ResNet50 and DenseNet121) in lung cancer detection using histopathological images.

2. Comparison of performance of the models in terms of accuracy, loss, precision and recall to find out the best performing architecture.

3. Implementing a weighted ensemble model to investigate the benefits of the combination of multiple deep learning models to enhance predictive accuracy.

Experimental scores confirm that VGG16 will always outperform the rest of the models, both in terms of validation accuracy and lower loss, and weighted ensemble competes but with a slightly worse performance. The findings can be used to continue supporting the use of deep learning in the analysis of medical imagery and support the importance of comparative model analysis in the creation of CAD systems that identify lung cancer.

2 Related Work

Deep learning (DL) architecture-based lung cancer detection and identification is a multi-stage process, the goal of each stage being specific. Image preprocessing is the first stage with the main objective of obtaining a better region of interest (ROI) in CT scan images of the lung. The most common are intensity normalization, which can be used to balance pixel values and image value truncation, which can be used to eliminate noisy and irrelevant data. The lung nodule segmentation follows the preprocessing step, and it is related to the process of outlining and refining the boundaries of the potential nodules. Segmentation should be done well because it directly influences subsequent analyses. It is then followed by feature extraction and selection. The features of salient properties to malignancy diagnosis are size, shape, texture, and intensity of segmented nodules. Using feature selection, dimensionality reduction, computational performance, and overall model performance are maximized later. The final step is classification in which the extracted and chosen features are used to classify the benign and malignant nodules. This step best suits DL architecture, which has the capability of learning high-end features. This part covers a wide range of advanced technology in terms of lung cancer detection.

Initially, the use of lung cancer was determined by machine learning (ML) models using handcrafted features and conventional classifiers. The Support Vector Machines (SVMs) and decision trees are some of the methods that were commonly used to classify the benign or malignant nodules. As an example, Zhu et al. [15] utilized an SVM on the radiomic feature derived by CT scan and achieved a 90.3 percent accuracy in identifying the malignancy. Similarly, Sluimer et al. [16] presented the conventional Computer-Aided Diagnosis (CAD) systems which depended on handcrafted features and image processing. Nonetheless, those approaches were limited by severe limitations. Features selected manually were not generalizable, they were likely to introduce redundancy, and they increased space complexity that affected consistency and efficiency negatively.

Deep learning-based CAD (DL-CAD) systems were developed, and this development became a shift in the analysis of medical images. Unlike traditional systems, DL-based models, and specifically Convolutional Neural Networks (CNNs), they can be trained on the raw data to learn discriminative features of the data. So nowadays, the DL models have become the dominant paradigm in the diagnosis of lung cancer. It has been trained in 2D and 3D DL and has a higher accuracy rate as compared to the traditional ML methods. In 2D techniques, such as, Roth et al. [17] suggested a multi-view CNN on the basis of 2D CT slices with a sensitivity of approximately 81%. Shah et al. [18] obtained an almost 95% accuracy in a similar architecture.

It was then followed by other developments with 3D CNN models that had the entire CT volume as input so that spatial and volumetric information would be captured. In a study by Anthimopoulos et al. [19], 3D CNN was used to detect lung nodules and it was said to be more precise and sensitive as opposed to the earlier models. Tyagi and Talbar [20] applied 3D CNN to detect and estimate malignancy of nodules. Their model achieves a Dice Similarity Coefficient (DSC) of 80.74% as compared to the LUNA16 dataset and a local dataset.

Other DL architecture has been investigated besides CNNs. Selvapandian et al. [21] proposed a lung nodule detection system based on the Generative Adversarial Network (GAN) with Sine Cosine SailFish (SCSF) algorithm. This method enhanced their detection accuracy with a 94.74 percent and 95.64 percent score at the first and second levels of classification, respectively, on the LIDC-IDRI dataset. Similarly, Chandrasekar et al. [22] presented Multivariate Ruzicka Regressed eXtreme Gradient Boosting Data Classification (MRRXGBDC) algorithm implemented with the Service-Oriented Architecture (SOA), which outperforms the rest of the models, LUNA16, by 9, 50, and 11 percentage points better accuracy, less false positives, and less time to make prediction, respectively. Gautam et al. [23] explored the ensemble deep transfer learning by using the weighted ensemble combination of ResNet152, DenseNet169, and EfficientNetB7. Their model based on the LIDC-IDRI data achieved an accuracy of 97.23% and a sensitivity of 98.60 since the best weight distributions resulted in current false negative being minimized. Rani et al. [24] suggested a four-step decision support system, where noise is eliminated by applying Hybrid CLAHE with Unsharp Masking, Radon Transform-based Single Seeded Region Growing is applied to segmentation, GLCM and GLRLM are used as feature extractors, and the classifier is AMPWSVM. Accuracy on this model was 93.3 and 90 percent on LIDC-IDRI and internal database respectively.

There are also other advanced and hybrid architectures which have been explored. Bushara et al. [25] suggested the LCD-CapsNet model that integrates CNNs and Capsule Networks and achieves 94 per cent accuracy and AUC = 0.989 on the LIDC-IDRI data. Navaneethakrishnan et al. [26] suggested a Deep CNN based on BDHOA that combined Bat Algorithm with the Deer Hunting Optimization and obtained the best scores of 92.43, 94.21 and 89.15 percent accuracy, sensitivity, and specificity, respectively, on LIDC-IDRI CT scans.

Problems still exist, particularly in segmenting small nodules, which is likely to be susceptible to low contrast and noise on the surrounding areas. To address this, Akila Agnes et al. [27] proposed Wavelet U-Net++ which integrates the portion of wavelet pooling into U-Net++ to enable it capture finer details. Their method had mean DSC of 0.936 and IoU of 0.878 on LIDC-IDRI. Venkatesan et al. [28] came up with a hybrid model Hybrid WPHT + DLBP to extract features, and Adaptive Harris-Hawk Optimization to get features and use it in OSVM classification with IW-BS parameter optimization. This scheme had very high performance on the LUNA16 dataset including 99.55 percent accuracy, 99.33 percent precision/recall, 99.66 percent specificity and an AUC of 99.5 per cent.

Overall, studies in lung cancer detection have developed since the early times of the ML-based CAD systems with hand-constructed features to the modern day of the DL architectures, which automatically extract features; and can utilize volumetric CT data. Although earlier ML algorithms provided a baseline, DL-based algorithms, especially CNNs, GANs, and hybrid ensembles have demonstrated significant gains in diagnostic accuracy, sensitivity and robustness on benchmark datasets.

3 Proposed Method

The suggested approach is based on two key stages, namely, (i) preprocessing of data including data augmentation and (ii) transfer learning using the already trained deep learning models such as ResNet50, VGG16, VGG16 and DenseNet121. The objective of the study is to classify the images of chest histopathology into five groups. The discussion of each of the phases is provided below.

3.1 Data Preprocessing and Augmentation

At this point, the problem of overfitting became addressed with the help of data augmentation procedures, as one of the main concerns of training deep pretrained models with relatively small datasets. Deep pretrained structures require an enormous number of samples and in case of smaller data, model generalization is impaired. Augmentation was used to produce more synthetic images to counter this problem. The operation enhances the diversity and strength of the dataset thereby increasing generalization especially in the analysis of histopathological images [30].

The process of augmentation involved three primary operations: resizing, flipping, and rotation. All images were resized to 224 × 224 × 3 pixels to conform to the input requirements of the pretrained networks. Random horizontal flipping was added for ensuring robustness to positional changes of histopathological structures. In addition, some images were augmented using a 15° rotation, creating new variations. More

importantly, augmentation was only used on the training dataset, not affecting validation and test sets to avoid any biased evaluation. These augmentation methods in aggregate improved the generalization ability of the suggested models.

3.2 Deep Neural Networks and Transfer Learning

The invention of Convolutional Neural Networks (CNNs) boosted the performance of image classification significantly. Deep neural networks, however, usually need large-scale datasets to provide maximum accuracy since their capability to automatically learn spatial and temporal features relies greatly on training data volume. In medical imaging, where datasets tend to be small, transfer learning is an effective alternative [31].

Transfer learning exploits knowledge acquired by models that were trained on huge benchmark datasets such as ImageNet [32]. In this approach pretrained models are adapted to new tasks and features representations are recycled. Two steps that are usually involved in the process are feature extraction and parameter tuning. The learned filters are then used by the pretrained model to extract useful features of the novel dataset in the process of feature extraction. Fine-tuning of the model to domain tasks is then done by parameter fine-tuning and architectural adjustments. This method reduces the cost of computations as well as the need to use large volumes of labelled data, and could be used in the classification of histopathological images.

This research used four pre-trained models, DenseNet121 [33], ResNet50 [34] and VGG16 [35]. They were first trained using the ImageNet dataset and thereafter fine-tuned using the chest histopathological images.

3.3 VGG16

VGG16 still is the most influential architecture in computer vision, as it is the one that won the 2014 ILSVRC competition [36]. In contrast to other models, which are based on the consideration of the impact of different hyperparameters, VGG16 is focused on the employment of the same-padding with max-pooling layers and the same-padding 3 x 3 convolutional filters exclusively. The architecture is the straightforward but rich sequential architecture as shown in Fig. 1; in this architecture the convolutional and pooling layers are arranged in a sequential fashion one on the top of the other, so as to acquire the higher level features sequentially. Many of the existing image classifier models have been developed based on VGG16 and it remains one of the most commonly used networks on non-ImageNet datasets.

3.4 ResNet50

ResNet50 ushers in the idea of residual learning that factors in performance reduction as we go about making networks so deep [37]. The conventional CNNs reduce their accuracy and efficiency depending on the vanishing gradient problem. ResNet can address this since it has skip connections (identity mappings) that can be learned to learn residual values, to be more closely aligned to its predictions to the actual results.

The design enables models to be trained in more detail without accuracy. The lower layers only identify the more sophisticated objects yet the structure of the plain structure

is chosen by the layers at the top. ResNet50 has a depth of 50 layers and has reached a compromise in performance and depth, and it does not degrade in performance as much as networks deeper than 30 layers do. Hence, it has been found to be an effective alternative to the more complicated image classification process, including histopathological image diagnosis.

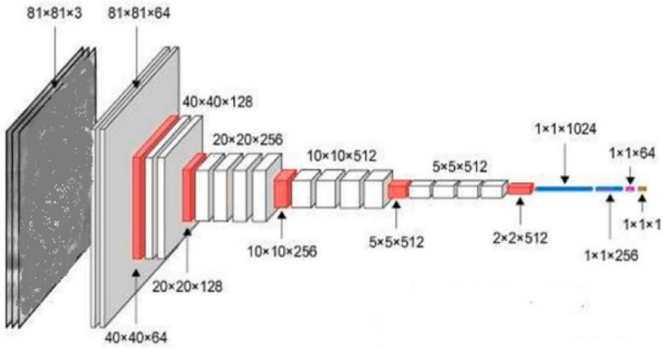


Fig. 1. VGG-19 architecture

3.5 DenseNet121

The denseNet architecture, suggested by Huang et al. [38], is premised on the idea of dense connectivity, in which each of the layers is connected directly with all the preceding layers [37]. Transition layers that are comprised of average pooling and convolution bridge dense blocks which are comprised of Convolution operations, ReLU activation and Batch Normalization. The high level of connectivity allows as much feature reuse as possible, which creates a small and computationally efficient model.

The denseNet121 has 121 layers, which consists of 116 convolutional layers, 4 dense blocks, four pooling layers, three transition layers and a final classification layer [39]. DenseNet makes a size of feature maps identical to boost gradient flow and feature propagation. The initial DenseNet121 architecture is represented in Fig. 2 [40].

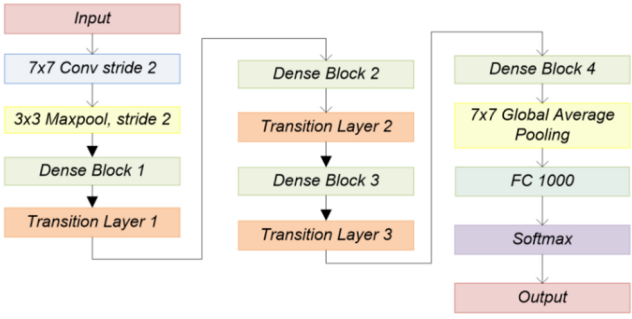


Fig. 2. Original Architectures of DenseNet-121

3.6 Ensemble Algorithms

Although the individual models such as VGG16, ResNet50 and DenseNet121 are good performing models, they may not be able to meet the demands of the critical medical uses in their original forms. This paper has applied ensemble learning techniques to enhance the degree of reliability and accuracy. An ensemble takes the forecasts of many models and hence variance is reduced and improved generalization is performed, particularly when small training sets are used.

Deep learning algorithms are nonlinear and highly changeable as it is vulnerable to training data. Probably, the prediction can be inconsistent when one model is applied. Ensemble methods can overcome this issue by averaging the output of more than one learner, which makes it more predictable and more exact. The max voting, average voting and weighted average are some of the most commonly used ensemble methods in the classification problems.

Voting Classifier Model

In max voting method, multiple models generate forecasts and the outcome is based on the most popularly voted category. The models are equally contributing and the most popular one by the votes is taken as the output.

Model Averaging Ensemble

The model averaging technique is used to compute an average of all the model predictions. The average that is taken of the predictions makes them less variable and less susceptible to generalization thus more robust than the one that a single model generated.

Weighted Average Ensemble

The weighted averaging approach gives results to models with distinct weights depending on their performance. This is unlike simple voting where, unlike simple voting, models with greater accuracy or reliability are pointed out as more influential in the ultimate predictions. This can be protracted to methods like stacking (stacked generalization) during which weights are optimized in a systematic fashion.

Voting, average and weighted averaging ensembles have been implemented using VGG16, ResNet50 and DenseNet121 in this study. To estimate ensemble performance and accuracy, ten samples were used. The outcomes of the experiment prove that ensemble methods are more successful in classification than independent models.

4 Result and Discussion

The three pretrained deep learning architectures VGG16, ResNet50, and DenseNet121 were comparatively evaluated on the LC25000 lung and colon histopathological image dataset as the current study. Training accuracy, validation accuracy, training loss and validation loss were used to evaluate the performance of the three models and then a weighted ensemble model combining the prediction of all the three architectures was developed.

4.1 Dataset

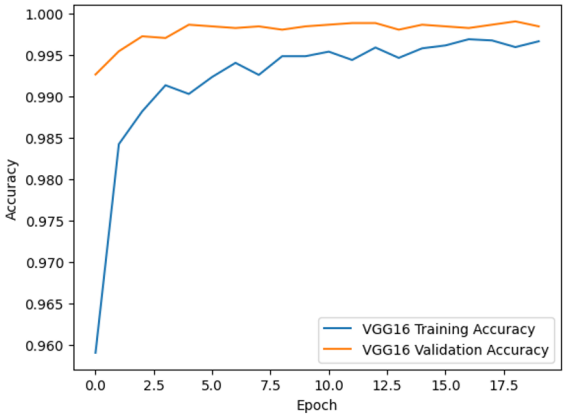
LC25000 database has 25,000 images of histopathology which are in five classes, and each image is of 768 x 768-pixels resolution and in JPEG format [30]. A database was created out of a set of original HIPAA-compliant and validated data comprising of 750 lung tissue images (250 benign, 250 adenocarcinomas, 250 squamous cell carcinoma), and 500 colon tissue images (250 benign, 250 adenocarcinomas). The Augmentor package was used to add these images to bring them to a total of 25,000 samples.

The five classes of the dataset, each with 5,000 images, are:

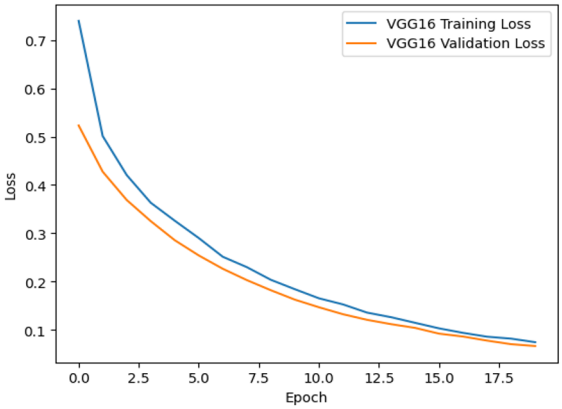
- Lung benign tissue
- Lung adenocarcinoma
- Lung squamous cell carcinoma
- Colon adenocarcinoma
- Colon benign tissue

4.2 VGG16 Performance

Among the individual models, the best performance was shown by VGG16, which had a training accuracy of 99.7% and validation accuracy of 99.88% at 20 epochs as given in Fig. 3. The training loss and validation loss were 0.0757 and 0.0660, respectively, which depict excellent generalization on unseen data. These findings indicate the efficacy of the model for histopathological image classification tasks.



(a) VGG16 Accuracy



(b) Loss graph

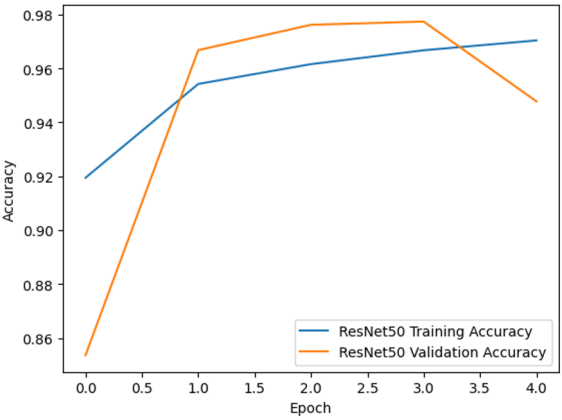
Fig. 3. (a) Accuracy and (b) Loss graph for VGG16 Model

4.3 ResNet50 Performance

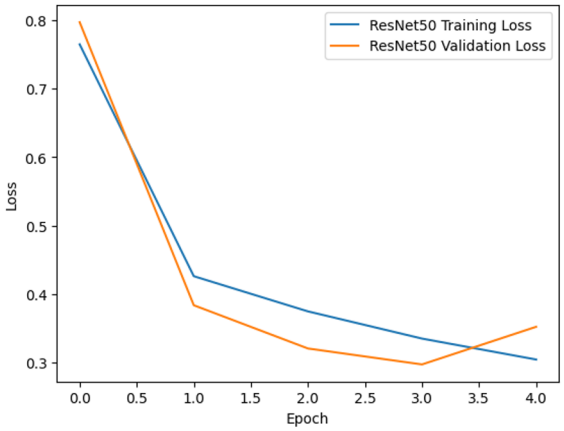
The ResNet50 model was able to obtain training accuracy of 97.05% and validation accuracy of 97.84% at just 5 epochs as shown in Fig. 4. Although a bit lower than VGG16, the validation loss (0.2580) was steady, indicating that ResNet50 can learn discriminative features very well even with a relatively small number of training epochs.

4.4 DenseNet121 Performance

With 97.46 and 99.52 training and validation accuracy respectively, denseNet121 showed good results as Fig. 5. The model was more likely to train and maintain lower validation loss rates (0.8435 and 0.7656, respectively) but achieve very close state-of-the-art accuracy, as feature reuse and effective gradient propagation is associated with DenseNet architecture.

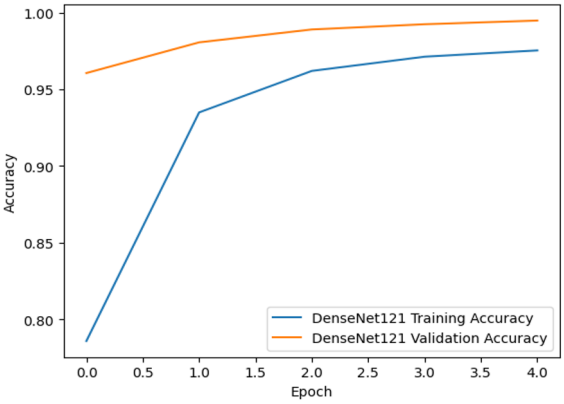


(a) Restnt50 accuracy

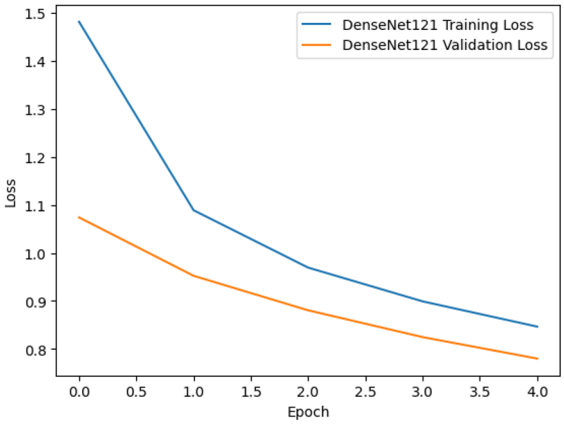


(b) Loss graph

Fig. 4. Accuracy and Loss graph for ResNet50 Model



(a) Accuracy



(b) Loss

Fig. 5. Accuracy and Loss graph for DenseeNet121 Model

4.5 Weighted Ensemble Model Performance

Given the power already possessed in the predictive model, the weighted ensemble model was created with the combination of VGG16, ResNet50 and DenseNet121 wherein 0.3, 0.2 and 0.5 were given as the weight of each model respectively. The precision of the ensemble playing was 96.88 percent and a loss of 0.2457. Although the ensemble made good predictions, the prediction of the best model (VGG16) did not decrease compared to the ensemble. This means that in as much as there is a tendency to add strength when ensembles are involved, performance of a single strong model may be affected adversely whereby it is combined with comparatively weaker models.

4.6 Comparative Analysis

The accuracy and loss plot of the training and validation depict the learning behavior of each model as shown in Table 1. VGG16 convergence was also stable and low overfitting, but the loss value of ResNet50 and DenseNet121 increased faster with fewer epochs, which is also an indicator of a noisy optimization. The ensemble approach equally played its part and emphasized the fact that there should be a wise allocation of weight in order to have full benefit of the complementary benefits of the respective models.

Table 1: Accuracy of all models

Model	Epochs	Training Accuracy	Training Loss	Validation Accuracy	Validation Loss
VGG16	20	0.9970	0.0757	0.9988	0.0660
ResNet50	5	0.9705	0.2895	0.9784	0.2580
DenseNet121	5	0.9746	0.8435	0.9952	0.7656
Weighted Ensemble	-	-	-	0.9688	0.2457

To conclude, it is evident that VGG16 is the strongest model that could be used to detect lung cancer on the LC25000 dataset as it is more accurate, with lower loss and higher generalization than ResNet50 and DenseNet121. Ensemble techniques enhance the resilience but the optimization of the model weights is crucial so that one is capable of making better model performance as compared to one having the best single network.

5 Conclusion

The present piece presented a comparative evaluation of three of the pretrained deep models, including VGG16, ResNet50, and DenseNet121, to detect lung cancer on LC25000 histopathological data. The experimental findings show that VGG16 outperformed all the remaining models with stable results, reporting a validation accuracy of 99.88% with minimal training loss and validation loss, and thus being the best model for this dataset. DenseNet121 was also very competitive, achieving 99.52% validation accuracy but with slightly greater loss values. ResNet50 also had slightly lower accuracy but underwent stable learning with fewer epochs. A weighted ensemble model of all three networks achieved 96.88% accuracy, but failed to break the highest individual model, suggesting that efficient ensemble techniques demand advanced weight optimization or sophisticated fusion methods to take full advantage of complementary model capabilities. Overall, the results show that deep learning, and VGG16 in particular, is promising for accurate and computerized lung cancer diagnosis from histopathological images. The study highlights the need for comparative model analysis and sheds light on the potential and pros and cons of using ensemble techniques in medical image analysis. Future research could consider: Fine-tuning weight strategy for ensembles, Adding more pretrained models, Verifying performance on large and varied datasets to further improve generalization in actual clinical environments.

References

1. **Wikipedia**: Cancer. Available at: https://en.wikipedia.org/wiki/Cancer. Accessed 30 Apr 2021.
2. Wadekar, S., Singh, D.K.: Enhanced residual network framework for robust classification of noisy lung cancer CT images. In: Proceedings of the International Conference on Advanced Network Technologies and Intelligent Computing (ANTIC 2023), CCIS, vol. 2093, pp. 21–35. Springer, Singapore (2023).
3. National Cancer Institute: Cancer statistics. Available at: https://www.cancer.gov/about-cancer/understanding/statistics. Accessed 30 Apr 2021.
4. American Cancer Society: CT scan for cancer. Available at: https://www.cancer.org/treatment/understanding-your-diagnosis/tests/ctscan-for-cancer.html. Accessed 30 Apr 2021.
5. Mayo Clinic: Lung cancer. Available at: https://www.mayoclinic.org/diseases-conditions/lung-cancer/symptoms-causes/syc-20374620. Accessed 30 Apr 2021.
6. John, J., Mini, M.G.: Multilevel thresholding based segmentation and feature extraction for pulmonary nodule detection. *Procedia Technology* **24**, 957–963 (2016).
7. NHS: Lung cancer. Available at: https://www.nhs.uk/conditions/lung-cancer. Accessed 30 Apr 2021.
8. Netto, S.M.B., Silva, A.C., Nunes, R.A., Gattass, M.: Automatic segmentation of lung nodules with growing neural gas and support vector machine. *Computers in Biology and Medicine* **42**(11), 1110–1121 (2012).
9. Wild, C.P., Stewart, B.W. (eds.): *World Cancer Report 2014*. World Health Organization, Geneva (2014).
10. UC Davis Medical Center: What is a CT scan? Available at: https://www.ucdmc.ucdavis.edu/radiology/UCDHS_CT_FAQ_v1.pdf. Accessed 30 Apr 2021.

11. Wadekar, S., Singh, D.K.: A modified convolutional neural network framework for categorizing lung cell histopathological images based on residual network. *Healthcare Analytics* **4**, 100224 (2023).
12. Swensen, S.J., Jett, J.R., Sloan, J.A., Midthun, D.E., Hartman, T.E., Sykes, A.M., Aughenbaugh, G.L., Zink, F.E., Hillman, S.L., Noetzel, G.R., Marks, R.S.: Screening for lung cancer with low-dose spiral computed tomography. *American Journal of Respiratory and Critical Care Medicine* **165**(4), 508–513 (2002).
13. Digital Pathology Association: Computer-aided diagnosis: the tipping point for digital pathology. Available at: https://digitalpathologyassociation.org/blog/computer-aided-diagnosis-the-tipping-point-for-digital-pathology. Accessed 30 Apr 2021.
14. Abdullah, D.M., Ahmed, N.S.: A review of most recent lung cancer detection techniques using machine learning. *International Journal of Science and Business* **5**(3), 159–173 (2012).
15. Zhu, Y., Tan, Y., Hua, Y., Wang, M., Zhang, G., Zhang, J.: Feature selection and performance evaluation of SVM-based classifier for differentiating benign and malignant pulmonary nodules by CT. *Journal of Digital Imaging* **23**, 51–65 (2010).
16. Sluimer, I., Schilham, A., Prokop, M., van Ginneken, B.: Computer analysis of computed tomography scans of the lung: A survey. *IEEE Transactions on Medical Imaging* **25**(4), 385–405 (2006).
17. Roth, H.R., Wang, Y., Yao, J., Lu, L., Burns, J.E., Summers, R.M.: Deep convolutional networks for automated detection of posterior-element fractures on spine CT. In: *Medical Imaging 2016: Computer-Aided Diagnosis*, SPIE, vol. 9785, pp. 165–171 (2016).
18. Shah, A.A., Malik, H.A.M., Muhammad, A., Alourani, A., Butt, Z.A.: Deep learning ensemble 2D CNN approach towards the detection of lung cancer. *Scientific Reports* **13**(1), 2987 (2023).
19. Anthimopoulos, M., Christodoulidis, S., Ebner, L., Christe, A., Mougiakakou, S.: Lung pattern classification for interstitial lung diseases using a deep convolutional neural network. *IEEE Transactions on Medical Imaging* **35**(5), 1207–1216 (2016).
20. Tyagi, S., Talbar, S.N.: CSE-GAN: A 3D conditional generative adversarial network with concurrent squeeze-and-excitation blocks for lung nodule segmentation. *Computers in Biology and Medicine* **147**, 105781 (2022).
21. Selvapandian, A., Prabhu, S.N., Sivakumar, P., Rao, D.B.J.: Lung cancer detection and severity level classification using sine cosine sailfish optimization based GAN with CT images. *The Computer Journal* **65**(6), 1611–1630 (2021).
22. Chandrasekar, T., Raju, S.K., Ramachandran, M., Patan, R., Gandomi, A.H.: Lung cancer disease detection using service-oriented architectures and multivariate boosting classifier. *Applied Soft Computing* **122**, 108820 (2022).
23. Gautam, N., Basu, A., Sarkar, R.: Lung cancer detection from thoracic CT scans using an ensemble of deep learning models. *Neural Computing and Applications* **36**(5), 2459–2477 (2023).
24. Rani, K.V., Sumathy, G., Shoba, L.K., Shermila, P.J., Prince, M.E.: Radon transform-based improved single seeded region growing segmentation for lung cancer detection using AMPWSVM classification. *Signal, Image and Video Processing* **17**(8), 4571–4580 (2023).
25. Bushara, A.R., Kumar, R.S.V., Kumar, S.S.: LCD-capsule network for the detection and classification of lung cancer on CT images. *Multimedia Tools and Applications* **82**(24), 37573–37592 (2023).
26. Navaneethakrishnan, M., Anand, M.V., Vasavi, G., Rani, V.V.: Deep fuzzy SegNet-based lung nodule segmentation and optimized deep learning for lung cancer detection. *Pattern Analysis and Applications* **26**(3), 1143–1159 (2023).
27. Agnes, S.A., Solomon, A.A., Karthick, K.: Wavelet U-Net++ for accurate lung nodule segmentation in CT scans. *Biomedical Signal Processing and Control* **87**, 105509 (2024).

28. Venkatesan, N., Pasupathy, S., Gobinathan, B.: Efficient lung cancer detection using optimal SVM and improved weight-based beetle swarm optimization. *Biomedical Signal Processing and Control* **88**, 105373 (2024).
29. Gupta, H., Singh, H., Kumar, A.: Texture and radiomics inspired data-driven cancerous lung nodules severity classification. *Biomedical Signal Processing and Control* **88**, 105543 (2024).
30. Borkowski, A.A., Bui, M.M., Thomas, L.B., Wilson, C.P., DeLand, L.A., Mastorides, S.M.: Lung and colon cancer histopathological image dataset (LC25000). *arXiv preprint arXiv:1912.12142* (2019).
31. Pan, S.J., Yang, Q.: A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering* **22**, 1345–1359 (2009).
32. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., et al.: ImageNet large scale visual recognition challenge. *International Journal of Computer Vision* **115**, 211–252 (2015).
33. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017)*, pp. 2261–2269 (2017).
34. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2016)*, pp. 770–778 (2016).
35. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In: *International Conference on Learning Representations (ICLR 2015)*, pp. 1–14 (2015).
36. Abas, M.A.H., Ismail, N., Yassin, A.I.M., Taib, M.N.: VGG16 for plant image classification with transfer learning and data augmentation. *International Journal of Engineering and Technology* **7**(4), 90–94 (2018).
37. Jegou, S., Drozdzal, M., Vazquez, D., Romero, A., Bengio, Y.: The one hundred layers Tiramisu: Fully convolutional DenseNets for semantic segmentation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 1175–1183 (2017).
38. Taylor, A.G., Mielke, C., Mongan, J.: Automated detection of moderate and large pneumothorax on frontal chest X-rays using deep convolutional neural networks. *PLOS Medicine* **15**(11), 1–15 (2018).
39. Remya, R., Rajini, N.H.: Efficient densely connected convolutional networks-based detection and classification of breast cancer using mammogram images. *Journal of Intelligent and Fuzzy Systems* **29**(4), 10663–10678 (2020).
40. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017)*, pp. 2261–2269 (2017).