Proceedings of the 10th World Congress on Electrical Engineering and Computer Systems and Sciences (EECSS'24)
Barcelona, Spain – August 19 - 21, 2024
Paper No. ICBES 142
DOI: 10.11159/icbes24.142

Improving Lung Disease Classification from Chest X-ray Images using an Efficient Clustering Approach

Aya Hage Chehade[1], Nassib Abdallah[1,2], Jean-Marie Marion[1], Karim Chéhadé[3], Mohamad Oueidat[4], Pierre Chauvet[1]
[1]LARIS, University of Angers, France
[2]LaTIM, INSERM UMR 1101, University of Brest, France
[3]ISEN Yncréa Ouest, Brest, France
[4]Faculty of Technology, Lebanese University, Lebanon
aya.hagechehade@etud.univ-angers.fr; nassib.abdallah@univ-angers.fr; marion@uco.fr;
karim.chehade@isen-ouest.yncrea.fr; mohoueidat@yahoo.com; pierre.chauvet@uco.fr

Abstract **-** Lung diseases are a major health problem and one of the leading causes of death worldwide. Chest X-ray (CXR) is one of the most common radiological examinations for screening thoracic diseases. Despite the existing methods that have achieved significant progress in the classification of thoracic diseases, none of the studies take into account the presence of artifacts such as wires or objects in the images. Based on the above problem, in this paper we present a novel methodology for clustering sharp images from images containing artifacts, and then perform the classification exclusively to the cluster containing sharp images without artifacts.
We selected CXR of pneumonia and normal cases from the ChestX-Ray14 dataset and performed histogram equalization as preprocessing technique. By applying the DenseNet-121 model exclusively to the cluster containing images without artifacts, we achieved a higher area under the curve (AUC) than the model applied to all images. Our approach thus achieved an AUC of 79.58% for pneumonia and normal images classification. To evaluate the effectiveness of our method, we conducted experiments on another disease, namely consolidation. The results demonstrated that our method is promising, highlighting its potential for broader applications in lung disease classification. This research highlights the importance of considering the presence of artifacts when diagnosing lung diseases from radiographic images. The code will be available upon request.

Keywords**:** Lung Disease, ChestX-Ray14 Dataset, X-ray Images, Image Processing, Histogram Equalization, Image Classification, Deep Learning, K-means Clustering.

1. Introduction

Lung diseases pose a major threat to global health, ranking as the third most leading cause of mortality worldwide [1], and causing around five million deaths annually [2]. Consequently, early diagnosis of lung diseases is essential for effective treatment and to reduce the risk of mortality [3]. Chest X-ray (CXR) is the most commonly imaging diagnostic technique used to identify lung diseases [4], including pneumonia [5] due to its simple, rapid and cost-effective procedure [6]. However, manual observation and interpretation of CXR is a time-consuming process. In addition, distinguishing different types of diseases from chest X-ray images is a difficult task for an expert, and can lead to missed detections and therefore life-threatening diagnostic errors due to the complex nature of chest X-rays [7].

Recently, with the advancements of technology, Computer Aided Diagnosis (CAD) systems were proposed to analyse chest radiographs in order to reduce the workload of radiologists and to improve clinical diagnosis [8]. More recently, Deep Learning (DL) has been widely investigated due to its general applicability to problems involving automated feature extraction and image classification tasks [9], [10].

Many researchers focused on developing deep learning-based techniques for disease detection and classification. Rajpurkar et al. [11] developed CheXNet, a 121-layer DenseNet model, for pneumonia detection from the ChestX-Ray14 dataset. This model exceeded the average performance of radiologists on the pneumonia detection task and achieved an area under the curve (AUC) of 76.8%. Souid et al. [12] proposed the classification and prediction of lung pathologies in frontal chest X-rays using the MobileNetV2 model, supplemented by CNN layers. Their approach obtained an AUC of 73.3% for the pneumonia classification. Ma et al. [13] proposed a cross-attention network scheme for thoracic disease classification,

using an attention loss that forced the model to focus more accurately on pathogenic areas. The model achieved an AUC of 72.2% for the classification of pneumonia.

However, none of the previous studies consider the presence of artifacts in CXR images. As illustrated in Figure 1, X-ray images include artifacts such as objects, medical devices, wires and electrodes, and may be low-quality. These artifacts may introduce noise and undermining the overall quality of the CXR image, since they can mask important anatomical areas as well as introduce irrelevant features within the X-ray image. This can impact the effectiveness of machine learning models and the reliability of medical diagnoses. Consequently, it is crucial to develop strategies that consider the presence of artifacts in X-ray images. This will improve the classification performance of deep learning models, contributing to the establishment of precise and reliable diagnoses and thus improving the quality of healthcare for patients.

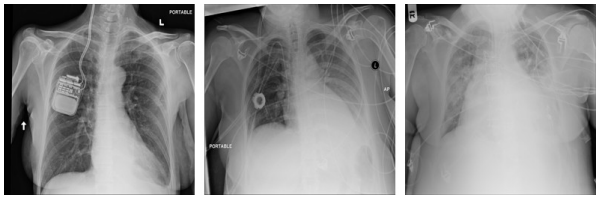


Fig. 1. Examples of chest X-ray images with artifacts and bad quality captures.

It is important to highlight that the issue of missing labels for images containing electronic components may occur in any database. This underscores the broader importance of clustering (unsupervised labelling) in efficiently processing image data, especially in the medical field. This contributes to enhanced data quality, resulting in more precise and meaningful results in the classification of medical images, thus reinforcing the efficiency of diagnostic analyses in the healthcare sector. For this purpose, our work's novelty is rooted in the use of a novel methodology for clustering sharp images from images containing artifacts, and then perform the classification exclusively to the cluster containing sharp images without artifacts. As a result, by eliminating images containing artifacts and preventing artifacts from introducing noise into the classification model, our method improves classification performance and reinforces the reliability of diagnoses based on chest X-ray images, which is essential for high-quality healthcare.

This paper is organized as follows: Section 2 presents the materials and methods used, including the dataset, the preprocessing method and the clustering technique, and the classification process. In Section 3 and 4, we present and discuss the results obtained using the ChestX-ray14 dataset. Finally, in section 5, we conclude our work and present the perspectives.

2. Material and methods

2.1. Dataset

The ChestX-Ray14 dataset [14] is extracted from the clinical PACS databases in the hospitals affiliated to National Institutes of Health Clinical Center and includes 112,120 frontal view chest X-ray images from 30,805 unique patients. Each CXR is labelled with binary labels for 14 different diseases. The CXR is labelled as "no finding" if none of these diseases were detected. All CXRs are in PNG format and have a size of 1024 × 1024. The dataset also contains other data such as the sex and age of the patient and the view position of the X-ray. In this database, 322 images are labelled as pneumonia and 60,361 images are no finding. To address class imbalance, we down-sampled only the no finding images to obtain 354 samples. Therefore, we conducted our experiments on a total of 676 images.

The database lacks labels for images containing artifacts, and the use of such images may introduce electronic noise in the classification process. To address this issue, it is important to implement clustering as a preliminary step. This involves identifying images with electronic components to effectively minimize the impact of electronic noise, leading to an improvement in the performance of the classification.

2.2. Overview

For the classification of pneumonia and normal X-ray images, a total of 676 X-ray images were used. As illustrated in Figure 2, our methodology comprises three major steps: preprocessing, clustering, and classification. In the

preprocessing phase, we applied the Histogram Equalization technique to amplify contrast and enhance the visibility of image details. Following this, features related to contours and texture were extracted. The subsequent step involved clustering clustering through the application of the K-means algorithm. This process yielded two clusters: the first comprising unclear unclear images with wires and electronic objects, and the second consisting of sharp images without artifacts. So, in the rest rest of our study, we worked only on the second cluster containing images without electronic objects to eliminate the noise noise effect caused by the other group containing images with artifacts. This improves model performance since it will be able to focus specifically on lung features that are relevant for disease classification, without being disturbed by interference related to electronic objects. Therefore, for the second cluster, images were resized to 224 × 224 pixels and normalized in an interval [0;1]. 20% of the data was used as test data, 10% as validation data and 70% was devoted to training the data (randomly chosen). Thereafter, the classification was performed using the DenseNet-121 model. The performance of our model was quantified by the Area Under the Curve (AUC). In the literature, AUC remains a predominant metric for evaluating the performance of lung disease classification models, which justifies its choice in the context of our study.

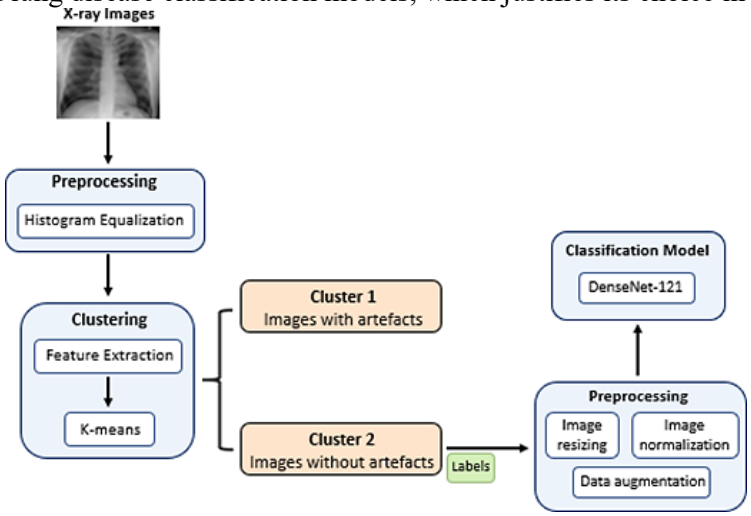
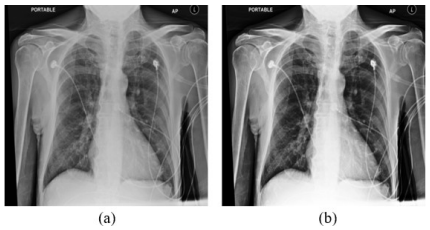


Fig. 2: Overview of the proposed methodology.

2.3. Preprocessing

In the preprocessing, Histogram Equalization is used for image contrast enhancement. The histogram equalization technique aims to achieve a uniform distribution of gray levels within an image, thereby adjusting the brightness and contrast of dark and low-contrast images, leading to an improvement in overall image quality [15], [16]. This contributes to the enhancement of image details, the sharpening of contours, and an overall improvement in image visual quality. Figure 3 visualizes the enhancement result of X-ray image using Histogram Equalization.



(a) (b)

Fig. 3: A sample of chest x-ray image: (a) original image, and (b) sharpened image using the Histogram Equalization technique.

2.4. Clustering

Following the preprocessing, since images with electronic components are unlabelled in the database and may introduce noise during classification, we proposed an image preprocessing module that will separate images into two clusters: one containing images with artifacts and the other containing images without artifacts. To achieve this, we initiated the process

by extracting relevant features related to texture and contours from the images. Subsequently, a K-means method was used to effectively separate the images into two clusters.

1- Feature extraction: Feature extraction, especially related to contours and texture of X-ray images, is essential for interpreting and understanding the information topology within the images. The identification of subtle details in contours and texture allows K-means to accurately differentiate between sharp images and those affected by artifacts. To achieve four characteristics were computed in this study, including:

- Variance of Gradients feature: This feature computes the horizontal and vertical partial derivatives of the image using the Sobel filter. These derivatives serve to quantify the intensity variations among pixels, offering valuable insights into how brightness changes in different directions across the image. The subsequent step involves combining these derivatives to calculate the magnitude of the gradient. Specifically, it assesses the variance of gradient magnitude values across the entire image. Regions with complex or rough textures may exhibit higher gradient variance compared to regions with more homogeneous textures. Consequently, the variance of gradients feature is sensitive to local texture variations. Therefore, this feature serves as a valuable indicator of the gradient variation across the entire image.
- Lung sharpness feature: This feature assesses the sharpness of lung contours in a X-ray image by using the Sobel filter to compute horizontal and vertical partial derivatives. These derivatives highlight variations in pixel intensity related to contours. The combination of these partial derivatives yields the gradient magnitude, representing the sharpness of image contours and measuring the precision of brightness changes. A higher gradient magnitude indicates sharper and more well-defined contours. To specifically focus on the lung region of interest, a mask is created to isolate this area, improving measurement efficiency by excluding unnecessary information outside the lungs. This mask is applied to the previously computed contours, ensuring that only contours within the lung region contribute to the sharpness assessment. A high value for this feature indicates well-defined and sharp lung contours, while reduced sharpness may indicate the presence of artifacts or unwanted objects affecting image quality.
- Black-white ratio feature: This feature evaluates the distribution of black and white pixels in an image by calculating the percentage of black pixels (indicating low intensity) and white pixels (indicating high intensity). Essentially, it quantifies the density of dark structures in comparison to light areas within the image. In the context of CXR images, this characteristic aids in distinguishing lung areas (typically darker) from non-lung areas (typically lighter), contributing to the identification of lungs and objects in the image. A high black ratio may suggest high tissue density, providing insights into the density of anatomical structures. Conversely, a high white ratio may be indicative of artifacts, overexposure, or the presence of unwanted objects that alter the brightness distribution.
- Histogram of Oriented Gradients (HOG) feature: HOG features offer a robust representation of gradient patterns in an image. This method involves dividing the image into cells and computing gradients in each cell, capturing local orientations of contours [17]. The process continues by calculating the histogram of oriented gradients, normalizing results block-wise, concatenating 2×2 grid cells, and generating the HOG descriptor at each grid location [18]. Essentially, HOG features analyse how changes in brightness occur in different directions within the image. Widely recognized for their effectiveness in object detection and pattern recognition [19]–[21], HOG features can discern specific gradient patterns characteristic of structures or objects. This makes them a valuable tool for identifying elements of interest within X-ray images.

2- K-means: Following the extraction of key features from the X-ray images, the next step involved normalization, and these normalized features served as input data for the unsupervised learning algorithm known as K-means clustering. The objective of K-means clustering is to group data by maximizing the similarity of features within groups and maximizing the differences between these groups [22], [23]. In our context, this unsupervised learning method uncovered hidden structures within our dataset, and efficiently separated images into two clusters: one containing sharp images and the other with images affected by artifacts and unwanted elements. This automated approach emerged as a valuable asset in our research, contributing to improve classification performance and, consequently, the precision of medical diagnoses based on X-ray images. In our study, k-means was applied to both no finding and pneumonia images, resulting in two clusters for each class. The distribution of pneumonia and no finding images in each cluster, obtained through the K-means clustering algorithm, is summarized in Table 1.

	Pneumonia	No finding	Total
Cluster 1: Images with artifacts	152	133	285
Cluster 2: Images without artifacts	170	221	391

Table 1: Number of images of pneumonia and no finding classes in each cluster obtained using k-means clustering.

In the rest of the study, we combined the clusters containing the sharp images without artifacts of pneumonia and no finding to have a total of 391 images. These images will be used for classification. Indeed, the elimination of images containing artifacts aims to mitigate the impact of noise induced by these images, thereby potentially improving the classification model's performance, which will no longer be disrupted by unwanted elements.

2.5. Classification

The DenseNet-121 model was used to perform the classification of normal and pneumonia images. The images were resized to 224 × 224 pixels and normalized before being inputted into the pre-trained DenseNet-121 model. During training, we initialized the DenseNet-121 with weights pre-trained on ImageNet. The weights of the lower convolutional layers were frozen, and we added two dense layers with 1024 and 512 hidden neurons, respectively. This was followed by a dropout layer with a rate of 0.5. Finally, the last fully connected layer was replaced with a fully connected layer having a 2-dimensional output and a sigmoid activation function to classify the image as pneumonia or normal. Training was conducted using mini-batches of size 16, and the ADAM optimizer was employed with a learning rate of 10^{-4}. To regularize the network, an early stopping strategy was implemented with a patience of five epochs to detect training convergence and prevent overfitting. The model was compiled using 'binary_crossentropy' as the loss function. To increase the size of the dataset and improve the model's generalization ability, data augmentation techniques were applied. Augmentation includes a rotation range of 10, horizontal flipping, a width shift range of 0.1, a height shift range of 0.1, a zoom range of 0.1 and setting the fill-mode to nearest.

3. RESULTS

In all our results, we obtained the highest AUC by applying the DenseNet-121 model on the cluster containing the sharp images only. With clustering (unsupervised labelling), we achieved an AUC of 79.58% for the DenseNet-121 model applied to the cluster without artifacts. To assess our unsupervised labelling approach, we manually labelled pneumonia and no finding images as either with or without artifacts, establishing a ground truth for evaluation. Subsequently, we computed the confusion matrix for each class, as depicted in Figure 4, and calculated performance metrics presented in Table 2. Also, when we applied our model on an additional disease namely consolidation, we achieved the highest AUC of 80.26% using clustering (unsupervised labelling) with only the cluster containing images without electronic objects. Table 3 presents the results of the binary classification of the two diseases by applying the DenseNet-121 model, with and without the clustering step. From Table 4, we can see that our model outperforms the literature for the classification of all the two lung diseases: pneumonia and consolidation, using the same dataset.

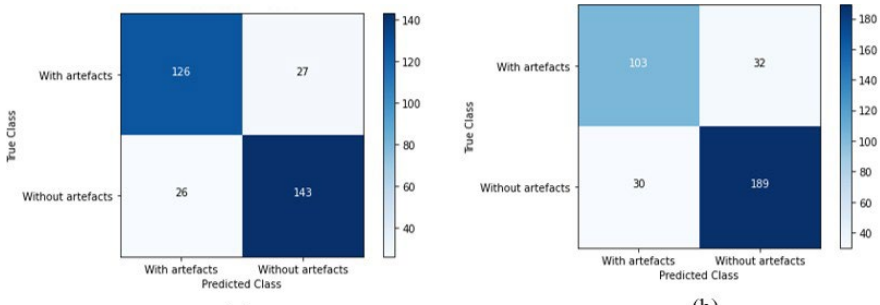


Fig. 4: The confusion matrix of clustering for: (a) pneumonia class and (b) no finding class.

	Pneumonia	No finding
Accuracy	83.54%	82.49%
Recall	84.62%	86.3%
Precision	84.12%	85.52%
F1-score	84.37%	85.91%

Table 2: Clustering performance metrics for the pneumonia and no finding classes.

	DenseNet-121		
	Without clustering	**With clustering (unsupervised labelling)**	**With clustering (manual labelling)**
Pneumonia	77.96% ±1.5%	79.58% ±1.6%	81.47% ±1.8%
Consolidation	78.52% ±1.4%	80.26% ±1.3%	-

Table 3: AUC of the classification of the two lung diseases with and without the clustering step.

Reference	Year	Pneumonia	Consolidation
Wang et al. [14]	2017	66.4%	77%
Rajpurkar et al. [11]	2017	76.8%	74.1%
Ma et al. [13]	2019	72.2%	75%
Souid et al. [12]	2021	73.3%	79%
Yang et al. [24]	2022	73.4%	75.6%
Jin et al. [25]	2023	73.9%	72.5%
Our proposed approach	2024	**79.58%**	**80.26%**

Table 4: Comparison of the achieved results with other existing methods in the literature for the classification of the two lung diseases using the same dataset.

4. Discussion

To address the potential introduction of noise during classification, especially from images with artifacts that lack labels in the dataset, we implemented an image preprocessing module. This module was specifically designed to distinguish between sharp images and those containing artifacts. In this process, we extracted various contour and texture features to serve as input data for the K-means clustering algorithm. The outcome, as outlined in Table 1, resulted in the creation of two clusters for each class: one cluster comprises sharp images without artifacts, while the other cluster consists of images affected by artifacts. This customized approach enhances the robustness of our classification process by effectively handling images with potential noise or artifacts.

To evaluate our labelling approach, we manually labelled the images of pneumonia and no finding as with and without artifacts, creating a reliable ground truth for evaluation. Then, as shown in Figure 4 and Table 2, we computed the confusion matrix and the performance metrics for each class, respectively. The clustering of pneumonia achieved an accuracy of 83.54% and a F1-score of 84.37%. Similarly, the clustering of no finding achieved an accuracy of 82.49% and a F1-score of 85.91%. Examining the confusion matrices (a) and (b) reveals that only 53 and 62 samples out of 322 and 354 images were misclassified, respectively. Therefore, the clustering (unsupervised labelling) proves to be highly accurate in discerning clear images from those containing artifacts.

Following that, we conducted two experiments using the DenseNet-121 model. In the first experiment, we applied the model to all images without clustering. In the second experiment, we applied the model after clustering, specifically to the cluster containing sharp images without artifacts, and excluded the cluster containing images with artifacts. The AUC was calculated for each case. To guarantee the stability of the results, we conducted ten different random splits of the dataset into training, validation, and testing sets. The results are presented in terms of mean and standard deviation. As presented in Table 3, without clustering, we obtained an AUC of 77.96% for the classification of pneumonia and no finding. With the application of clustering (unsupervised labelling) and utilizing the model exclusively on the cluster containing sharp images without artifacts, we achieved an AUC of 79.58% for pneumonia classification. In comparison, the DenseNet-121 model applied to the cluster without artifacts obtained through manual labelling attained an AUC of

81.47%. The marginal difference between the AUC values of the model with unsupervised labelling and manual labelling indicates the high efficiency of our clustering method in distinguishing clear images from those containing artifacts. This approach is not only efficient but also requires less time and effort compared to manual labelling by a physician, making it applicable to a larger number of images.

Following this, we generalized our approach for the classification of consolidation and no finding using 2,751 images. We used all the consolidation images from the ChestX-ray14 dataset, totalling 1,310 images marked as consolidation. For the no finding class, we down-sampled the images to obtain 1,441 samples for classification. Subsequently, we applied clustering (unsupervised labelling) using the k-means algorithm on the proposed features extracted from the images. This process resulted in two clusters for each class: the first cluster containing sharp images and the second containing images with artifacts. As a result, we obtained a total of 1,328 consolidation and no finding images that are clear without artifacts. The results, as presented in Table 3, indicate that the classification outcomes with the use of clustering (unsupervised labelling) outperform those without clustering. This is confirmed by the AUC values; the DenseNet-121 model applied to all images without clustering resulted in an AUC of 78.52%. However, after clustering, the model achieved an improved AUC of 80.26%. Therefore, as clearly illustrated from the results in Table 3, applying the model on images that do not contain artifacts leads to improve the AUC. Moreover, as indicated in Table 4, our proposed approach demonstrates superior performance compared to the literature in the classification of two lung diseases, namely pneumonia and consolidation, using the same dataset.

Therefore, in the diagnosis of lung diseases based on radiographic images, it is essential to account for artifacts. These artifacts, stemming from technical issues in X-ray equipment, unwanted interference, or other factors, can significantly impact the quality of medical images. In a field where precision is critical, these artifacts can distort the interpretation and detection of lung anomalies, potentially leading to inaccurate diagnoses and inappropriate treatment decisions. Through the implementation of clustering and the prevention of noise introduction through artifacts, our method improves the classification performance of deep learning models. By eliminating unwanted noise, these models can focus on relevant diagnostic features, thereby enhancing their performance. This results in a more precise detection of lung diseases, with potential significant implications for clinical outcomes. In the medical context, this methodology empowers radiologists to optimize the detection of abnormalities, contribute to more effective treatment decisions, and, ultimately, elevate the overall quality of diagnoses.

5. Conclusion

This paper focuses on clustering (unsupervised labelling) to separate sharp images from those containing artifacts before performing lung disease classification. Features related to contours and texture are extracted from images, and the K- means clustering algorithm is used to create two clusters of images based on their extracted features. Then, we performed the classification of lung disease exclusively on the cluster containing sharp images without artifacts, using the DenseNet-121 model. The experimental results showcase the effectiveness of our approach, achieving an AUC of 79.58% and 80.26% for pneumonia and consolidation classification, respectively. In contrast, when considering all images with and without artifacts, the AUC values drop to 77.96% for pneumonia and 78.52% for consolidation classification. This confirms the vital importance of taking into account the presence of artifacts when interpreting radiographic images. It improves model performance since it will be able to focus specifically on lung features that are relevant for disease classification, without being disturbed by interference related to unwanted objects such as wires or electronic objects. Consequently, it enables ensure accurate treatment, minimizes diagnostic errors and guarantee a consistent results interpretation. In the future, it is planned to explore other feature extraction techniques to improve artifact identification, and to process images containing artifacts to improve the performance of the model for lung disease classification.

References

[1] Han, T., Nunes, V., Souza, L., Marques, A., Silva, I., Junior, M., Sun, J. & Reboucas Filho, P. Internet of medical things—based on deep learning techniques for segmentation of lung and stroke regions in CT scans. IEEE Access. 8 pp. 71117-71135, 2020.

[2] Tobias, R., De Jesus, L., Mital, M., Lauguico, S., Guillermo, M., Sybingco, E., Bandala, A. & Dadios, E. CNN-based deep learning model for chest X-ray health classification using tensorflow. 2020 RIVF International Conference On Computing And Communication Technologies (RIVF). pp. 1-6, 2020.
[3] Bhandary, A., Prabhu, G., Rajinikanth, V., Thanaraj, K., Satapathy, S., Robbins, D., Shasky, C., Zhang, Y., Tavares, J. & Raja, N. Deep-learning framework to detect lung abnormality–A study with chest X-Ray and lung CT scan images. Pattern Recognition Letters. 129 pp. 271-278, 2020.
[4] Chen, L., Mao, T. & Zhang, Q. Identifying cardiomegaly in chest x-rays using dual attention network. Applied Intelligence. 52, 11058-11067, 2022.
[5] Rahman, T., Chowdhury, M., Khandakar, A., Islam, K., Islam, K., Mahbub, Z., Kadir, M. & Kashem, S. Transfer learning with deep convolutional neural network (CNN) for pneumonia detection using chest X-ray. Applied Sciences. 10, 3233, 2020.
[6] Kabir, S., Farrokhvar, L. & Dabouei, A. A Weakly supervised approach for thoracic diseases detection. Expert Systems With Applications. 213 pp. 118942, 2023.
[7] Islam, K., Wijewickrema, S., Collins, A. & O'Leary, S. A Deep Transfer Learning Framework for Pneumonia Detection from Chest X- ray Images. VISIGRAPP (5: VISAPP). pp. 286-293, 2020.
[8] Shamrat, F., Azam, S., Karim, A., Ahmed, K., Bui, F. & De Boer,F. High-precision multiclass classification of lung disease through customized MobileNetV2 from chest X-ray images. Computers In Biology And Medicine. 155 pp. 106646, 2023.
[9] Al Mamlook, R., Chen, S. & Bzizi, H. Investigation of the performance of machine learning classifiers for pneumonia detection in chest X-ray images. 2020 IEEE International Conference On Electro Information Technology (EIT). pp. 098-104, 2020.
[10] Sharma, S. & Guleria, K. A Deep Learning based model for the Detection of Pneumonia from Chest X-Ray Images using VGG-16 and Neural Networks. Procedia Computer Science. 218 pp. 357-366, 2023.
[11] Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., Ding, D., Bagul, A., Langlotz, C., Shpanskaya, K. & Others Chexnet: Radiologist- level pneumonia detection on chest x-rays with deep learning. ArXiv Preprint ArXiv:1711.05225, 2017.
[12] Souid, A., Sakli, N. & Sakli, H. Classification and predictions of lung diseases from chest x-rays using mobilenet v2. Applied Sciences. 11, 2751, 2021.
[13] Ma, C., Wang, H. & Hoi, S. Multi-label thoracic disease image classification with cross-attention networks. International Conference On Medical Image Computing And Computer-assisted Intervention. pp. 730- 738, 2019.
[14] Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., and Summers, R. 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 (CVPR), pages 2097-2106, 2017.
[15] Veluchamy, M. & Subramani, B. Image contrast and color enhancement using adaptive gamma correction and histogram equalization. Optik. 183 pp. 329-337, 2019.
[16] Rahman, T., Khandakar, A., Qiblawey, Y., Tahir, A., Kiranyaz, S., Kashem, S., Islam, M., Al Maadeed, S., Zughaier, S., Khan, M. & Others Exploring the effect of image enhancement techniques on COVID-19 detection using chest X-ray images. Computers In Biology And Medicine. 132 pp. 104319, 2021.
[17] Jawahar, M., Prassanna, J., Ravi, V., Anbarasi, L., Jasmine, S., Manikan- dan, R., Sekaran, R. & Kannan, S. Computer-aided diagnosis of COVID- 19 from chest X-ray images using histogram-oriented gradient features and Random Forest classifier. Multimedia Tools And Applications. 81, 40451-40468, 2022.
[18] Ho, T. & Gwak, J. Multiple feature integration for classification of thoracic disease in chest radiography. Applied Sciences. 9, 4130, 2019.
[19] Li, B., Cheng, K., Yu, Z. & Others Histogram of oriented gradient based gist feature for building recognition. Computational Intelligence And Neuroscience. 2016.
[20] Salau, A. & Jain, S. Feature extraction: a survey of the types, techniques, applications. 2019 International Conference On Signal Processing And Communication (ICSC). pp. 158-164, 2019.
[21] Ayalew, A., Salau, A., Abeje, B. & Enyew, B. Detection and classification of COVID-19 disease from X-ray images using convolutional neural networks and histogram of oriented gradients. Biomedical Signal Processing And Control. 74 pp. 103530, 2022.
[22] Putri, F., Wibowo, N. & Mustofa, H. Clustering of Tuberculosis and Normal Lungs Based on Image Segmentation Results of Chan-Vese and Canny with K-Means. Indonesian Journal Of Artificial Intelligence And Data Mining. 6, 18-28, 2023.
[23] Darwis, M., Hasibuan, L., Firmansyah, M., Ahady, N. & Tiaharyadini, R. Implementation of K-Means clustering algorithm in mapping the groups of graduated or dropped-out students in the Management Department of the National University. JISA (Jurnal Informatika Dan Sains). 4, 1-9, 2021.

[24] Yang, M., Tanaka, H. & Ishida, T. Performance improvement in multi- label thoracic abnormality classification of chest X-rays with noisy labels. International Journal Of Computer Assisted Radiology And Surgery. pp. 1-9, 2022.
[25] Jin, Y., Lu, H., Zhu, W. & Huo, W. Deep learning based classification of multi-label chest X-ray images via dual-weighted metric loss. Computers In Biology And Medicine. 157 pp. 106683, 2023.