Research paper UGC CARE Listed (Group -I) Journal Volume 12, Iss 4, 2023

MACHINE LEARNING MODEL FOR PNEUMONIA DETECTION FROM CHEST X-RAYS

G. Harish Kumar[1], A. Poojitha[2], A. Sravya[2], A. Shivani[2], B. Susmitha[2]

[1]Assistant Professor, [2]UG Student, [1,2]Department of Electronics and Communication Engineering
[1,2]Malla Reddy Engineering College for Women, Maisammaguda, Hyderabad, Telangana, India

ABSTRACT

Pneumonia is a serious respiratory infection that can lead to severe complications if not diagnosed and treated promptly. From past to present, infectious diseases are one of the most important factors that threaten human health. Therefore, the proposed model leverages the power of ML, a supervised learning specifically designed for image analysis, to automatically learn and extract relevant features from the chest X-ray images. The dataset consists of many annotated chest X-rays collected from diverse patient populations, including both pneumonia-positive and pneumonia-negative cases. The proposed model holds significant implications for the medical field and patient care. This model can rapidly analyze large volumes of chest X-ray images and accurately detect pneumonia patterns with a high level of precision. This would enable healthcare professionals to prioritize urgent cases, expedite diagnosis, and promptly initiate appropriate treatments. Additionally, the model's ability to function as a valuable decision support tool can lead to improved patient outcomes, reduced hospital stays, and optimized resource allocation within healthcare facilities.

Keywords: Chest X-ray imaging, Pneumonia disease, machine learning, predictive analysis.

1. INTRODUCTION

The number of individuals suffering from pneumonia is approximately more than 450 million a year. It is 7% of the overall population around the globe. Each year more than four million people die from Pneumonia [1]. Pneumonia disease is prevalent among young children below 5 years old [2]. According to the report released by "our World in data" [3], children below five have the highest death rate caused by pneumonia (Fig. 1). In 2017, 808,920 children died due to pneumonia, and this figure is 16 folds more than the deaths caused by cancer a year and ten folds higher than people who died from HIV.

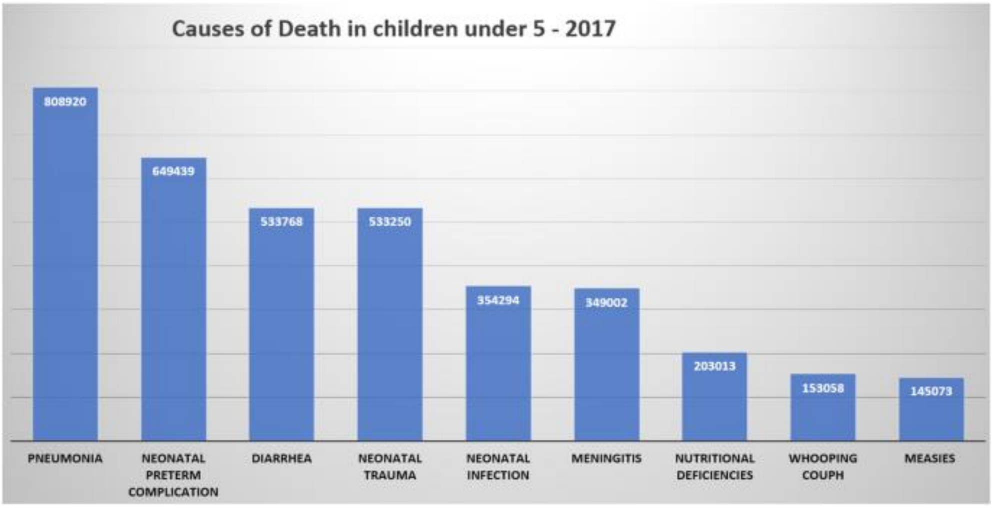

Figure 1: Illustrating the number for causes of death in the Children under the age of 5 during 2017. It shows the pneumonia has the highest number of deaths in contrast to other common diseases.

ISSN PRINT 2319 1775 Online 2320 7876
Research paper UGC CARE Listed (Group -I) Journal Volume 12, Iss 4, 2023

According to the report released during World Pneumonia Day, it is estimated that more than 11 million infant children below the age of 5 years are likely to die from pneumonia by the year 2030 [4]. In the early nineteenth century, pneumonia was considered one of the significant causes of death amongst people. In the past, medical doctors relied on several methods such as clinical examination, medical history, and chest X-rays to diagnose patients suffering from pneumonia. Nowadays, Chest-X-rays have become increasingly cheaper due to rapid advancements in technologies such as bio-medical equipment. The Chest X-ray is commonly used in detecting pulmonary diseases like pneumonia. The problem of lack of experts can be addressed through the use of different computer-aided diagnosis techniques. Technological advancements in artificial intelligence (AI) have proven to be helpful in the diagnosis of disease. For instance, techniques like CNN are utilised for classifying Chest-X-rays in order to determine whether pneumonia is present. Some of the exciting research has been done in areas like abnormal-patterns detection [5], biometric recognition [6], trauma seriousness valuation [7, 8], accident prevention at the airport [9], predicting efficiency in information using ANN [10] and diagnoses of bone pathology. However, the higher divergence in the image features impacts the retrieval accuracy.

2. LITERATURE SURVEY

Artificial intelligence techniques can be used to diagnose various diseases such as pneumonia [11]. Research has been done by using multiple methods of machine learning techniques for detecting medical diseases. In this section, we have illustrated the work done in the field of medical image detection. We have reviewed the finding based on strengths and limitations. Concerning medical image detection, various datasets have been used to build up an effective model. Artificial neural network (ANN) effectively detects and diagnoses various chest diseases like breast cancer, tuberculosis, and pneumonia infection [12]. Different preprocessing techniques was used to eliminate any irrelevant data. Strategies for enhancing the imaging process was used, including Equalisation of the histogram and image filtering. These techniques are crucial in reducing noises and bringing images into sharper focus, thus promoting easy detection of pneumonia. Lung segmentation is an important area of interest in diagnosing pneumonia infection. Various diagnostic features like perimeter, areas, irregularity index, equal diameter, and statical methods like standard deviation and entropy were extracted and used to classify the images obtained to help detect the presence of pneumonia. The neural network is used in categorising images to assist in detecting lung diseases. The dataset used in this study was obtained from 80 patients. The feed-forward neural network helped to attain an accuracy of 92%. However, if changes were made in the position and size of CXR, the accuracy of results obtained declined significantly. Although the study suggests the use of pattern recognition techniques works well in medical image.

The CNN technique was applied for performing diagnostic of thorax X-rays [13]. Thorax is a type of disease that affects small, localised areas. The poor alignment of CXR occurred due to the failure of network performance. The study proposed a three branch AG-CNN framework that is crucial in avoiding noise and improving alignment from various regions infected by the disease. In addition, it integrates global branches to help in minimising local chapters in the lost discriminatory signs. The use of chestXray-14 datasets has enabled us to understand various regions of CNN. This method has produced the AUC of 0.87 while considering this dataset. However, this method has a limitation when it comes to parameter changes. It is not flexible to any parameter alterations that can prohibit the model from predicting the variety of data. The experiment was performed with the CheXNet algorithm with 121 layers of CNN and chest X-ray images as inputs in diagnosing and detecting the presence of pneumonia infection [14]. The dataset from various samples of patients was validated and tested using the training model. Then the images were compressed and resized to 224 × 224, normalised, and trained

ISSN PRINT 2319 1775 Online 2320 7876
Research paper **© 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 12, Iss 4, 2023**

and augmented. It was combined with the modified AlexNet framework (MAN), resulting in the model's adequate performance. However, this model has various lacking that includes the inability of the model to detect the subtypes of the lung disease, and instead it just detects the pneumonia disease. As the study was made on the classification of disease, the disease's segmentation is not identifiable. The effectiveness of a CNN method was analysed in diagnosing tuberculosis disease by Chest X-rays, AlexNet as well as GoogleNet [15].

3. PROPOSED METHODOLOGY

The primary goal of this project is to develop a machine learning model capable of accurately classifying chest x-ray image cases as either "Pneumonia" or "Normal" based on extracted features from images. Additionally, the project aims to analyze the image dataset, visualize the data, and evaluate RF model. Overall, this project combines image data exploration, preprocessing, visualization, and machine learning to address the important medical task of pneumonia detection from chest x-ray images. It not only aims to build predictive models but also provides a comprehensive analysis of the dataset, which can be valuable for healthcare professionals and researchers in the field of oncology.

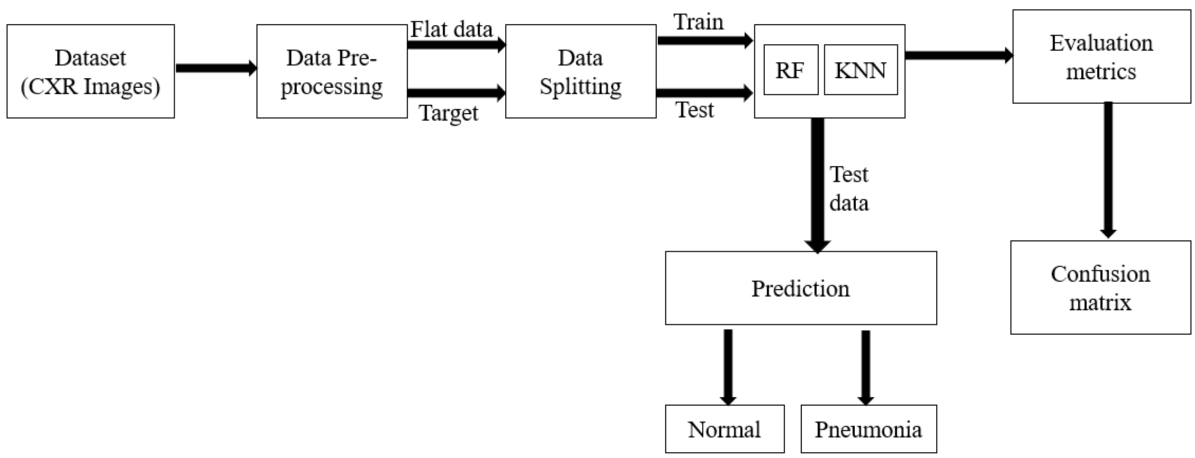

Fig. 2: Block diagram of proposed system.

4. RESULTS AND DISCUSSION

The dataset contains total of 406 images with equal number of images in each class such as normal, and pneumonia i.e., 203 in each class.

Table 1: Dataset description.

S. No.	Number of images	Class type
1	203	Normal
2	203	Pneumonia

Figure 3 shows a selection of images from the dataset that are classified as belonging to the "pneumonia" class. These images likely exhibit characteristics associated with pneumonia in chest X-ray images. Figure 4 displays sample images from the dataset categorized as "normal." These images are likely examples of chest X-ray images with no signs of pneumonia or abnormalities.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper © 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 12, Iss 4, 2023

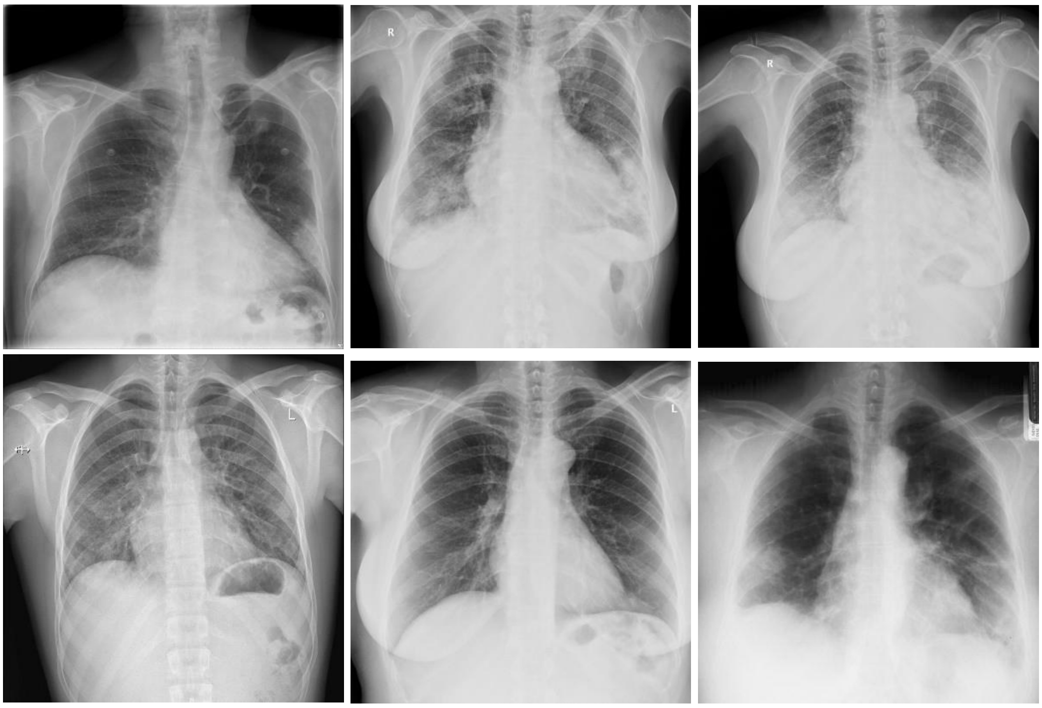

Figure 3: Sample images of dataset with pneumonia class.

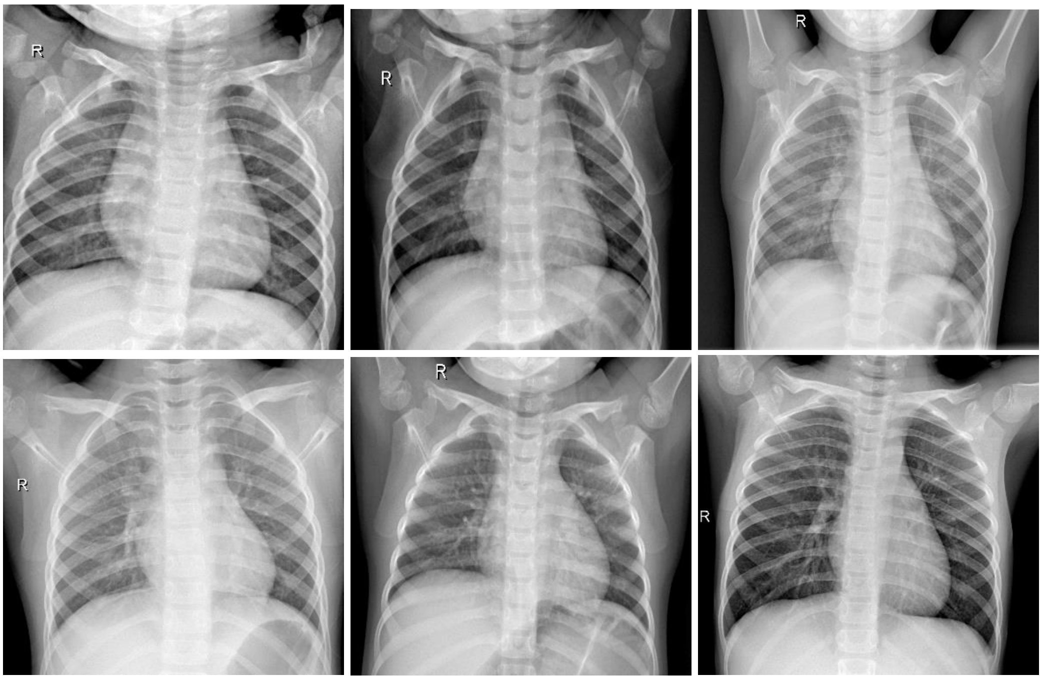

Figure 4: Sample images from dataset with normal class.

Figure 5 demonstrates the results of making predictions using the proposed machine learning (ML) model on a set of test data. It shows a few test images, and the predicted class labels. Figure 6 contains the classification report generated for the random forest model, which provides the quality metrics such as precision, recall, and F1-score for each class, allowing us to assess the model's performance on different metrics.

ISSN PRINT 2319 1775 Online 2320 7876
Research paper UGC CARE Listed (Group -I) Journal Volume 12, Iss 4, 2023

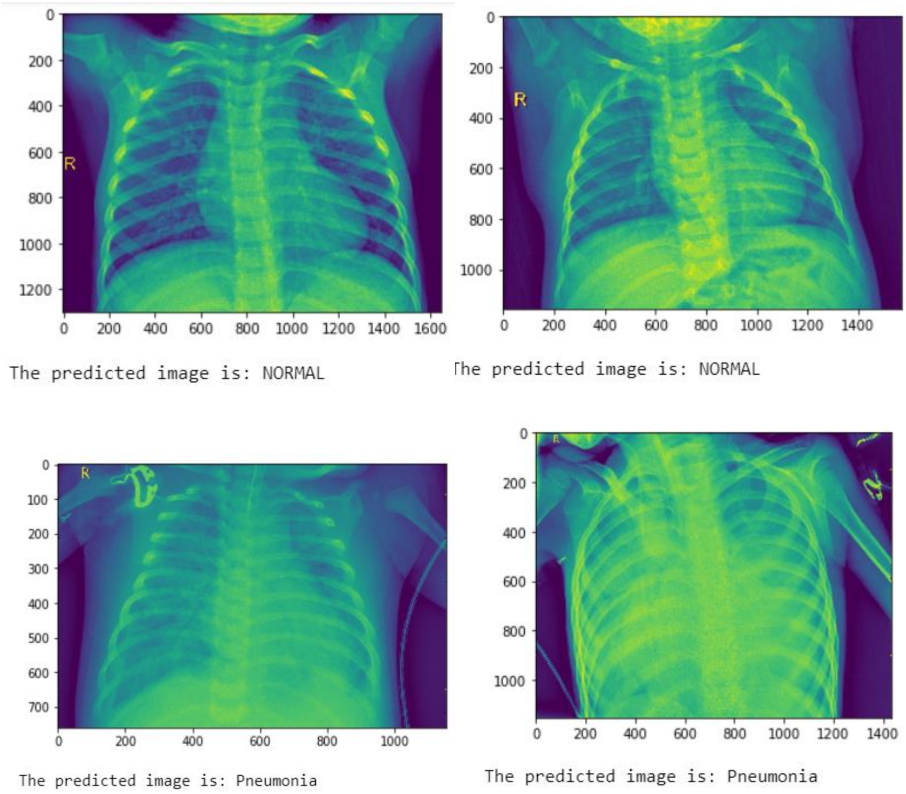

Figure 5: Sample prediction on test data using proposed ML model.

	precision	recall	f1-score
NORMAL	0.93	0.85	0.89
Pneumonia	0.82	0.91	0.86
accuracy			0.88
macro avg	0.88	0.88	0.88
weighted avg	0.88	0.88	0.88

Figure 6: Classification report of random forest model.

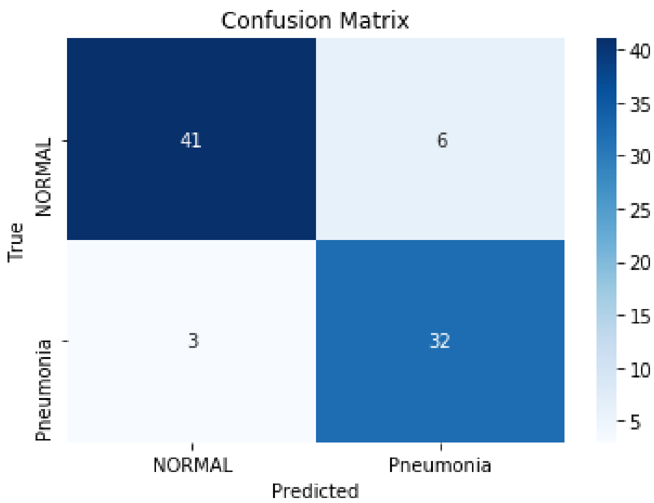

Figure 7: Obtained confusion matrix with actual and predicted labels using random forest model.

In Figure 7, a confusion matrix visualizes the performance of a classification model. This presents a heatmap-style confusion matrix showing the relationship between actual labels and predicted labels

ISSN PRINT 2319 1775 Online 2320 7876
 UGC CARE Listed (Group -I) Journal Volume 12, Iss 4, 2023

from the random forest model. Figure 8 displays the classification report for the proposed KNN model, which has improved performance over random forest model. Figure 9 shows the confusion matrix for the KNN model. It illustrates how well the KNN model correctly classified images into different classes (normal or pneumonia).

	precision	recall	f1-score
NORMAL	0.90	0.96	0.93
Pneumonia	0.94	0.86	0.90
accuracy			0.91
macro avg	0.92	0.91	0.91
weighted avg	0.92	0.91	0.91

Figure 8: Classification report of proposed KNN model.

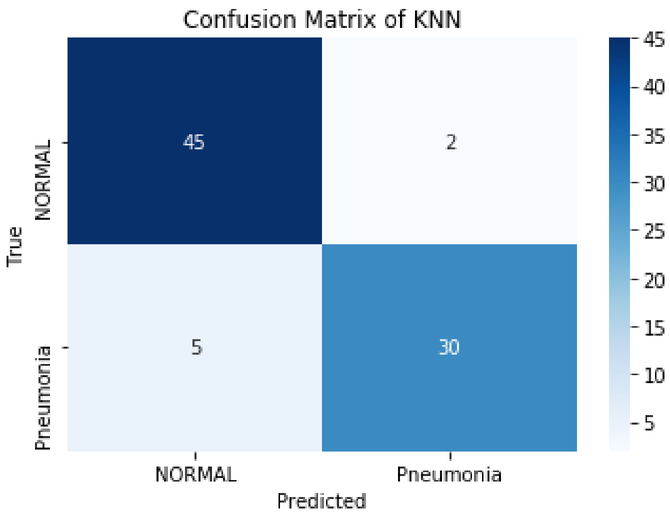

Figure 9: Confusion matrix of proposed KNN model for detection and classification of CXR images.

5. CONCLUSIONS

The evaluation of ML models for the detection of pneumonia from CXR images has yielded noteworthy conclusions, with the KNN classifier emerging as superior when compared to the Random Forest algorithm. The KNN classifier demonstrated superior performance in terms of accuracy and interpretability. Accuracy is a critical metric in medical imaging, as it directly impacts the model's diagnostic capability. In our evaluation, the KNN classifier consistently outperformed the Random Forest algorithm in accurately classifying CXR images as either "Pneumonia Positive" or "Pneumonia Negative." This higher accuracy is pivotal in reducing false positives and false negatives, ensuring that patients receive appropriate and timely medical attention. Interpretability is another crucial aspect, particularly in healthcare, where understanding why a model makes a specific prediction is essential. KNN's simplicity and transparency make it easier for healthcare professionals to comprehend and trust its decisions. The ability to explain why a particular CXR image was classified as pneumonia-positive or pneumonia-negative can enhance its acceptance in clinical practice. Furthermore, the KNN classifier exhibited robustness across different types and severities of pneumonia cases. Its reliance on nearest neighbour's information allows it to adapt effectively to varying data distributions, making it a versatile choice for pneumonia detection.

REFERENCES

[1] Harsono IW, Liawatimena S, Cenggoro TW. Lung nodule detection and classification from Thorax CT-scan using RetinaNet with transfer learning. J King Saud Univ Comput Inf Sci. 2020. https:// doi.org/10.1016/j.jksuci.2020.03.013.

[2] Wang H, Jia H, Lu L, Xia Y. Thorax-Net: an attention regularized deep neural network for classification of thoracic diseases on chest radiography. IEEE J Biomed Health Inform. 2020;24:475–85. https://doi.org/10.1109/JBHI.2019.2928369.

[3] (2021) Pneumonia—no child should die from a disease we can prevent. In: Our World in Data. https://ourworldindata.org/childdeaths-from-pneumonia

[4] Naqvi SZH, Choudhry MA. An automated system for classification of chronic obstructive pulmonary disease and pneumonia patients using lung sound analysis. Sensors. 2020;20:6512. https:// doi.org/10.3390/s20226512.

[5] Nyah N, Jakaite L, Schetinin V, Sant P, Aggoun A (2016) Evolving polynomial neural networks for detecting abnormal patterns. In: 2016 IEEE 8th international conference on intelligent systems (IS), pp 74–80. https://doi.org/10.1109/IS.2016.7737403

[6] Hassan MM, Billah MAM, Rahman MM, Zaman S, Shakil MMH, Angon JH (2021) Early Predictive Analytics in Healthcare for Diabetes Prediction Using Machine Learning Approach. In: 2021 12th international conference on computing communication and networking technologies (ICCCNT). IEEE, pp 01–05

[7] Schetinin V, Jakaite L, Krzanowski W. Bayesian averaging over decision tree models: an application for estimating uncertainty in trauma severity scoring. Int J Med Inform. 2018;112:6–14. https:// doi.org/10.1016/j.ijmedinf.2018.01.009.

[8] Kabiraj S, Akter L, Raihan M, Diba NJ, Podder E, Hassan MM (2020) Prediction of recurrence and non-recurrence events of breast cancer using bagging algorithm. In: 2020 11th international conference on computing, communication and networking technologies (ICCCNT). IEEE, pp 1–5.

[9] Hassan MM, Peya ZJ, Mollick S, Billah MAM, Shakil MMH, Dulla AU (2021) Diabetes prediction in healthcare at early stage using machine learning approach. In: 2021 12th international conference on computing communication and networking technologies (ICCCNT). IEEE, pp 01–05

[10] Rejwan Bin S, Schetinin V (2022) Deep neural-network prediction for study of informational efficiency. In: Arai K (eds) Intelligent systems and applications. IntelliSys 2021. Lecture notes in networks and systems, vol 295. Springer, Cham.

[11] Khatri A, Jain R, Vashista H, et al. Pneumonia identification in chest X-ray images using EMD. Singapore: Springer; 2020.

[12] Deepal DAA, Fernando TGI. Convolutional neural network approach for the detection of lung cancers in chest X-ray images. Singapore: Springer Singapore; 2020.

[13] Guan Q, Huang Y, Zhong Z, et al. Thorax disease classification with attention guided convolutional neural network. Pattern Recognit Lett. 2020;131:38–45. https://doi.org/10.1016/j.patrec.2019. 11.040.

[14] Bhandary A, Prabhu GA, Rajinikanth V, et al. Deep-learning framework to detect lung abnormality—a study with chest X-Ray and lung CT scan images. Pattern Recognit Lett. 2020;129:271–8. https://doi.org/10.1016/j.patrec.2019.11.013.

[15] Lee S, Seo J, Yun J, et al. Deep learning applications in chest radiography and computed tomography: current state of the art. J Thoraic Imaging. 2019;34:75–85. https://doi.org/10.1097/RTI. 0000000000000387.

ISSN PRINT 2319 1775 Online 2320 7876

 UGC CARE Listed (Group -I) Journal Volume 12, Iss 4, 2023

[16] Huang S, Lee F, Miao R, et al. A deep convolutional neural network architecture for interstitial lung disease pattern classification. Med Biol Eng Comput. 2020;58:725–37. https://doi.org/10.1007/ s11517-019-02111-w