

TO ANALYZE THE LUNGS X-RAY IMAGES USING MACHINE LEARNING ALGORITHM: AN IMPLEMENTATION TO PNEUMONIA DIAGNOSIS

Saroj Agrawal and Yogesh Kumar Gupta*

Received: August 16, 2023; Revised: March 07, 2024; Accepted: March 25, 2024

Abstract

Introduction: Respiratory diseases, particularly pneumonia, pose a significant threat to human life. Pneumonia affects the respiratory function in the human body and is a dangerous lung disease. This study aims to propose a model for detecting pneumonia in chest XR images. By utilizing statistical-based features, relevant and informative features are extracted from lung X-ray images. **Objective:** The objective is to obtain high accuracy in pneumonia identification; the target of this work is to generate a model that can precisely recognize the presence of pneumonia by evaluating chest X-ray pictures. **Method:** The Method follows a three-phase approach: preprocessing, categorization, and extraction of features. Preprocessing is the stage when various filters are applied to the chest X-ray images to enhance their eminence and eradicate noise. The feature extraction phase involves extracting statistical-based features from the preprocessed images. These features capture relevant information regarding a pneumonia diagnosis. Finally, in the classification phase, algorithms for machine learning are employed to use the retrieved features to categorize the X-ray pictures as infected or uninfected. **Result:** The proposed model successfully detects the presence of pneumonia accurately. By leveraging advanced machine learning algorithms, the model achieves accurate X-ray image classification for the chest. **Conclusion:** This study concludes by presenting a model for detecting pneumonia by examining chest X-ray pictures. To accurately classify infected and non-infected lungs, the proposed model makes use of image dispensation methods and machine learning algorithms. The model's high accuracy in pneumonia detection can significantly contribute to early diagnosis and treatment.

Keywords: GLCM; Homomorphic filter; Lungs; Pneumonia; SVM

Department of Computer Science and Engineering, Faculty of Computer Science and Engineering, Poornima University, Sitapura, Jaipur – 303905, Rajasthan, India. E-mail: sarojagr708@gmail.com; gyogesh@banasthali.in

* Corresponding author

DOI: <https://doi.org/10.55766/sujst-2024-02-e02133>

Suranaree J. Sci. Technol. 31(2):030178(1-13)

Introduction

In the past few years, respiratory diseases have been reported to be causing danger to the lives of individuals in large numbers. Pneumonia, being one of the hazardous lung diseases, has a proven impact on the respiratory functioning in the human body. Pneumonia is a respiratory infection caused by exposure to anti-agents like viruses, bacteria, and fungi. It is reported to impact individuals of all ages and has proven to be a significant cause of fatality around the globe. Infections such as Pneumonia spread easily and at a faster rate due to direct contact with infected people. Individuals contacted have knowledge of pneumonia, painful inhaling, and consuming less oxygen as a result of contamination of alveoli, i.e., small sacs in the lungs that take the air inside at the time of breathing. Complications due to pneumonia, such as sepsis (uncontrolled inflammation in the body), respiratory failure, and further infections, may lead to fatal conditions if not immediately attended medically. People with co-morbidities such as impaired immune systems, asthma, smokers, and infants, and the risk of having this severe inflammation in the elderly is increased by bacteria, viruses, and fungus (George *et al.*, 2023).

Numerous people have died, and severe crises have been triggered by epidemics and chronic diseases throughout history that took years to resolve (Saraiva *et al.*, 2019; Dhiman *et al.*, 2022). They could be applied to the classification of lung diseases, lung segmentation, and other extremely difficult visual tasks related to medical imaging. Recent advancements in DL have made it possible to match or even outperform human performance in a variety of fields. The results of therapies like cancer treatment and chevalier investigations can also be ascertained using DL. The categorization of thoracic illnesses using an X-ray imaging modality has shown good results when labelled data and DL-based algorithms are combined. In the past, deep neural network (DNN) models were created and tested by human experts using a method called continuous trial-and-error, which requires time, resources, and knowledge (Sourab and Kabir, 2022).

Additionally, according to WHO's 2019 report, pneumonia claimed the lives of 14% of children globally under the age of five. The DL-based model was expanded in order to forecast pneumonia early. By encouraging people to act, offering early treatment, and lowering the growing pneumonia death rate, many lives can be saved. In epidemiological research and public health, CXR

are a crucial apparatus for detecting pneumonia since they are easy to use and extremely effective. It is a difficult undertaking that necessitates the presence of an experienced radiologist to recognize pneumonia on CXR images. The transmission of bacteria, viruses, and fungus in the air we breathe is what causes respiratory illnesses with a connection to the chest (Abiyev and Ma'aitah, 2018). The underlying cause of these illnesses is breathing issues. The primary factor increasing death is pneumonitis. Children under five years are more likely to die from pneumonia than from any other cause. For the most part, the method for identifying pneumonia is an X-ray. It is more thorough and less expensive than other common diagnostic techniques. To find pneumonia, radiologists meticulously classify and assess X-ray pictures. This takes time, and it may cause disagreements among radiologists about diagnosing diseases. As a result, this problem gives rise to pneumonia detection techniques that can be practical in the medical field for accurate, immediate pneumonia diagnosis. In infection diagnosis systems, machine learning (ML), deep learning (DL), texture, and statistical approaches are tremendously effective tools (Sharma *et al.*, 2021; Singh and Tripathi, 2022). Too many diseases, injuries, or other health issues. Most of the incidents, it seems, are connected. In contrast to pandemics, outbreaks are more localized and less likely to arouse public alarm. Pneumonia is a difficult situation that results in a number of health issues and is one of the previous pandemics. Infectious illness outbreaks within communities can be categorized as pandemics or eruptions. Pneumonia alone accounts for more than a million hospital admissions as well as close to 500,000 fatalities annually in the United States. (Rajaraman *et al.*, 2018).

Causality and Severity

The severity of pneumonia can also be determined by the causality of the condition. Viral pneumonia is less severe and occurs gradually as compared to bacterial pneumonia, especially in the case of children (Bouch and Williams, 2006). However, the condition may become severe if both viral and bacterial pneumonia co-exist. The proliferation of such conditions in multiple lobes of the lungs requires hospitalization (Saravagi *et al.*, 2022). Fungal pneumonia is another form that weakens the immune system, due to which recovery becomes slow and difficult.

With evidential reports of the severity, incidence, and prevalence rate of the hazardous disease, research in the field of developing technology-based methods and techniques that may contribute to effective diagnosis and prevention of pneumonia in all ages become essential and is the need of the hour (Varshni *et al.*, 2019). Medical science is seen to be dependent upon computer-assisted systems for better diagnosis. The digital images procured with the help of automated systems are found to be accurate and more effective in recognizing intricate features required to diagnose specific diseases.

Pneumonia and Normal

A spectrum of highlights is used by radiologists in an x-beam image to detect Pneumonia. The ground glass design (GGO) is generally seen in Pneumonia carriers. GGO becomes more significant during the later stages of pneumonia compared to the initial stages. In radiography, murkiness is used to signify aspiratory amalgamation, this contrasts with the whitish areas shown in Figure 1(a), a radiograph of a normal patient, and Figure 1(b), a radiograph of a patient with classic pneumonia.

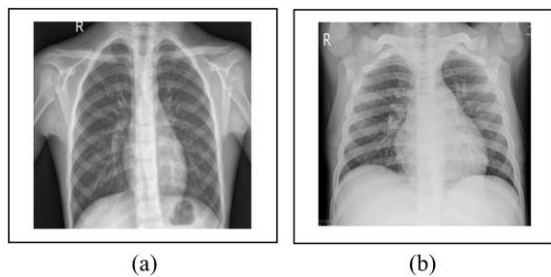


Figure 1. (a) Normal Image, (b) Pneumonia Image

Related Work

Radiologists, doctors, and others can now diagnose diseases more accurately and earlier thanks to advances in machine intelligence approaches in the medical industry. Using X-ray pictures to diagnose respiratory disorders has become increasingly common during the past ten years. We investigate machine learning algorithms for the categorization of PNEUMONIA infection using X-RAY pictures, driven by early successes of radiological research. On X-RAY pictures, PNEUMONIA features are successfully identified using a quadratic discriminant technique. The sternness of COVID-19 lung infections can be assessed using the CNN Model Tool (pneumonia in general), which can also be used to compare the sizes of the lungs of healthy people and patients with pneumonia using the ACM model20. To separate highlights from a given image, Sourab and Kabir (2022) created a convolutional neural system model without any prior planning. The summary of various previous research papers on the related subject with an accuracy of their model is given below:

Table 1 summarizes the proposed algorithms and their corresponding accuracies for diagnosing pneumonia from chest XR images in various studies. The review reveals that several deep learning algorithms, including VGG16, CNN, DenseNet-121, customized CNN, an ensemble of pre-trained CNNs, ResNet18, Xception, InceptionV3, DenseNet121, and MobileNetV3, have been used in the past. Transfer learning with deep CNNs and together possessing characteristics AlexNet, VGG-16, and VGG-19 have also been implemented. The accuracies of these algorithms range from 80% to 98.43%.

The proposed algorithm in this work is Quadratic Discriminant Analysis (QDA), which

Table 1. Comparative Literature Survey

S. No.	Author	Proposed Algorithm	Accuracy
1	Ayan and Ünver (2019)	VGG16	84.50%
2	Abiyev and Ma'aitah (2018)	CNN	92.40%
3	Cohen <i>et al.</i> (2019)	DenseNet-121	80%
4	Stephen <i>et al.</i> (2019)	CNN	93.73%
5	Saraiva <i>et al.</i> (2019)	CNN	94.40%
6	Rajaraman <i>et al.</i> (2018)	Customized CNN	96.20%
7	Chouhan <i>et al.</i> (2020)	Ensemble of pre-trained CNNs	96.39%
8	Toğaçar <i>et al.</i> (2020)	Combined features of pre-trained AlexNet, VGG-16 and VGG-19	96.84%
9	Rahman <i>et al.</i> (2020)	Transfer learning with deep CNN	98.00%
10	Hashmi <i>et al.</i> (2020)	ResNet18, Xception, InceptionV3, DenseNet121, and MobileNetV3	98.43%
11	Varshni <i>et al.</i> (2019)	CNN	80.02%
12	Khatri <i>et al.</i> (2020)	EMD	Near 80%
13	This Work (2024)	Quadratic Discriminant Analysis,	98.60%

achieves an accuracy of 98.60% in diagnosing pneumonia from chest X-ray images. Compared to the other studies reviewed, this algorithm demonstrates a higher accuracy rate. QDA is a classification algorithm that seeks to identify patterns in data and has been previously applied in image analysis. The study adds to the existing literature by proposing a new algorithm with a high point of accuracy, which may contribute toward the early diagnosis and treatment of pneumonia.

Present State-of-the-Art

This segment narrates research completed to predict pneumonia in CXR pictures by pre-trained and constructed models. A DL model for forecast pneumonia using CXR imaging that integrates the DenseNet169, MobileNetV2, and Vision Transformer models. In order to extract features from photos, three different models were utilized, and the resulting features were then used in the experimental procedure. (Stephen *et al.*, 2019).

To predict pneumonia in CXR data, the researcher39) created a 121-layer CheXNet model. The calculated F1 rating compared to the 0.387 average observed F1 score of the model. Compared to the CheXNet algorithm, this is substantially lower. Zhang (2021) researchers concluded that the Inf-Net algorithm can perform significantly better after thoroughly examining several AI-based methods for pneumonia identification. With the best outcomes compared to earlier design techniques, to identify pneumonia, Khatri *et al.* (2020) the authors used three CNN frameworks: Res Net, Dense Net, and Google Net. The study used two datasets and produced accurate findings for the Radiological Society of North America (RSNA) and Carmany datasets of 87.02% and 98.8%, respectively. Pun, and Agarwal (2021) studied a Deep CNN framework that was developed by combining the ResNet50V2 and XceptionNet models to predict outcomes in terms of efficiency. The results had a peak accuracy of 91.4% in comparison to alternative models that were already in use. Singh and Tripathi (2022) presented a mixed ensemble model to forecast pneumonia. Using the Adam optimizer and an 8-batch size, this design was produced. Using the Adam optimizer and an 8-batch size, this model was created. Along with F1 score values, the results are summarized in recall and precision. Lowitz (1983) the author has a collection of hybrid DL systems that combine SVM, RBFs, and logistic regression with three dissimilar classification techniques. H. A fully interconnected layer for classifying images. The weights needed to extract features from the image were decided in the second stage. Finally, he categorized the CXR images using a computer-aided model. The CXR

X-ray image data set was subjected to feature extraction using the DCNN approach by the author (Haralick *et al.*, 1973). AUCs, which describe the severity of the patient's health challenge, were created from the data. In this study, Shangjie Yao *et al.* (2021) showed their CNN method of diagnosing pneumonia and argued that the model used may overestimate the prognostication of pneumonia in the real world. An AUC of 0.931 is obtained from the experiment. To identify pneumonia, a CNN algorithm was created (Hashmi *et al.*, 2020).

Research Gap

The research gap addressed in this study is the need for an accurate and efficient process for detecting pneumonia from chest XR images. This is a serious problem because pneumonia is a dangerous lung condition that puts people's lives at risk. Despite several deep learning algorithms being used in the past for detecting pneumonia, there is still a need for a reliable and precise method that can help with the early detection and prevention of illness. The proposed algorithm in this study, Quadratic Discriminant Analysis (QDA), attains a great degree of accuracy, which may fill the research gap in developing an accurate and efficient way for using chest X-ray pictures to identify pneumonia.

- Statistical feature analysis not done on pneumonia-infected X-ray images.
- Previous model working on direct images using deep learning.
- No research was found on mathematical analysis of pneumonia detection.

Research Methodology

This section discusses the methodology adopted for better categorization of pneumonia XR images. In the first subsection, dataset depiction is given, which depicts the source of primary data being collected for the research. The second part describes the fundamental structure of the suggested model, and describes the frameworks utilized for image preparation and enhancement. In the second subsection, Following this, feature extraction, classification process, and performance analysis are provided.

Phase 1: Preprocessing In this phase, various filters are adopted for image preparation to enhance the eminence of chest XR images. The pre-processing step is essential as it helps to remove the noise and artifacts from the images, which

may impact how accurately the next stages are performed. The feature extraction stage then starts with the pre-processed images as their source material.

Phase 2: Feature Extraction Multiple statistical and texture-based features are used to extract pertinent and useful features that combine lungs' X-ray pictures from several sources. The phase of classification is then initiated using the extracted features as input.

Phase 3: Classification At this stage, the extracted features of infected and non-infected lungs X-ray images are used to determine whether pneumonia has infected the person or not. Many algorithms are used to classify the images, including the proposed Quadratic Discriminant Analysis (QDA). A combination of all the features is used to obtain the best results. The generated dataset is used to evaluate the system, which contains around 356 X-ray images of lungs, including 201 normal lungs X-ray images and 155 lung-infected images.

By using chest X-ray pictures, the suggested methodology seeks to create a reliable and effective method for identifying pneumonia. The method utilizes pre-processing techniques, feature extraction, and classification algorithms to achieve the best results. The work advances knowledge by proposing a novel algorithm with excellent accuracy that could help with pneumonia therapy and early detection.

Objectives

- Statistical-based feature extraction of lungs infected X-ray images.
- A system review of literature to gain insight view of the subject.
- Proposing a model to examine chest X-ray images and achieve the highest accuracy.
- Validating the proposed model.

Materials and Methods

Dataset Depiction

The data source for this investigation was made up of 356 chest X-ray pictures in JPEG format, with two classifications: normal lungs (201 images) and infected lungs with pneumonia (155 images). The images were sourced from Kaggle.com, and all images were physically evaluated. The researchers used a training dataset consisting of 284 images, with 119 Pneumonia X-ray pictures and 165 Non-Pneumonia XR pictures. The dataset is essential in evaluating the planned method's accuracy and efficiency in diagnosing pneumonia from CXR images.

The employ of physically evaluated images from multiple sources makes the study a meaningful contribution to the body of literature by improving the validity and dependability of its findings.

Approach to our Model

The section explains the methodology applied in this research work. The model proposes three stages of classification. Figure 2 provides a thorough description of the model. There are 4 stages in the proposed model:

1. Preprocessing stage
2. Feature-extraction stage
3. Classification stage
4. Performance Analysis

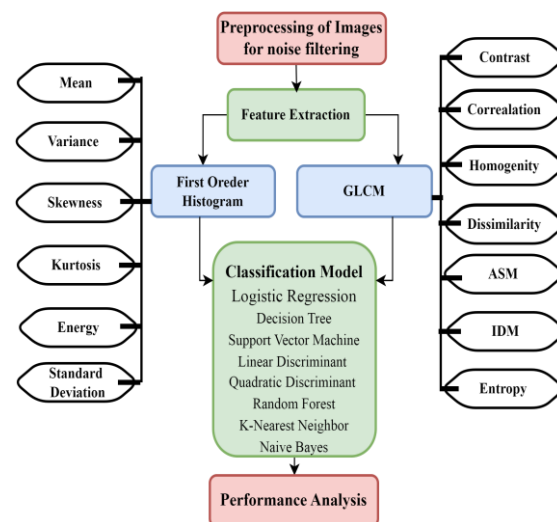


Figure 2. Research flow diagram

The Pre-Processing Stage

Two popular filtering techniques, one Homomorphic and the other CLAHE are employed for digital picture improvement, chiefly when the input picture suffers from low elucidation conditions. In various imaging applications, like medical, biometric, and robotic vision, these filtering approaches are used.

- CLAHE- To augment the visibility level of foggy pictures or videos, (CLAHE) Contrast limited adaptive histogram equalization is used. The current research utilized the CLAHE enhancement technique for enhancing the X-ray image quality. Homomorphic filtering was used to further enhance the picture quality. This mechanism works in the frequency area and applies a high-pass filter to decrease the importance of less frequency components (Arya *et al.*, 2018).

- Homomorphic filtering methods operate on the illumination-reflectance model. This model is classified by two primary elements viewed as the picture. The main component is the source light that was incident on the scene in question ($i(x,y)$). The ($r(x,y)$) reflection component of the scene's objects is the second component. Utilizing this reality, the homomorphic filtering procedure seeks to decrease the importance of the first element by lowering the image's low-frequency content. By imposing the filtering approach in the frequency

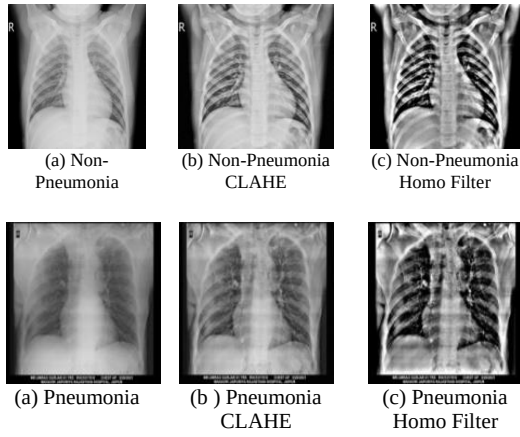


Figure 3. Normal and filtered images

range, this can be accomplished. A picture in the frequency domain is essentially translated from the spatial domain to the frequency area using this procedure. For this, one may employ the Fourier transform (Arya *et al.*, 2018).

Feature Extraction

An abstraction of a picture is used to describe and quantify the elements of a picture. The quantification can be a normal number, a real, an integer, or a binary number. In Other words, a feature vector is a list of numbers that constitute a picture. After preprocessing the next step comes to be feature extraction, which uses statistically based approaches to extract the features. Consequently, in the initial step of an image analysis, the pictures were categorized as either infectious or normal images. In the current research, the features are clipped using a first-order histogram, GLCM. Five features of the gray image are taken out using a first-order histogram: mean, standard deviation, variance, skewness, and kurtosis. The eight texture features from GLCM are energy, correlation, homogeneity, entropy, and contrast. ASM, IDM, and dissimilarity. A total of 13 features are taken out statistically and texture-based.

- **The first-order histogram**
- **GLCM**

➤ The first-order histogram. It is used for extracting statistical features. The two variables, x and y , are represented as matrix space. A picture is represented as a matrix form, which explains the intensity distribution of the level. However, as it takes more time to process, the problem can be overcome by converting images into a gray-scale image. Five features are extracted with the help of a first-order histogram- mean, kurtosis, variance, entropy, and skewness of the grey image. A histogram provides statistical information on the picture. The characteristics of first-order histograms can be explained by the equations given below (Arya *et al.*, 2018).

$$Mean = \sum_{i=0}^{N-1} (P_{ij}) \quad (1)$$

The mean value of normal is lower than infected because, in the infected X-ray images, white pixels are more due to a white cloud (Arya *et al.*, 2018).

$$Variance = \sum_{i=0}^{N-1} (1 - \mu^2) (P_{ij})^2 \quad (2)$$

The variance value of infected X-ray images is lesser than normal X-ray images because, in an infected X-ray image, there is a high variation of intensity around means. (Arya *et al.*, 2018).

$$Skewness = \sum_{i=0}^{N-1} (1 - \mu^3) * P_{ij} \quad (3)$$

The symmetry of the histogram is represented via skewness w.r.t the mean value. The skewness value is higher than infected XR images for non-infected XR images because the tail indicated at the right part is lengthier than the left part. Infected XR images it is like the normal XR images but smaller than the normal X-ray images. (Arya *et al.*, 2018).

$$Entropy = \sum_{i=0}^{N-1} (P_{ij}) * \log P_{ij} \quad (4)$$

The entropy value of the normal X-ray images is greater than infected X-ray images as intensity distribution in the normal X-ray images shows greater uniformity than images of infected X-rays (Arya *et al.*, 2018).

Table 2. First-order histogram features of normal and Pneumonia X-ray images

	Image	Mean	Std	Variance	Skewness	Kurtosis	Category
1	IMAGE 1	110.46	93.74	8,787.10	0.26	-1.38	Non Pnuemonia
2	IMAGE 2	106.83	85.02	7,229.00	0.25	-1.17	Non Pnuemonia
3	IMAGE 3	116.76	89.72	8,049.37	0.10	-1.33	Non Pnuemonia
4	IMAGE 4	105.74	92.59	8,572.53	0.26	-1.36	Non Pnuemonia
1	IMAGE 1	93.39	66.88	4,473.51	0.28	-0.60	Pnuemonia
2	IMAGE 2	101.89	69.13	4,778.94	0.18	-0.73	Pnuemonia
3	IMAGE 3	93.98	71.65	5,134.09	0.23	-0.84	Pnuemonia
4	IMAGE 4	114.18	90.57	8,203.20	-0.04	-1.44	Pnuemonia

Table 3. GLCM features of normal and Pneumonia X Ray

S.No	Path	Contrast	Dissimilarity	Homogeneity	Energy	Correlation	ASM	Entropy	IDM	Category
1	Image 1	7.97	1.82	0.52	0.25	0.88	0.05	4.36	0.53	Non-Pnuemonia
2	Image 2	2.31	0.96	0.64	0.24	0.96	0.06	4.13	0.61	Non-Pnuemonia
3	Image 3	2.15	0.89	0.66	0.25	0.97	0.06	4.00	0.64	Non-Pnuemonia
4	Image 4	2.13	0.84	0.69	0.31	0.97	0.10	3.79	0.67	Non-Pnuemonia
1	Image 1	1.63	0.66	0.72	0.22	0.95	0.04	3.93	0.68	Pneumonia
2	Image 2	2.12	0.89	0.65	0.18	0.94	0.03	4.23	0.60	Pneumonia
3	Image 3	1.65	0.61	0.75	0.25	0.96	0.06	3.81	0.71	Pneumonia
4	Image 4	1.51	0.59	0.77	0.31	0.98	0.09	3.70	0.71	Pneumonia

$$Kurtosis = \sum_{i,j=0}^{N-1} (1-\mu)^4 * (P_{ij}) - 3 \quad (5)$$

The probability distribution is represented via kurtosis. If the value is more than zero, then a peak that is smaller in size is seen around the mean value and is rectangular round in shape. If this value is negative or is less than zero, then the peak is around the mean is high, and the tails are flat. Kurtosis with a high value in infected X-ray pictures shows a sharper peak than normal X-ray images.

Table 2 indicates the values for normal and Pneumonia X-ray images that are chosen from the data set randomly. The first four results indicate the values of non-infected images, and the enduring four indicate the value of infected images. The table presents the first-order histogram features of eight X-ray images; the remaining four are classified as pneumonia, while four are non-pneumonia. The characteristics include skewness, kurtosis, skewness, variance, and mean. The mean values of the non-pneumonia images range from 105.74 to 116.76, while those of the pneumonia images range from 93.39 to 114.18. The standard deviation values for non-pneumonia images range from 85.02 to 93.74, while those of pneumonia images range from 66.88 to 90.57. The variance values for non-pneumonia images range from 7229.00 to 8787.10, while those of pneumonia images range from 4473.51 to 8203.20. The skewness values for all images are positive, indicating that the distribution of pixel values is skewed towards higher values. The kurtosis values for all images are negative, which denotes a platykurtic distribution

that is flatter than a normal distribution. The features can be used for developing an image classification model that can distinguish between X-ray images of pneumonia and non-pneumonia.

➤ GLCM (Gray-Level Co-Occurrence Matrix)

GLCM Arya *et al.* (2018) extracted statistical features. In the GLCM matrix, rows and columns equal the gray levels in the picture. This matrix is defined by G_d, θ i, j, where d is the distance and θ is the orientation of the 'i' and 'j' values in the image. In this study, Table 3 shows the extracted numerical values using the GLCM method for normal and Pneumonia X-ray pictures, and arbitrary selection from the dataset; in the table, the first four results are non-infected images, and the residual four are infected images. The equations given below represent.

The eight features, namely Energy, correlation, homogeneity, entropy, and contrast, along with ASM, IDM, and dissimilarity. Dissimilarity serves as a representation of the separation between adjacent pairs of objects (pixels) in the pertinent domain.

The GLCM features of the normal and pneumonia X-ray pictures are shown in the table. Contrast, Dissimilarity, Homogeneity, Energy, Correlation, ASM, Entropy, and IDM are among the characteristics of the GLCM. The table displays four photos of normal and four images of pneumonia. The values for each GLCM feature are provided for each image along with the corresponding category. The results show that the GLCM features differ between normal and pneumonia X-ray images, with higher values of Contrast, Dissimilarity, and

Entropy and lower values of Homogeneity, Energy, Correlation, and IDM for pneumonia X-ray images. These findings suggest that GLCM features can be useful in distinguishing normal and pneumonia X-ray images.

$$Contrast = \sum_{i,j=0}^{N-1} (P_{ij}) * (i - j)^2 \quad (6)$$

Contrast is the measurement of Intensity contrast between pixels and their neighbors across an X-ray image; it determines the variation among the greatest and the smallest values of a neighboring pixel set.

$$Energy = \sum_{i,j=0}^{N-1} (P_{ij})^2 \quad (7)$$

Energy is the summation of squared data in the GLCM, and it shows texture uniformity. Correlation is the measurement of the joint probability of the grey scale pixel pairs (Arya *et al.*, 2018).

$$Homogeneity = \sum_{i,j=0}^{N-1} \frac{P_{ij}}{1 + (i - j)^2} \quad (8)$$

Homogeneity is the feature that is used to calculate the proximity of the distribution of grey scale pixels in the GLCM (Arya *et al.*, 2018).

$$Entropy = \sum_{i,j=0}^{N-1} (P_{ij}) * \log P_{ij} \quad (9)$$

Entropy is in image processing; discrete entropy depicts the number of bits (Arya *et al.*, 2018).

$$Angular\ Second\ Moment\ (ASM) = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (P_{ij})^2 \quad (10)$$

The homogeneity of an image can be measured using ASM, also known as energy. In a homogeneous scene, GLCM will have a few $p(i, j)$ with relatively high values due to few gray levels, which will enhance the sum of squares.

$$IDM = \frac{\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (i, j) p(i, j) - \mu_x \mu_y}{\sigma_x \sigma_y} \quad (11)$$

The local homogeneity of an image is measured using Inverse difference moment (IDM). Closeness in gray-scale values increases the

incidence of co-occurrence of pixel pairs and thus enhances the IDM value (Arya *et al.*, 2018).

Classification

The current model proposed in the study classifies X-ray images into two classifications: Normal and statistical-based features indicating Pneumonia infection. Abnormal images are examined and tested further. The proposed method is tested using Logistic Regression, Decision Tree, SV Machine, Linear Discriminant Analysis, Naive Bayes, Random Forest, K-Nearest Neighbors, and Quadratic Discriminant Analysis.

➤ Logistic Regression: includes predictive modeling such as regression, hence termed as logistic regression. This is classified under the classification algorithm due to the reason that it classifies samples (Stokes *et al.*, 2021).

➤ Decision Tree: It includes classification and regression issues. This is mainly chosen for solving categorization issues. This can be compared to a tree-structured analyzer, with leaf nodes standing in for results and interior nodes symbolizing the distinctiveness of the data set (Stephen *et al.*, 2019).

➤ Support Vector Machine: This is a classifier that constructs a high-dimensional hyper-plane to analyze data into suitable modules using utmost margins. Kernel functions transform low-dimensional data to peak-dimensional data for better performance (Stephen *et al.*, 2019).

➤ Linear Discriminant Analysis: It is a Normal discriminant or discriminant Function Analysis technique that is usually adopted for supervised classification tasks. Used to model group differences, such as used to distinguish between two or more classes. A low-dimensional space can benefit from features that are projected from a high-dimensional region (Stephen *et al.*, 2019).

➤ Quadratic Discriminant Analysis: It is similar to linear discriminant analysis but relaxes the assumption that all classes have equal means and covariance. Therefore, it must be calculated separately (Yang *et al.*, 2017).

➤ Random Forest: It is a supervised learning method based on ensemble learning that is used to tackle complicated issues and enhance model performance. It may be applied to challenges involving categorization and regression. According to the description, “a random forest is a classifier that averages a set of decision trees across different subsets of a given dataset to enhance the accuracy of predictions of that dataset” (Rajaraman *et al.*, 2018).

➤ K-Nearest Neighbor: A non-parametric and based on the Supervised Learning technique, it stores the available information and Identifies by similarity of the new data point, which can be

further classified into a well-suited category. It is helpful in regression and classification. It is mostly applied to classification-related problems (Ieracitano *et al.*, 2022).

➤ Naive Bayes: This method is founded on the Bayes principle. Used to organize classification issues. NB is a probabilistic classifier and predicts the probability of an object (Kallianos *et al.*, 2019; Ieracitano *et al.*, 2022).

Performance Analysis

Accuracy, sensitivity, and specificity were used as performance measures. Accuracy refers to the proportion of data appropriately analyzed as true values and false values. H. Her X-ray of infected lungs in the infected category and the default image in the Non-infected class. Sensitivity refers to the proportion of accurate classifications of true data associated with infection pictures in this study. Specificity refers to the percentage of standard images correctly mapped to false cues. They are given by: (Sirazitdinov *et al.*, 2019)

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN}$$

$$\text{Sensitivity} = \frac{TP}{TP+FN}$$

$$\text{Specificity} = \frac{TN}{FP+TN}$$

Here, TP denotes true-positive values, referring to images that were correctly identified as infected, and FP denotes false-positive values, which are images that were incorrectly recognized as normal X-ray images. TN stands for true negative values, signifying an appropriately identified image as a non-infected image, and false-negative results, referring to an incorrectly identified image as an infected image.

Results and Discussion

Model Explanation

The proposed model demonstrated the categorization of pneumonia and (Normal) non-pneumonia through several X-ray images.

The model was built from scratch, separating it from previous models, which depend excessively on the machine learning approach. The research can be subsequently expanded to identify and categorize radiographic pictures showing pneumonia as a differentiation of the two of significant concern. The model proposes four stages of classification:

- Preprocessing stage
- Feature-extraction stage
- Classification stage
- Performance Analysis

Validation and Vérification of Model

In order to categorize the X-ray images of lungs as infected and non-infected and to project different stages of infection by using the proposed model. All the dataset was converted to data frames using feature extraction. This data frame was evaluated using eight classifications of machine learning algorithms. To create a feature database of the composed medical images along with extorting the related information. In order to classify the dataset, a performance Analysis consisting of Accuracy, Sensitivity, and Specificity was done (Sirazitdinov *et al.*, 2019). The output of this classification phase provided the best accuracy model.

Table 4 shows the Performance Analysis consisting of Accuracy, Sensitivity and Specificity of all classification algorithms used on our test dataset.

The table displays the analysis of the performance of various machine learning models on X-ray images for the identification of pneumonia. Logistic Regression, Decision Tree, Support Vector Machine, Linear Discriminant Analysis, Quadratic Discriminant Analysis.

Random Forest, K-Nearest Neighbors, and Bayes are among the models. Accuracy, Sensitivity, and Specificity are the evaluation criteria that are used to compare the models. The model with the highest accuracy, sensitivity, and specificity is found to be quadratic discriminant analysis, with

Table 4. First-order histogram features of normal and Pneumonia X-ray images

S.No.	Model	Accuracy	Sensitivity	Specificity
1	Logistic Regression	93.06%	97.83%	84.62%
2	Decision Tree	93.06%	95.65%	88.46%
3	Support Vector Machine	83.33%	97.83%	57.69%
4	Linear Discriminant	93.06%	97.83%	84.62%
5	Quadratic Discriminant	98.61%	97.83%	100.00%
6	Random Forest	95.83%	97.83%	92.31%
7	K-Nearest Neighbors	70.83%	78.26%	57.69%
8	Naïve Bayes	83.33%	93.48%	65.38%

values of 98.61%, 97.83%, and 100.00%. The research's conclusions are consistent with those of earlier studies that have demonstrated the effectiveness of machine learning models in identifying pneumonia from X-ray pictures (Rahman *et al.*, 2020; Zhang, 2021).

Evaluation of the Proposed Model

The parameter evaluation metrics for measuring the performance of best ML algorithms are discussed in the following part. The results of the validation and verification of the model depict that the Quadratic Discriminate Analysis and its parameters are found to be the most accurate among the other models (Yang *et al.*, 2017). The low false positive rate and high true positive rate in the provided table serve as evidence of the model's accuracy. The false positive rate means that there are fewer chances of false detection of the normality of the patient, which means that the patient having pneumonia has fewer chances of being diagnosed as normal. On the other hand, assessing the true positive rate it can be understood that the model is appropriately diagnosing and confirming the existence of the disease. Thus, the results adequately support the effectiveness of the Quadratic Discriminate Analysis Algorithm.

The confusion matrix shows the classification concert of a binary classification model for pneumonia diagnosis.

The predicted and actual conditions are classified as positive or negative. FP (False Positive) embodies the number of negative cases that were incorrectly classified as positive, TP (True Positive) represents the number of positively classified cases that were properly categorized, and TN (True Negative) represents the number of negatively classified cases that were correctly classified. In this study, the model QDA correctly classified 45 pneumonia-positive cases as positive (TP), misclassified one pneumonia-negative case as positive (FP), correctly classified 26 pneumonia-negative cases as negative (TN), and did not misclassify any pneumonia-positive cases as negative (FN). This indicates that the model QDA has a high sensitivity (TP rate) and a low False positive rate (FP rate). Figure 4 is a comparative analysis of all classification models and the resulting confusion matrices of the Quadratic Discriminant Algorithm, i.e., a best-performing model for binary classification in this research, shown in Figure 5.

Medical Sensation

The research will be very useful in detecting lung infection by extracting the features from

images and analyzing them by image processing. It will help the doctors and patients in a big way. The findings of this research will make it simpler for radiologists and medical professionals to identify mathematical numbers that represent a pneumonia infection.

The principle of the current work is that machine learning methods were used to categorize X-ray pictures of pneumonia. 356 X-ray pictures from Kaggle.com were used as the dataset for this study. First-order histogram and GLCM features were retrieved from the photos and added to the training and testing sets of the dataset. The photos were classified using a variety of machine learning methods, and the effectiveness of each approach was assessed using accuracy, sensitivity, and specificity metrics.

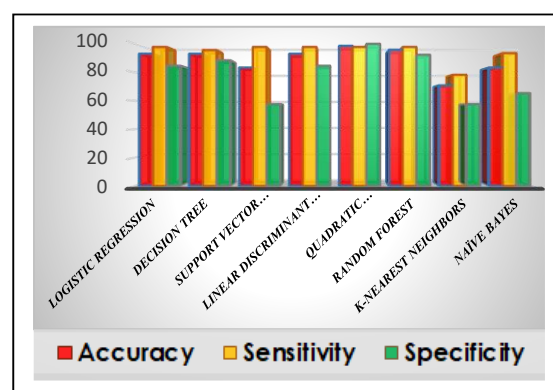


Figure 4. Comparative study

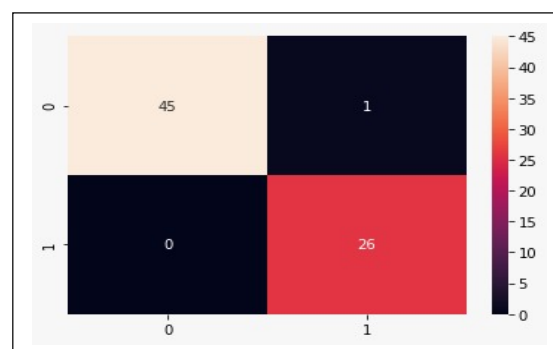


Figure 5. Confusion Matrix

The quadratic discriminant analysis (QDA) technique achieved the highest accuracy (98.61%), sensitivity (97.83%), and specificity (100%), according to the results. The QDA algorithm's confusion matrix revealed 45 true positive cases and 26 true negative cases with no false negatives and one false positive case.

This study has significant ramifications for employing machine learning algorithms to diagnose pneumonia accurately. Because of the QDA algorithm's excellent sensitivity and accuracy, it is possible to diagnose pneumonia using X-ray pictures. The research shows how machine learning algorithms could help radiologists identify pneumonia and other illnesses.

Discussion

Eight machine learning models' effectiveness is discussed in an experimental study employing X-ray picture datasets. The study created a method to investigate chest X-ray images and look for pneumonia. Multiple statistical-based features are used to extract pertinent and informative features using X-ray images of lungs, obtained from different sources. This research includes three phases: Preprocessing, Feature extraction, and Classification. Various filters are adopted for image pre-processing. Many algorithms like SVM, Quadratic Discriminant, etc. Yang *et al.* (2017) used to classify and determine whether pneumonia had infected the person or not by the extracted features. The combination of all the features helped in obtaining the best results. Overall it was found that Quadratic Discriminant Analysis has an accuracy of 98.6%, significantly better than all other classifiers.

The present research aimed to explore using machine learning methods to distinguish between normal and pneumonia chest X-ray pictures. The research involved the extraction of first-order histogram features, and GLCM features from the images, followed by the use of various machine-learning models for classification. The results showed that the GLCM features were more effective than the first-order histogram features for classification, with the Quadratic Discriminant Analysis (QDA) model performing the best among all the tested models.

However, the performance of the models varied widely, with some models achieving high accuracy and sensitivity, but low specificity, and vice versa. The confusion matrix showed that the QDA model had 45 true positives, one false positive, and no false negatives, indicating its high performance for correctly classifying pneumonia cases. The present research has several implications for medical practice, as it highlights the potential use of machine-learning approaches to classify chest X-ray pictures accurately and effectively. This could aid medical professionals in diagnosing and treating pneumonia cases more effectively. However, the study's flaws, especially the limited

sample size and the lack of diversity in the dataset, must be considered.

Conclusions

With sufficient review analysis, Pneumonia is difficult to diagnose without effective radiological mechanisms, and there is insufficient access to this diagnostic mechanism by the population at large. Image data processing utilizing Chest X-ray reports can be effectively using machine learning to accomplish based prediction models. The proposed model developed using algorithms in Python language can be an effective mode of detection of hazardous disease. The Quadratic discriminant Analysis is found to be better than other classifying methods with 98.6% accuracy.

In the Python language, efficient models are built using deep learning algorithms. This model helps doctors easily identify pneumonia. The projected framework was differentiated with state-of-the-art machine learning techniques and demonstrated to be more effective in forecasting. Additionally, the utilization of both statistical-based features from first-order histograms and texture-based features from GLCM in Mathematical values enhances the characterization of image content in this research.

Moreover, the integration of various machine learning algorithms for classification on different parameters for best classification, and is done on the basis of Mathematical values, showcases a comprehensive approach in medical image analysis, potentially leading to more accurate and efficient diagnosis of pneumonia from X-ray images. Other ML algorithms were also prosecuted, and in-depth mathematical analysis has been exercised in pursuit of better accuracy and precision.

In conclusion, this research aimed to classify X-ray images as normal or pneumonia using machine learning techniques. The research used first-order histogram and GLCM features to extract information from images, and then applied various machine learning algorithms, including logistic regression, decision tree, support vector machine, linear discriminant analysis, quadratic discriminant analysis, random forest, K-nearest neighbors, and Bayes, to classify the images. The results showed that quadratic discriminant analysis had the highest accuracy rate of 98.61%. The research also analyzed the performance of each algorithm using accuracy, sensitivity, and specificity measures. Random forest had the highest accuracy rate of 95.83%, while quadratic discriminant analysis had the highest sensitivity rate of 97.83%. However, the support

vector machine had a low specificity rate of 57.69%. Finally, the research presented a confusion matrix to show the true positive, false positive, false negative, and true negative rates for the pneumonia classification. The matrix indicated that there were 45 true positives, one false positive, and zero false negatives, indicating that the algorithms correctly identified pneumonia cases in most instances. Overall, this research demonstrated that machine learning algorithms can accurately classify X-ray images as normal or pneumonia, with quadratic discriminant analysis being the most effective algorithm in this study as they suggest that machine learning can be used to accurately identify and diagnose the condition.

Future Scope

In future research, it will be imperative to investigate the scalability and generalizability of models developed or trained on small datasets when applied to larger datasets. Furthermore, future investigations may delve into techniques for optimizing model performance on larger datasets, including but not limited to fine-tuning model architectures, incorporating transfer learning methodologies, or exploring ensemble methods to leverage the increased data volume effectively.

References

- Abiyev, R.H. and Ma'aitah, M.K.S. (2018). Deep convolutional neural networks for chest diseases detection. *Journal of healthcare engineering*. <https://doi.org/10.1155/2018/4168538>
- Arya, M., Mittal, N., and Singh, G. (2018). Texture-based feature extraction of smear images for the detection of cervical cancer. *IET Computer Vision*, 12(8):1049-1059. <https://doi.org/10.1049/iet-cvi.2018.5349>
- Ayan, E. and Ünver, H. Murat (2019). Diagnosis of pneumonia from chest X-ray images using deep learning. In: 2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT); Istanbul, Turkey, p. 1-5. Ieee. <https://doi.org/10.1109/EBBT.2019.8741582>
- Bouch, C. and Williams, G. (2006). Recently published papers: Pneumonia, hypothermia and the elderly. *Crit. Care*, 10(5): 167. <https://doi.org/10.1186/cc5049>
- Chouhan, V., Singh, S.K., Khamparia, A., Gupta, D., Tiwari, P., Moreira, C., Damaševičius, R., and de Albuquerque, V.H.C. (2020). A novel transfer learning based approach for pneumonia detection in chest X-ray images. *Applied Sciences*, 10(2). <https://doi.org/10.3390/app10020559>
- Cohen, J.P., Bertin, P., and Frappier, V. (2019). Chester: A web delivered locally computed chest X-ray disease prediction system. *arXiv preprint arXiv:1901.11210*. <https://doi.org/10.48550/arXiv.1901.11210>
- Dhiman, G., Chang, V., Singh, K.K., and Shankar, A. (2022). ADOPT: automatic deep learning and optimization-based approach for detection of novel coronavirus covid-19 disease using X-ray images. *Journal of Biomolecular Structure and Dynamics*, 40(13):5836-5847. <https://doi.org/10.1080/07391102.2021.1875049>
- George, G.S., Mishra, P.R., Sinha, P., and Prusty, M.R. (2023). COVID-19 detection on chest X-ray images using Homomorphic Transformation and VGG inspired deep convolutional neural network. *Biocybernetics and Biomedical Engineering*, 43(1):1-16. <https://doi.org/10.1016/j.bbe.2022.11.003>
- Haralick, R.M., Shanmugam, K., and Dinstein, I. (1973). Textural features for image classification. In: *IEEE Transactions on Systems, Man, and Cybernetics*, 3(6):610-621. <https://doi.org/10.1109/TSMC.1973.4309314>
- Hashmi, M.F., Katiyar, S., Keskar, A.G., Bokde, N.D., and Geem, Z.W. (2020). Efficient pneumonia detection in chest xray images using deep transfer learning. *Diagnostics*, 10(6):417. <https://doi.org/10.3390/diagnostics10060417>
- Ieracitano, C., Mammone, N., Versaci, M., Varone, G., Ali, A.-R., Armentano, A., Calabrese, G., Ferrarelli, A., Turano, L., Tebala, C., Hussain, Z., Sheikh, Z., Sheikh, A., Sceni, G., Hussain, A., and Morabito, F.C. (2022). Fuzzy-enhanced deep learning approach for early detection of Covid-19 pneumonia from portable chest X-ray images. *neurocomputing*. 48(7):202-215. <https://doi.org/10.1016/j.neucom.2022.01.055>
- Kallianos, K., Mongan, J., Antani, S., Henry, T., Taylor, A., Abuya, J., and Kohli, M. (2019). How far have we come? Artificial intelligence for chest radiograph interpretation. *Clinical Radiology*, 74(5):338-345. <https://doi.org/10.1016/j.crad.2018.12.015>
- Khatr, A., Jain, R., Vashista, H., Mittal, N., Ranjan, P., and Janardhanan, R. (2020). Pneumonia identification in chest x-ray images using EMD. In: Sarma, H., Bhuyan, B., Borah, S., and Dutta, N. (eds) *Trends in Communication, Cloud, and Big Data. Lecture Notes in Networks and Systems*, Springer, Singapore, 99:87-98. https://doi.org/10.1007/978-981-15-1624-5_9
- Lowitz, G.E. (1983). Can a local histogram really map texture information? *Pattern Recognit*, 16(2):141-147. [https://doi.org/10.1016/0031-3203\(83\)90017-1](https://doi.org/10.1016/0031-3203(83)90017-1)
- Punn, N.S. and Agarwal, S. (2021). Automated diagnosis of COVID-19 with limited posteroanterior chest X-ray images using fine-tuned deep neural networks. *Applied Intelligence*, 51(5):2689-2702. <https://doi.org/10.1007/s10489-020-01900-3>
- Rahman, T., Chowdhury, Muhammad, E.H., Khandakar, A., Islam, K.R., Islam, K.F., Mahbub, Z.B., Kadir, M.A., and Kashem, S. (2020). Transfer learning with deep convolutional neural network (CNN) for pneumonia detection using chest X-ray. *Applied Sciences*, 10(9):3233. <https://doi.org/10.3390/app10093233>
- Rajaraman, S., Candemir, S., Kim, I., Thoma, G., and Antani, S. (2018). Visualization and interpretation of convolutional neural network predictions in detecting pneumonia in pediatric chest radiographs. *Applied Sciences*, 8(10):1715. <https://doi.org/10.3390/app8101715>
- Saraiva, A.A., Santos, D.B.S. Costa, N.J.C., Sousa, J.V.M., Ferreira, N.M.F., Valente, A., and Soares, S. (2019). Models of learning to classify X-ray images for the detection of pneumonia using neural networks. *Bioimaging (Bristol. Print)*, 76-83. <https://doi.org/10.5220/0007346600760083>
- Saravagi, D., Agrawal, S., Saravagi, M., and Rahman, M.H. (2022). Diagnosis of lumbar spondylolisthesis using a pruned CNN model. *Computational and Mathematical Methods in Medicine*. <https://doi.org/10.1155/2022/2722315>
- Sharma, S., Khanra, P., and Ramkumar, K.R. (2021). Performance analysis of biomass energy using machine and deep learning approaches. *Journal of Physics:*

- Conference Series, 2089(1):012003. <https://doi.org/10.1088/1742-6596/2089/1/012003>
- Singh, S. and Tripathi, B.K. (2022). Pneumonia classification using quaternion deep learning. *Multimedia Tools and Applications*, 81:1743-1764. <https://doi.org/10.1007/s11042-021-11409-7>
- Sirazitdinov, I., Kholiavchenko, M., Mustafaev, T., Yixuan, Y., Kuleev, R., and Ibragimov, B. (2019). Deep neural network ensemble for pneumonia localization from a large-scale chest X-ray database. *Computers & Electrical Engineering*, 78:388-399. <https://doi.org/10.1016/j.compeleceng.2019.08.004>
- Sourab, S.Y. and Kabir, M.A. (2022). A comparison of hybrid deep learning models for pneumonia diagnosis from chest radiograms. *Sensors International*, 3:100167. <https://doi.org/10.1016/j.sintl.2022.100167>
- Stephen, O., Sain, M., Maduh, U.J., and Jeong, D.-U. (2019). An efficient deep learning approach to pneumonia classification in healthcare. *Journal of Healthcare Engineering*, Article ID 4180949 <https://doi.org/10.1155/2019/4180949>
- Stokes, K., Castaldo, R., Franzese, M., Salvatore, M., Fico, G., Pokvic, L.G., Badnjevic, A., and Pecchia, L. (2021). A machine learning model for supporting symptom-based referral and diagnosis of bronchitis and pneumonia in limited resource settings. *Biocybernetics and Biomedical Engineering*, 41(4):1288-1302. <https://doi.org/10.1016/j.bbe.2021.09.002>
- Toğaçar, M., Ergen, B., Cömert, Z., and Özyurt, F. (2020). A deep feature learning model for pneumonia detection applying a combination of mRMR feature selection and machine learning models. *IRBM*, 41(4):212-222. <https://doi.org/10.1016/j.irbm.2019.10.006>
- Varshni, D., Thakral, K., Agarwal, L., Nijhawan, R., and Mittal, A. (2019). Pneumonia detection using CNN based feature extraction. In: 2019 IEEE International Conference On Electrical, Computer And Communication Technologies (ICECCT), Coimbatore, India, IEEE, p. 1-7. <https://doi.org/10.1109/ICECCT.2019.8869364>
- Yang, F., Hamit, M., Yan, C.B., Yao, J., Kutluk, A., Kong, X.M., and Zhang, S.X. (2017). Feature extraction and classification on esophageal X-ray images of Xinjiang Kazak nationality. *Journal of Healthcare Engineering*, Article ID 4620732 <https://doi.org/10.1155/2017/4620732>
- Yao, S., Chen, Y., Tian, X., and Jiang, R. (2021). Pneumonia detection using an improved algorithm based on Faster R-CNN. *Computational and Mathematical Methods in Medicine*, Article ID 8854892. <https://doi.org/10.1155/2021/8854892>
- Zhang, F. (2021). Application of machine learning in CT images and X-rays of COVID-19 pneumonia. *Medicine*, 100(36):e26855. <https://doi.org/10.1097/MD.00000000000026855>