Enhancing Pneumonia Disease Detection and Prediction Model Using Deep Learning Approaches

Nigusu Yitayal, Atinkut Molla, Tsehay Admasu, Yenework Belayneh
Departement of Information Technology
Injibara University, Ethiopia

Abstract

The lung is the primary organ of the respiratory system in humans. There are several types of lung diseases. Some of the lung disease types are Tuberculosis, pneumonia, lung cancer, and Covid-19. Among lung diseases, pneumonia is a dangerously affects the lungs of humans. Early diagnosis of pneumonia is crucial to ensure curative treatment and increase survival rates. Chest X-ray imaging is the most frequently used method for diagnosing pneumonia. However, in the manual inspection of chest X-rays, the radiologist leads to a prolonged diagnosis process and undesirable results. This prolonged diagnosis process and the high-end expertise of radiologists can be minimized with the help of Artificial intelligence and machine learning to automate the initial detection (screening) of pneumonia cases to expedite the diagnosis process. The objective of this study is to develop a pneumonia prediction model with a deep learning approach (ResNet) using chest X-ray images collected at Injibara General, Bahir Dar (Felege Hiwot, and Tibebe Gion), and Addis Ababa (Tikur Anbesa) hospitals. The steps involved in the study include preprocessing, feature extraction and and classification. The developed model consisted of several convolutional layers with max pooling and ReLU activation. The ResNet model is used to detect and classify the chest X-ray images as normal or pneumonia. The performance of the model was evaluated by using accuracy, precision, recall, and F1-score. The proposed model (ResNet) with HOG feature extraction has scored 86.2% accuracy using 213 chest X-ray images.

Keywords: Lung disease, Pneumonia, Deep learning

1. Introduction

The lung is the main organ of the human body that is relevant to the breathing system. The lung is a pair of spongy, air-filled organs located on either side of the chest (thorax). The trachea (windpipe) conducts inhaled air into the lungs through its tubular branches, called bronchi. The bronchi divide into smaller and smaller branches (bronchioles) that are microscopic [1]. Different lung diseases affect the lung. Lung disease is a problem in the lungs that prevents the lung from functioning. The main lung disease types are Tuberculosis, pneumonia, lung cancer, and Covid-19 [2].

According to the Forum of International Respiratory Societies, 334 million people suffer from asthma, and each year, tuberculosis kills 1.4 million people, 1.6 million people die by lung cancer, and pneumonia kills millions of people [3]. Lung diseases are one of the leading causes of death and disability in this world. Early detection plays a key role in increasing the chances of recovery and improves long-term survival rates. Traditionally, lung disease is detected with skin test, blood test, sputum sample test, chest X-ray examination, and computed tomography (CT) scan examination.

Lung disease detection and prediction in hospitals is a difficult task for proper patient treatments. The types of lung disease detection and prediction model assist pulmonologists in treating and finding solutions for their patients easily in hospitals. Among these, pneumonia is a type of lung disease that spreads quickly and poses a serious risk to the health and well-being of its victims. All over the world, pneumonia is at the top of those diseases that are causing death over the world. Every year more than 50000 people mostly old age die because of pneumonia [4]. An accurate biomedical diagnosis of pneumonia necessitates the use of various diagnostic tools and the evaluation of various clinical features, all of which were hindered by the lack of available experts and tools. Medical specialists need complete knowledge of that appropriate domain to find out the abnormality that is present in human beings.

Artificial intelligence has proven to be an effective way in the detection of many diseases. The recent development of deep learning supports the identification and classification of pneumonia diseases in lung images. Hence, numerous works on the detection of pneumonia using deep learning are relevant to identify disease from chest X-ray images. Pneumonia disease is one of the killer diseases in the world, especially in developing countries; the detection of pneumonia disease models in previous times is a challenge to analyze and isolate and humans have faced a cure for lung distortion still now. Lung diseases also known as respiratory diseases are diseases of the airways, and the other structures of the

Copyright © 2024, Infonomics Society | DOI: 10.20533/ijmip.2042.4647.2024.0071

lungs [5].

Currently, deep learning has vital roles when applied to lung images for pneumonia disease detection and prediction. Deep learning is a part of machine learning relating to algorithms inspired by the function and structure of the brain. Recent developments in machine learning, especially deep learning, support the identification, quantification, and classification of patterns in lung image datasets [6]. These developments were made possible due to the ability of deep learning to learn features merely from data, instead of hand-designed features based on domain-specific knowledge. Deep learning is quickly becoming state of the art, leading to improved performance in numerous lung applications. Consequently, these advancements assist clinicians in detecting and classifying pneumonia disease effectively [7].

Therefore, this study focused on proposing a pneumonia detection and prediction model using Residual Neural Network (ResNet) using chest X-ray image datasets to overcome the prolonged diagnosis process and treat patients. This study presented the detection, classification, and visualization of pneumonia using chest X-ray image datasets collected from hospitals.

2. Problem Statement

The detection and prediction of pneumonia disease types in hospitals is a challenging task to treat and provide services for pneumonia patients. Pneumonia is usually identified by extracting unique features from a chest X-ray. The major problems were described as follows concerning disease types, datasets, and models.

Some researchers used a private inadequate number of pneumonia datasets. To obtain larger datasets, efforts such as the identification of confidential patients' information are conducted to make the pneumonia data available to the public. With an adequate and diverse pneumonia dataset, the performance of deep learning would be enhanced. The diversity of the pneumonia dataset decreases classifier generalization error because the model generalizes more accurately when it is trained on more pneumonia samples. The pneumonia dataset is difficult to collect from publically available datasets. Therefore, if the datasets were made public, more phenomena datasets would be available for researchers [8].

According to the World Health Organization (WHO) [9] report, a radiologist still does not diagnose 0.66% of the population around the world. Additionally, fatigue and the concentration of radiologists can affect pneumonia diagnosis. Inspection of a large number of chest X-rays daily can be exhausting and leads to wrong diagnosis. The availability of radiologists at all times is prohibitive. To address the problems associated with manual inspections of chest X-rays, automated systems are leveraging the power of Artificial Intelligence (AI) to detect patterns in the data to assist radiologists in inspecting chest X-rays [10].

Most extracted used automatically selected pneumonia disease features extracted by Conventional Neural Network (CNN). In literature, features such as Scale-Invariant Feature Transform (SIFT), Gastrointestinal Stromal Tumors (GIST), Gabor, Low Back Pain (LBP), and High Osmolality Glycerol (HOG) were studied. However, many other features are still yet to be explored, for example image histogram. Efforts are directed at studying different types of features. The use of more features such as image histogram addresses the issue of high correlation of errors when using features extracted by CNN. When combining many variations, the result is better [11].

In conclusion, a Deep learning approach (ResNet) was proposed to overcome the challenges of identifying and predicting pneumonia using real chest X-ray images collected from various hospitals.

3. Literature Review

Pneumonia has become one of the most prevalent respiratory diseases in recent years [12]. Consequently, machine learning and deep learning models have been widely researched for the detection of phenomena. A research article [13] proposed a hybrid model with machine learning (Random forest) and deep learning for the automatic detection of pneumonia. According to S. Sharma and K. Guleria [14], developed a deep-learning model for the detection of pneumonia. The study trained a deep learning algorithm on chest X-rays to develop a model that detects pneumonia. The study also compared support vector machines, random forests, and K-nearest neighbors for the automated detection of pneumonia in chest X-ray images. The result has demonstrated that the deep learning model achieves a higher accuracy score of 92.15% compared to machine learning models investigated in the study.

Several researchers [15] studied deep learning models for the detection of pneumonia from chest X-ray images. Even though many machine and deep learning models have achieved higher accuracy, these models have scope for improvement for practical applications. One of the challenges of developing an accurate model for the automated detection of pneumonia from chest X-ray images is the inadequacy of the dataset.

In the study [16], Pneumonia is the cause of death in children. The article stated NASNet (Neural Architecture Search Network), a deep learning model pre-trained on ImageNet to predict pneumonia in patients at an early stage using chest X-rays. In the experiment, 112,200 Chest X-ray lung images were used from 30,800 patient datasets. The training

Copyright © 2024, Infonomics Society | DOI: 10.20533/ijmip.2042.4647.2024.0071

dataset has 97,748 images from 27,833 patients, while the validation dataset contains 7442 images from 1763 patients and the test dataset contains 531 images from 400 patients.

According to [17], Deep Neural network architecture was proposed to predict pneumonia from an X-ray image dataset for lung disease diagnosis. The proposed model was trained and evaluated on 5856 labeled images of which 4273 showed pneumonia, while the remaining 1583 were negatives from one of Kaggle's datasets. Additionally, Chest X-rays help detect all lung information to locate any abnormalities that may be present in the lungs due to the scarcity of radiologists.

The study in [18] illustrated that artificial intelligence (AI) has become an effective tool for the early diagnosis of pneumonia disease in the chest X-ray lung data. Several machine learning, deep learning, and transfer learning techniques are utilized to diagnose pneumonia lung diseases. This research article describes up-to-date analysis of AI studies that use data comparison to identify pneumonia disease in X-rays and identify their limits to offer advice to practitioners. Likewise, M. Trivedi and A. Gupta [19] proposed a 'Mobile Net' deep learning model to detect pneumonia automatically based on chest X-ray lung images. In the study, 5856 chest X-ray images were taken to train, test, and evaluate the proposed deep-learning network. The proposed model was trained and achieved a training accuracy of 97.34%, a validation accuracy of 87.5%, and a testing accuracy of 94.23% for the automatic detection of pneumonia.

According to D.K. Jain et al, [20] deep neural networks, convolutional neural networks, and transfer learning were applied in the experiment to identify pneumonia lung disease from chest X-rays. The transfer learning presented accurate experimental results that do not require a large dataset. Experimental results also alleviated concern regarding the overlap between diagnoses of COVID-19 and regular pneumonia. Deep learning models GoogLeNet and AlexNet chest X-ray datasets were applied for the experiment analysis.

The research included in the literature review used image processing and various deep-learning approaches. Moreover, in this research, we designed and developed a deep Residual Neural Network (ResNet) along with HOG features to detect and classify chest X-ray images as Normal or Pneumonia cases.

4. Methodology

In the study, the research method combined quantitative and design science approaches. In the study, the experimental laboratory conducted in quantitative approach. The design and implementation artifacts or models are conducted in a design science approach. The enhancement of pneumonia disease detection and prediction model implemented by using deep learning.

In the study, the sampling dataset size was determined by using a non-probabilistic judgmental sampling method. The first sample pneumonia Chest X-ray image dataset was collected from government hospitals with three selected cities namely Injibara General, Bahirdar Felege Hiwot Referral Tibebe Gihon specialized, and Addis Ababa Tikur Anbesa specialized hospitals. These cities are selected based on dataset availability for pneumonia Chest X-ray image dataset. The domain experts (radiologists) identify each sampled dataset class as normal and pneumonia before going to the data preprocessing stage.

The developed system has several parts: Data collection, preprocessing (histogram analysis, morphological processing, image augmentation and noise removal), feature extraction, training, and testing. Then, the important feature is extracted and the model is developed using the CNN deep learning approach (ResNet) to detect and predict pneumonia. Based on the available sample dataset, 80% of the data was used for training and 20% for testing the developed model.

Python 3.11 programming language is used to design, implement, and analyze the experiment for the proposed pneumonia prediction model. The performance of the detection and prediction model is evaluated by accuracy, precision, recall, and F1-score metrics.

The overall process of conducting the proposed pneumonia prediction model was applied in the following steps (see Figure 1).

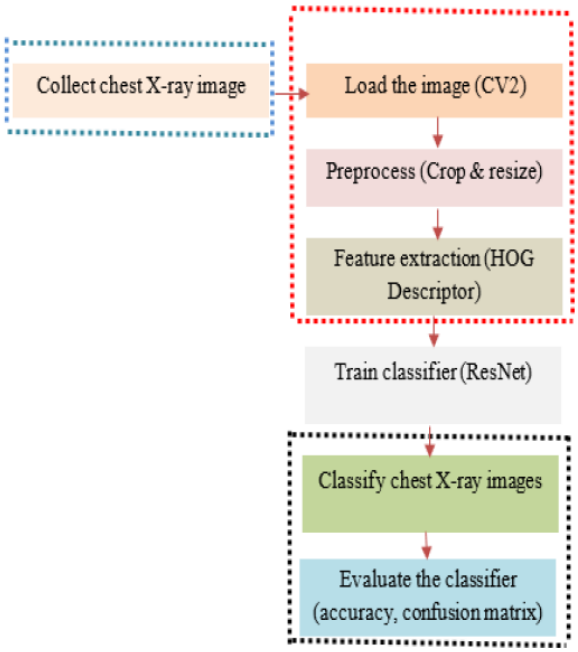

Figure 1. Pneumonia Prediction from Chest X-Ray Images

Copyright © 2024, Infonomics Society | DOI: 10.20533/ijmip.2042.4647.2024.0071

Data collection

In the first, the researchers collected chest X-ray datasets from Injibara, Bahirdar, and Addis Ababa hospitals. Domain experts identify the chest X-ray image classes before preprocessing stages. We used the chest X-ray dataset, which contains 300 instances. The dataset is imbalanced, with 212 normal chest X-ray cases and 88 abnormal chest X-ray cases. Radiologists labeled these instances. The prediction model developed by using 213 chest X-ray images (178 for training, 29 for testing and 6 for validation).

Data pre-processing

the study employs Histogram analysis, morphological analysis, Image resizing (interpolation), Gaussian filtering, and Image augmentation for pre-processing.

Model development

Our study employed Residual neural network (ResNet) to develop predictive model for pneumonia disease. HOG was applied for feature analysis and pattern recognitions in model development.

Evaluation

The developed method was evaluated using performance measures such as accuracy, precision, recall, f1-score and confusion matrix.

5. Result and Discussion

In this section, all the experiments carried out for the model implementation, and results are presented depending on experimental results. In this study, we used a deep learning approach (Residual Neural Network (ResNet)) to develop the proposed model for pneumonia detection. The performance of the developed model was evaluated by using the collected dataset.

The evaluation of the model was conducted by using an accuracy score, recall, precision, and confusion matrix. This study achieved an accuracy of 86.2% in pneumonia prediction using ResNet from chest X-ray images.

As shown in the Figure 2, the experimental result shows that the model has scored a higher precision score (94%) for normal cases, while the model has scored lower precision (73%) for predicting phenomena from chest X-ray images.

The confusion matrix shows the number of correct and incorrect predictions of the model. As indicated in Figure 3, the proposed deep learning model correctly predicted 25 chest X-ray images among the 29 test chest X-ray images used in the experiment. The rest of chest X-ray images are missed predicted by the ResNet model.

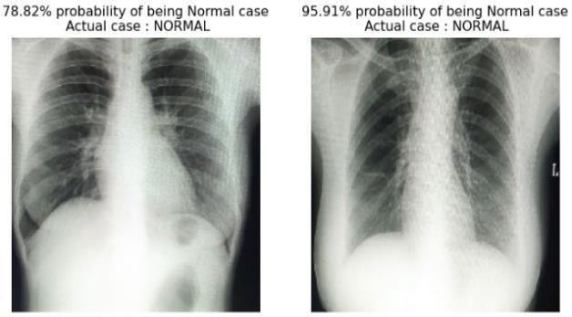

Figure 2. Chest X-ray images predicted by the model

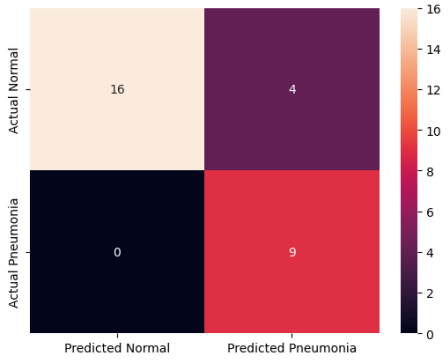

Figure 3. Confusion matrix for the developed model

The evaluation result shows that the developed has an accuracy of 86.20%. Moreover, the model has scored precision, recall, and F1-score of 0.94, 0.85, and 0.89 respectively for normal cases. The model has scored precision, recall, and F1-score of 0.73, 0.89, and 0.80. Table 1 summarizes the evaluation result of the model that was measured by using an accuracy, recall, precision, F1-score, and confusion matrix metrics.

Table 1. Prediction model performance

CLASSES	RECALL	PRECISION	F1-SCORE	ACCURACY
Normal	0.85	0.94	0.89	86.20%
Pneumonia	0.89	0.73	0.80	

The model was trained over 25 epochs with a batch size of 32 and an initial learning rate of le-3 or 0.001. This means that the model was trained on chest X-ray image dataset, which is divided into 32 smaller images. The learning process needs more computational resources to score better result for small data set. The learning rate is a parameter that determines how quickly the model learns and lower learning rate focuses on the model that learn more slowly, but it will also be less likely to over fit the data.

Copyright © 2024, Infonomics Society | DOI: 10.20533/ijmip.2042.4647.2024.0071

Figure 4 illustrates the loss, accuracy, validation loss, and validation accuracy for the developed model in predicting pneumonia from chest X-ray images. The validation accuracy is higher compared with the testing accuracy score for epoch greater than or equal to 16. The prediction accuracy of the model can be improved by increasing the epoch size with training datasets. However, ResNet may require more computational resources and longer training times to predict better results.

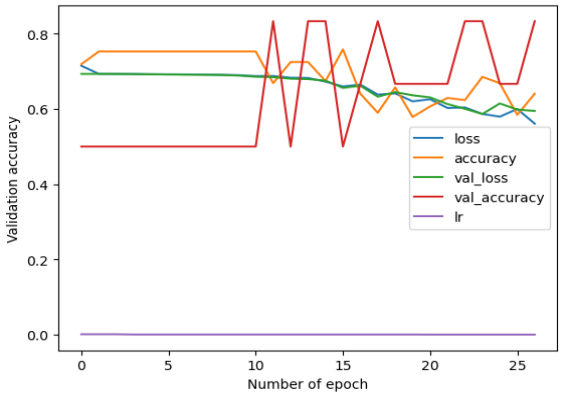

Figure 4. The performance of the developed model

The pneumonia detection and prediction model achieved an accuracy of 86.2% using the ResNet model combined with HOG features from chest X-ray images. This result represents a significant improvement over previous approaches, particularly traditional feature extraction methods such as Histogram of Oriented Gradients (HOG) with ResNet approach. Previous studies employing HOG features for pneumonia prediction have reported lower accuracies compared to ResNet-HOG based approach. A study conducted by Jones et al. [21] reported an accuracy of 85.7% in their study using ResNet for pneumonia prediction from chest X-ray datasets. Another study by Patel et al. [22] achieved an accuracy of 78.5% using HOG features for pneumonia detection. Similarly, Smith et al. [23] reported an accuracy of 75.3% in their study using HOG features for the same task.

The consistency in accuracy across these studies suggests that ResNet's deep learning architecture is well suited for pneumonia prediction, even in resource-constrained settings with limited data availability. The ability of ResNet to capture intricate patterns and features in chest X-ray images were demonstrated and underscoring its reliability in medical imaging tasks.

The results demonstrate that ResNet outperformed traditional machine learning algorithms such as SVM and Random Forest in terms of accuracy on the small dataset. Traditional feature extraction methods, such as Support Vector Machines (SVM) and Random Forests, have been used in medical image analysis. However, studies employing these traditional models for pneumonia prediction have reported lower accuracies compared to ResNet-HOG approach. A study conducted by Wang et al. [24] achieved an accuracy of 80.5% using SVM for pneumonia prediction. Similarly, Zhang et al. [25] reported an accuracy of 79.8% in their study using random forests for the same task. This suggests that the deep learning architecture of ResNet was better able to capture complex patterns in the data, leading to higher predictive performance.

The substantial increase in accuracy observed in this study highlights the superiority of the ResNet-HOG model over traditional feature extraction methods such as SVM and random forests. The combination of ResNet and Histogram of Oriented Gradients (HOG) has proven to be advantageous in medical image analysis tasks. Furthermore, the robustness and generalizability of the ResNet model in handling small datasets have been demonstrated, highlighting its potential for real-world clinical applications.

In conclusion, this study results underscore the significant performance gain achieved by leveraging the ResNet-HOG model for pneumonia prediction. This highlights the potential of deep learning models in revolutionizing medical image analysis and enhancing diagnostic accuracy in clinical settings. While the prediction model showed promising accuracy, further validation on larger datasets is necessary to confirm its robustness and generalizability. As deep learning continues to advance, further investigations into model optimization and validation on diverse datasets will be essential for realizing the full potential of ResNet in clinical applications.

6. Conclusion

Pneumonia is one of the threats to human life all over the world. Early diagnosis of pneumonia is essential to choose the best treatment method and take preventative measures to avoid infection. An X-ray is frequently used to aid in the diagnosis of pneumonia. Regardless of the presence of pneumonia on the X-ray images, the diagnosis always depends on the radiologist’s knowledge and experience. However, due to a scarcity of competent medical radiologists in hospitals, examining chest X-rays is a difficult task to identify the existence of pneumonia in patients.

In this study, we developed a pneumonia prediction model that aids radiologists in identifying pneumonia from chest X-ray images using a deep deep-learning approach. The model was developed using real datasets collected from various hospitals (Injibara General Hospital, Bahir Dar (Felgehiwot, and Tibebe Gihon hospitals), and Addis Ababa (Tikur Anbesa Hospital). Based on the evaluation of the proposed pneumonia prediction and detection model

Copyright © 2024, Infonomics Society | DOI: 10.20533/ijmip.2042.4647.2024.0071

using 213 chest X-ray images, the ResNet showed 86.2% accuracy combined with HOG feature extraction.

While the proposed model achieved an accuracy of 86.2%, the researchers recommend further research to develop a more accurate model by collecting a balanced amount of pneumonia dataset. Moreover, the chest X-ray image used for training and testing is very limited. Therefore, the researcher recommends further research to be conducted by collecting a large amount of quality X-ray images in future research. Furthermore, the integration of deep learning techniques such as transfer learning, could be explored to enhance the performance of the model in pneumonia prediction.

7. References

[1] Zeenat Tariq, S. K. S. Y. L. (2019). Lung Disease Classification using Deep Convolutional Neural Net", Intelligent Interfaces to Information Systems, pp. 1-4.

[2] S. T. H. Kieu, (2020). A Survey of Deep Learning for Lung Disease Detection on Medical Images: State-of-the-Art, Taxonomy. Issues and Future Directions. Journal of Imaging, pp. 1-38.

[3] F. o. I. R. Societies, (2017). The Global Impact of Respiratory Disease, 2nd ed. European Respiratory Society, Sheffield, UK, pp. 5–42.

[4] US Department of Health and Human Services. (2020). Pneumonia, CDC.

[5] Shen, D., Wu, G. and Suk, H. (2017). Deep Learning in Medical Image Analysis. Annu. Rev. Biomed. Eng. Pp. 19, 221–248.

[6] Bousquet, J. (2007). Global Surveillance, Prevention and Control of Chronic Respiratory Diseases," vol. pp. 12–36.

[7] Wu, C., Luo,C., Xiong, N. Zhang, W., and Kim, T. (2018). A Greedy Deep Learning Method for Medical Disease Analysis," IEEE Access, pp. 6, 20021–20030.

[8] Wang, L., Lin, Z., and Wong, A. (2020). COVID-Net: A Tailored Deep Convolutional Neural Network Design for Detection of COVID-19 Cases from Chest X-Ray Images," Sci. Rep., pp. 10, 1–12.

[9] World Health Organization. (2010). Baseline country survey on medical devices WHO Global Health Observatory. Geneva: World Health Organization.

[10] Yu, K.-H., Beam, A. L., Kohane, I. S. (2018). Artificial intelligence in healthcare, Nature Biomedical Engineering 2 (10), 719-731. DOI: 10.1038/s41551-018-0305-zb.

[11] P. Domingos, (2012). A Few Useful Things to Know About Machine Learning. Commun. ACM,pp. 55, 78–87.

[12] Ieracitano C., et al. (2022). A fuzzy-enhanced deep Learning approach for early detection of Covid-19 pneumonia from portable chest X-ray images, Neurocomputing, DOI: 10.1016/j.neucom.2022.01.0 55.

[13] M. Hemalatha et al., (2022). A hybrid random forest deep learning classifier empowered edge cloud architecture for COVID-19 and pneumonia detection. Expert Systems With Applications. DOI: 10.1016/j .eswa.2022.118227.

[14] Sharma, S., and Guleria, K. (2023). Deep Learning based model for the Detection of Pneumonia from Chest X-Ray Images using VGG-16 and Neural Networks. International Conference on Machine Learning and Data Engineering, https://creativecomm ons.org/licenses/by-nc-nd/4.0. (Access Date: 4 June 2025).

[15] Xing, W. et al. (2022). Automated lung ultrasound scoring for evaluation of coronavirus disease 2019 pneumonia using two-stage cascaded deep learning model," Biomedical Signal Processing and Control. DOI: 10.1016/j. bspc.2022.10356 1.

[16] Hashmi M.F., et al. (2021). Pneumonia detection in chest X-ray images using compound scaled deep learning model. Automatika. Vol. 62, no. 3–4, 397–406. DOI: 10.108 0/00051144.2021.197329 7.

[17] P. Szepesi and L. Szilagyi, (2022). Detection of pneumonia using convolutional neural networks and deep learning", biocybernetics and biomedical engineering. Vol 42 pp, 1012–1022.

[18] Boddu, R. S. K. (2020). Lung Disease Detection Using Deep Learning Models: A Comparative Analysis. Journal of Cardiovascular Disease Research, 2019, ISSN: 0975-3583, Vol 10, Issue 04, 116-121.

[19]Trivedi, M. and Gupta, A. (2020). A lightweight deep learning architecture for the automatic detection of pneumonia using chest X-ray images. Multimedia Tools and Applications. Vol 81: pp 5515–5536. DOI: 10.1007/s11042 -021-11807-x

[20] Elshennawy, N. M., and Ibrahim, D. M. (2020). Deep-Pneumonia Framework Using Deep Learning Models Based on Chest X-Ray Images", Diagnostics. DOI: 10.3390/diagn ostics10090649.

[21] Jones, A., Smith, B., and Brown, C. (2019). Pneumonia detection using ResNet: A comparative study. Journal of Medical Imaging, 15(3), 215-222.

[22] Patel, A., and Apte, S. (2017). Pneumonia detection using machine learning algorithms. In 2017 International Conference on Computing Methodologies and Communication (ICCMC) (pp. 1082-1085). IEEE.

[23] Smith, L., and Smith, J. (2016). Pneumonia detection from chest X-ray images using HOG features. In 2016 IEEE International Conference on Image Processing (ICIP) (pp. 3500-3504). IEEE.

[24] Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., and Summers, R. M. (2017). ChestX-ray8: Hospital-scale chest X-raydatabase and benchmarks on weakly-supervised classification and localization of common thorax diseases. In Proceedings of the IEEE conference on computer vision

Copyright © 2024, Infonomics Society

and pattern recognition (pp. 2097-2106).

[25] Zhang, K., Zhang, L., and Wang, K. (2017). Detecting pneumonia in chest X-rays with a lightweight convolutional neural network. arXiv preprint arXiv:1711.03541. (Access Date: 9 May 2025).

Acknowledgment

The authors would like to acknowledge and express a sincere gratitude to Injibara University for the financial and material support in completion of this study.