World Academy of Science, Engineering and Technology
International Journal of Health and Medical Engineering
Vol:17, No:4, 2023

Automatic Classification of Lung Diseases from CT Images

Abobaker Mohammed Qasem Farhan, Shangming Yang, Mohammed Al-Nehari

***Abstract*—**Pneumonia is a kind of lung disease that creates congestion in the chest. Such pneumonic conditions lead to loss of life due to the severity of high congestion. Pneumonic lung disease is caused by viral pneumonia, bacterial pneumonia, or COVID-19 induced pneumonia. The early prediction and classification of such lung diseases help reduce the mortality rate. We propose the automatic Computer-Aided Diagnosis (CAD) system in this paper using the deep learning approach. The proposed CAD system takes input from raw computerized tomography (CT) scans of the patient's chest and automatically predicts disease classification. We designed the Hybrid Deep Learning Algorithm (HDLA) to improve accuracy and reduce processing requirements. The raw CT scans are pre-processed first to enhance their quality for further analysis. We then applied a hybrid model that consists of automatic feature extraction and classification. We propose the robust 2D Convolutional Neural Network (CNN) model to extract the automatic features from the pre-processed CT image. This CNN model assures feature learning with extremely effective 1D feature extraction for each input CT image. The outcome of the 2D CNN model is then normalized using the Min-Max technique. The second step of the proposed hybrid model is related to training and classification using different classifiers. The simulation outcomes using the publicly available dataset prove the robustness and efficiency of the proposed model compared to state-of-art algorithms.

***Keywords*—**CT scans, COVID-19, deep learning, image processing, pneumonia, lung disease.

I. Introduction

AUTOMATED systems using deep learning have been developed to help with the detection of pneumonic conditions around the lungs using chest CT images. In this article, we will discuss the various approaches to automated classification of lung diseases from chest CT images. Lung complaints caused by various elements have introduced a higher mortality rate for a couple of years [1]. Nonetheless, certain individuals experienced extreme pneumonic conditions in their lungs that brought about death likewise. A large portion of the cases who succumbed to the Coronavirus had experienced high chest blockage (pneumonia) as a huge decrease in oxygen level and subsequently significant cardiovascular failure [2]. On the opposite side, pneumonia is likewise a sort of lung sickness that prompts irritation in the tiny air sacs inside the lungs of the human body [3]. It can be overflowing with fluids, which makes it difficult to relax. Pneumonia can be brought about by different reasons like viral contaminations (like COVID-19, bacterial influenza, or viral pipe), normal cold, and bacterial diseases. Because of the appearance of COVID-19 illness, it is an effective assignment for clinical professionals to identify lung diseases from various chest scans like X-rays and CT [4]. Subsequently, our concentration in this article is the early recognition and classification of lung disorders from the CT images for suitable treatment to lower the death rate.

An alternate sort of lung infection that has a huge danger to people is lung disease. The World Health Organization (WHO) asserts that roughly 8 million individuals experienced lung malignant growth till today. To treat lung cancer, a few studies have successfully introduced for its earliest expectancy, methods employing computer vision processes and delicate processing. The location of lung disease has been performed utilizing methods like X-ray, Magnetic Resonance Imaging (MRI), Computed Tomography (CT), and isotope. Among these, CT and X-ray chest imaging strategies are much of the time utilized for the recognition of different lung illnesses. The X-ray and CT images are used by radiologists and doctors to find lung sicknesses [5]. The X-ray and CT images are practical with comparative sorts of results contrasted with the MRI check. Thus, many specialists suggested the chest CT and X-ray scan for the examination of lung infections, particularly during the COVID-19 period [6]. The CT method has been utilized to analyze irregularities in the areas of the human body like the chest, skull, bones, teeth, and so on. For a long time, clinical specialists have utilized the CT strategy to dissect and investigate the different irregularities in the human body organs. In the investigation of the CT image of the patient, radiologists recognize a few conditions like pneumonia, nodule, pleurisy, radiation, invasion, fractures, pneumothorax, pericarditis, and so on [6].

The detection of pneumonic conditions around the lungs using chest CT images is viewed as an intricate interaction for radiologists. Therefore, it got huge consideration from the scientists for programmed lung infection identification. Since the previous decade, numerous CAD frameworks were presented using CT images for diagnosis purposes. The recent progress of automatic mechanisms of disease classification using deep learning connected with Internet of Things (IoT) [7]-[9] motivates the proposal of the framework in this paper. The current challenge of using the deep learning methods is related to the higher features learning time and delayed classification outcome. To end this, we proposed the deep learning model for the automatic classification of lung diseases from the chest CT images called HDLA. The novelty of this model lies in a

Abobaker Mohammed Qasem Farhan and Shangming Yang are with School of Information and Software Engineering, University of Electronic Science and Technology of China, Chengdu, China (e-mail: mabobaker957@gmail.com).

Mohammed Al-Nehari is with Wrocław University of Science and Technology, Public university in Wrocław, Poland.

simplified mechanism for the features extraction and classification independently. We have designed the 2D CNN layers with optimal features size rather than the 3D model to reduce the processing time of automatic CNN features extraction.

II. Related Works

Recently several studies introduced automatic pneumonic disease detection and classification from the X-ray or CT images using the deep learning models. This section presents some recent deep learning methods that are similar to the proposed approach.

In [10], authors have designed an artificial intelligence (AI)-based CAD system to detect COVID-19 induced pneumonia and classify it from other types of pneumonia. They designed an AI-based system to automatically assist physicians and radiologists with the fast diagnosis of lung diseases using CT scans. A CNN-based automatic CAD system was proposed in [11] for COVID-19 disease classification from CT images using different CNN pre-trained models. They have designed a transfer learning mechanism for high accuracy. A recent study on describing and experimentally evaluating the open-source CT images dataset related to chest infections is proposed in [12]. They described a dataset that consists of 1521 pneumonia patients and 130 features and SARS-CoV-2 status. In [13], authors have designed a COVIDNet-CT system to detect lung diseases from CT images using the CNN model. In [14], authors have designed an AI-based system from CT images. The AI-based mechanism was hypothesized to extract the CT scan features for the diagnosis. Another AI model proposed in [15] uses ResUNet to analyze the radiologist's performance for classifying the pneumonia disease from the CT images. A similar study was proposed in [16] to classify COVID-19 and other pneumonia using the CT images. In [17], authors have proposed Inf-Net (Infection segmentation deep Network) for automatic lung infection segmentation from chest CT scans. The parallel partial decoder was exploited to aggregate the high-level features and produce the global map. In [18], a weakly supervised algorithm of deep learning was proposed for COVID-19 disease detection and lesion localization from 3D CT images. A novel CAD system for pneumonic lung disease detection was proposed in [19] using CNN and super-resolution reconstructed scans. They designed SRGAN neural network for reconstructing the images of super-resolution from the chest CT scans. In [20], authors have designed another study for automatic lung pneumonia infection detection and segmentation from the CT scans using the deep learning model.

From the above studies, we have noticed that detection of lung diseases from CT images is still a challenging problem using deep learning models. The recent methods [10]-[20] have failed to address the CT image quality improvement and hence it limits the CAD system reliability. The deep learning-based models [10]-[20] were designed using the 3D CT scans which required higher time and space requirements. In the above state-of-art methods, the high-dimensional features vector leads to significant scale variations. Therefore, it leads to a complex and error-prone training process. To end this, we proposed an HDLA model using the robust 2D CNN model for the features extraction, features scaling, and soft computing algorithms for the classification.

III. Proposed Methodology

The proposed model of automatic lung diseases classification from the CT images is shown in Fig. 1. The proposed architecture has demonstrated both training and testing functions for lung diseases detection using pre-processing, CNN features extraction, and classification. For quality improvement, each chest CT image was first pre-processed using optimal filtering and contrast adjustment techniques. The improved chest CT image was further fed to the robust CNN model for the estimation of the features using the pre-trained ResNet50 model. The pre-trained ResNet50 model effectively assists the automatic features learning from the pre-processed chest X-ray images. The testing phase shows the outcome of the pre-processing step and CNN features extraction. The result of the proposed CNN model is the features vectors of size 1 x 512. After automatic features extraction, the classification phase was launched to classify the features of input X-ray image either of classes. In the classification phase, training and test features vectors are optimized using the min-max scaling technique to overcome the challenges discussed earlier. After the features optimization, different underlying classifiers have been applied to produce their trained models and classification outcome. We considered the publicly available chest CT images in four different classes such as Health, Bacteria Induced Pneumonia (BVIP), COVID-19 Induced Pneumonia (CIP), and Viral Induced Pneumonia (VIP).

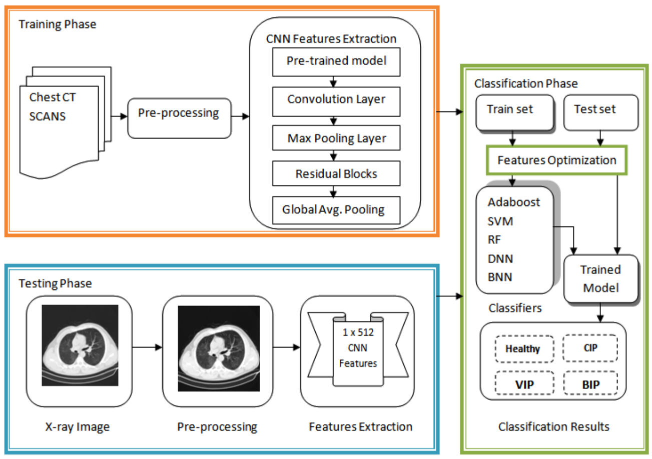

Fig. 1 Proposed CAD system for automatic CT-image based lung disease classification

A. Pre-processing

First, we standardize each input chest CT image by transforming it to a grayscale image and resizing it to 512 x 512. To balance the trade-off among the image quality improvement with minimum data loss, we applied the three functions such as contrast adjustment, Wiener filtering, and histogram equalization on input 2D X-ray image x. However, it leads to artifacts and noises in the outcome of the contrast adjustment

World Academy of Science, Engineering and Technology
International Journal of Health and Medical Engineering
Vol:17, No:4, 2023

function. Therefore, we applied the 2D Wiener filtering to produce the filtered image. We tried other filtering techniques, but the Wiener filtering produced the effective outcomes based on quality metrics Peak to Signal Noise Ratio (PSNR), Structural Similarity Index Matrix (SSIM), and Root Mean Square Error (RMSE). The Wiener filtering shows better outcomes compared to other techniques as it is an adaptive noise suppression technique. We applied Wiener filtering using the default neighborhood size N as [3, 3].

$$x2\,(i,j) = wiener2\{x\,(i,j) | (i,j) \in N\} \tag{1}$$

The $wiener2$ function first computes the local variance and mean around each pixel and then applies pixel-wise Wiener filtering to produce the filtered image x2. After that, we focus on improving the final contrast estimation of the $x2$ image by applying the histogram equalization function to produce pre-processed chest CT image $x3$.

B. Automatic Features Extraction

Considering the 2D pre-processed input chest CT image of size 512 x 512, we built the 2D CNN model based on the 2D ResNet50 pre-trained model. The ResNet50 is a CNN model with 50 deep layers trained on the ImageNet. The ResNet50 is more powerful for automatic features extraction with fast processing speed and minimum memory requirements compared to other CNN models. We transformed the input pre-processed image into 224 x 224 size as the ResNet50 takes the image of the same size as the input. To prevent the challenges of overfitting, faster features extraction, and facilitating training, we designed 2D CNN model in this paper. The detailed structure of the proposed 2D CNN model is shown in Fig. 2. It shows Max Pooling Layer (MLP) and four residual layers.

The proposed CNN takes input $x3$ and performs the squasing function using (2):

$$x^{2D}{}_j^l = tanh\left(pool_{max}\left(ReLU\left(\sum_i y_j^{l-1}(x3) * k_{ij}\right) + b_j^l\right)\right) \tag{2}$$

$$x^{CNN} = gap\,(x^{2D}{}_j^l / \,4) \tag{3}$$

where, $x^{2D}{}_j^l$ is result of 2D CNN features extraction using convolutional layer l of j^{th} input, y_j^{l-1} represents the previous convolutional layer features maps of $x3$, k_{ij} represents i^{th} trained convolutional kernels, b_j^l represents the additive bias, $\tanh(.)$ represents the activation function, $pool_{max}(.)$ represents the operation of max pooling for features extraction, $ReLU(.)$ represents the operation of the ReLU layer, and $gap\,(.)$ represents the global average pooling layer to estimate the final feature vector x^{CNN}.

Further, we applied min-max normalization technique that scales each feature into the range 0 to 1. The min-max normalization applied on x^{CNN} to produce the scaled feature vector x^{scaled} as:

$$x^{\text{scaled}} = \frac{(x^{CNN} - \min(x^{CNN}))}{(max(x^{CNN}) - \min(x^{CNN}))} \tag{4}$$

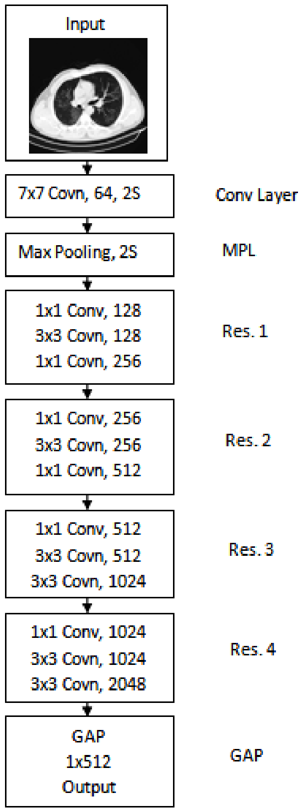

Fig. 2 Structure of proposed 2D CNN model

C. Classification

After automatic features extraction, we applied the different classifiers shown in Fig. 1 for the classification purpose. The outcome of classifiers for each input test CT scan image is either Healthy, VIP, BIP, or CIP. Each classifier was trained on 70% training dataset and used to classify the 30% test dataset. The training and testing datasets contain the scaled features for each chest CT image. We have collected the CT lung infection images from the publicy available dataset called COVIDx CT-2 of size 3900 scans [21]. It consists of 750 healthy CT images (from 80 patients), 1150 CIP CT scans (from 180 patients), 1000 BIP CT scans (from 140 patients), and 1000 VIP CT scans (from 158 patients).

IV. Simulation Results

We implemented the proposed model using the MATLAB tool with state-of-art techniques. We implemented all methods on the system equipped with Intel® Xeon® Gold 6130 processor, Central Processing Unit (CPU) of 2.10 GHz, RAM of 8 GB, and Graphical Processing Unit (GPU) of Titan RTX graphics card. We applied the proposed HDLA using different classifiers on both datasets and measure their performances in terms of accuracy, recall, precision, F1-score, and specificity parameters. Table I shows the simulation results using each classifier Random Forest (RF), Deep Neural Network (DNN), Support Vector Machine (SVM), AdaBoost, and Backpropagation Neural Network (BNN). According to Table I: (1) among all the classifiers, the DNN classifier produced the higher lung disease classification performances, and (2) the optimized CNN features achieved performance improvements compared to the raw CNN features. For all the classifiers, the proposed HDLA model has produced improved classification performances consistently ranging from 95% to 99%. Table II shows the comparative study with similar methods from [13], [16], [17], [19], and [20]. Compared to all state-of-art methods, the proposed model shows higher accuracy with minimum processing time [22]. The key reasons for such improvements are the image quality enhancement without loss of data, highly effective CNN features extraction model using the transfer learning and pre-trained ResNet50 model, and features scaling techniques.

TABLE I
SIMULATION RESULTS

Accuracy					
Classifiers	RF	SVM	BNN	Adaboost	DNN
CNN Features	93.19	93.29	92.28	94.28	96.05
Scaled CNN Features	96.91	96.09	95.31	96.49	98.29
Precision					
Classifiers	RF	SVM	BNN	Adaboost	DNN
CNN Features	94.45	93.99	92.92	94.97	96.51
Scaled CNN Features	97.29	96.68	95.89	96.98	98.34
Recall					
Classifiers	RF	SVM	BNN	Adaboost	DNN
CNN Features	95.45	95.18	94.87	96.01	97.29
Scaled CNN Features	97.48	97.01	96.84	97.29	98.81
Specificity					
Classifiers	RF	SVM	BNN	Adaboost	DNN
CNN Features	94.02	93.59	93.29	95.01	97.01
Scaled CNN Features	97.14	96.25	95.49	96.72	98.47
F1-score					
Classifiers	RF	SVM	BNN	Adaboost	DNN
CNN Features	94.97	94.48	93.64	95.38	96.72
Scaled CNN Features	97.41	96.72	96.25	97.03	98.41

TABLE II
COMPARATIVE ANALYSIS OF STATE-OF-ART METHODS

Methods	Overall Accuracy (%)	Processing Speed (Seconds)
Gunraj et al. [13]	93.14	5378
Bai et al. [16]	94.22	5982
Fan et al. [17]	95.12	6019
Tan et al. [19]	94.15	6356
Shiri et al. [20]	95.59	5413
HDLA-DNN	98.29	4782

V. CONCLUSION

In this paper, the framework HDLA was proposed for lung disease detection using deep learning and machine learning techniques from the chest CT images. The robust CNN model was designed using the pre-trained model for the 2D input image and produced the 1D feature vector with optimum kernel design. To reduce the processing speed and the memory requirements, we applied the features scaling mechanism integrated with the proposed 2D CNN model. To detect lung disease from the input chest X-ray images, we employed the different underlying classifiers. The hybrid approach of the deep learning model for automatic features extraction and classification overcomes the research problems of the state-of-art methods.

REFERENCES

[1] Elibol, E. Otolaryngological symptoms in COVID-19. *Eur Arch Otorhinolaryngol* (2020). https://doi.org/10.1007/s00405-020-06319-7.

[2] Padda, I., Khehra, N., Jaferi, U. *et al.* The Neurological Complexities and Prognosis of COVID-19. *SN Compr. Clin. Med.* 2, 2025–2036 (2020). https://doi.org/10.1007/s42399-020-00527-2.

[3] Pagliano, P., Sellitto, C., Conti, V. *et al.* Characteristics of viral pneumonia in the COVID-19 era: an update. *Infection* 49, 607–616 (2021). https://doi.org/10.1007/s15010-021-01603-y.

[4] Sun, W., Zheng, B., & Qian, W. (2016). Computer aided lung cancer diagnosis with deep learning algorithms. Medical Imaging 2016: Computer-Aided Diagnosis. doi:10.1117/12.2216307.

[5] Peng, S., Chen, J., Zhang, W., Zhang, B., Liu, Z., Liu, L., … Lv, F. (2021). The role of chest CT quantitative pulmonary inflammatory index in the evaluation of the course and treatment outcome of COVID-19 pneumonia. Scientific Reports, 11(1). doi:10.1038/s41598-021-87430-5.

[6] Shirani, F., Shayganfar, A. & Hajiahmadi, S. COVID-19 pneumonia: a pictorial review of CT findings and differential diagnosis. Egypt J Radiol Nucl Med 52, 38 (2021). https://doi.org/10.1186/s43055-021-00415-2.

[7] Mahajan, H.B., Badarla, A. & Junnarkar, A.A. CL-IoT: cross-layer Internet of Things protocol for intelligent manufacturing of smart farming. J Ambient Intell Human Comput 12, 7777–7791 (2021). https://doi.org/10.1007/s12652-020-02502-0.

[8] Mahajan, H.B., Badarla, A. Cross-Layer Protocol for WSN-Assisted IoT Smart Farming Applications Using Nature Inspired Algorithm. Wireless Pers Commun (2021). https://doi.org/10.1007/s11277-021-08866-6.

[9] Uke, N., Pise, P., Mahajan, H.B., et.al. (2021). Healthcare 4.0 Enabled Lightweight Security Provisions for Medical Data Processing. Turkish Journal of Computer and Mathematics (2021), Vol. 12, No. 11. https://doi.org/10.17762/turcomat.v12i11.5858.

[10] Zhang, K., Liu, X., Shen, J., Li, Z., Sang, Y., Wu, X., … Wang, G. (2020). Clinically Applicable AI System for Accurate Diagnosis, Quantitative Measurements, and Prognosis of COVID-19 Pneumonia Using Computed Tomography. Cell, 182(5), 1360. doi:10.1016/j.cell.2020.08.029.

[11] Zhao, W., Jiang, W. & Qiu, X. Deep learning for COVID-19 detection based on CT images. Sci Rep 11, 14353 (2021). https://doi.org/10.1038/s41598-021-93832-2

[12] Ning, W., Lei, S., Yang, J., Cao, Y., Jiang, P., Yang, Q., … Wang, Z. (2020). Open resource of clinical data from patients with pneumonia for the prediction of COVID-19 outcomes via deep learning. Nature Biomedical Engineering, 4(12), 1197–1207. doi:10.1038/s41551-020-00633-5

[13] Gunraj, H., Wang, L., & Wong, A. (2020). COVIDNet-CT: A Tailored Deep Convolutional Neural Network Design for Detection of COVID-19 Cases from Chest CT Images. Frontiers in Medicine, 7. doi:10.3389/fmed.2020.608525

[14] Wang, S., Kang, B., Ma, J., Zeng, X., Xiao, M., Guo, J., … Xu, B. (2021). A deep learning algorithm using CT images to screen for Corona virus disease (COVID-19). European Radiology. doi:10.1007/s00330-021-07715-1

[15] Yang, Y., Lure, F. Y. M., Miao, H., Zhang, Z., Jaeger, S., Liu, J., & Guo, L. (2020). Using artificial intelligence to assist radiologists in distinguishing COVID-19 from other pulmonary infections. Journal of X-Ray Science and Technology, 1–17. doi:10.3233/xst-200735

[16] Bai, H. X., Wang, R., Xiong, Z., Hsieh, B., Chang, K., Halsey, K., … Liao, W.-H. (2020). Artificial Intelligence Augmentation of Radiologist Performance in Distinguishing COVID-19 from Pneumonia of Other Origin at Chest CT. Radiology, 296(3), E156–E165. doi:10.1148/radiol.2020201491

[17] Fan, D.-P., Zhou, T., Ji, G.-P., Zhou, Y., Chen, G., Fu, H., … Shao, L. (2020). Inf-Net: Automatic COVID-19 Lung Infection Segmentation from CT Images. IEEE Transactions on Medical Imaging, 39(8), 2626–2637. doi:10.1109/tmi.2020.2996645

[18] Wang, X., Deng, X., Fu, Q., Zhou, Q., Feng, J., Ma, H., … Zheng, C. (2020). A Weakly-supervised Framework for COVID-19 Classification and Lesion Localization from Chest CT. IEEE Transactions on Medical Imaging, 1–1. doi:10.1109/tmi.2020.2995965

[19] Tan, W., Liu, P., Li, X., Liu, Y., Zhou, Q., Chen, C., … Zhang, Y. (2021). Classification of COVID-19 pneumonia from chest CT images based on reconstructed super-resolution images and VGG neural network. Health

Information Science and Systems, 9(1). doi:10.1007/s13755-021-00140-0

[20] Shiri, I., Arabi, H., Salimi, Y., et.al. (2021). COLI-Net: Deep learning-assisted fully automated COVID-19 lung and infection pneumonia lesion detection and segmentation from chest computed tomography images. Int J Imaging Syst Technol. 2021;1–14. 10.1002/ima.22672.

[21] https://www.kaggle.com/hgunraj/covidxct

[22] Mahajan, H.B., Uke, N., Pise, P. et al. Automatic robot Manoeuvres detection using computer vision and deep learning techniques: a perspective of internet of robotics things (IoRT). Multimed Tools Appl (2022). https://doi.org/10.1007/s11042-022-14253-5.