Comparison of CNN models in Non-small Lung Cancer Diagnosis

Haoxiang Xu
The University of British Columbia
BC, Canada
xhx1998@student.ubc.ca

***Abstract*—Since the traditional manual method of classifying lung cancer CT photos is very time-consuming, a new automated, highly accurate classification method is urgently needed. This has led to the use of the Convolutional Neural Network (CNN) for image classification and recognition as the best choice in the medical field. However, due to the problem of vanishing gradient when the layer depth of these models is too deep, the gradient will be vanishingly small and this leads to extremely low learning efficiency, which may even be close to 0. In this paper, a reimplemented CNN model based on ResNet is used to improve low learning efficiency. Three different CNNs are also compared and illustrate that the deeper neural networks have better learning efficiency and higher accuracy for classifying lung CT images. Specifically, LeNet, AlexNet, and ResNet are reimplemented, where the first two CNNs represent traditional models. By comparing the accuracy of the three models, we conclude that the traditional model is sufficient for the task of classifying lung cancer models and has an acceptable accuracy rate when the number of training sessions reaches a certain level, but it is not competitive in learning efficiency and accuracy for the same number of training sessions compared to the subsequent models with more neurons.**

Keywords—Non-small cell lung cancer; Convolutional Neural Network; Medical Diagnosis; ResNet

I. INTRODUCTION

According to statistics, 20,900 Canadians died from lung cancer in 2015, accounting for 20,600 new cases of the disease [1-2]. This represents 7.9 percent of all fatalities in Canada that year. In Canada, lung cancer is the most prevalent cancer type and the leading cause of cancer-related fatalities. The extremely high mortality rate of lung cancer and the enormous number of patients make it urgent to develop a system that can improve the effectiveness and efficiency of treatment.

In conventional treatment, the use of CT screening to diagnose the type of lung cancer is time-consuming and inefficient, which has led to the development of methods for automatic classification of digital chest radiographs and computed tomography (CT) scans as an active area of research from around 2000 onwards [3-7]. To date, Convolutional Neural Network (CNN) - an automatic classification system is widely used in the medical field and has achieved remarkable results in the detection of skin lesions, tumours, colon cancer, etc. [8-11] Although CNNs have now become the most dominant image classification method in the medical field, being applied to lung cancer identification are still traditional CNN models including AlexNet [12], LeNet [13], etc. Eri Matsuyama et al [14] and Mesut Toğaçar et al [15] achieved considerable accuracy in identifying lung cancer in 2018 and 2019 respectively using a modified version of AlexNet and a hybrid structure of the LeNet model.

Traditional models are simple in structure and have fewer neurons. But at the same time, their disadvantages are also obvious: their learning efficiency in training is not only slower than later models. For example, ResNet, but also due to the problems of vanishing gradient and overfitting, which can occur in CNN models might lead to a sudden drop in accuracy during training. In contrast, ResNet [16] is easier to optimize neural networks with high depth and can obtain higher accuracy from more hidden layers, meaning that ResNet can better establish a positive correlation between neural network depth and accuracy. In 2021, the team of Guangyu et al. modified the structure based on ResNet to use this model for the detection of COVID-19 base on chest CT images. In their study, they compared other mainstream diagnostic COVID-19 models including VGG16 [17], DensNet121 [18], SqueezeNet1.0 [19], etc., and concluded that the modified ResNet outperformed other models in terms of accuracy, sensitivity, precision, robustness, and class activation map [20]. It has been shown that ResNet not only addresses all the major shortcomings of traditional CNNs but has been used in similar cases with promising results.

Therefore, in this paper, a reimplemented CNN model based on ResNet is used to improve low learning efficiency and sudden drop in accuracy that usually happened in traditional CNN models. LeNet and AlexNet are also reimplemented for the purpose of comparing traditional and later CNN architectures. The structure of our ResNet contains 5 sets of alternating convolution and identity blocks, with a total of 53 layers, including 48 solution layers along with 1 MaxPool, 1 Average Pool layer, 1 flatten layer, 1 fully connected layer and 1 output layer. This makes the number of layers embraced by ResNet nearly four times that of AlexNet and seven times that of LeNet. By the above experimental setup, ResNet achieved an accuracy of 98.86% and a loss rate below 0.1 after 32 epochs of training, and the rising trend was stable. In contrast, AlexNet and LeNet's accuracy rose slowly and fluctuated, reaching maximum accuracy of only 53.83% and 31.32%, respectively, and their loss rates were always higher than 1.0. It was not until the number of training epochs reached about 256 that the accuracy of AlexNet and LeNet reached an acceptable level, with 95.56% and 92.42% respectively.

The experimental results demonstrate that the conventional model is adequate for classifying lung cancer models and has an acceptable accuracy rate when the number of training sessions reaches a certain level, but it is not competitive at all in terms of learning efficiency and accuracy for the same number of training sessions compared to the subsequent model with more neurons. Our analysis illustrates that for a task like classifying lung CT images, neural networks with deeper layer depths possess superior learning efficiency and higher accuracy rates. To compare traditional and later models in terms of layer depth, ResNet ends up being one of the best options since it enables the training of deeper neural networks by establishing jumping linkages between layers and preventing the issue of disappearing gradients. This allows the layer depth of ResNet to reach several or even tens of times the traditional CNN.

II. METHOD

This section meticulously introduces the three CNN models used in this study and illustrates the structural differences between the traditional and later models. In this search, data from the database are preprocessed first to increase the diversity of the data to avoid overfitting (Sec. A). Secondly, three different CNN models are reconstructed, including LeNet, AlexNet, and ResNet (Sec. B). Finally, the above models are compared on the dimensions of accuracy, and loss rate, to contrast the learning efficiency of the traditional and later models (Sec. C).

A. Data Pre-processing

The database used in this study [24] contains a total of about 1000 chest CT images, Fig. 1 shows a sampling of the images in the database. All images were taken of the lungs at the same angle and with the same type of computed tomography scanner, which means that all data in the database contain variables that differ only between patients. This paper focuses on the classification of non-small cell lung cancers, "Fig. 2" shows three subclasses belonging to non-small cell lung cancer with one normal class for comparison. To avoid the overfitting problem that happens commonly in CNN models, all data in the database are horizontally flipped, scaled, and shifted to increase data diversity.

B. Model Implementation

LeNet, AlexNet and ResNet are used for comparing traditional and later CNN architectures. We give detailed descriptions of these three models in the following.

1) LeNet Implementation

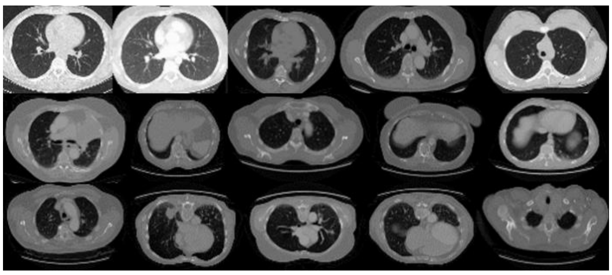

Fig. 1. A random sampling of non-small cell lung cancer CT image database

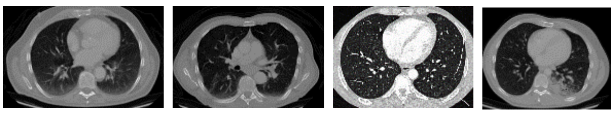

Fig. 2. Images from left to right belong to the class of Adenocarcinoma. Large Cell Carcinoma, Normal, Squamous Cell Carcinoma

LeNet was proposed by Yann LeCun et al. in 1989 [13]. For example, Mesut Toğaçar et al. [15] used LeNet to perform the classification of lung cancer in their 2019 study, and after they modified LeNet's original Adam optimization method to RMSprop. They achieved a classification accuracy that is 81.20%. They also compared the modified LeNet with VGG16, an architecture proposed in 2013 [17], and found that LeNet was more accurate than VGG-16 with the same training epoch, which implies that LeNet is one of the earliest CNNs but still has certain competitiveness in today's lung cancer applications.

In this paper, a modified version of the LeNet that references the study [15] is reimplemented and shown in Fig. 3. LeNet has 2 sets of convolutional layer and average pooling layer group, along with a flatten layer, two fully connected layers and an output layer, for a total of 8 layers. Unlike the original LeNet, the optimizer here is not Adam but RMSprop. Formulas for RMSprop are defined in (1) and (2), respectively. The choice of the optimizer is crucial for this comparison experiment since the optimizer determines the learning efficiency of the model and the RMSprop optimizer is just right for online and non-stationary situations which also fits the current work scenario. Moreover, when LeNet is with Adam optimizer and LeNet is with RMSprop optimizer simultaneously for lung cancer classification based on chest CT images, the accuracy of Adam optimizer LeNet was 73.82%, while RMSprop optimizer LeNet was 81.20% after the same 100 training epochs.

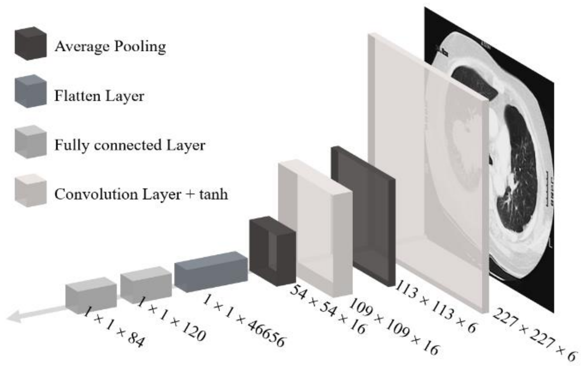

Fig. 3. LeNet structure

$$E[g^2]_t = 0.9[g^2]_{t-1} + 0.1g_t^2 \tag{1}$$

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{E[g^2]_{t+\epsilon}}} g_t \tag{2}$$

2) AlexNet Implementation

AlexNet has more layers and no longer uses Tanh as an activation function as LeNet. In a study [14], AlexNet is modified based on the original AlexNet structure. Besides, the chest CT images are processed with wavelet transformation before inputting them into the neuron network, and the resulting model reaches an accuracy close to around 99% after nearing 350 training epochs. In comparison, the original AlexNet is only 84.4% accurate. Study [15] finds that using Daubechies order 2 wavelet basis function (db2) in Fig. 4 to preprocess chest CT images can let AlexNet achieve significantly higher accuracy in the same training period.

Wavelet transform is widely used in the medical field because of its data compression, image enhancement and noise removal capabilities [21-22]. For lung cancer classification, the main purpose of the wavelet used is to extract the subject information from the images and remove the distracting noise.

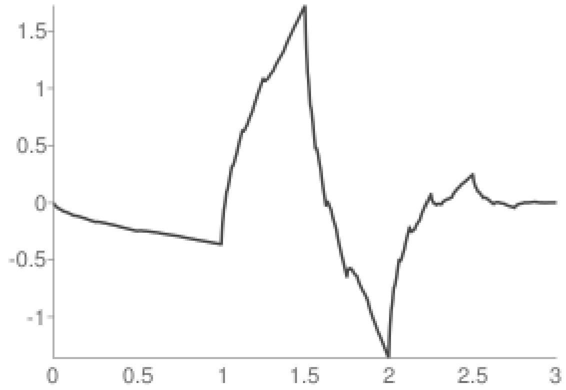

Fig. 4. Wavelet Daubechies order2 (db2)

In this paper, a modified AlexNet is reconstructed which is referenced in Eric Matsuyama et al.'s paper [14]. Unlike the original AlexNet, the input image will be passed through a db2 wavelet transformation layer before all other layers to better extract the subject content in the image. The AlexNet used in this paper is shown in Fig.5. AlexNet has more layers than LeNet. The activation function used in AlexNet is no longer the Tanh Hidden Layer Activation Function defined in (3), but the Rectified Linear Activation Function defined in (4). This activation function can better solve the vanishing gradient problem which causes the CNN to have more layers.

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (3)$$

$$f(x) = \max(0, x) \quad (4)$$

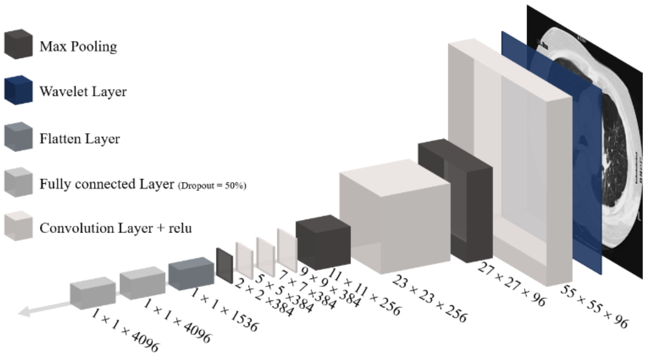

Fig. 5. AlexNet structure

3) *ResNet Implementation*

ResNet [16] makes it possible to train deeper neural networks by creating jump links between groups of layers, avoiding the problem of vanishing gradient. The ResNet used in this paper has 50 layers of which 48 Convolution layers along with 1 Max Pooling and 1 Average Pooling layer, known as ResNet50.

The overall CNN model is shown in Fig. 6. Pre-training ResNet50 is used for model construction in this paper, and all neurons in the pre-trained model were set to be trainable.

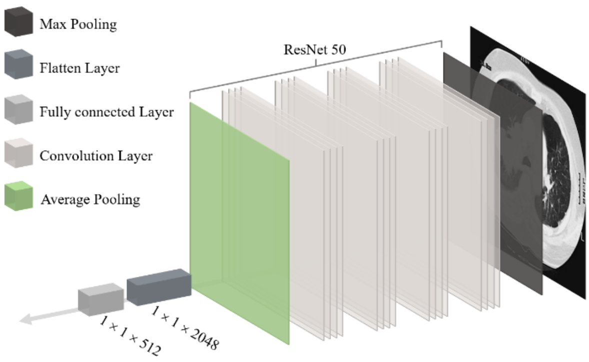

Fig. 6. ResNet structure

C. Method Comparison

To compare the learning efficiency of the traditional and later models to demonstrate that neural networks with more complex structures and deeper layers are more competitive for lung cancer image classification, this paper compares the models in two dimensions. Firstly, accuracy after different training epochs will be recorded for cross-sectional comparison between models. Accuracy is the correct rate of CNN prediction for data other than training data after a certain number of training epochs. Secondly, the loss values are compared between the models. The loss values imply how well the models perform after each iteration. This paper records the model accuracy curves and loss value curves from 0 to 256 training epochs. To ensure that the accuracy of all models is close to saturation, and the loss value is as close to 0 as possible.

III. EXPERIMENTAL SETTINGS

To compare the learning efficiency, including accuracy and loss rate, of the traditional and later models for image classification of lung cancer, all models are trained based on the same dataset. Also, because the two traditional models LeNet, and AlexNet in this paper are derived from previous research, the modified versions of these two traditional models have only minor structural differences from the original ones but have better performance in the lung cancer field than the original ones.

This section introduces the data set used for the experiments in (Sec. A) and the coordinates used for the experimental comparisons in (Sec. B). Furthermore, the detailed experimental parameters of the referenced models are presented (Sec. C), as well as the setup parameters of the ResNet representing the later model. Finally, detailed experimental procedures will be introduced (Sec. D).

A. Dataset Description

The dataset used in this paper is from Kaggle [24]. It contains one thousand non-identical sizes chest CT images. 70% of them are training set, 20% are test set, and the remaining 10% are used for validation. We classify subsets of non-small cell lung cancer so that this dataset is classified according to the categories under non-small cell lung cancer including Adenocarcinoma, Large cell carcinoma, Squamous cell carcinoma, and another category that indicates normal cell. Fig. 7 shows the percentage of all categories in the database. Fig. 8 shows the weight of the training set, test set, and validation set in the whole dataset.

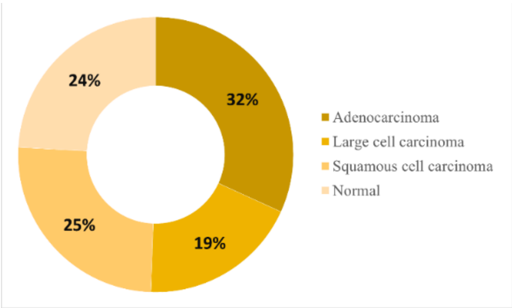

Fig. 7. Weight of each category in the Chest CT-Scan images Dataset

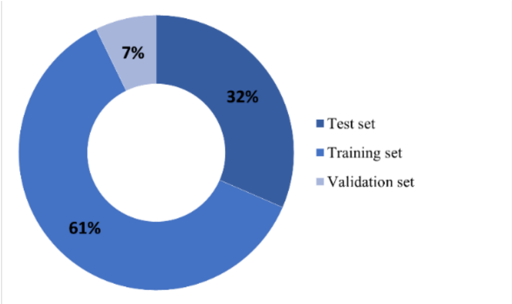

Fig. 8. Weight of each data subset

B. *Evaluation Indicators*

We use learning efficiency, loss and effectiveness to make detailed comparisons among the models. Accuracy represents the percentage of correct predictions made by the model after a certain number of iterations. The loss rate represents the model learning effect for each iteration. These two values are used as vertical coordinates and the training epochs are used as the horizontal coordinates. We observe the curves of these two charts. The training accuracy by training epochs chart will visualize the model learning efficiency between training epochs, while the loss rate by training epochs chart will explain the model learning effectiveness between training epochs.

The learning efficiency is the number of learning epochs it takes for a model to reach an acceptable accuracy rate. A model has a higher learning efficiency when it achieves the same accuracy rate as other models in a shorter learning epoch. A higher learning efficiency allows the model to be applied at a lower cost, including economic cost and time cost. The time efficiency is defined as (5).

$$Learning\ Efficiency = \frac{Accuary}{training\ epochs} \qquad (5)$$

C. *Model Settings*

1) *Input layer dimension*

The images in the database are uniformly sized before entering the input layer, which has a dimension of 227×227×3. All images are scaled to a square of 227 pixels and split into three layers based on 3 RGB channels.

2) *Model loss function and optimizer*

The loss function used in all models is categorical crossentropy which is shown in (6), and this equation is used for multi-category classification tasks. The following sum is computed by the categorical crossentropy loss function to determine the loss of an example [25].

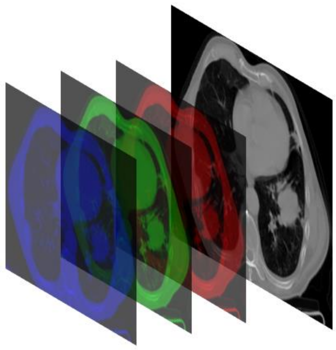

Fig. 9. The original chest CT image is splinted into 3 RGB channels, each channel only records one colour value

$$Loss = -\sum_{i=1}^{output\ size} y_i \cdot \log \hat{y}_i (6)$$

The optimizers of the reference model used in this paper are SGD and RMSprop, respectively. The loss function and optimizer used in models are shown in Table I.

TABLE I. MODEL SETTINGS

Model	Loss Function	Optimizer
LeNet	Categorical Crossentropy	RMSprop
AlexNet	Categorical Crossentropy	SGD
ResNet	Categorical Crossentropy	SGD

3) *Ouput layer dimension*

The output layer of all models used in this paper is a dense layer with four neurons, which correspond to three subclassifications of non-small cell lung cancer and one normal cell classification.

D. *Experimental Procedures*

LeNet, AlexNet and ResNet are trained until all models reach learning saturation, i.e., the loss rate of all models floats less than 0.2 concerning the before and after iterations, and out of this state for more than 12 training epochs. In the above method, the highest accuracy rate that can be achieved by all models after the growth period will be obtained. Also, the accuracies and loss rates of all models are recorded to observe the learning efficiency of the models as well as the learning effectiveness.

IV. RESULT AND DISCUSSION

In this section, we compare the accuracy and loss rates among all models in (Sec. A), the learning efficiency between models in (Sec. B), and the learning effects in (Sec. C).

A. *Comparison of Accuracy and Loss rate*

In this study, we first compare the highest accuracy rates that the three models can achieve within 32 learning epochs. As shown in TABLE II, ResNet can achieve a relatively impressive accuracy rate within 32 learning epochs. While on the contrary, the two CNNs representing traditional models, LeNet and AlexNet, their accuracy curves rise sluggishly and are not competitive compared to ResNet as shown in Fig. 10.

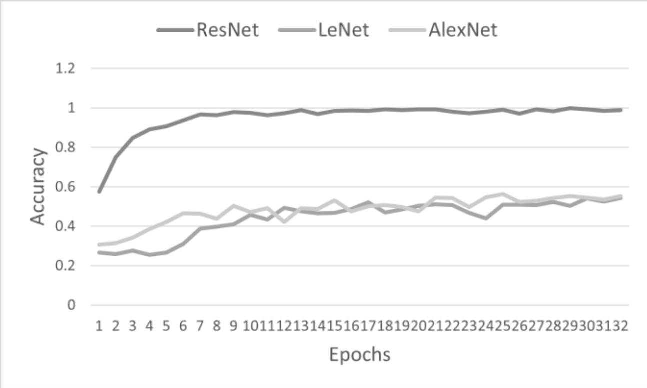

Fig. 10. Line chart with the number of training epochs as the x-axis and accuracy as the y-axis. only the first 32 epochs are recorded

TABLE II. MODEL ACCURACY COMPARISON

	ResNet	LeNet	AlexNet
Highest accuracy in 1-32 epochs	99.84%	54.32%	56.28%
Highest accuracy in 1-256 epochs	100.00%	63.78%	90.21%

Next, we compare the training epochs. In this stage, each model has trained 256 epochs, and as shown in Table II, we record the highest test set prediction accuracy in these 256 epochs. As shown in Fig. 11, ResNet stays in a high accuracy state after a rapid growth period in the early stage, while AlexNet maintains a steady upward trend during the whole training process. Meanwhile, the accuracy of LeNet has been fluctuating between 50% and 60% in the middle and late stages.

The loss rates of the models are shown in Fig.12, where the loss rate of ResNet has dropped to around zero in less than 50 training epochs, which implies that 50 training epochs can already make the ResNet to be in the learning saturation state. The loss rate of AlexNet shows an approximate linear decreasing trend throughout the training process, which implies that the improvement of the model accuracy is approximate and positive. The loss rate of LeNet is kept around 1 until the end after a short drop, with a significant increase in the middle, probably due to overfitting caused by the vanishing gradient problem. LeNet also enters a state of learning saturation in the middle, which means that it is difficult to make LeNet's accuracy continue to improve with more learning epochs.

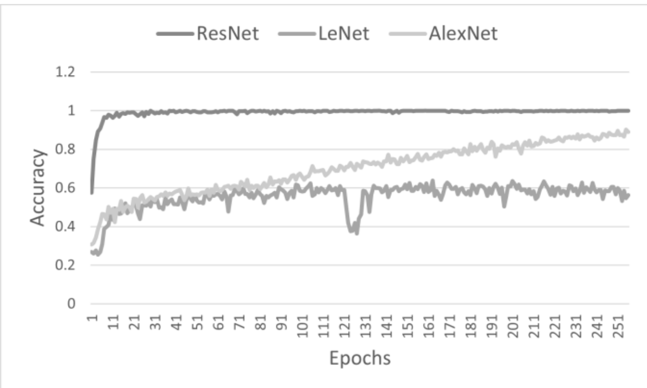

Fig. 11. Line chart with the number of training epochs as the x-axis and accuracy as the y-axis. Including records for the whole training process.

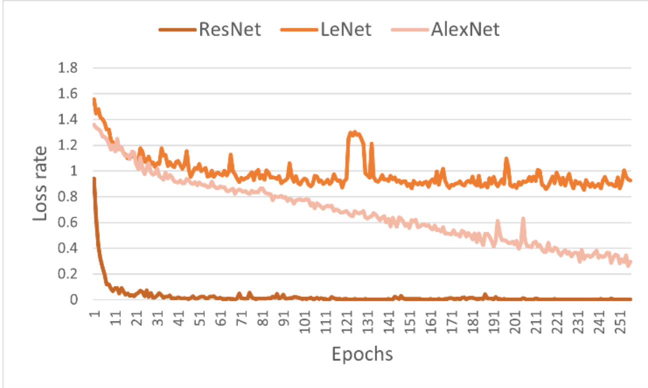

Fig. 12. Line chart with the number of training epochs as the x-axis and loss rate as the y-axis. Including records for the whole training process.

B. *Comparison of Learning efficiency*

We use learning efficiency to describe the efficiency of model learning, which is the differential of the model accuracy curve. As shown in Fig.13, the learning efficiency curves for each of the three models are shown, where the ResNet starts at 0.6, indicating that the accuracy rate improves from 0% to 60% in the first learning epoch, and the subsequent curve decreases exponentially. Meanwhile, the AlexNet and LeNet curves also maintain an exponential decrease. The learning efficiency of LeNet and AlexNet always trails that of ResNet in the learning epochs from 0 to 32. Because the model that has not reached learning saturation must have a higher learning efficiency compared with the saturated model, and because ResNet has reached learning saturation in the first 32 epochs, it is meaningless and misleading to compare the learning efficiency curves of the models after 32 epochs.

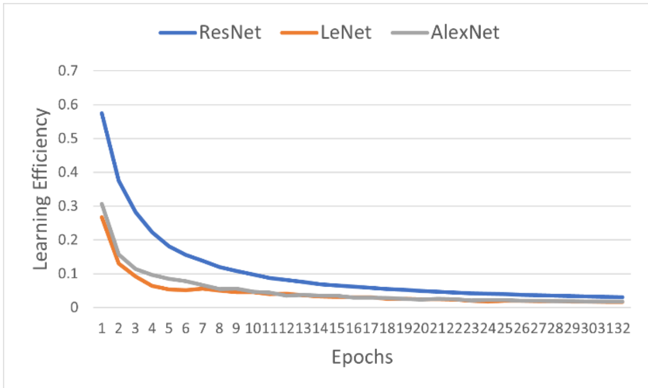

Fig. 13. Line chart with the number of training epochs as the x-axis and model learning efficiency as the y-axis. Including records for only the first 32 epochs.

V. CONCLUSION

In this paper, we compare LeNet and AlexNet in non-small cell lung cancer diagnosis tasks with the accuracy, loss rate and learning efficiency of Resnet. In a total of 256 training epochs, ResNet not only has the highest accuracy rate and very low loss rate but also has the highest learning efficiency before reaching learning saturation. This indicates that ResNet is more competitive than LeNet and AlexNet, which are two models representing traditional CNNs. It also proves that the traditional models are not overwhelmingly competitive in the field of lung cancer identification as it is believed, or even not competitive.

In future research, we will explore in more depth the effect of CNN model structure on image classification accuracy in the medical field.

REFERENCES

[1] Information from: https://www.canada.ca/en/public-health/services/chronic-diseases/cancer/lung-cancer.html

[2] Bromfield, G., D. Dale, and P. De. "Canadian Cancer Society's Advisory Committee on Cancer Statistics. Canadian cancer statistics 2015." (2015).

[3] Matsuyama, Eri, and Du-Yih Tsai. "Automated classification of lung diseases in computed tomography images using a wavelet-based convolutional neural network." *Journal of Biomedical Science and Engineering* 11, no. 10 (2018): 263.

[4] Uchiyama, Yoshikazu, Shigehiko Katsuragawa, Hiroyuki Abe, Junji Shiraishi, Feng Li, Qiang Li, Chao-Tong Zhang, Kenji Suzuki, and Kunio Doi. "Quantitative computerized analysis of diffuse lung disease in high-resolution computed tomography." *Medical Physics* 30, no. 9 (2003): 2440-2454.

[5] Sluimer, Ingrid, Arnold Schilham, Mathias Prokop, and Bram Van Ginneken. "Computer analysis of computed tomography scans of the lung: a survey." *IEEE transactions on medical imaging* 25, no. 4 (2006): 385-405.

[6] Song, Yang, Weidong Cai, Yun Zhou, and David Dagan Feng. "Feature-based image patch approximation for lung tissue classification." *IEEE transactions on medical imaging* 32, no. 4 (2013): 797-808.

[7] Teramoto, Atsushi, Hiroshi Fujita, Katsuaki Takahashi, Osamu Yamamuro, Tsuneo Tamaki, Masami Nishio, and Toshiki Kobayashi. "Hybrid Method for the Detection of Pulmonary Nodules Using Positron Emission Tomography/Computed Tomography: A Preliminary Study." *International Journal of Computer Assisted Radiology and Surgery* 9, no. 1 (2013): 59–69.

[8] Nasr-Esfahani, E., S. Samavi, N. Karimi, S.M.R. Soroushmehr, M.H. Jafari, K. Ward, and K. Najarian. "Melanoma Detection by Analysis of Clinical Images Using Convolutional Neural Network." *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2016.

[9] Arevalo, John, Fabio A. González, Raúl Ramos-Pollán, Jose L. Oliveira, and Miguel Angel Guevara Lopez. "Representation learning for mammography mass lesion classification with convolutional neural networks." *Computer methods and programs in biomedicine* 127 (2016): 248-257.

[10] Sirinukunwattana, Korsuk, Shan E. Ahmed Raza, Yee-Wah Tsang, David RJ Snead, Ian A. Cree, and Nasir M. Rajpoot. "Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images." *IEEE transactions on medical imaging* 35, no. 5 (2016): 1196-1206.

[11] Sarvamangala, D. R., and Raghavendra V. Kulkarni. "Convolutional Neural Networks in Medical Image Understanding: A Survey." *Evolutionary Intelligence* 15, no. 1 (2021): 1–22.

[12] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." *Advances in neural information processing systems* 25 (2012).

[13] LeCun, Yann, Léon Bottou, Yoshua Bengio, and Patrick Haffner. "Gradient-based learning applied to document recognition." *Proceedings of the IEEE* 86, no. 11 (1998): 2278-2324.

[14] Matsuyama, Eri, and Du-Yih Tsai. "Automated Classification of Lung Diseases in Computed Tomography Images Using a Wavelet-Based Convolutional Neural Network." *Journal of Biomedical Science and Engineering* 11, no. 10 (2018): 263–74.

[15] Toğaçar, Mesut, Burhan Ergen, and Zafer Cömert. "Detection of Lung Cancer on Chest CT Images Using Minimum Redundancy Maximum Relevance Feature Selection Method with Convolutional Neural Networks." *Biocybernetics and Biomedical Engineering* 40, no. 1 (2020): 23–39.

[16] He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Deep residual learning for image recognition." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770-778. 2016.

[17] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." *arXiv preprint arXiv:1409.1556* (2014).

[18] Huang, Gao, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q. Weinberger. "Densely connected convolutional networks." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4700-4708. 2017.

[19] Iandola, Forrest N., Song Han, Matthew W. Moskewicz, Khalid Ashraf, William J. Dally, and Kurt Keutzer. "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size." *arXiv preprint arXiv:1602.07360* (2016).

[20] Jia, Guangyu, Hak-Keung Lam, and Yujia Xu. "Classification of Covid-19 Chest X-Ray and CT Images Using a Type of Dynamic CNN Modification Method." *Computers in Biology and Medicine* 134 (2021): 104425.

[21] Matsuyama, Eri, Du-Yih Tsai, Yongbum Lee, Masaki Tsurumaki, Noriyuki Takahashi, Haruyuki Watanabe, and Hsian-Min Chen. "A modified undecimated discrete wavelet transform based approach to mammographic image denoising." *Journal of digital imaging* 26, no. 4 (2013): 748-758.

[22] Daubechies, Ingrid. "Ten Lectures on Wavelets," 1992.

[23] Information from: http://wavelets.pybytes.com/wavelet/db2/

[24] Information from: https://www.kaggle.com/datasets/mohamedhanyyy/chest-ctscan-images

[25] Information from: https://peltarion.com/knowledge-center/documentation/modeling-view/build-an-ai-model/loss-functions/categorical-crossentropy