

Intelligent Deep Learning System for Enhanced Pulmonary Disease Diagnosis Through Five-Class Mode



Bahaa D. Jalil^{1*}, Mohammed A. Noaman Al-Hayanni²

¹Control and Systems Engineering Department, University of Technology-Iraq, Baghdad 10066, Iraq ²Electrical Engineering Department, University of Technology-Iraq, Baghdad 10066, Iraq

Corresponding Author Email: bahaa.d.jalil@uotechnology.edu.iq

Copyright: ©2024 The authors. This article is published by IIETA and is licensed under the CC BY 4.0 license (http://creativecommons.org/licenses/by/4.0/).

https://doi.org/10.18280/ria.380413

ABSTRACT

Received: 26 December 2023 Revised: 20 February 2024 Accepted: 31 March 2024 Available online: 23 August 2024

Keywords:

chest X-ray, CNNs, COVID-19, pulmonary diseases, deep learning, classification

The respiratory system's diseases, including many disorders that impair lung function and cause respiratory distress, are a significant public health issue. Many other etiological variables contribute to these disorders, such as genetic predisposition, smoking, infections, and exposure to environmental risks. To lessen the effects of these illnesses, prompt diagnosis and efficient therapy approaches are essential. This work presents a sophisticated lung disease diagnosis system based on the latest Deep-Learning (DL) models. The Gerry model, which use a Convolutional Neural Network (CNN) classification model, is being expanded to include four classes for lung disease. The proposed methodology demonstrates a substantial enhancement in accuracy, ranging from 0.432% to 1.621%, while concurrently reducing loss by 100% to 138%. CNN extends the procedure to incorporate a five-class model, which effectively differentiates between COVID-19, lung fibrosis, lung opacity, normal cases without anomalies, and pneumonia. We use a 22,851 Chest X-ray (CXR) image dataset to train, validate, and test the model. The resulting model has an impressive 92% overall accuracy. The following are the reported f1-scores, precision, and recall for each class: 91%, 89%, and 93% for lung opacity; 92%, 96%, and 93% for standard cases; 85%, 73%, and 78% for lung fibrosis; and 96%, 99%, and 97% for pneumonia. By this diagnostic method and with the aid of precise detection and categorization of various lung disorders, patient outcomes, and clinical decision-making can be potentially improved.

1. INTRODUCTION

Millions of deaths annually occur because of Respiratory illnesses. CXR scans are frequently employed as a first step in the early detection and treatment procedure. Artificial Intelligence (AI) making a revolution in the medical field, particularly in dealing with medical image analysis. Deep Learning (DL), which is one of the AI's techniques, has proven to be effective in pattern detection and feature extraction from images. This advances the diagnosing of diseases efficiently.

Transfer learning has shown success in the field of medical image analysis by utilizing DL models which are previously trained on a large dataset and then adapting them to specific tasks. It strongly addresses the challenge caused by limited labeled data. As continuous progression in AI, its potential to enhance outcomes and make a revolution in the healthcare field is significant [1-4].

Convolutional Neural Network (CNN) are a kind of DL algorithm. It is efficient in dealing with visual tasks. CNN showed success in many applications like image classification, natural language processing, and object detection [5, 6]. CNN architecture consists of three main layers: the convolution layer, the pooling layer, and the connected layer [7]. For a specific classification, settings of CCN layers with experiments and evaluation by using validation datasets. Utilizing learning rate schedules, weight initialization

procedures, and batch normalization for improved outcomes can improve the accuracy and generalization of the CNN model.

In addition, the diversity of respiratory disorders can create challenges for diagnosis, research, public health planning, and treatment options. Thus, designing a model with more classes that cover more diseases is a good option to solve this obstacle, such as Gerry's four-class model [8].

To cover these limitations this study satisfies two contributions.

1. Enhance CNN Gerry's four-class model from a previous study.

2. Design a model to classify types of CXR images, including COVID-19, lung opacity, pneumonia, and normal cases. Additionally, we include a class for identifying lung fibrosis, resulting in five categories.

2. RELATED WORK

Classifying lung CXR images automatically has proven challenging due to the complexity of identifying infectious and inflammatory lung diseases.

In a study conducted by Saeed and Alwawi [9], they successfully created a COVID-19 classification DL model using CNN technology. This model was trained on CXR

images, and it demonstrated an accuracy rate of 96.57% for the training dataset while achieving 92.29% accuracy for the validation dataset. Alshmrani et al. [10] employed a VGG19 model that had been pre-trained. They then utilized three blocks of CNN to categorize CXR scans into different groups, including pneumonia, lung cancer, tuberculosis, lung opacity, and COVID-19. Their approach showed an accuracy rate of 96.48%. Xu et al. [11] developed a 3D deep learning model to distinguish 618 images into three groups: COVID-19, Influenza A Viral Pneumonia (IAVP), and healthy an accuracy of 86.7% had achieved. Farhan and Yang [12] introduced a study in which a Hybrid Deep Learning Algorithm (HDLA) framework was developed and CXR images were categorized into normal, pneumonia, and COVID-19. Alwawi and Abood [13] employed a DL CNN structured model for classifying a dataset of CXR images into infected or uninfected with COVID-19. With the addition of an expanded dataset, the model achieved a 93.8% accuracy rate for the training set while 92.1% achieved for the validation set. To identify Chronic Fibrosing Interstitial Lung Diseases (CF-ILD) accurately from a set of chest radiograph images, Nishikiori et al. [14] proposed and tested a learning algorithm that resulted in a detection accuracy of 97.9%. KPL et al. [15] utilized machine learning algorithms to develop a web-based application. This application tested CXR scan images to predict the presence of COVID-19, tuberculosis, pneumonia, or COPD. Bharati et al. [16] introduced a learning framework. They combined VGG, data augmentation, and Spatial Transformer Network (STN) with CNN in that framework. When they applied the framework to the NIH CXR images dataset, 73% validation accuracy was achieved. Priyadarsini et al. [17] proposed sequential, functional, and transfer deep learning models. They utilized these three DL models to classify a dataset of three classes (tuberculosis, cancer, and pneumonia). The sequential model achieved a recall rate of 96.33%, an F1-score of 98.55%, and an accuracy rate of 98.43% as a result of pneumonia classification. Similarly, the sequential model fared well in the classification of tuberculosis with a recall rate of 98.88%, an F1-score of 97.99%, and an accuracy rate of 99.4%, whereas the functional model showed accuracy with a rate of 99.9% for cancer classification.

It is highlighted in the related work mentioned above that most models tailed specific kinds of lung diseases depending on the adopted datasets. While in this work, five different diseases have been classified using five classes deep learning model. Moreover, two datasets have been combined to produce high band learning facility that leads to obtain more accurate classification amongst five different diseases. Therefore, the most outperformed point of the proposed model is the accurate classification among very close result diseases.

3. PROPOSED LUNG DISEASE DIAGNOSING SYSTEM

The proposed system for classifying lung diseases is meant to help organize lung diseases according to CXR images. The system's implementation makes use of the transfer learning technique. The structure of the system, the dataset, and the DL model used will all be covered in this section.

3.1 System structure

The overall methodology for the proposed deep learning

CNN model for classifying five classes of lung diseases begins by acquiring as many authentic, accurately labeled CXR images as possible. Then, partition them into three parts: training, validation, and testing datasets. After that, the model is trained by the training and validation datasets so it can learn based on these datasets. Once the model completes the learning process, it will be ready to be fed by the testing dataset to evaluate its learning and produce its results, which is the ability to classify CXR images into one of the five classes: COVID-19 positive cases, Normal (healthy), Viral Pneumonia, Lung Opacity, and fibrosis). The general steps of the system work are outlined in Figure 1.



Figure 1. Overall system structure

3.2 Dataset

This research used digital CXR images obtained from two repositories on Kaggle. First, (Data-A) is a comprehensive labeled CXR image developed by researchers from Qatar University, Dhaka University, Pakistan, and Malaysia in collaboration with experts. This extensive dataset covers four categories: COVID-19 cases, Normal cases, Viral Pneumonia cases, and Lung Opacity cases. The dataset has been updated many times, with the recent update two years ago [18]. Second, the National Institutes of Health (NIH) Chest X-rays labeled dataset was introduced by NIH [19].

This study involves two phases. Phase 1 is concerned with improving Gerry's four-class lung disease classification model [8], and Data-A is used in the training process. In contrast, Phase 2 expands the model to include a fifth class, fibrosis, by extracting all CXR fibrosis images from the NIH chest X-ray dataset. This expanded dataset, Data-B, is then used to train the proposed five-class classification model.



Figure 2. Chest X-ray images

Data-A consists of 3616 images (each with dimensions of 299×299 pixels) related to COVID-19, 6012 images showcasing lung opacity, 10192 images representing normal cases, and 1345 images depicting viral pneumonia. In total, there are 21165 images distributed among these four categories. On the other hand, Data-B encompasses all the Data-A images plus 1686 high-resolution images (with dimensions of 1024×1024 pixels) showcasing fibrosis. This results in a combined dataset of 22851 images across the five classes. Figure 2 presents samples of CXR images of the dataset, and Table 1 shows the five categories of this study and how many images are in each category.

Table 1. Number of images in each class

Class Name	No. of Images
COVID	3616
Lung Opacity	6012
Normal	10192
Viral Pneumonia	1345
Fibrosis	1345

As a preprocessing step, all images are resized to 128×128 and distributed randomly along three datasets: training, validation, and testing. Since the work passes through two phases, one of four classes is called Phase 1, and the other of five classes is called Phase 2. In phase 1, the training dataset consists of 19048 images. And, the validation and testing datasets comprised 1059 and 1058, respectively. While, phase 2 includes 20565 images for the training dataset, 1143 for the validation dataset, and 1143 for the testing dataset. Table 2 illustrates the number of all dataset images in each part of training, validation, and testing, as well as the ratio of each to the total number of images in the dataset. As a preprocessing step, all the images are resized to 128×128 pixels and then randomly spread into three datasets: training, validation, and testing. The number of images in each varied depending on the two phases of work. For Phase 1, 19048 images belong to the training dataset, while the validation and testing datasets had 1059 and 1058 images, respectively. In contrast, phase 2 datasets are 20565 images in the training dataset and 1143 in both validation and testing datasets.

Table 2.	Division	of training,	validation,	and testing	datasets
----------	----------	--------------	-------------	-------------	----------

	Phase 1	Phase 2
Total dataset	21165	22851
Training dataset	19048	20565
Training dataset ratio	90%	90%
Validation dataset	1059	1143
Validation dataset ratio	5%	5%
Testing dataset	1058	1143
Testing dataset ratio	5%	5%

3.3 Deep learning model

This system uses a transfer learning approach by utilizing a pre-trained InceptionResNetV2 Keras model as the foundation for its implementation. Transfer learning is one of the machine learning techniques that utilizes knowledge gained by learning one previous problem to improve the performance of a new, similar one. A common approach in DL involves leveraging pre-trained model layers, preserving their acquired knowledge by freezing them, introducing additional trainable layers to tackle the specific task, and then training these new layers using the target dataset. Keras Applications provide users with pre-trained DL models with their respective weights. These models make it easier to perform tasks such as prediction feature extraction and fine-tuning for various applications [20, 21].

The InceptionResNetV2 model, a Keras model used for image classification, allows the option to load pre-trained weights from ImageNet if desired. These weights were obtained through pre-training on the ImageNet dataset [22]. InceptionResNetV2's size is 215 megabytes, contains 55.9 million parameters, and possesses a topological depth of 449 layers, encompassing activation and batch normalization layers [21]. The DL model proposed in this study utilizes the InceptionResNetV2 architecture without its top layer. As shown in Figure 3, the model takes three channels of CXR images as input data. We resized all dataset images during preprocessing to a standardized dimension of 128×128×3. The pre-trained, top layer excluded InceptionResNetV2 model is utilized for feature extraction. Then, a set of layers is incorporated for classification purposes. These include a BatchNormalization layer, a layer containing 256 neurons with ReLU activation, a dropout layer with a rate of 0.45, and a dense layer consisting of 5 neurons with Softmax activation. Each neuron corresponds to one of the classes in the classification task.

The model uses an Adamax optimizer during the training and validation process. It starts with a learning rate of 0.002, adjusted dynamically as the learning process progresses. Try and error technique is used to tune hyperparameters till get the values mentioned.

The suggested model has 54,737,637 parameters, partitioned into trainable parameters (54,674,021 parameters) and non-trainable parameters (63,616 parameters). Details on the tunable parameters of the model are provided in Table 3.



Figure 3. Architecture of the model

Table 3. The proposed model's tuned parameters

Parameters	Values
Dropout rate	0.45
Learning rate	0.002
Optimizer	Adamax
Total params	54,737,637
Trainable params	54,674,021
Non-trainable params	63,616

4. RESULTS

Python 3 and the Keras framework have been utilized in the implementation of the work of this study by using Acer Nitro AN515-58, with Core (TM) i9-12900H 2.90 GHz Intel(R) CPU, 32 GB RAM, and 8 GB RTX 4060 Laptop GPU.

This study involves two phases. The first phase (Phase 1) is concerned with enhancing Gerry's 4-class lung disease classification model [8], whereas Phase 2 is concerned with the proposed 5-class classification model. Thus, this section shows the results obtained from the two phases.

4.1 Phase 1

We suggest improvements for Gerry's 4-class lung disease DL classification model in this phase. The suggested improvements include modifying the first fully connected layer to 512 neurons and removing the dropout layer yielded a 0.5% to 1% gain in the loss and accuracy performance of the model at the end of 15 epochs. Figure 4 shows the loss and accuracy performance curves for both Gerry's and the enhanced model. A statistical comparison of the two models is presented in Table 4, which shows the improvements in accuracy in our enhanced model, 0.432%, 1.621%, and 0.999% for training, validation, and testing datasets respectively. In addition, our model decreases the loss ratio by 108%, 138.5%, and 100% for training, validation, and testing datasets respectively.



Figure 4. Loss and accuracy for Gerry's model and the enhanced model

 Table 4. Comparison of the two models (Loss & Accuracy)

	Training Dataset	Validation Dataset	Testing Dataset
Loss (Gerry's model)	3.08%	7.285%	5.909%
Loss (enhanced model)	1.48%	3.054%	2.945%
Accuracy (Gerry's model)	99.533%	93.201%	94.05%
Accuracy (enhanced model)	99.963%	94.712%	94.99%
Accuracy improvement ratio	0.432%	1.621%	0.999%
Loss improvement ratio	108%	138.5%	100%

Regarding precision, recall, and f1-score metrics, Table 5 compares the two models numerically, whereas Figure 5 shows the confusion matrix.

 Table 5. Precision, Recall, and F1-score of the two four-class models (GM refers to Gerry's Model and EM refers to the Enhanced model)

Disease	Precision	Recall	F1-Score
COVID-19 (GM)	97%	96%	97%
COVID-19 (EM)	99%	99%	99%
Lung Opacity (GM)	90%	93%	91%
Lung Opacity (EM)	90%	93%	92%
Normal (healthy) (EM)	95%	94%	94%
Normal (healthy) (GM)	96%	95%	96%
Viral Pneumonia (GM)	100%	97%	98%
Viral Pneumonia (EM)	96%	96%	96%
Average (GM)	96%	95%	95%
Average (EM)	95%	96%	96%
Improvement ratio of Average	-1%	1%	1%
Improvement ratio of COVID-19	2%	3%	2%
Improvement ratio of Lung Opacity	0%	0%	1%
Improvement ratio of Normal (healthy)	1%	1%	2%
Improvement ratio of Viral Pneumonia	-4%	-1%	-2%



(b) Enhanced model



Table 5 shows the calculations of the average for Gerry's model and our enhanced model. There are improvements in the Recall and F1-score with a ratio of 1% for each. While there is a drawback in precision with a ratio of -1%. In addition, the calculations show improvements in all performance metrics for COVID-19, Lung Opacity, and Normal (healthy). While the drawback in performance appears in the ratio of Viral Pneumonia.

The model shows progress in accuracy, loss ratios, precision, recall, and F1-score for three classes. However, it struggles with a specific class, indicating potential issues like a lack of training data or class distribution imbalance. And highlighting the need for continuous research and development. Future research should focus on data augmentation, model architecture enhancement, and advanced ensemble techniques. Integrating transfer learning findings and domain adaptation could improve the model's flexibility and effectiveness across other instructional domains. Addressing this issue involves incorporating methodologies like data augmentation, model architecture adjustments, and hyperparameter optimization.

4.2 Phase 2



Figure 6. Proposed model performance

This phase marked the implementation of the proposed 5class lung diseases CNN model. Data-B was employed for training, validation, and testing purposes. Despite using only 30 epochs and a batch size of 40, the model achieved remarkable results with a dynamically adjusting learning rate starting at 0.002. The performance of the model was evaluated using loss and accuracy for training, validation, and testing datasets, as well as precision, recall, and F1-score. Furthermore, a confusion matrix was generated to evaluate the classification performance of the proposed model. Figure 6 and Table 6 show the loss and accuracy for the datasets, whereas Table 7 shows the recall, f1-score, and precision, and Figure 7 shows the confusion matrix.

Table 6. Loss and accuracy obtained

	Training Dataset	Validation Dataset	Testing Dataset
Loss	1.98%	4.3986%	4.395%
Accuracy	99.93%	92.30%	92.04%

 Table 7. Performance metrics values

	Precision	Recall	F1-Score
COVID-19	98%	96%	97%
Fibrosis	85%	73%	78%
Lung Opacity	91%	89%	90%
Normal (healthy)	92%	96%	93%
Viral Pneumonia	96%	99%	97%



Figure 7. Proposed model confusion matrix

5. CONCLUSION

Advanced learning and other forms of intelligence methods such, as CNN have become tools for categorizing diseases especially in the field of medical imaging. Using a CNN model based on the trained InceptionResNetV2 structure has shown excellent results in identifying 5 different lung diseases from CXR images. The model achieved a training accuracy of 99.932%. Despite a number of trainings epochs, the model maintained validation and testing accuracies at 92.301% and 92.04% respectively. These results highlight the potential of learning techniques in the diagnosis of lung diseases by providing effective X ray analyses. Expanding the evaluation of the model by testing it on datasets, including CT scans and utilizing datasets to boost its learning capabilities are promising line for future research. Additionally exploring models within Keras could lead to enhancements, in the performance of the model.

Our improved CNN models show encouraging results in controlled environments; however, they face several obstacles when applied in real-world scenarios. These factors encompass managing the fluctuation and intricacy of realworld data, guaranteeing the ability to efficiently process large amounts of data, and addressing ethical and societal concerns related to fairness, transparency, and privacy. In addition, preserve model performance through ongoing monitoring and adjustment. Addressing these problems necessitates doing rigorous testing to assess the model's robustness, optimizing its efficiency, adhering to ethical norms, and maintaining it consistently to assure its reliability, efficacy, and ethical integrity in real-world deployment settings.

REFERENCES

- Goldberg, J.E., Rosenkrantz, A.B. (2019). Artificial intelligence and radiology: A social media perspective. Current Problems in Diagnostic Radiology, 48(4): 308-311. https://doi.org/10.1067/j.cpradiol.2018.07.005
- [2] Alzubaidi, L., Fadhel, M.A., Al-Shamma, O., Zhang, J., Duan, Y. (2020). Deep learning models for classification of red blood cells in microscopy images to aid in sickle cell anemia diagnosis. Electronics, 9(3): 427. https://doi.org/10.3390/electronics9030427
- [3] Alzubaidi, L., Al-Amidie, M., Al-Asadi, A., Humaidi, A.J., Al-Shamma, O., Fadhel, M.A., Duan, Y. (2021). Novel transfer learning approach for medical imaging with limited labeled data. Cancers, 13(7): 1590. https://doi.org/10.3390/cancers13071590
- [4] Alzubaidi, L., Al-Shamma, O., Fadhel, M.A., Farhan, L., Zhang, J., Duan, Y. (2020). Optimizing the performance of breast cancer classification by employing the same domain transfer learning from hybrid deep convolutional neural network model. Electronics, 9(3): 445. https://doi.org/10.3390/electronics9030445
- [5] Madhavi, M., Supraja, P. (2022). COVID-19 infection prediction from CT scan images of lungs using Iterative Convolution Neural Network model. Advances in Engineering Software, 173: 103214. https://doi.org/10.1016/j.advengsoft.2022.103214
- [6] Purwono, P., Ma'arif, A., Rahmaniar, W., Fathurrahman, H.I.K., Frisky, A.Z.K., Haq, Q.M.U. (2023). Understanding of convolutional neural network (CNN): A review. International Journal of Robotics and Control Systems, 2(4): 739-748. https://doi.org/10.31763/ijrcs.v2i4.888
- [7] Uthaib, M.A., Croock, M.S. (2021). Multiclassification of license plate based on deep convolution neural networks. International Journal of Electrical and Computer Engineering, 11(6): 5266-5276. http://doi.org/10.11591/ijece.v11i6.pp5266-5276
- [8] Gerry. Lung Disease F1 score ~91%. (2023). https://www.kaggle.com/code/gpiosenka/lung-diseasef1-score-91/.
- [9] Saeed, R.S., Alwawi, B.K.O.C. (2023). A binary classification model of COVID-19 based on convolution neural network. Bulletin of Electrical Engineering and

Informatics, 12(3): 1413-1417. https://doi.org/10.11591/eei.v12i3.4832

- [10] Alshmrani, G.M.M., Ni, Q., Jiang, R., Pervaiz, H., Elshennawy, N.M. (2023). A deep learning architecture for multi-class lung diseases classification using chest Xray (CXR) images. Alexandria Engineering Journal, 64: 923-935. https://doi.org/10.1016/j.aej.2022.10.053
- [11] Xu, X., Jiang, X., Ma, C., Du P., Li X., Lv S., Yu, L., Ni, Q., Chen Y., Su, J., Lang, G., Li Y., Zhao H., Liu J., Xu K., Ruan L., Sheng J., Qiu Y., Wu W., Liang T., Li L. (2020). A deep learning system to screen novel coronavirus disease 2019 pneumonia. Engineering, 6(10): 1122-1129. https://doi.org/10.1016/j.eng.2020.04.010
- [12] Farhan, A.M.Q., Yang, S. (2023). Automatic lung disease classification from the chest X-ray images using hybrid deep learning algorithm. Multimedia Tools and Applications, 82: 38561-38587. https://doi.org/10.1007/s11042-023-15047-z
- [13] Alwawi, B.K.O.C., Abood, L.A. (2021). Convolution neural network and histogram equalization for COVID-19 diagnosis system. Indonesian Journal of Electrical Engineering and Computer Science, 24(1): 420-427. http://doi.org/10.11591/ijeecs.v24.i1.pp420-427
- [14] Nishikiori, H., Kuronuma, K., Hirota, K., Yama, N., Suzuki, T., Onodera, M., Onodera, K., Ikeda, K., Mori, Y., Asai, Y., Takagi, Y., Honda, S., Ohnishi, H., Hatakenaka, M., Takahashi, H., Chiba, H. (2023). Deeplearning algorithm to detect fibrosing interstitial lung

disease on chest radiographs. European Respiratory Journal, 61: 2102269. http://doi.org/10.1183/13993003.02269-2021

- [15] KPL, S., CS, N., AM, J., Pavani P. (2023). Detection and classification of lung diseases using machine and deep learning techniques. Journal of Computer Science and Software Development, 2: 1-10.
- [16] Bharati, S., Podder, P., Mondal, M.R.H. (2020). Hybrid deep learning for detecting lung diseases from X-ray images. Informatics in Medicine Unlocked, 20: 100391. https://doi.org/10.1016/j.imu.2020.100391
- [17] Priyadarsini, M.J.P., Rajini, G.K., Hariharan, K., Raj, K.U., Ram, K.B., Indragandhi, V., Pandya, S. (2023). Lung diseases detection using various deep learning algorithms. Journal of Healthcare Engineering, 2023. https://doi.org/10.1155/2023/3563696
- [18] Viradiya, P. (2023). COVID-19 Radiography Dataset. https://www.kaggle.com/datasets/preetviradiya/covid19 -radiography-dataset/.
- [19] National Institutes of Health. (2023). Random Sample of NIH Chest X-ray Dataset. https://www.kaggle.com/nihchest-xrays/sample.
- [20] Keras. (2023). Transfer Learning & Fine-Tuning. https://keras.io/guides/transfer_learning/.
- [21] Keras. (2023). Keras Applications. https://keras.io/api/applications/.
- [22] Keras. (2023). InceptionResNetV2. https://keras.io/api/applications/inceptionresnetv2/.