# Detection and classification of pneumonia using the Orange3 data mining tool

**Muneera Altayeb[1], Areen Arabiat[1], Amani Al-Ghraibah[2]**
[1]Communications and Computer Engineering Department, Faculty of Engineering, Al-Ahliyya Amman University, Amman, Jordan
[2]Medical Engineering Department, Faculty of Engineering, Al-Ahliyya Amman University, Amman, Jordan

**Article Info**

***Article history:***

Received Jun 8, 2024
Revised Jul 17, 2024
Accepted Aug 6, 2024

***Keywords:***

Mel frequency cepstral coefficient
Adaptive boosting
Decision trees
Gradient boosting
Random forest
Orang3 data mining tool

**ABSTRACT**

A chest X-ray can convey a lot about a patient's condition. However, it requires a specialized and skilled doctor to determine the type of lung disease with high accuracy. Here comes the role of deep learning techniques (DL) and artificial intelligence (AI) in accelerating the process of detecting lung diseases and classifying them with high precision, which saves time and effort for the patient and the doctor alike. This work presents a proposed model for a machine learning (ML) and AI system to analyze chest X-ray images and categorize them into four cases normal, viral pneumonia, bacterial pneumonia, and coronavirus disease 2019 (COVID-19). The system relies on extracting Mel frequency cepstral coefficient (MFCC) features from a dataset consisting of 4,800 chest X-ray images, and then these features are used to train four basic classifiers based on the data mining tool Orange3, which are adaptive boosting (AdaBoost), decision trees (DTs), gradient boosting (GB), and random forest (RF). The model was tested and evaluated, where the AdaBoost classifier excelled with an accuracy of 100%, followed by RF with an accuracy of 99.5%. Finally, GB and DTs came with a classification accuracy of 98.5%, and 97.2%, respectively.



***Corresponding Author:***

Muneera Altayeb
Department of Communications and Computer Engineering, Faculty of Engineering, Al-Ahliyya Amman University
Al-Salt, Amman, Jordan
Email: m.altayeb@ammanu.edu.jo

## 1. INTRODUCTION

Pneumonia is one of the most dangerous types of infection that affects one or both lungs. It is contagious, as it can spread from one person to another [1]. The main cause of this disease is infection resulting from bacteria, fungi, or viruses [2]. Pneumonia often causes the alveoli in the lungs to fill with fluid or pus. The symptoms of this disease range from mild to serious, and the severity of pneumonia depends on the age of the infected person, his general health, and the cause of the infection [3], [4]. When the clinical signs are identical, it is difficult to differentiate between the forms of pneumonia and whether it is coronavirus disease 2019 (COVID-19), bacterial pneumonia, or viral pneumonia. Therefore, X-ray diagnostic imaging techniques have become one of the most accurate methods for distinguishing between the vital signs of the type of pneumonia infection [5]–[7]. Chest X-ray images can reveal a great deal about a patient's condition [8]. However, this technology requires doctors and experts to diagnose and detect the disease in its early stages. Therefore, the importance of artificial intelligence (AI) and machine learning techniques (ML) has emerged in the field of accurate diagnosis of this type of disease in the early stages [9], [10]. Deep learning (DL) models are effective techniques for extracting features from X-ray images and classifying them for early diagnosis of pneumonia infection [11]–[13].

In the last few years, researchers have employed DL, AI, and ML to analyze chest X-ray to enhance the accurate diagnosis of lung diseases. Darici *et al.* [13], attempted to provide a model to analyze a chest X-ray dataset and classify pneumonia into two categories, either bacterial pneumonia or viral pneumonia. The multiclass classification was performed based on ensemble learning and convolutional neural networks (CNN). The classification model was built from scratch rather than pre-trained models to verify the effectiveness of the two methods on medical data. An average accuracy of 95% was obtained for each model. Reddy *et al.* [14] developed a DL model that involves three steps, preprocessing, feature extraction using visual geometry group16 (VGG16), Inception v3, and residual neural networks (ResNet 50), and the last step is COVID-19 classification based on radiological images. The classification accuracy of their proposed model reached 98.80%. Shea *et al.* [15] described DL algorithms for classifying three lung diseases: pleural effusion, B-lines, and lung consolidation. Classification algorithms were developed based on CNN and long short-term memory (LSTM). The performance of these algorithms was also verified using a large and diverse data set that includes 22,400 lung scans, where the accuracy of the proposed model was verified, reaching 90%. Akgundogdu [16] presented a study using ML techniques to classify and diagnose pneumonia in children using chest X-ray images. The proposed system first used a 2D discrete wavelet transform process to extract features from the dataset. The extracted features were then used in the classification process based on the RF algorithm. Also, a 10-fold cross-validation method was used to evaluate the model performance. The results showed that the introduced method achieved great success with an accuracy of 97.11%.

Menshawy *et al.* [17] proposed a multi-class classification model based on XGBoost and AdaBoost algorithms to detect COVID-19 based on chest X-ray images. The presented model achieved a recall score of 94.0% and 96.2% for XGBoost and AdaBoost, respectively. Darici [18] presented a new technique for classifying chest X-ray images, where the proposed technique combines CNN with the AdaBoost machine learning algorithm. The basic features were extracted from the X-ray images using CNN and ResNet-152 algorithm. The classification process was then carried out using the AdaBoost algorithm, and the results showed that the classification process based on the features extracted using CNN achieved better performance with 93% F1-score, 94.5% accuracy, 94% recall, and 93% precision. Amin *et al.* [19] proposed a system that relied on pre-processing chest X-ray images and then extracting basic features from them. After that, the process of classifying basic lung diseases, coronavirus, pneumonia, and tuberculosis was carried out by training algorithms RF, K-nearest neighbor (KNN), logistic regression (LR) support vector machine (SVM), extreme gradient boosting (XGBoost), decision trees (DTs), naïve Bayes (NB), AdaBoost, and ensemble model (EM). The experimental results depicted that the proposed system achieved a classification accuracy of 98%. On the other hand, Elshennawy and Ibrahim [20] developed four different models to detect and classify pneumonia based on an LSTM, ResNet152V2, CNN, and MobileNetV2. All models are implemented and evaluated using Python. The results demonstrated the proposed framework achieved an accuracy of 99.22% for the ResNet152V2 model, while the other models CNN, MobileNetV2, and LSTM-CNN achieved an accuracy of 91%.

In our work, we present a unique and different model combining the Orange3 data mining tool used to build several classification models and MATLAB to extract features from chest X-ray in an innovative way that relies on performing several pre-processing steps. Then the images are converted into signals to extract Mel frequency cepstral coefficient (MFCC) features which are used to train the basic classifiers in this work adaptive boosting (AdaBoost), decision trees (DTs), gradient boosting (GB), and random forest (RF).

## 2. PROPOSED METHOD

The primary phases in this study began with extracting MFCC features from the dataset. of chest X-ray images which consist of different categories (normal, viral pneumonia, bacterial pneumonia, and COVID-19 cases) after applying image preprocessing using MATLAB to improve the quality of features. Then utilize the feature groups to develop an automated model using the Orange3 data mining tool which classifies the input data as normal, COVID-19, bacterial pneumonia, or viral pneumonia. Figure 1 shows the structure of the suggested model. The classifiers used in this study included AdaBoost, DTs, GB, and RF where the dataset was split into two groups: 70% as a training dataset, and 30% as a testing dataset, the assessing was conducted utilizing a 10-fold cross-validation process. The assessment of the classification process was done regarding accuracy, sensitivity, precision, and F-measure. Figure 2 shows the training model using Orange3.

### 2.1. Dataset

The open Kaggle dataset that was used in this research consists of 4,800 chest X-rays that were performed as part of the clinical care of patients at Guangzhou Medical Center [21]. Figure 3(a) shows the normal chest X-ray depicts clean lungs without any spots of aberrant opacification while Figure 3(b) shows a COVID-19 lung X-ray. Figure 3(c) shows bacterial pneumonia, which is often characterized by localized lobar consolidation, whereas viral pneumonia in Figure 3(d) shows a more diffuse interstitial appearance in both lungs.

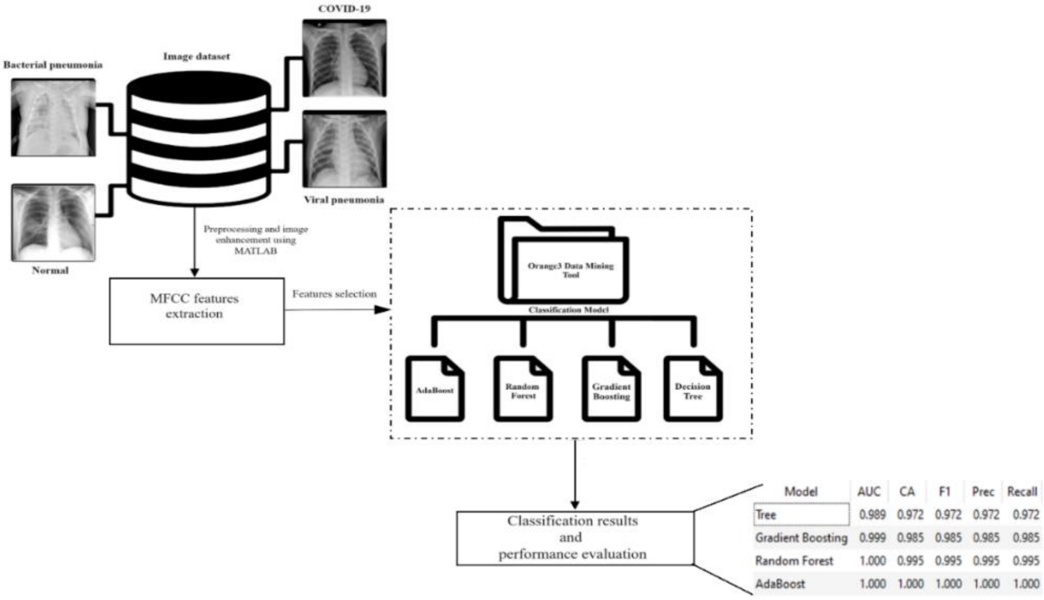


Figure 1. Proposed model structure

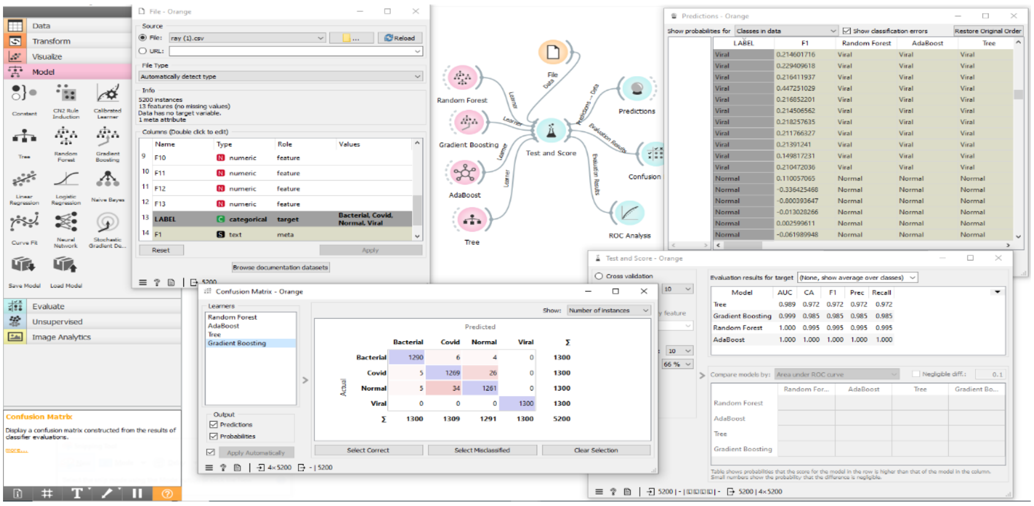

Figure 2. Orange3 training model

(a) (b) (c) (d)

Figure 3. Chest X-ray images (a) normal, (b) COVID-19, (c) bacterial Pneumonia, and (d) viral Pneumonia [21]

### 2.2. Feature extraction

Feature extraction is the process of extracting an image's visual components which in turn contributes to the image analysis and classification process [22]. There are many methods to extract features from images, this work relied on a set of image pre-processing operations using MATLAB, such as thresholding, sharpening, and applying several enhancing techniques including the Median filter and Prewitt filters to improve the extracted features. In the next stage, the enhanced images are converted into 1D signals, after which frequency-dependent features are extracted using the MFCC technique.

#### 2.2.1. Mel-frequency cepstral coefficients

MFCCs are cepstral representations of signals in which frequency bands are distributed according to a Mel-scale [23]. In the proposed model, several steps are implemented to extract MFCC coefficients starting from converting the chest X-ray into a signal followed by framing it where the signal is divided into a group of overlapping frames. The number of frames in our model is 40 of $N$=2000 samples, with successive frames separated by $L$=512 samples where adjacent frames overlap with $N - L$ samples, i.e. about 60%. After that Hamming windowing is done to smooth the edges and improve the harmonics at the beginning and end of the frame as depicted in (1):

$$x_i(n) = x(n)w(n - m_i), \quad i = 0,\ldots,K - 1 \tag{1}$$

where $w(n)$ represents the Hamming window, $x(n)$ is the input signal, $K$ is the number of frames, $m_i$ is the number of samples by which the window is shifted to produce the $i^{th}$ frame then performing the fast Fourier transform (FFT) of the resulting signal $x_i(n)$ followed by Mel filter banks, taking the logarithm of all filter bank energies to generate Mel frequency coefficients using one of the most popular formulas described in (2) [24], and then applying inverse fast Fourier transform (IFFT). Figure 4 summarizes the steps of the MFCC feature extraction process.

$$f_{mel} = 2595\, log_{10}\,(1 + \frac{f}{700}) \tag{2}$$

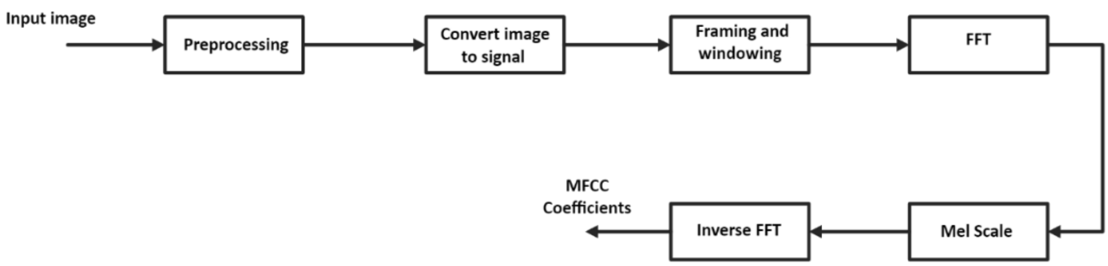


Figure 4. Steps of MFCC feature extraction

### 2.3. Classification methods

Classification and pattern recognition are one of the crucial functions in data mining and machine learning. Deep learning and machine learning are commonly used to analyze data, classify processes, and make educated decisions based on the patterns and visions extracted from the dataset [25]. In this work, the classification model was created using the data mining tool Orange3, where several supervised classification methods were used AdaBoost, DTs, GB, and RF. The following sections explain each classification method used in this research.

#### 2.3.1. Adaptive boosting (AdaBoost)

The AdaBoost algorithm is an ensemble method that combines the training of several weak classifiers based on a supervised learning method to produce a robust classifier that optimally separates positive and negative situations [26], [27]. Using partitioned or single-level decision trees, AdaBoost combines weak classifiers into a strong classifier. Data points are given equal weights in this method, with greater weights given to points with imprecise classifications. As the model continues to train until a lower error is obtained, points with higher weights are assigned greater importance in subsequent models [28], [29]. Figure 5 shows the basic principle of AdaBoost. After each stage, several new models are created according to the sampling error rate of the previous model. The sample weight is changed according to the sampling error rate. By increasing the weight of some unacceptable samples to stand out enough to be noticed in the sample guarantee. According to the above-mentioned interaction, weak learners are acquired through iterative preparation and after each model is generated this is how models 1,2,3…, and $N$ are individual models which can be known as choice trees. Finally, a weighted combination is performed to obtain a powerful learner [30].

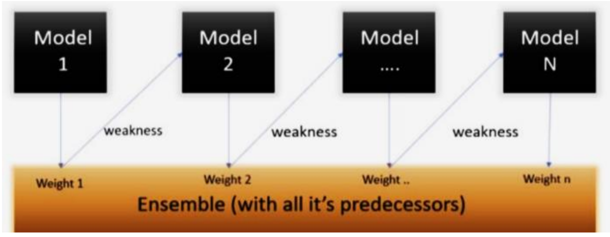


Figure 5. The basic principle of AdaBoost [31]

### 2.3.2. Decision trees (DTs)

A decision tree is a supervised learning method used for classification and regression. It trains a given sample when the outcome is already known. As depicted in Figure 6, the decision tree comprises mostly of the root node, decision node, and leaf node. The sample begins at the root node and is classified based on the rules of each layer; therefore, it cannot be separated. Decision trees require conditions to achieve the highest information gain or the lowest Gini index. Decision Trees require conditions to achieve the highest information gain or lowest Gini index [32]. The last leaf will have a label that indicates whether it is phishing or genuine based on the impurity testing, and just one tree structure flowchart will be created. In decision rules, the decision tree may be normalized as follows:

– If condition 1, condition 2, and condition 3 are met, then the result.
– The Gini impurity may be expressed in (3). Here, $k$ represents the total number of classes, and $p(i)$ shows the likelihood of receiving any class of data [33].

$$Gini\ (D) = 1 - \sum_{i=1}^{k} \sum p(i)) \quad (3)$$

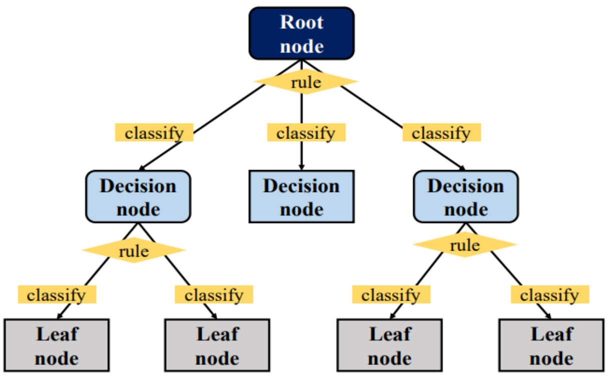


Figure 6. Decision trees structure [34]

### 2.3.3. Gradient boosting

Using an appropriate cost function, gradient boosting is a machine-learning approach that trains several weak classifiers, such as decision trees, as shown in Figure 7, by integrating more trees and resolving faults in its prior base models, the GB tree approach can improve prediction accuracy. to produce a robust model for regression and classification tasks. This improves accuracy and speed. Using a combination of many weak models, gradient boosting is a machine-learning strategy that enhances performance in regression and classification. To provide high-precision predictions, mean square error (MSE) is used as a cost function to decrease bias error and maximize prediction accuracy [35]. The GB model is described in (4):

$$Fm(x) = \sum_{m=1}^{M} \gamma_m \; h_m(x) \quad (4)$$

where $h_m(x)$ represent weak learners and $\gamma_m$ denotes the learning rate [36].

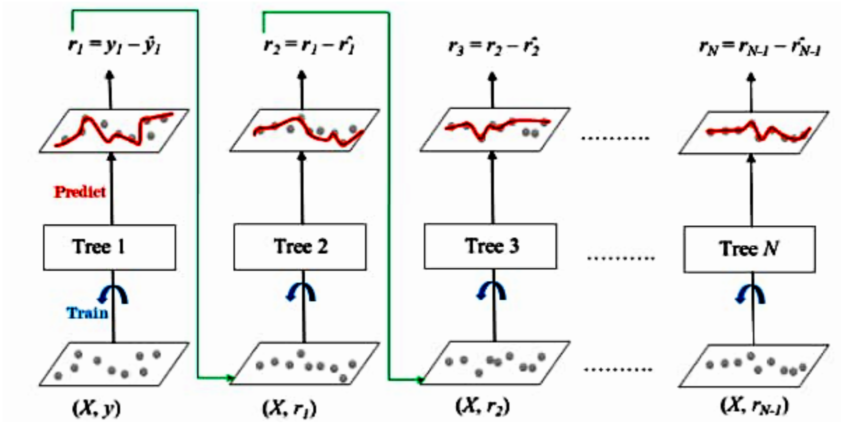


Figure 7. Structure of gradient boosting [37]

#### 2.3.4. Random forest (RF)

The RF method is a tree-predictor setup used for unsupervised learning, regression, and classification. It is a modified form of bagged trees. It employs branches of random data and randomized regression trees, and it is based on binary decision trees in an ensemble-classifier system. The "out of bag" (OOB) data is used by the algorithm to assess general faults. The quantity of terminal nodes, trees, and variables required for RF development [38]. The final decision is made using the voting process as described in Figure 8. The expected results of most trees are divided into features, each with a different information gain and entropy. Compared to a single D-tree, RF is more stable and dependable [39].

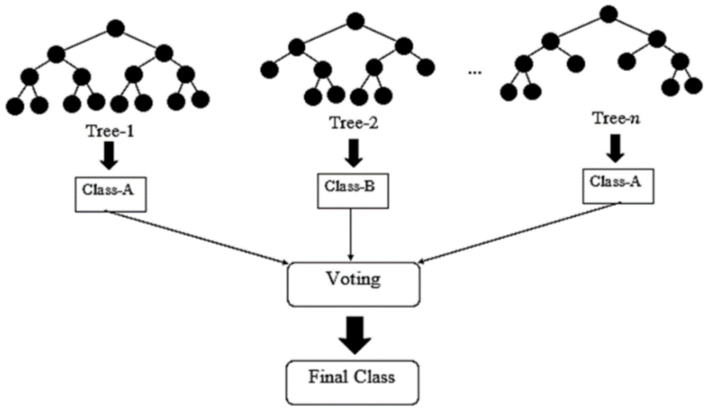


Figure 8. Random forest classifier illustration [40]

## 3. CLASSIFICATION RESULTS

Analyzing the effectiveness of the suggested model in forecasting is important after testing the validation of the main model assumptions. Accordingly, evaluation measures were used to assess the adequacy of the proposed models. For this purpose, the confusion matrix is an effective tool for determining how well the classification and prediction algorithms perform. It calculates the amounts of true positive (TP), true negative (TN), false positive (FP), and false negative (FN) in accuracy rate calculations [41]. In this work, the efficiency of the suggested model was evaluated utilizing F-measure, accuracy, precision, and sensitivity as shown in Table 1.

In this research, several algorithms were used to classify chest X-ray images, including AdaBoost, DTs, GB, and RF. Performance evaluation for each algorithm is carried out using a confusion matrix as depicted in Figure 9, where Figure 9(a) represents the confusion matrix for RF, and Figure 9(b) represents the confusion matrix for AdaBost, while Figure 9(c) describes the confusion matrix of DTs and the confusion matrix for GB is expressed in Figure 9(d). What can be observed from these results is that AdaBoost excelled with a classification accuracy of 100%, while the classification accuracy for RF reached 99.5%, followed by GB and DTs with a classification accuracy of 98.5% and 97.2%, respectively. On the other hand, it can be observed that AdaBoost outperforms all classifiers by 100% for all performance metrics; accuracy, precision, sensitivity, and F-measure as shown in Table 2. An analysis was conducted to compare the performance of all classifiers, and the results are shown in Figure 10. On the other hand, the result accuracy demonstrated in this paper is superior to previous research. In [13], the accuracy rate was 95%, in [14], the accuracy rate was 98.80%, and in [15], the accuracy rate was 90%. However, in this study, the accuracy rate reached 100%.

Table 1. Performance evaluation metrics [42]

| Metrics | Equations |
|---|---|
| Accuracy | $\frac{TP+TN}{TP+TN+FP+FN}$ |
| Sensitivity | $\frac{TP}{TP+FN}$ |
| Precision | $\frac{TP}{TP+FP}$ |
| F-measure | $\frac{2*\text{Precision}*\text{Sensitivity}}{\text{Precision}+\text{Sensitivity}}$ |

Table 2. Classification results and performance evaluation

| Model | AUC | Accuracy | F-measure | Precision | Sensitivity |
|---|---|---|---|---|---|
| Decision tree | 0.989 | 0.972 | 0.972 | 0.972 | 0.972 |
| Gradient boosting | 0.999 | 0.985 | 0.985 | 0.985 | 0.985 |
| Random forest | 1.000 | 0.995 | 0.995 | 0.995 | 0.995 |
| AdaBoost | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |


Figure 9. Confusion matrix evaluation (a) confusion matrix of RF, (b) confusion matrix AdaBoost, (c) confusion matrix DTs, and (d) confusion matrix GB

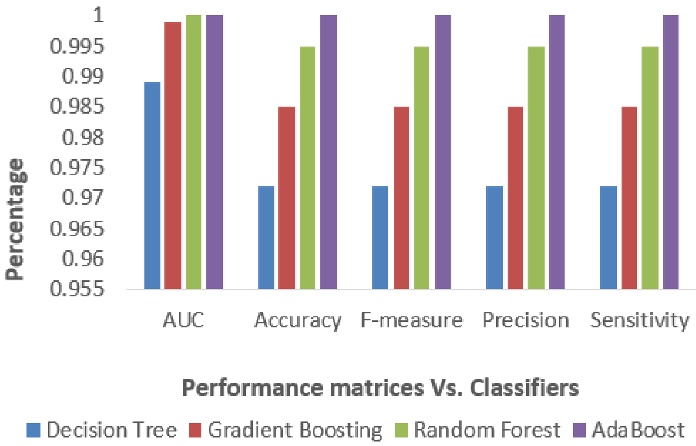


Figure 10. performance matrices vs classifiers

## 4. CONCLUSION

this work reviewed and tested several classification methods for detecting and classifying pneumonia based on 4,800 chest X-ray images obtained from Kaggle. This study depended on using MATLAB to extract MFCC features from images. After that, a classification model was built from 4 basic classifiers AdaBoost, DTs, GB, and RF based on the data mining tool Orange3. the results indicated the efficiency of the features retrieved from images in classifying X-ray images into four categories: Covid-19, bacterial pneumonia, viral pneumonia, and normal. A comparative analysis was conducted to determine the best classifier among the four methodologies, and the performance comparison revealed that the AdaBoost classifier obtained the best accuracy of 100%, followed by RF, GB, and DTs with accuracy of 99.5%, 98.5%, and 97.2%, respectively.

## REFERENCES

[1] M. F. Hashmi, S. Katiyar, A. G. Keskar, N. D. Bokde, and Z. W. Geem, “Efficient pneumonia detection in chest xray images using deep transfer learning,” *Diagnostics*, vol. 10, no. 6, Jun. 2020, doi: 10.3390/diagnostics10060417.

[2] G. J. Chowdary *et al.*, “Impact of machine learning models in pneumonia diagnosis with features extracted from chest X-rays using VGG16,” *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, vol. 12, no. 5, pp. 1521–1530, Apr. 2021, doi: 10.17762/turcomat.v12i5.2119.

[3] I. A. Ahmed, E. M. Senan, H. S. A. Shatnawi, Z. M. Alkhraisha, and M. M. A. Al-Azzam, “Multi-techniques for analyzing X-ray images for early detection and differentiation of pneumonia and tuberculosis based on hybrid features,” *Diagnostics*, vol. 13, no. 4, p. 814, Feb. 2023, doi: 10.3390/diagnostics13040814.

[4] S. Singh, M. Kumar, A. Kumar, B. K. Verma, K. Abhishek, and S. Selvarajan, “Efficient pneumonia detection using vision transformers on chest X-rays,” *Scientific Reports*, vol. 14, no. 1, Jan. 2024, doi: 10.1038/s41598-024-52703-2.

[5] S. P. Koyyada and T. P. Singh, “A systematic survey of automatic detection of lung diseases from chest X-ray images: COVID-19, pneumonia, and tuberculosis,” *SN Computer Science*, vol. 5, no. 2, Jan. 2024, doi: 10.1007/s42979-023-02573-8.

[6] R. Prince, Z. Niu, Z. Y. Khan, M. Emmanuel, and N. Patrick, “COVID-19 detection from chest X-ray images using CLAHE-YCrCb, LBP, and machine learning algorithms,” *BMC Bioinformatics*, vol. 25, no. 1, Jan. 2024, doi: 10.1186/s12859-023-05427-5.

[7] M. V. Sanida, T. Sanida, A. Sideris, and M. Dasygenis, “An advanced deep learning framework for multi-class diagnosis from chest X-ray images,” *J*, vol. 7, no. 1, pp. 48–71, Jan. 2024, doi: 10.3390/j7010003.

[8] T. Rajasenbagam, S. Jeyanthi, and J. A. Pandian, “Detection of pneumonia infection in lungs from chest X-ray images using deep convolutional neural network and content-based image retrieval techniques,” *Journal of Ambient Intelligence and Humanized Computing*, Mar. 2021, doi: 10.1007/s12652-021-03075-2.

[9] S. Goyal and R. Singh, “Detection and classification of lung diseases for pneumonia and Covid-19 using machine and deep learning techniques,” *Journal of Ambient Intelligence and Humanized Computing*, vol. 14, no. 4, pp. 3239–3259, Sep. 2023, doi: 10.1007/s12652-021-03464-7.

[10] G. M. M. Alshmrani, Q. Ni, R. Jiang, H. Pervaiz, and N. M. Elshennawy, “A deep learning architecture for multi-class lung diseases classification using chest X-ray (CXR) images,” *Alexandria Engineering Journal*, vol. 64, pp. 923–935, Feb. 2023, doi: 10.1016/j.aej.2022.10.053.

[11] M. Constantinou, T. Exarchos, A. G. Vrahatis, and P. Vlamos, “COVID-19 classification on chest X-ray images using deep learning methods,” *International Journal of Environmental Research and Public Health*, vol. 20, no. 3, Jan. 2023, doi: 10.3390/ijerph20032035.

[12] R. R. H. Khan, R. Zaki, A. Zaki, J. Najjar Mojarrab, and A. Zahid, “Transient arrhythmia in a patient with human granulocytic anaplasmosis: an uncanny presentation,” *Cureus*, Feb. 2021, doi: 10.7759/cureus.13241.

[13] M. B. Darici, Z. Dokur, and T. Olmez, “Pneumonia detection and classification using deep learning on chest x-ray images,” *International Journal of Intelligent Systems and Applications in Engineering*, vol. 8, no. 4, pp. 177–183, Dec. 2020, doi: 10.18201/ijisae.2020466310.

[14] A. S. K. Reddy *et al.*, “Multi-modal fusion of deep transfer learning based COVID-19 diagnosis and classification using chest X-

ray images," *Multimedia Tools and Applications*, vol. 82, no. 8, pp. 12653–12677, Sep. 2023, doi: 10.1007/s11042-022-13739-6.
[15] D. E. Shea *et al.*, "Deep learning video classification of lung ultrasound features associated with pneumonia," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, Jun. 2023, vol. 2023-June, pp. 3103–3112, doi: 10.1109/CVPRW59228.2023.00312.
[16] A. Akgundogdu, "Detection of pneumonia in chest X-ray images by using 2D discrete wavelet feature extraction with random forest," *International Journal of Imaging Systems and Technology*, vol. 31, no. 1, pp. 82–93, Oct. 2021, doi: 10.1002/ima.22501.
[17] L. Menshawy, A. H. Eid, and R. F. Abdel-Kader, "Enhanced COVID-19 classification using ensemble meta-algorithms on chest X-ray images," in *Engineering Solutions Toward Sustainable Development*, Springer Nature Switzerland, 2024, pp. 485–497.
[18] M. B. Darici, "Performance analysis of combination of CNN-based models with AdaBoost algorithm to diagnose COVID-19 disease," *Politeknik Dergisi*, vol. 26, no. 1, pp. 179–190, Mar. 2023, doi: 10.2339/politeknik.901375.
[19] S. U. Amin, S. Taj, A. Hussain, and S. Seo, "An automated chest X-ray analysis for COVID-19, tuberculosis, and pneumonia employing ensemble learning approach," *Biomedical Signal Processing and Control*, vol. 87, Jan. 2024, doi: 10.1016/j.bspc.2023.105408.
[20] N. M. Elshennawy and D. M. Ibrahim, "Deep-pneumonia framework using deep learning models based on chest X-ray images," *Diagnostics*, vol. 10, no. 9, Aug. 2020, doi: 10.3390/diagnostics10090649.
[21] P. Mooney, "Chest X-ray images (Pneumonia)," *Kaggle*, 2018. https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia (accessed Mar. 09, 2024).
[22] G. Dhiman *et al.*, "Multi-modal active learning with deep reinforcement learning for target feature extraction in multi-media image processing applications," *Multimedia Tools and Applications*, vol. 82, no. 4, pp. 5343–5367, Feb. 2023, doi: 10.1007/s11042-022-12178-7.
[23] M. Altayeb and A. Al-Ghraibah, "Classification of three pathological voices based on specific features groups using support vector machine," *International Journal of Electrical and Computer Engineering*, vol. 12, no. 1, pp. 946–956, Feb. 2022, doi: 10.11591/ijece.v12i1.pp946-956.
[24] A. S. Haq, M. Nasrun, C. Setianingsih, and M. A. Murti, "Speech recognition implementation using MFCC and DTW algorithm for home automation," *International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*, vol. 7, pp. 78–85, Oct. 2020, doi: 10.11591/eecsi.v7.2041.
[25] A. Arabiat and M. Altayeb, "An automated system for classifying types of cerebral hemorrhage based on image processing techniques," *International Journal of Electrical and Computer Engineering*, vol. 14, no. 2, pp. 1594–1603, Apr. 2024, doi: 10.11591/ijece.v14i2.pp1594-1603.
[26] J. Zheng, D. Lin, Z. Gao, S. Wang, M. He, and J. Fan, "Deep learning assisted efficient AdaBoost algorithm for breast cancer detection and early diagnosis," *IEEE Access*, vol. 8, pp. 96946–96954, 2020, doi: 10.1109/ACCESS.2020.2993536.
[27] Y. Xie, T. Wang, M. K. Jeong, and G. T. Lee, "Multi-source adaptive thresholding AdaBoost with application to virtual metrology," *International Journal of Production Research*, vol. 62, no. 17, pp. 6344–6359, Feb. 2024, doi: 10.1080/00207543.2024.2314151.
[28] T. R. Mahesh *et al.*, "AdaBoost ensemble methods using K-Fold cross validation for survivability with the early detection of heart disease," *Computational Intelligence and Neuroscience*, vol. 2022, pp. 1–11, Apr. 2022, doi: 10.1155/2022/9005278.
[29] S. Usha and S. Kanchana, "Revived ant colony optimization-based AdaBoost algorithm for heart disease and diabetes (HDD) prediction," *Journal of Theoretical and Applied Information Technology*, vol. 101, no. 4, pp. 1552–1567, 2023.
[30] J. Sinaga and B. Sinaga, "Data mining classification of filing credit customers without collateral with K-Nearest neighbor algorithm (Case study: PT. BPR Diori Double)," *Journal Of Computer Networks, Architecture and High Performance Computing*, vol. 2, no. 2, pp. 204–210, Jan. 2020, doi: 10.47709/cnapc.v2i2.401.
[31] P. Thamilselvan, "Lung cancer prediction and classification using AdaBoost data mining algorithm," *International Journal of Computer Theory and Engineering*, vol. 14, no. 4, pp. 149–154, 2022, doi: 10.7763/IJCTE.2022.V14.1322.
[32] F. E. Botchey, Z. Qin, and K. Hughes-Lartey, "Mobile money fraud prediction—a cross-case analysis on the efficiency of support vector machines, gradient boosted decision trees, and naïve Bayes algorithms," *Information*, vol. 11, no. 8, Jul. 2020, doi: 10.3390/info11080383.
[33] N. H. A. Malek, W. F. W. Yaacob, Y. B. Wah, S. A. Md Nasir, N. Shaadan, and S. W. Indratno, "Comparison of ensemble hybrid sampling with bagging and boosting machine learning approach for imbalanced data," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 29, no. 1, pp. 598–608, Jan. 2023, doi: 10.11591/ijeecs.v29.i1.pp598-608.
[34] C. Wang, S. Xu, and J. Yang, "AdaBoost algorithm in artificial intelligence for optimizing the IRI prediction accuracy of asphalt concrete pavement," *Sensors*, vol. 21, no. 17, Aug. 2021, doi: 10.3390/s21175682.
[35] R. Sibindi, R. W. Mwangi, and A. G. Waititu, "A boosting ensemble learning based hybrid light gradient boosting machine and extreme gradient boosting model for predicting house prices," *Engineering Reports*, vol. 5, no. 4, Nov. 2023, doi: 10.1002/eng2.12599.
[36] O. Lyashevska, F. Malone, E. MacCarthy, J. Fiehler, J. H. Buhk, and L. Morris, "Class imbalance in gradient boosting classification algorithms: application to experimental stroke data," *Statistical Methods in Medical Research*, vol. 30, no. 3, pp. 916–925, Dec. 2021, doi: 10.1177/0962280220980484.
[37] N. Aksoy and I. Genc, "Predictive models development using gradient boosting based methods for solar power plants," *Journal of Computational Science*, vol. 67, Mar. 2023, doi: 10.1016/j.jocs.2023.101958.
[38] Q. B. Pham *et al.*, "Predicting landslide susceptibility based on decision tree machine learning models under climate and land use changes," *Geocarto International*, vol. 37, no. 25, pp. 7881–7907, Oct. 2022, doi: 10.1080/10106049.2021.1986579.
[39] I. Shanmugam, J. Rethnaraj, and G. Mani, "Selection of crop varieties and yield prediction based on phenotype applying deep learning," *International Journal of Electrical and Computer Engineering*, vol. 13, no. 6, pp. 6806–6816, Dec. 2023, doi: 10.11591/ijece.v13i6.pp6806-6816.
[40] G. D. Apriliana, T. Siswantining, D. Sarwinda, and A. Bustamam, "Analysis of data mining for classification of Obstructive Sleep Apnea in chronic obstructive pulmonary disease patients," in *AIP Conference Proceedings*, 2020, vol. 2242, doi: 10.1063/5.0007884.
[41] D. Krstinić, M. Braović, L. Šerić, and D. Božić-Štulić, "Multi-label classifier performance evaluation with confusion matrix," in *Computer Science & Information Technology*, Jun. 2020, pp. 01–14, doi: 10.5121/csit.2020.100801.
[42] A. Almomani *et al.*, "Ensemble-based approach for efficient intrusion detection in network traffic," *Intelligent Automation and Soft Computing*, vol. 37, no. 2, pp. 2499–2517, 2023, doi: 10.32604/iasc.2023.039687.

**BIOGRAPHIES OF AUTHORS**

**Muneera Altayeb** earned her B.Sc. in computer engineering in 2007, and her M.Sc. in communications engineering from the University of Jordan in 2010. She has been working as a lecturer in the Department of Communications and Computer Engineering at Al-Ahliyya Amman University since 2015, in addition to her administrative experience as assistant dean of the Faculty of Engineering during the period (2020-2023). Her research interests focus on the following areas digital signals and image processing, machine learning, robotics, and artificial intelligence. She can be contacted at email: m.altayeb@ammanu.edu.jo.

**Areen Arabiat** obtained B.Sc. in computer engineering in 2005 from al Balqa Applied University (BAU), and her M.Sc. in intelligent transportation systems (ITS) from Al-Ahliyya Amman University (AAU) in 2022. She is currently a computer lab supervisor at the Faculty of Engineering/Al-Ahliyya Amman University (AAU) since 2013. Her research interests are focused on the following areas machine learning, data mining, artificial intelligence, and image processing. She can be contacted at email: a.arabiat@ammanu.edu.jo.

**Amani Al-Ghraibah** earned her B.Sc. in biomedical engineering in 2008 from Jordan University of Science and Technology/Jordan. She got her Ph.D. and MSc in electrical and computer engineering from New Mexico State University/USA in 2012 and 2015, respectively. She is currently an assistant professor in the Medical Engineering Department at Al-Ahliyya Amman University. Her research interest is in the fields of digital signals and images processing and pattern recognition. She can be contacted at email: a.ghraibah@ammanu.edu.jo.