

LUNG-CENTRIC FEATURE ANALYSIS FOR ACCURATE PNEUMONIA DETECTION IN CHEST X-RAY IMAGES

Najah Alsubaie¹, Tahani Alqahtani¹, and Syarifah Bahiyah Rahayu^{2,3}

¹Department of Computer Sciences, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University (PNU), Saudi Arabia

²Department of Defence Science, Faculty of Defence Science and Technology, National Defense, Malaysia

³Cyber Security and Digital Industrial Revolution Centre, National Defense, University of Malaysia, Malaysia

ABSTRACT

Pneumonia, a lung inflammation and consolidation disorder, poses diagnostic challenges necessitating accurate detection. This paper introduces an innovative automated approach using segmented lung morphology and texture attributes from Chest X-ray (CXR) images. Unlike conventional methods analyzing the entire CXR, our focus narrows to segmented lung regions. Discriminative ranking of extracted features enhances the categorization of CXR images into pneumonia and normal cases. Diverse machine learning classifiers are evaluated, yielding a compelling 86% accuracy—validating our method's efficacy in distinguishing pneumonia from normal cases. This study offers a robust and efficient diagnostic avenue for improved pneumonia differentiation.

KEYWORDS

lung, chest X-ray, machine learning, pneumonia, morphology, texture

1. INTRODUCTION

Pneumonia is a common respiratory infection that causes inflammation of the air sacs in one or both lungs. The survival rate for pneumonia can vary depending on factors such as patient age, overall health, as well as the severity of the infection [23, 34]. Common symptoms of pneumonia include cough, fever, chest pain, and difficulty breathing. Early diagnosis of pneumonia is important to prevent potential complications and improve survival rates. Treatment for pneumonia usually involves antibiotics to clear the infection and supportive care to manage symptoms such as fever and cough. In severe cases, hospitalization may be necessary to provide more intensive treatment and support [23, 34, 5, 3, 2]. X-ray imaging is the most common diagnostic tool for the inspection of any inflammation or fluid in the lungs. The X-ray images of pneumonia patients normally show an increased density of the lungs, either in a lobar or diffuse pattern, due to inflammation and fluid buildup. In some cases, this may appear as a "ground-glass" appearance or a consolidation of lung tissue. White patches in the chest X-ray are normally an indication of the infection. In contrast, healthy X-ray images show clear lungs without any signs of inflammation or fluid accumulation, see Figure 1.

One of the primary difficulties radiologists encounter during pneumonia diagnosis is the high variability in the appearance of pneumonia in X-ray images. Pneumonia can present in various patterns, densities, and sizes, making it challenging to develop standard diagnostic criteria. Moreover, various confounding factors such as suboptimal imaging conditions and overlapping structures such as blood vessels and underlying lung structures can further complicate the diagnosis process. Another significant challenge in pneumonia diagnosis from X-ray images is the subjective interpretation of the radiologists. Interpretation of the images can vary depending on individual experience, knowledge, and expectations of the radiologists. The differing opinions of radiologists can lead to delayed or inaccurate diagnoses, impacting patient outcomes [25]. Automated detection of pneumonia could improve the diagnosis process, reduce the time and effort required by radiologists and provide consistent results. Moreover, the automated methods can generalize to wide variations within the X-ray images. Hence, provides a more reliable and accurate outcome [46].

Several approaches have been proposed for chest X-ray classification. Yee *et al.*[51] investigated three models for pneumonia detection on chest x-ray images: Inception V3 [43, 52], CNN, and SVM. The SVM model outperformed the others, boasting an impressive 93.1% accuracy. Mathew *et al.*[24], used the Histogram of Oriented Gradients (HOG) to classify and detect COVID-19 cases from chest X-rays using three classification algorithms: SVM, KNN, and RF. The results demonstrated that SVM had maximum accuracy.

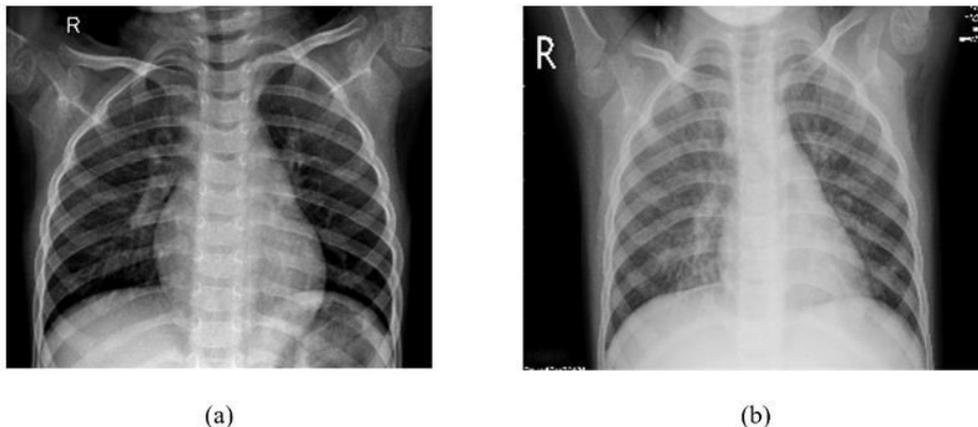


Figure. 1. Two chest x-ray images for (a) Normal case, and (b) Pneumonia [21, 2].

of 98.12%. Chandra *et al.*[7], used multilayer perceptron, random forests, and logistic regression using a publicly available dataset named Chest X-ray14 [1] to classify X-rays as either normal or pneumonia. The logistic regression classifier achieved the highest accuracy of 95% on the test set. Mamlook *et al.*[4], experimented with RF, KNN, DT, Xgboost, Adaboost and CNN for the classification of chest X-ray images as normal or abnormal. They illustrated that the deep learning model outperforms the other machine learning techniques. The CNN model achieved the highest accuracy of 98.46%. Hussain *et al.* [18] extracted several morphological and textural features to classify chest X-ray images into COVID-19, bacterial pneumonia, viral pneumonia and normal. They compared several conventional machine learning methods for the best classification accuracy.

Deep-learning algorithms have been widely used for chest X-ray image classification. For instance, Ibrahim *et al.* [19], compared several deep learning models for the classification of COVID-19, pneumonia, and lung cancer from a combination of chest X-ray and Computed Tomography (CT) images. Four architectures were considered: VGG followed by CNN, ResNet

followed by Gated Recurrent Unit (GRU), ResNet, and ResNet followed by Bidirectional GRU (Bi-GRU) [12]. They found that VGG19 + CNN model performs better than the other three suggested models [29]. Chouhan et al. [9], adopted transfer learning by extracting features using the ImageNet-trained AlexNet [36, 49], Dense Convolutional Network (DenseNet121) [17], Inception V3 [43, 52] GoogLeNet [42], and ResNet18 [15, 40] architectures. An ensemble model consisting of all five pre-trained models outperformed all other models. The ensemble model achieved 96.4% accuracy and 99.62% recall on unseen data from the Guangzhou Women and Children's Medical Center dataset [21]. Kundu *et al.*[22], devolved a weighted average ensemble of three convolutional neural network models, GoogLeNet [42], ResNet [15], and DenseNet [17] to classify chest X-rays into pneumonia vs normal. Using a five-fold cross-validation scheme, they evaluated the proposed method on two publicly available pneumonia X-ray datasets, provided by Kermany *et al.* [21] and the Radiological Society of North America (RSNA) [50], respectively. They obtained 98.81% and 86.85% accuracy rates on the Kermany and RSNA datasets, respectively. Mun *et al.*[30] trained several deep learning models: Xception [8], DenseNet, ResNet [15], InceptionResNetV2 [41], and VGG16 [45]. The models were trained using the Guangzhou Women and Children's Medical Centre dataset [2]. The result of the majority voting achieved an accuracy of 97.56% and 99.14% on train and test datasets, respectively.

The literature predominantly employs two classification methodologies: deep learning-based and hand-crafted-based approaches. While deep learning techniques are powerful in image classification tasks, their applicability is constrained due to their requisites of substantial datasets and considerable computational resources. This may limit the scalability and accessibility of the models for some users or scenarios [16, 47]. For the hand-crafted-based classifiers, the literature has a common drawback in which the features are extracted from the entire X-ray images. Therefore, irrelevant parts may be involved during feature calculations, such as the neck, shoulder, and abdomen. Extracting features from unrelated body parts may cause noisy data and incorrect feature representation.

In this work, we aim to overcome the limitations by extracting features solely from the segmented lungs. This approach will ensure that only the left and right lungs are considered during feature calculation while ignoring the irrelevant parts within the X-ray image. In addition, morphological and textural features are analysed to find the most representative features for the classification of pneumonia and normal X-ray images.

The literature predominantly employs two classification methodologies: deep learning-based and hand-crafted-based approaches. While deep learning techniques are powerful in image classification tasks, their applicability is constrained due to their requisites of substantial datasets and considerable computational resources. This may limit the scalability and accessibility of the models for some users or scenarios [16,47]. For the hand-crafted-based classifiers, the literature has a common drawback in which the features are extracted from the entire X-ray images. Therefore, irrelevant parts may be involved during feature calculations, such as the neck, shoulder, and abdomen. Extracting features from unrelated body parts may cause noisy data and incorrect feature representation.

In this work, we aim to overcome the aforementioned limitations by extracting features solely from the segmented lungs. This approach will ensure that only the left and right lungs are considered during feature calculation while ignoring the irrelevant parts within the X-ray image. In addition, morphological and textural features are analysed to find the most representative features for the classification of pneumonia and normal X-ray images.

2. MATERIALS AND METHODS

The proposed methodology encompasses four distinctive stages, outlined as follows:

1. **Applying a Pre-trained Deep Learning Model:** A pre-trained deep learning model was employed as the first step to facilitate lung segmentation, thereby affecting the extraction of pulmonary structures from the chest X-ray images.
2. **Extraction of Morphological and Texture Features:** In the subsequent stage, morphological and textural features were systematically extracted from both the left and right lung components within each individual X-ray image.
3. **Feature Relevance Selection:** The third phase entailed a meticulous process of feature selection, guided by a battery of statistical tests, to identify and retain the most relevant attributes.
4. **Evaluation of Machine Learning Classifiers:** The fourth stage involved a comprehensive evaluation of multiple machine learning classifiers using the chosen feature sets, with the overarching objective of attaining the highest achievable classification score.

Finally, we validate the clinical significance of the selected features.

The four stages are depicted in Figure 2.

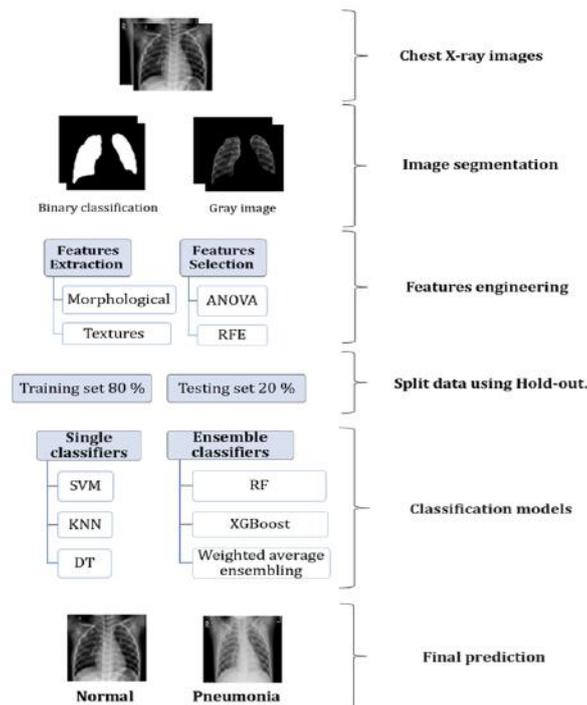


Figure. 2. Proposed framework.

2.1. Dataset

Chest X-ray images (anterior-to-posterior) were selected from retrospective cohorts of one to five-year-old paediatric patients at the Guangzhou Women and Children's Medical Centre, Guangzhou [21]. The dataset is divided into two subsets for each class (Normal/Pneumonia). There are 2797 X-ray images of the same size, 256×256. There are 1565 images for the normal

class and 1232 for the pneumonia class. All chest X-ray imaging was performed as part of routine clinical care for the patients. For the analysis of chest X-ray images, all chest radiographs were screened for quality control by removing any scans that were of poor quality or unreadable. The image diagnoses were then graded by two expert physicians [2].

2.2. Image Segmentation

We employed a pretrained U-Net model for image segmentation to extract the lungs out of the X-ray image [33]. Weights for the pre-trained model were used to separate the segmented regions into left and right lungs. Since our dataset is interior- to posterior, we labelled the segmented regions as left lung and right lung, see Figure 3.

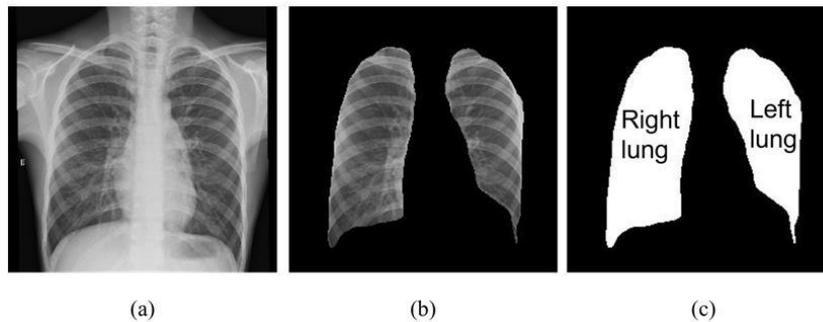


Figure. 3. Lung segmentation. (a) original image, (b) segmented lungs (c) segmented mask and lung labels.

2.3. Feature Extraction

Two types of features were extracted: morphological and texture features.

2.3.1. Morphological features of the lung encompass the structural characteristics of the lung. We extract the following morphological features for each segmented region: area, perimeter, eccentricity, convex area, bounding box area, extent, solidity, orientation, major axis length, and minor axis length.

2.3.2. Textural features related to pixel intensities and can be utilized to quantify changes in lung tissue density. This is useful in diagnosing conditions like consolidation in pneumonia X-ray images. The texture features were extracted using the GLCM techniques which measure how often pairs of pixels with specific grey levels, distances, and orientations occur in the image [14, 48]. The extracted GLCM features for both the left and right lungs are contrast, energy, homogeneity, correlation, dissimilarity, and Angular Second Moment (ASM).

The total number of features extracted per segmented lung is 16. Therefore, 32 features are computed for each X-ray image.

2.4. Features Selection

We employed two statistical methods for feature selection: ANOVA [11] and Recursive Feature Elimination (RFE) [10]. ANOVA utilises the F-test to assess variance between two groups,

specifically normal and pneumonia cases [26, 44]. The feature exhibiting the highest F-statistic is considered relevant for predicting the target variable, as illustrated by the following equation.

$$F = \frac{\sum_{i=1}^k n_i \frac{(\bar{x}_i - \bar{x})^2}{K-1}}{\sum_{i=1}^k \sum_{j=1}^{N-K} \frac{x_{ij}^2}{N-K}} \quad (1)$$

Where \bar{x}_i is the mean of i^{th} group, n_i is the number of samples in the i^{th} group. \bar{x} is the mean of all samples in the dataset, K is the number of groups, x_{ij} is the i^{th} feature of the j^{th} group. Table 3.1 shows the features with their scores according to each feature selection method. The Recursive Feature Elimination (RFE) method is designed to systematically reduce the feature set by iteratively assessing smaller subsets of features. You can find the features along with their corresponding scores for each feature selection method in Table 1.

We inspected the correlation within the computed features as shown in Figure 4. When two features display a strong correlation, it suggests the potential for predicting one variable based on another. In these instances, the model can effectively utilize just one of the correlated features, as the second one doesn't contribute supplementary information [10]. For instance, Figure 4 shows a significant positive correlation between the features, such as L_axis_minor_length with L_perimeter and L_axis_major_length. In contrast, there is a negative correlation between the features such as L_bbox_area with L_area and L_area_convex, see Figure 4.

In order to further understand the distribution of each selected feature, we plotted each selected feature per class, see Figure 5. In Figure 5 (a, b), there is a significant difference between the right-axis minor length and left-axis major length between normal and pneumonia classes. In the normal class, the median value is greater than the median value of the pneumonia class. In Figure 5 (C, D), right ASM and left ASM are notably different in the normal and pneumonia classes. In the normal class, the median value is less than the median value of the pneumonia class.

The feature selection process started by first selecting the top-ranked features. Subsequently, we systematically eliminated any feature displaying a strong correlation with previously chosen ones. In parallel, we validated our choices by subjecting the selected features to the ANOVA test. Ultimately, the set of features incorporated into our model consisted of those that excelled in both ANOVA and RFE analyses while maintaining minimal mutual correlation. These encompassed the following features: right-axis minor length, left-axis major length, extent, ASM, eccentricity, correlation, and solidity for both the left and right lung datasets. The list of selected features is described in Table 2.

Table 1. Lung features from the highest to lowest scores in both the ANOVA test and the RFE method. L and R refer to the left and right lungs, respectively. Highly correlated features were not selected.

ANOVA	Scores	RFE	Scores
L axis minor length	158.59	L energy	17
L dissimilarity	156.22	L dissimilarity	16
L homogeneity	147.16	R dissimilarity	15
R extent	141.51	R asm	14
L asm	141.48	L homogeneity	13
L energy	138.10	L asm	12
R axis major length	120.54	R energy	11
R bbox area	112.04	L contrast	10

R perimeter	107.17	L solidity	9
L area_convex	104.92	R solidity	8
L bbox_area	104.51	L bbox_area	7
L area	102.88	R homogeneity	6
L perimeter	90.81	R extent	5
L contrast	90.77	R axis minor length	4
R orientation	61.97	L perimeter	3
L axis major length	59.15	R contrast	2
R eccentricity	53.92	L orientation	1
R area_convex	45.49	L eccentricity	1
R area	38.55	L axis minor length	1
L orientation	30.86	L correlation	1
L correlation	26.63	L extent	1
R homogeneity	22.55	R area	1
R asm	19.81	L axis major length	1
R axis minor length	19.31	L area_convex	1
R dissimilarity	18.63	R area_convex	1
R energy	18.36	R correlation	1
R correlation	15.24	R orientation	1
R solidity	12.44	R eccentricity	1
L extent	11.89	R axis major length	1
L eccentricity	8.27	R perimeter	1
L solidity	2.29	R bbox_area	1
R contrast	1.87	L area	1

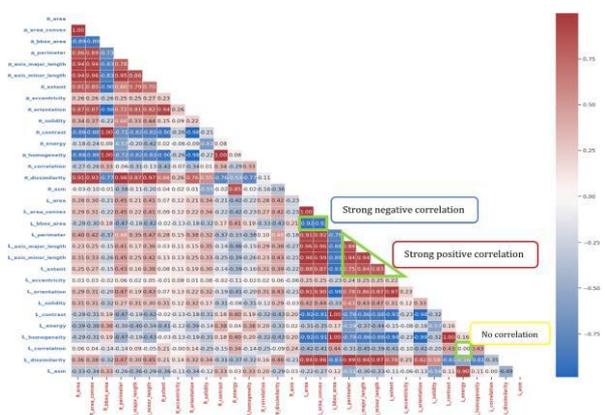


Figure. 4. Correlation coefficient for the calculated features. Strong negative correlation, strong positive correlation and no correlation are highlighted.

2.5. Classification

An imbalanced dataset, characterized by an unequal distribution of class labels, presents a common challenge in data analysis. To tackle this issue within our dataset, we adopted a random under-sampling technique [27]. This approach yielded a well-balanced dataset comprising 1232 CXR images for both the normal and pneumonia classes. The number of samples in training and

testing sets are shown in Table 3. Following this, we applied a hold-out sampling technique, randomly partitioning the dataset into distinct training and test sets for further analysis [20].

The study utilized supervised learning methodologies to classify data into different groupings. We employed various machine learning algorithms, namely Random Forest (RF) [20], Decision Trees [37], K-Nearest Neighbours (KNN) [20], and Support Vector Machine (SVM) [37]. A weighted average probability ensemble was experimented with three different machine learning models: RF, SVM, and KNN. The weighted ensemble method [22] assigns weights to multiple models based on their ability to make correct predictions. We compute the performance measures for each model (i). $A_i = precision_{(i)}$, $Recall_{(i)}$, and $F1-score_{(i)}$ of the model's predictions (x) such that i is RF, SVM, and KNN. The $Tanh$ of these metrics is then summed to derive a weight for each model, as shown in equation 2.

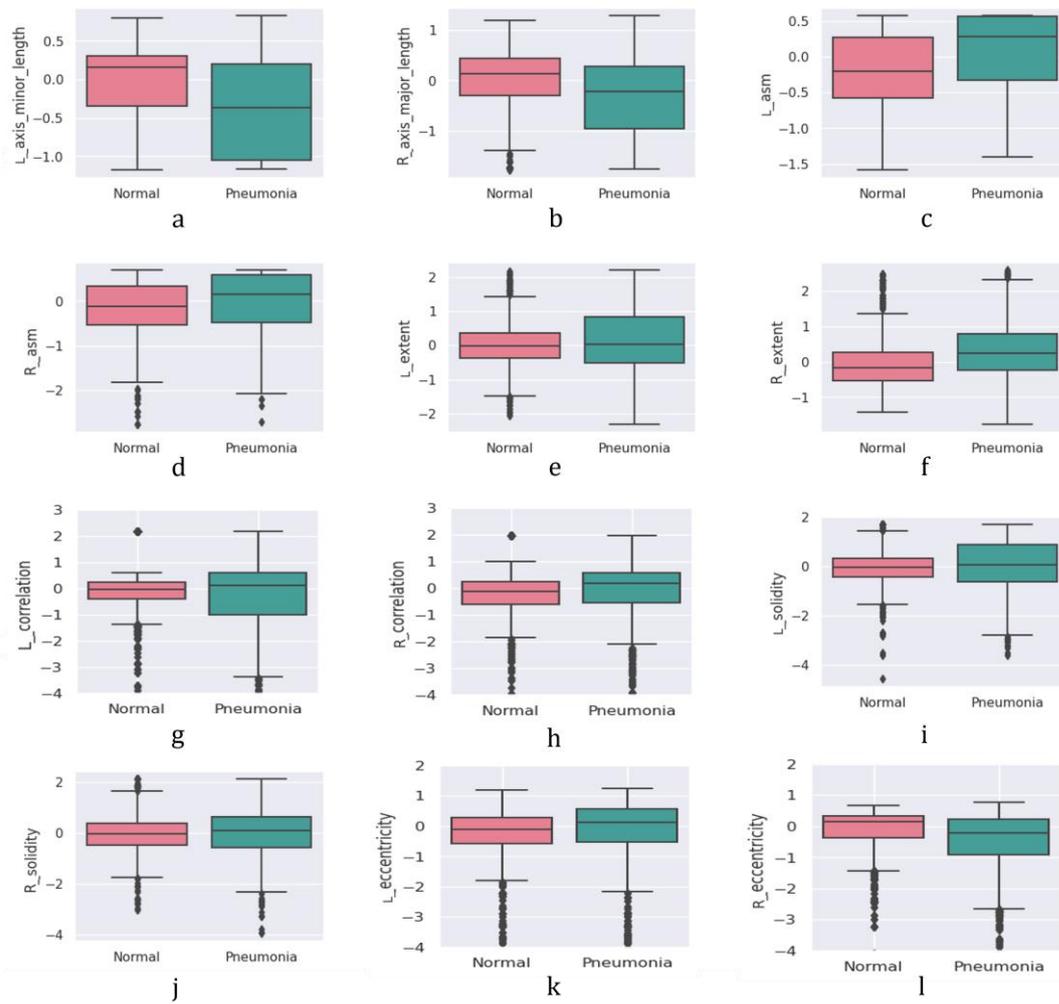


Figure. 5. Boxplot for the most important features. The boxplots (i, h, k and l) were rescaled to allow visible distribution of features in each class.

Table 2. Brief description for the selected features. [13, 28, 18]

Feature	Description
Axis major length	The length of the major axis of the best-fitting ellipse for the lung region.
Axis minor length	The length of the minor axis of the best-fitting ellipse for the lung region.
Solidity	The solidity of a region determines the pixel ratio of a convex hull.
Extent	The ratio of the lung area to the bounding box area.
Correlation	Correlation is a statistical measure of the linear relationship between the grey-level pairs in an image.
Eccentricity	The eccentricity of an ellipse is the length of its major axis divided by the distance between its foci.
Angular Second Moment (ASM)	ASM measures the uniformity of pixels.

Table 3. Training and validation split.

Hold out	Training set (80%)	Testing set (20%)	Total
Normal	992	240	1232
Pneumonia	979	253	1232
Total	1971	493	2464

The weighted sum of the predicted probabilities for each model gives the ensemble probabilities, which are then used to make the final predictions.

$$w^{(i)} = \sum_{x \in A^{(i)}} \tanh(x) = \sum_{x \in A^{(i)}} \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (2)$$

The model weights are thus allocated based on each model's ability to correctly identify the classes in the test data, allowing the ensemble to give more importance to the better-performing models. Furthermore, we employed ensemble learning techniques, such as AdaBoost [38], Gradient Boosting [6], and Extreme Gradient Boosting (XGBoost) [6].

2.6. Experiments

Several performance metrics, including accuracy, recall, precision, and F-measure, have been used to evaluate the performance of predictive models such as RF, DT, KNN, and SVM. Accuracy: The ratio of the total number of correctly classified instances to the total number of instances in the test set [20].

$$\text{Accuracy} = \frac{\text{Number of correct prediction}}{\text{Total number of prediction}} \quad (3)$$

Precision: It is the percentage between the actual number of true positive instances and the overall number of instances labelled as positive [20].

$$\text{Precision} = \frac{\text{True positives}}{\text{True positives} + \text{False positives}} \quad (4)$$

Recall: It is the percentage of true positive instances to the number of instances that belong to the positive class [20].

$$\text{Recall} = \frac{\text{True positives}}{\text{True positives} + \text{False negatives}} \quad (5)$$

F₁-measure: The mean of the measures of precision and recall [20].

$$F_1 \text{ measure} = 2 \times \frac{(\text{Precision} \times \text{Recall})}{(\text{Precision} + \text{Recall})} \quad (6)$$

Each model executes 10 times and then calculates the average metrics to ensure the consistency of the results, Table 4. We tune the parameters for each classifier using hyperparameter tuning techniques [32, 16]. Different combinations of hyperparameters were systematically searched and evaluated to find the best configuration for each model. The GridSearch[47] was utilized to identify the optimal hyperparameters for our models. The performance of each configuration was evaluated using the F1 score as the scoring metric. The hyperparameters considered for tuning varied for each model and included parameters such as the number of estimators, maximum depth, minimum sample split, learning rate, and kernel type.

The highest accuracy achieved by RF is 86% due to having typically robust outliers and the ability to manage them in an automated manner [20]. The XGBoost classifier exhibits a performance that ranks second and bears a resemblance to the gradient-boosting classifier. This could be due to its basis on the gradient boosting framework, which involves the sequential addition of new models aimed at fixing errors made by previous models [6]. The Gradient Boosting classifier exhibits similarity to RF and XGBoost classifiers in its capability of learning complex associations between features and classes [39]. The result of a weighted ensemble was affected by the low classification accuracy of the SVM, which is 78%, and KNN is 81%, Figure 6.

Table 4. Performance results of the predictive models

Classifier	Class	Average Precision	Average Recall	Average F1-score	Average Accuracy
DT	Normal	71	82	69	75
	Pneumonia	81	69	75	
AdaBoost	Normal	72	77	75	75
	Pneumonia	78	73	75	
SVM	Normal	75	81	78	78
	Pneumonia	82	76	79	
KNN	Normal	75	92	82	81
	Pneumonia	91	72	80	
Weighted average	Normal	88	74	80	83
	Pneumonia	79	91	85	
Gradient Boosting	Normal	81	88	84	85
	Pneumonia	88	81	85	
XGBoost	Normal	82	87	84	85
	Pneumonia	88	83	85	
RF	Normal	83	89	86	86
	Pneumonia	89	84	86	

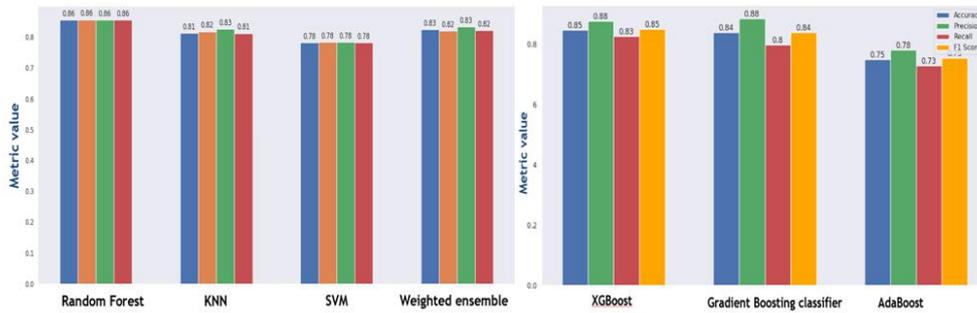


Figure. 6. Comparison of metrics for classifiers RF, KNN, SVM, Weighted ensemble, XGboost gradient boosting and Adaboost.

3. DISCUSSION

In this section, we inspect the selected features and their clinical interpretation using chest X-ray images. We noticed that the entire lung in pneumonia cases is smaller than that of healthy cases. For instance, the left-lung minor-axis length was one of the highest score features in the feature selection stage (score 195 using the ANOVA test). The box-blot Figure 5 (A) shows that the median for pneumonia cases is lower than the healthy cases. This observation refers to the loss of the normal borders between thoracic structures in the affected lungs, which appear as white spots on the X-ray images [39, 35], see Figure 7 (a,b). Similarly, the left-lung major-axis length). The box-blot 5 (B) shows that there is a significant difference in the median value in pneumonia cases compared to healthy cases. Figure 7 (c,d) shows sample images of pneumonia and healthy cases. lung tissue in the left.

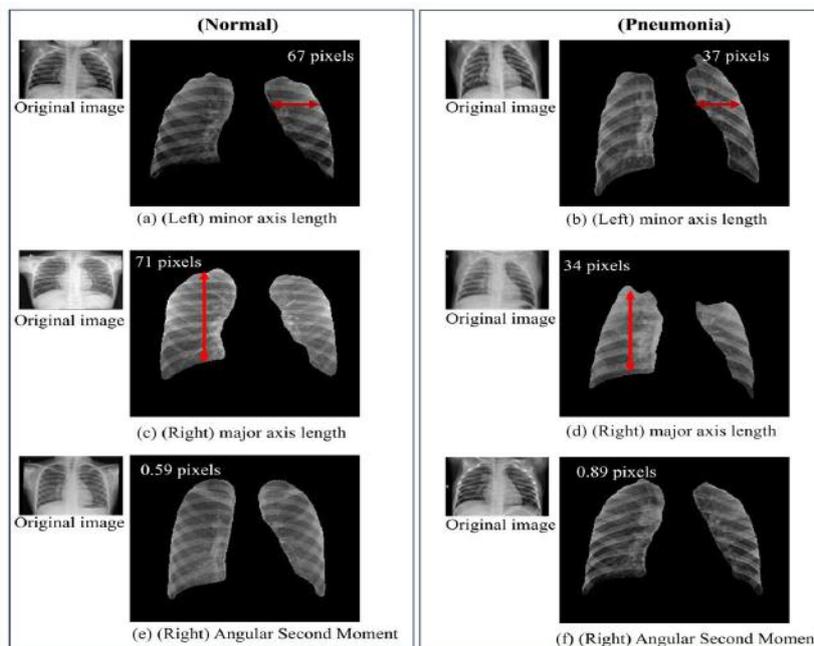


Figure. 7. Sample results for the Normal/Pneumonia classification with the top selected features highlighted. (a,b) left lung minor axis, (c,d) Right lung major axis, (c,d) Right lung Angular second moment.

Another example is the ASM (score 141 ANOVA test). The homogeneity of the grey level distribution in the image is represented by the angular second moment. When an image has excellent homogeneity or when its pixels are substantially similar, the angular second moment is high [28]. This is reflected in the values computed for pneumonia and healthy lungs, Figure 7 (c,d). The affected lungs have lower ASM compared to healthy lungs. This is due to the high density developed in the affected lungs due to the inflammatory exudate and pus replacing the air in the alveoli. This causes a high casting of shadow compared to the healthy lungs where the X-ray casts less of a shadow. Therefore, the distribution of gray levels in affected lungs is much less than the healthy lungs due to more uniformity of the gray levels.

4. CONCLUSIONS

Our research aimed to leverage the power of machine learning techniques and statistical analysis to enhance the accuracy of chest X-ray image classification, explicitly distinguishing between normal and pneumonia cases. The U-Net architecture was utilized for image segmentation and applied statistical analysis to identify discriminative features in chest X-ray images. Additionally, popular machine learning algorithms were employed for classification and ensemble learning techniques to improve predictive performance further. The highest accuracy is 86%, achieved by the RF algorithm, closely followed by XGBoost, with an accuracy of 85%. These outcomes indicate our approach's effectiveness in distinguishing between normal and pneumonia CXR images. Nevertheless, it is important to acknowledge certain limitations. Our study focuses solely on normal and pneumonia cases and did not include other lung pathologies. In future work, we aim to experiment with an alternative dataset to enhance our outcomes and validate our findings. Other possible future work could be conducting an experimental deep learning analysis and comparing the outcomes with the traditional classification results. Other future work might consider expanding the scope to encompass a broader range of lung abnormalities for a more comprehensive diagnostic framework.

ACKNOWLEDGEMENTS

We thank Princess Nourah bint Abdulrahman University for funding this project.

REFERENCES

- [1] Nih chest x-ray dataset - machine learning datasets. (n.d.). machine learning datasets. <https://datasets.activeloop.ai/docs/ml/datasets/nih-chest-x-ray-dataset/>. accessed Jan. 11, 2023.
- [2] 2. "pneumonia," mayo clinic. <https://www.mayoclinic.org/diseases-conditions/pneumonia/diagnosis-treatment/drc-20354210>. accessed Jun. 12, 2023.
- [3] Radiologyinfo.org. <https://www.radiologyinfo.org/en/info/chestrad>. accessed Jun. 12, 2023.
- [4] Rabia Emhamed Al Mamlook, Shengfeng Chen, and Hanin Fawzi Bzizi. Investigation of the performance of machine learning classifiers for pneumonia detection in chest x-ray images. In 2020 IEEE International Conference on Electro Information Technology (EIT), pages 098–104. IEEE, 2020.
- [5] A. L. Association. Pneumonia. <https://www.lung.org/lung-health-diseases/lung-disease-lookup/pneumonia>. accessed Jun. 12, 2023.
- [6] Candice Bent'ejac, Anna Csorgo, and Gonzalo Mart'inez-Mun'oz. A comparative analysis of gradient boosting algorithms. *Artificial Intelligence Review*, 54(3):1937–1967, aug 2020.
- [7] Tej Bahadur Chandra and Kesari Verma. Pneumonia detection on chest x-ray using machine learning paradigm. In *Proceedings of 3rd International Conference on Computer Vision and Image Processing: CVIP 2018, Volume 1*, pages 21–33. Springer, 2020.
- [8] Fran¸ois Chollet. Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1251–1258, 2017.

- [9] Vikash Chouhan, Sanjay Kumar Singh, Aditya Khamparia, Deepak Gupta, Prayag Tiwari, Catarina Moreira, Robertas Damaševičius, and Victor Hugo C De Albuquerque. A novel transfer learning based approach for pneumonia detection in chest x-ray images. *Applied Sciences*, 10(2):559, 2020.
- [10] Jim Frost. Gupta, a. (2020, october 10). feature selection techniques in machine learning (up- dated 2023). *Analytics Vidhya*. <https://www.analyticsvidhya.com/blog/2020/10/feature-selection-techniques-in-machine-learning/>. accessed Jan. 11, 2023.
- [11] 11. Jim Frost. Understanding analysis of variance (anova) and the f-test. *The Minitab Blog*. Pridobljeno, 2016, <https://blog.minitab.com/en/adventures-in-statistics-2/understanding-analysis-of-variance-anova-and-the-f-test>. accessed Jan. 11, 2023.
- [12] Antonio Gulli and Sujit Pal. *Deep learning with Keras*. Packt Publishing Ltd, 2017.
- [13] Masoom A Haider, Alireza Vosough, Farzad Khalvati, Alexander Kiss, Balaji Ganeshan, and Georg A Bjarnason. Ct texture analysis: a potential tool for prediction of survival in patients with metastatic clear cell carcinoma treated with sunitinib. *Cancer Imaging*, 17(1):1–9, 2017.
- [14] Toto Haryanto, Adib Pratama, Aniati Murni, Heru Suhartanto, Jan Pidanic, and Aniati Murni Arymurthy. Multipatch-glm for texture feature extraction on classification of the colon histopathology images using deep neural network with gpu acceleration. *Journal of Computer Science*, 16(3):280–294, 2020.
- [15] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [16] Zhengyu He. Deep learning in image classification: A survey report. In *2020 2nd International Conference on Information Technology and Computer Application (ITCA)*, pages 174–177. IEEE, 2020.
- [17] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017.
- [18] Lal Hussain, Tony Nguyen, Haifang Li, Adeel A Abbasi, Kashif J Lone, Zirun Zhao, Mahnoor Zaib, Anne Chen, and Tim Q Duong. Machine-learning classification of texture features of portable chest x-ray accurately classifies covid-19 lung infection. *BioMedical Engineering OnLine*, 19:1–18, 2020.
- [19] Dina M Ibrahim, Nada M Elshennawy, and Amany M Sarhan. Deep-chest: Multi-classification deep learning model for diagnosing covid-19, pneumonia, and lung cancer chest diseases. *Computers in biology and medicine*, 132:104348, 2021.
- [20] John D Kelleher, Brian Mac Namee, and Aoife D’arcy. *Fundamentals of machine learning for predictive data analytics: algorithms, worked examples, and case studies*. MIT press, 2020.
- [21] Daniel Kermany, Kang Zhang, Michael Goldbaum, et al. Labeled optical coherence tomography (oct) and chest x-ray images for classification. *Mendeley data*, 2018.
- [22] Rohit Kundu, Ritacheta Das, Zong Woo Geem, Gi-Tae Han, and Ram Sarkar. Pneumonia detection in chest x-ray images using an ensemble of deep learning models. *PLoS one*, 16(9):e0256630, 2021.
- [23] Lionel A Mandell, Richard G Wunderink, Antonio Anzueto, John G Bartlett, G Douglas Campbell, Nathan C Dean, Scott F Dowell, Thomas M File Jr, Daniel M Musher, Michael S Niederman, et al. Infectious diseases society of america/american thoracic society consensus guidelines on the management of community-acquired pneumonia in adults. *Clinical infectious diseases*, 44(Supplement 2):S27–S72, 2007.
- [24] Cina Mathew and P Asha. Detection of covid-19 from chest x-ray scans using machine learning. In *AIP Conference Proceedings*, volume 2463. AIP Publishing, 2022.
- [25] Ashley ND Meyer, Traber D Giardina, Lubna Khawaja, and Hardeep Singh. Patient and clinician experiences of uncertainty in the diagnostic process: current understanding and future directions. *Patient Education and Counseling*, 104(11):2606–2615, 2021.
- [26] Rupert G Miller Jr. *Beyond ANOVA: basics of applied statistics*. CRC press, 1997.
- [27] Roweida Mohammed, Jumanah Rawashdeh, and Malak Abdullah. Machine learning with oversampling and undersampling techniques: overview study and experimental results. In *2020 11th international conference on information and communication systems (ICICS)*, pages 243–248. IEEE, 2020.
- [28] P Mohanaiah, P Sathyanarayana, and L GuruKumar. Image texture feature extraction using glm approach. *International journal of scientific and research publications*, 3(5):1–5, 2013.

- [29] Jojo Moolayil and Jojo Moolayil. An introduction to deep learning and keras. *Learn Keras for Deep Neural Networks: A Fast-Track Approach to Modern Deep Learning with Python*, pages 1–16, 2019.
- [30] Ng Weng Mun, Mahmud Iwan Solihin, Li Sze Chow, and Affiani Machmudah. Pneumonia identification from chest x-rays (cxr) using ensemble deep learning approach. In *Proceedings of the 6th International Conference on Electrical, Control and Computer Engineering: InECCE2021*, Kuantan, Pahang, Malaysia, 23rd August, pages 1139–1151. Springer, 2022.
- [31] M Muthuraman. An exhaustive analysis of feature extraction techniques for identifying diseased lungs from ct images. *Journal of Positive School Psychology*, pages 7226–7235, 2022.
- [32] Kizito Nyuytiymbiy. Parameters and hyperparameters in machine learning and deep learning. *Towards Data Science*, 2022.
- [33] 33. Nikhil Pandey. N. (2019, january 22). lung segmentation from chest x-ray dataset. Kaggle. <https://www.kaggle.com/code/nikhilpandey360/lung-segmentation-from-chest-x-ray-dataset/notebook>. accessed Jan. 11, 2023.
- [34] Mathias W Pletz, Francesco Blasi, James D Chalmers, Charles S Dela Cruz, Charles Feldman, Carlos M Luna, Julio A Ramirez, Yuichiro Shindo, Daiana Stolz, Antoni Torres, et al. International perspective on the new 2019 american thoracic society/infectious diseases society of america community-acquired pneumonia guideline: a critical appraisal by a global expert panel. *Chest*, 158(5):1912–1918, 2020.
- [35] M. E. Prep. (2019, december 20). diagnosing pneumonia on chest x-ray. *medical exam prep*. <https://www.medicalexamprep.co.uk/diagnosing-pneumonia-on-chest-x-ray/>. accessed Jan. 11, 2023.
- [36] Hariharan Ravishankar, Prasad Sudhakar, Rahul Venkataramani, Sheshadri Thiruvenkadam, Pavan Annangi, Narayanan Babu, and Vivek Vaidya. Understanding the mechanisms of deep transfer learning for medical images. In *Deep Learning and Data Labeling for Medical Applications: First International Workshop, LABELS 2016, and Second International Workshop, DLMIA 2016, Held in Conjunction with MICCAI 2016, Athens, Greece, October 21, 2016, Proceedings 1*, pages 188–196. Springer, 2016.
- [37] Susmita Ray. A quick review of machine learning algorithms. In *2019 International conference on machine learning, big data, cloud and parallel computing (COMITCon)*, pages 35–39. IEEE, 2019.
- [38] A. Saini. Master the adaboost algorithm: Guide to implementing understanding adaboost. *analytics vidhya*.(2021, september 15). <https://www.analyticsvidhya.com/blog/2021/09/adaboost-algorithm-a-complete-guide-for-beginners/>. accessed Jan. 11, 2023.
- [39] Silhouette Sign. Basic signs in chest radiology: Silhouette, hilar overlay, and hilar convergence signs. *Chest Imaging*, page 23, 2019.
- [40] Hyojoo Son, Hyunchul Choi, Hyeonwoo Seong, and Changwan Kim. Detection of construction workers under varying poses and changing background in image sequences via very deep residual networks. *Automation in Construction*, 99:27–38, 2019.
- [41] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alexander Alemi. Inception-v4, inception- resnet and the impact of residual connections on learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 31, 2017.
- [42] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.
- [43] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826, 2016.
- [44] Barbara G Tabachnick and Linda S Fidell. *Experimental designs using ANOVA*, volume 724. Thomson/Brooks/Cole Belmont, CA, 2007.
- [45] Srikanth Tammina. Transfer learning using vgg-16 with deep convolutional neural network for classifying images. *International Journal of Scientific and Research Publications (IJSRP)*, 9(10):143–150, 2019.
- [46] Xiao Tang, Rong-Hui Du, Rui Wang, Tan-Ze Cao, Lu-Lu Guan, Cheng-Qing Yang, Qi Zhu, Ming Hu, Xu-Yan Li, Ying Li, et al. Comparison of hospitalized patients with ards caused by covid-19 and h1n1. *Chest*, 158(1):195–205, 2020.
- [47] Huan Tian, Tianqing Zhu, Wei Liu, and Wanlei Zhou. Image fairness in deep learning: problems, models, and challenges. *Neural Computing and Applications*, 34(15):12875–12893, 2022.

- [48] Khin Mya Mya Tun and Khaing Aung Soe. Feature extraction and classification of lung cancer nodule using image processing techniques. *International Journal of Engineering Research and Technology*, 3(3), 2014.
- [49] Risheng Wang, Tao Lei, Ruixia Cui, Bingtao Zhang, Hongying Meng, and Asoke K Nandi. Medical image segmentation using deep learning: A survey. *IET Image Processing*, 16(5):1243–1267, 2022.
- [50] Xiaosong Wang, Yifan Peng, Le Lu, Zhiyong Lu, Mohammadhadi Bagheri, and Ronald M Summers. Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2097–2106, 2017.
- [51] Sara Lee Kit Yee and Wong Jee Keen Raymond. Pneumonia diagnosis using chest x-ray images and machine learning. In *proceedings of the 2020 10th international conference on biomedical engineering and technology*, pages 101–105, 2020.
- [52] Ye Zhang, Yi Hou, Kewei OuYang, and Shilin Zhou. Multi-scale signed recurrence plot based time series classification using inception architectural networks. *Pattern Recognition*, 123:108385, 2022.

AUTHORS

Najah Al-Subaie is an Assistant Professor at the Department of Computer Sciences at PNU. She received her PhD in 2019 from the University of Warwick in the UK. During her PhD, she worked with the Tissue Image Analytics (TIA) Center and collaborated with the Department of Pathology at Coventry University Hospitals in developing machine learning-based methods to analyse abnormality in digitised tissue samples.

Tahani Alqahtani is a computer scientist who completed her master's degree in data science. Her research focused on developing algorithms for automated x-ray image analysis, aimed at improving disease diagnosis. After earning her master's degree with distinction, she joined Princess Nourah bint Abdulrahman University as a lecturer in the computer science department.

Dr. Syarifah Bahiyah Rahayu is an academic and researcher. Currently serving as the Head of the Department of Science Defense within the Faculty of Defense Science and Technology at the National Defense University of Malaysia. Dr. Rahayu is also a Research Fellow at the Department of Knowledge Technology & Security, where her expertise contributes significantly to the ongoing research initiatives at the Cyber Security & Digital Industrial Revolution Centre. Her work underscores her commitment to enhancing national security through cutting-edge research and technological advancements.