Comparative Study of Deep Features for Lung Disease Classification in Chest X-ray Image

Ei Ei Khaing
FCST
Universtiy of Computer Studies
Taungoo, Myanmar
eekhaingct@gmail.com

Thu Zar Aung
Pro Rector
University of Technologies
(Yatarpon Cyber City),Pyin Oo Lwin
whiteangel.tza@gmail.com

Abstract— **As pulmonary disease spreads around the world, an accurate method of diagnosis is required. Since the virus spreads so rapidly, it can be difficult for medical professionals to diagnose lung disease. This is particularly challenging part of diagnosing COVID-19 and pneumonia in patients. The moment of, the main goal of this technology is to create an approach for detecting respiratory problems early and preventing the virus from spreading rapidly. Medical image classification relies heavily on image processing techniques to provide a decision-support system. In image classification, pre-trained Neural Networks are also utilized to extract deep characteristics. The paper, several deep feature extraction methods are analyzed for a challenging medical image analysis assignment involving the evaluation of chest X-ray images. In our feature analysis, we present how pre-trained Alex Net, ResNet-101, and VGG19 Net can accurately label the Chest X-ray image for lung disease classification. In the proposed framework, three diseases are distinguished: COVID-19, viral pneumonia, and normal. In experiments, accuracy, true positive rate, and false negative rate are used to gauge the effectiveness of the classifier.**

Keywords— **Pre-trained Neural Networks, Deep feature, Lung Disease Classification, Alex Net, ResNet-101, and VGG19 Net**

I. Introduction

Medical imaging, geographic information systems, satellite television, and astronomy are a few examples of real-world applications for digital image processing and deep feature extraction. Using the layer activations as features, a pre-trained neural network is employed as a feature extractor. The extracted activations serve as the features for training another machine learning model, such as SVM, DT, KNN, etc. Feature analysis is a crucial stage in image classification due to its predominance over the classification step. Artificial intelligence's subfield of machine learning consists of a collection of techniques that can handle varied information settings and complex input, generalize knowledge from the past, and produce new ideas. [1].

CT scans were used to swiftly identify COVID-19 using a dual-sampling focus system. It was suggested that a 3D convolutional neural network (CNN) using a focus function via the internet be utilized to focus on illness locations inside the lungs when generating diagnostic judgments [2]. The CNN, SVM and Sobel filter fusion were suggested for identifying of COVID-19 in X-ray images. A new series of X-ray images was acquired, and they were then put through a high-pass filter that used a Sobel filter to recover the image contours. An SVM classifier using a ten-fold validation set approach is utilized to identify the images after they have been fed into a CNN deep learning model. The Sobel filter and CNN-SVM combination improved CNN's effectiveness [3].A dual-sampling attention network was developed to rapidly identify COVID-19 from CT images. An online attention module with a 3D convolutional network (CNN) was suggested to focus on illness locations inside the lungs when making diagnosis judgments [4]. To detect COVID-19 using X-ray images, CNN, SVM, and Sobel filter fusion were suggested. A high pass filter that used a Sobel filter to recover the image contours was applied to a fresh collection of X-ray pictures. This was done to address computational challenges in identifying pictures and health-care smart systems. Machine learning approaches have been employed to choose the proper attributes and forecast COVID-19 patients after the features were developed. To extract location-specific characteristics, all pictures were standardized to get disease and lung field fragmentation. The infection-size-aware random forest (iSARF) method was created for automatically dividing individuals into sections according to the lengths of the impacted carcinomas. Random forests were subsequently applied to categorize each group of subjects dataset and experimental findings section, and a conclusion.

II. Deep Feature

The classification of lung diseases is one of the key elements in the advancement of medicine and has an effect on global health. They can easily diagnose these ailments and quickly regain health if lung disease is discovered in its early stages. Manual lung disease monitoring is labor intensive, necessitates specific medical knowledge, and takes a long time to process. In computer vision and image processing, a good feature is essential for precisely capturing the characteristics of an ordinary picture with a constrained set of variables. The recommended system analyses deep features for lung illness diagnosis in chest X-ray images using image processing techniques, deep feature analysis, and SVM algorithm. The suggested system has two basic phases: deep feature steps of extraction and classification. Deep feature analysis uses pre-trained AlexNet, ResNet-101 Net, and VGG19 Net to extract deep features, which are then compared to highlight the classification performance of each feature in classifying lung diseases.

A. AlexNet

It has eight learned layers, including five convolutional and three fully connected layers. Figure 1 depicts the architecture of AlexNet.

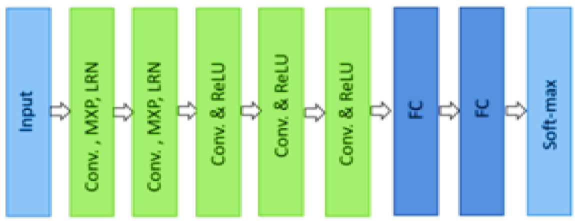

Fig. 1. Architectre of AlexNet

B. Residual Network (ResNet-101)

The architecture of ResNet-101 is illustrated in figure2. A typical forward-feeding network with a residual link is called ResNet. The following equation [5] can be used to define the residual unit's final output, xl:

$$xl = \mathcal{F}\,(xl{-}1) + xl{-}1$$

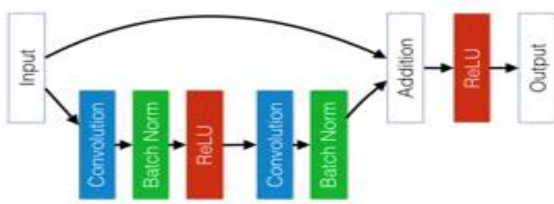

Fig. 2. Architecture of ResNet-101

C. Visual Geometry Group Network (VGG19 Net)

The VGG19 architecture has five distinct building components.. The architecture of VGG19 Net is shown in figure 3[5].

Fig. 3. Architecture of VGG19

III. Deep Features in Lung Disease Classification

The deep feature. Additionally, the classifier's testing and training periods are crucial for correctly identifying lung diseases in chest X-ray images. Linear discriminant analysis (LDA), Support Vector Machine (SVM), Decision Tree (DT), and K-nearest Neighbour (KNN) are examples of traditional machine learning classification techniques.

The system, linear discriminant analysis (LDA) machine learning methods are used to conduct an experimental examination of deep features. Deep features in image classification are also highlighted by the classification performance measurements. The overall system design for comparative study of deep features is shown in figure4.

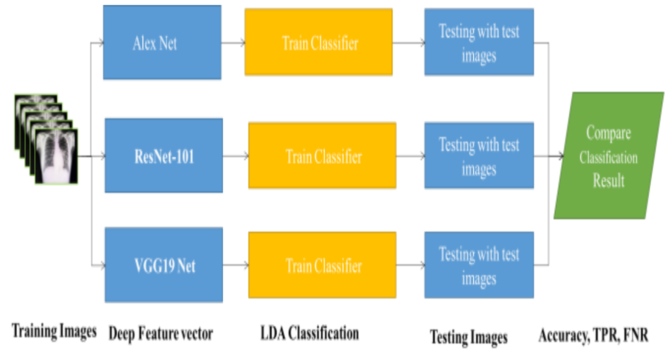

Fig. 4. Overview System of Comparative Study of Different Deep Features for Lung Disease Classification

A. Linear discriminant analysis (LDA)

It is simpled and effective. The idea is simple and widely accepted, thus there are many expansions and adaptations to it. There are several uses for it, such as the linear separation of data points, the classification of multi-featured data, the differentiation among various features of a dataset, etc. LDA is used in technologies for consumer and marketing recognition, medical care analysis, and machine vision. LDA is a quick and accurate classification technique with many real-world uses [5].

IV. Dataset

An open access database containing more than seventy thousand chest X-ray Images for the COVID-19, Viral Pneumonia, and Normal. Doctors marked up many of the collection's imges. Table 1 contains the total amount of photos for each label.

Table I. COVID-19_Radiography_Dataset

No.	Label	No. of images
1	COVID-19	3616
2	Viral Pneumonia	1345
3	Normal	10192
Total		**15153**

V. Experimental Result

The effectiveness of the classifier for deep features is evaluated in this research using classification methods. The Dataset is divided into two sets called training and testing. The images in training set are used to extract deep features and train classifier to get classifier. The testing step is used to demonstrate the classifier's classification accuracy by dividing the number of data in a particular class that were properly classified by the full count of data in that

$$\text{Classification Accuracy} = \frac{\text{Number of correctly classified}}{\text{Total number of test data}}$$

The quantity of information elements that were accurately identified as belonging to the class is measured by the true positive rate. This is appropriately determined are those with a high TPR.

$$\text{True Positive Rate} = \frac{\text{Number of True Positives}}{\text{Number of True Positives} + \text{Number of False Negatives}}$$

False Negative Rate, also known as Missed Rate, counts the number of data that are mistakenly excluded from the class.

$$\text{False Negative Rate} = \frac{\text{Number of False Negatives}}{\text{Number of True Positives} + \text{Number of False Negatives}}$$

Table II. Evaluation measures of classifier according to training 60% and testing 40% of COVID-19_Radiography_Dataset

Classification	Classification Algorithm	Classification Accuracy	True Positive Rate	False Negative Rate
LDA	AlexNet	**91.81%**	**85.17%**	**14.83%**
	ResNet101	85.9%	84.0%	16.0%
	VGG19 Net	87.9%	87.1%	12.9%

Table III. Evaluation measures of classifier according to training 80% and testing 20% of COVID-19_Radiography_Dataset

Classification	Classification Algorithm	Classification Accuracy	True Positive Rate	False Negative Rate
LDA	AlexNet	**97.33%**	97.2%	2.8%
	ResNet101	90.81%	90.4%	9.6%
	VGG19 Net	92.56%	90.9%	9.1%

A. Result and Discussion

In table 2 and table3, according to classification accuracy, Alex Net can correctly classify 91.81% and 97.33% of test results, and the category of lung diseases is still incorrect. According to True Positive Rate and False Negative Rate, Alex Net feature can relevantly classified 85.17% and 97.2% of test data and 14.83% and 2.8% of a positive categorization test outcome in tables 2 and 3 is most likely inaccurate.

Our suggested feature continues to the best categorization and could produce meaningful different texture feature values for each label because classification accuracy value is larger true positive rate regarding any tests. The form nature has no bearing on the proposal feature since our proposed feature based on deep feature analysis approach. The statistical column chart of the classification performance measures of different deep features according to training 60% and testing 40% of dataset is shown in figure 6 and training 80% and testing 20% is in figure 5.

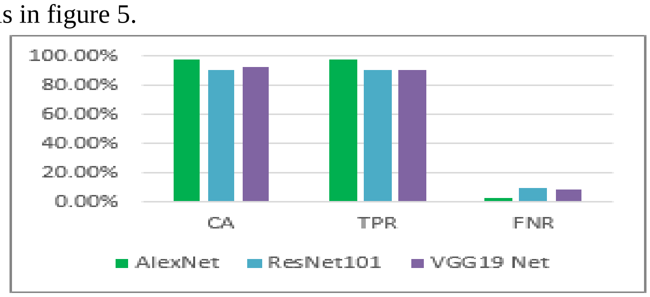

Fig. V. Column Chart for Evaluation measures of classifier according to training 80% and testing 20% of COVID-19_Radiography_Dataset

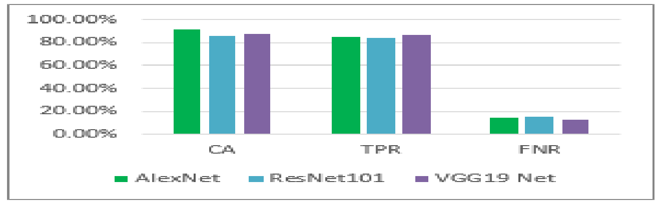

Fig. VI. Column Chart for evaluation measures of classifier according to training 60% and testing 40% of COVID-19_Radiography_Dataset

VI. CONCLUSIONS

This study examined various deep learning techniques for identifying lung diseases in chest X-ray images. LDA, which uses several deep features taken from pre-trained Alex Net, RestNet101, and VGG19Net, is the main classification algorithm for identifying lung diseases. In order to provide a satisfactory result with being able to categorize lung disease for chest X-ray images, the study analyses and suggests various deep features. The research study uses classification metrics to demonstrate the benefits of combining deep learning and machine learning for the identification of lung diseases. One of the greatest machine learning algorithms, LDA can work in conjunction with deep transfer learning to improve classification results for the recognized lung illness issue of classification. Our test demonstrates the benefits and bearing of deep features in various categorization algorithms. In order to correctly identify the enormous amounts of unlabeled train data in supervised learning, we will have to use more machine learning algorithms or other dimensional reduction strategies in the future.

ACKNOWLEDGMENT

I want to publicly thank Dr. Thu Zar Aung, ProRector of the University of Technology in Yatanarpon Cyber City, who served as my research supervisor for enabling me to conduct this study and for her essential guidance. She has greatly impressed me with her enthusiasm, determination, sincerity, and drive. She has shared with me the techniques for conducting the research and clearly explaining the outcomes of the study.

REFERENCES

[1] M de Bruijne, “Machine learning approaches in medical image analysis: From detection to diagnosis,” vol. 33, pp. 94–97.

[2] Ouyang, Xi, et al. "Dual-sampling attention network for diagnosis of COVID-19 from community acquired pneumonia." IEEE Transactions on Medical Imaging 39.8 (2020): 2595-2605.

[3] Sharifrazi, Danial, et al. "Fusion of convolution neural network, support vector machine and Sobel filter for accurate detection of COVID-19 patients using X-ray images." Biomedical Signal Processing and Control 68 (2021): 102622.

[4] Wan X, Zhang X, Liu L (2021) An improved VGG19 transfer learning strip steel surface defect recognition deep neural network based on few samples and imbalanced datasets. Appl Sci11:2606. https://doi.org/10.3390/app11062606

[5] Priyankur Sarkar, “What is LDA: Linear Discriminant Analysis for Machine Learning”, 27th Dec, 2022, https://www.knowledgehut.com/blog/data-science/linear-discriminant-analysis-for-machine-learning.