ISSN 1870-4069

Integrating Radiograph Normalization Preprocessing and Discriminative Feature Selection for Efficient and Automated Pneumonia Detection

Salvador E. Ayala-Raggi, Angel Ernesto Picazo-Castillo, Aldrin Barreto-Flores, José Francisco Portillo-Robledo

Benemérita Universidad Autónoma de Puebla,
Facultad de Ciencias de la Electrónica,
Mexico

{saraggi,a.picazo.2505}@gmail.com,{aldrin.barreto,
francisco.portillo}@correo.buap.mx

Abstract. This paper introduces a technique for the detection of Viral Pneumonia using automatic localization, followed by pose and scale normalization of the specific region of interest (lungs) in chest radiographs. This method employs PCA and weighted K-NN regression. Our proposed approach includes estimating corner positions within the region of interest through interpolation, then mapping the image within that identified region onto a standardized fixed-size template. The primary goal is to achieve uniformity among training images in terms of position, angular pose, scale, and contrast, effectively aligning them. Subsequently, the eigenfaces method is employed to extract a reduced set of principal features from the normalized images. Among these PCA-derived features, those exhibiting the highest between-class discrimination capability are chosen using the Fisher criterion. Our results highlight the effective synergy achieved by integrating our lung region alignment technique with the meticulous selection and weighting of the most discriminative PCA features. This synergy is sufficient to achieve peak accuracies of 95.6% and 97.3% in classifying Viral Pneumonia radiographs using conventional classifiers, specifically weighted K-NN and MLP, respectively. Notably, our findings demonstrate that, in contrast to convolutional neural networks, a simpler technique can yield comparable classification results.

Keywords: Image classification, fisher discriminant, viral pneumonia, K-Nearest neighbors, multilayer perceptron.

1 Introduction

Pneumonia is a lung disease caused by bacteria and viruses; a person can be infected through the air, saliva, or mucus. Furthermore, children and the

elderly are at a higher risk of contracting it, according to [6]. Currently, various methods exist for detecting this disease, such as tomography, chest X-rays, and ultrasounds. However, tomography is more expensive than an X-ray, and ultrasound is not always available or affordable. Hence, X-rays prove to be a more common detection method [1,2,23,38,27].

Presently, there exist datasets accessible containing labeled radiographs, which can be employed to train diverse machine learning algorithms [2]. The establishment of these repositories has been a cooperative endeavor involving establishments and domain-specialist medical professionals [33,30,26]. Nevertheless, the obstacle lies in the absence of consistency in the area of interest (pulmonary region) within these images. Some radiographs encompass redundant or unrelated data for categorization, such as supplementary bodily components or objects obscuring the thoracic area. This can negatively impact the precision metrics of categorization algorithms [5,10].

In this work, we aim to demonstrate the hypothesis that aligning the region of interest in both the training images and the test image, such that the anatomical structures within the lungs are positionally consistent across all images, can enable simple and conventional classification methods like K-NN or MLP to achieve better accuracy results, provided that a reliable feature reduction method like PCA is employed in conjunction with a feature selection process based on their discriminatory capability.

To this end, we propose applying two consecutive processes. The first process involves the detection and normalization of the lung region, ensuring that the images within the lung region exhibit the same alignment, location, scale, and improved contrast as much as possible. In the second process, the "Eigenfaces" method (PCA) will be applied to the aligned regions to obtain a reduced set of statistically independent features. Finally, based on the Fisher criterion [35], we propose performing a selection of the features that best discriminate between classes. Using this set of optimal features and a traditional classifier such as K-NN or MLP, the classification accuracy will be measured.

This work is divided into four parts. Part 1 discusses the related work and the utilized database. Part 2 describes and presents the algorithm called "Lung Finder Algorithm" (LFA) for the normalization procedure. Part 3 presents the theory of "Eigenfaces" and Fisher linear discriminant, as well as the feature weighting applied in our analysis for the normalized image features. Finally, in Part 4, the precision metrics are compared when utilizing our methodology with the weighted K-Nearest Neighbors (K-NN) classifier and the Multilayer Perceptron [9].

1.1 Related Work

Currently, various methodologies have been developed for classification of chest radiographs, as evidenced in previous studies [17,13,28,11,4,31,37]. These methodologies make use of deep learning algorithms or traditional machine learning classifiers [7,8], and have reported high levels of classification accuracy, greater than 96%. However, the architectures employed in these algorithms still

Table 1. Comparison of the different preprocessing methods from related works.

Authors	Image Normalization	Features selection	Classifier	Accuracy
Changawala et al., (2021)	Not Used	Not Used	MLP (Involution)	98.31%
Liu et al., (2023)	Not Used	Used	SVM	100%
Park et al.,(2015)	Not Used	Used	SVM	93.5%
Lv et al.,(2022)	Not Used	Used	KNN	96.14%
Gadermayr et al., (2017)	Used	Not Used	SVM	97%
Kociolek et al.,(2020)	Used	Used	SVM	96%

face challenges in achieving a reliable classification of COVID-19 [32], as their accuracy decreases when tested with other datasets different from those used for training. This raises the need of exploring new proposals for normalizing and aligning the lungs region before classifying, instead of just facing the problem by training classifiers like CNNs with a large number of of different datasets, to cope with the bias imposed by a particular one.

Efficient non CNN-based works have been proposed too, as in [3], where a Multilayer Perceptron (MLP) and an architecture based on image involution were used, which proposes kernels similar to CNNs but shares their weights dynamically in all dimensions, thus reducing the number of multiplications necessary for the calculations. The former obtained a maximum classification accuracy of 98.31%. Feature selection has proven to be effective in increasing classification accuracy in other works, as observed in a study on [20] which used support vector machines to recognize the orbit axis of the sensors, as in another study [29] where it was also possible to classify the frequencies of an encephalogram. Furthermore, in a work carried out by Chengzhe et al. [21], the K-NN algorithm was applied successfully.

Several studies have shown that image normalization improves classification results. In a study on kidney radiographs [10], the best results were obtained using CNN and image normalization techniques. Also, in another [19] work, different normalization techniques were used on different types of radiographs to improve image classification. It is important to highlight that the results of our work are not intended to devalue CNNs in image classification, but rather to present an alternative option, and to demonstrate that image alignment and a proper feature selection technique can produce results comparable to the most commonly used algorithms. in the state of the art. In the table 1 we show the comparison of the different pre-processing methods used in some published works [3,20,29,21,10,19].

1.2 Data Set of Radiographic Images

The database used for this work was "COVID-19 Radiography Database" [4,31] from kaggle. This data set was selected because it has been used in other similar works[25,15]. The content of this data set is 6012 images already labeled as pulmonary opacity (other lung diseases), 1345 as viral pneumonia, 10192 as normal, and finally 3616 as COVID-19.

2 Overview of the Lung Finder Algorithm (LFA)

The goal of this algorithm is to locate the lungs in the radiographs, and it consists of a training and testing stage, as shown in figure 1. During the training stage, 400 images from the Pneumonia, COVID-19, and Normal classes were randomly selected from the data set. Histogram equalization (HE) [12,24] was applied to all images and regions of interest were manually labeled by placing 4 provisional landmarks easily located by a human user. It was agreed that two of them would be located, one in the middle of the cervical vertebrae just at the upper limit of the lungs, and the other also on the spine but below where the lung region ends. The other two provisional landmarks are forced to the user to place them on a imaginary straight line perpendicular to the spine that intersects it just in the middle of the two previous landmarks.These last two landmarks are located in the left and right sides of lung region. Finally, and by using these 4 provisional positions, we compute 4 final and permanent landmarks at the corners of the rectangular lung region. On the other hand, ten new images randomly rotated and displaced were then generated for each labeled image to increase the data set and have an *augmented dataset*. Next, a dimensionality reduction to this set of 4400 images was applied using the "Eigenfaces" method based on Principal Component Analysis (PCA) [39,18].

During the test stage, and after a contrast improvement (H.E.), a new image is projected to the "Eigenfaces" linear subspace in order to convert it to a compact few dimensions vector which is compared via euclidean distance with each of the 4000 examples contained within the augmented dataset to find k nearest neighbors $k - NN$. The landmarks associated with these k most similar images from the augmented dataset are used to estimate the 4 landmarks of the test image by interpolation. These predicted landmarks are the coordinates of the corners of the lung ROI which can be used to warp the inside region to a standard template of fixed size.

2.1 Coordinates Labeling for the LFA Training Stage

Each of the images selected for this stage requires a manual labeling where the region of interest of the lungs is delimited by a set of coordinates. These points or landmarks become the labels used by a regression weighted K-NN to predict the corner coordinates of the novel image. The coordinates the lung region are shown in figure 2, and consist of four points: Q1(x1,y1), Q2(x2,y2), Q3(x3,y3)

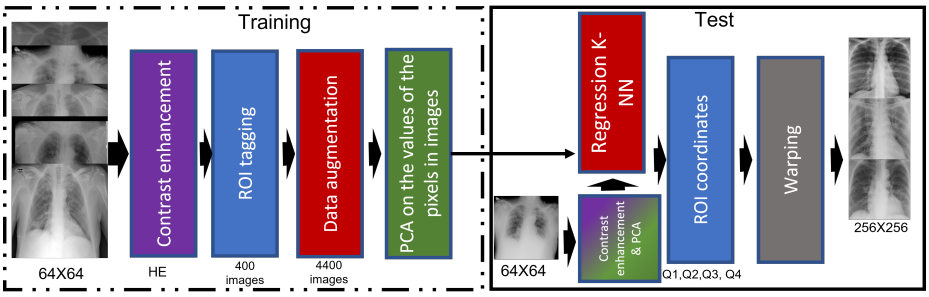

Fig. 1. Lung Finder Algorithm description. During the training phase, 400 images were tagged with their coordinates. PCA was applied to reduce the dimensionality of the images. In the testing phase, an example radiograph is provided as input, and the algorithm extracts the region of interest as the output. During the test phase, the test image is compared with its nearest neighbors to interpolate its coordinates. Finally, the algorithm outputs the extracted region of interest in a new image.

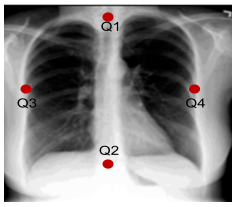

Fig. 2. Example of an array of coordinates Q1, Q2, Q3, and Q4 on a radiograph.

and Q4(x4,y4). Q1 and Q2 represent the length of the lungs, while Q3 and Q4 represent their width. In total, 400 images were labeled manually.

The labeling process is shown in figure 3. First, the Q1 point at the top of the lungs is manually located, using the spine as reference. The Q2 point is then placed at the bottom of the lungs. When the points Q1 and Q2 are placed, a straight line connecting them automatically appears, and at the midpoint of this line a perpendicular line is drawn containing the points Q3 and Q4. These last two points are constrained to be placed by the user only along the perpendicular line, and may have a different distance from the midpoint of the Q1Q2 line, due to the fact that the lungs are not symmetrical to each other.

2.2 Data Augmentation

Data augmentation is used in various machine learning tasks, such as image classification, to expand a limited database and avoid overfitting [22,34,19]. In the case of our algorithm, we have used a large dataset [4,31]. However, in order to have a set with ROI coordinates sufficiently varied we decided to generate artificial examples based on a randomly selected set, 400 images extracted from original set. The additional artificial images were generated by producing

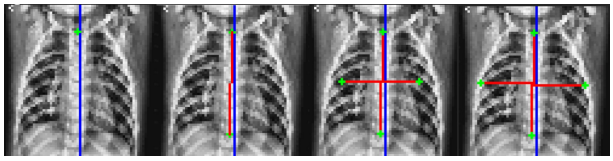

Fig. 3. Sequential placement of the points Q. First Q1 is placed, then Q2 so that Q3 and Q4 appear on the perpendicular line that crosses the midpoint of the line Q1Q2. Finally, Q3 and Q4 are adjusted.

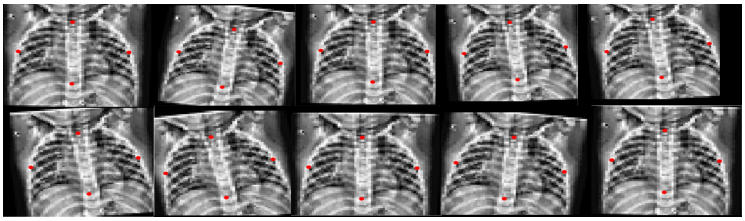

Fig. 4. Example of artificial images during data augmentation, applying translation and rotation operations.

random translations and rotations of the original images. Ten additional artificial images were created from each of the original 400, resulting in a total of 4400 images. First, it was necessary to define the range of operations on the images. For rotation, we set a range of -10 to 10 degrees, suggested by [31], and for translation a range from -5 to 5 pixels. These values were calculated by analyzing the coordinates of the 400 manually labeled images. In summary, the LFA training set contains 4400 images where the coordinates of the landmarks are normal distributed. Figure 4 shows an example of artificial images with their corresponding landmarks.

2.3 Estimating the Corner Coordinates of the Lung Region by Regression

As shown in figure 1, in the test stage a new image is introduced from which it is desired to obtain its region of interest. Contrast enhancement and feature reduction are automatically applied to the test image by projecting it onto the "Eigenfaces". The weights obtained in this projection are used in the "weighted regression K-NN" algorithm to find the most similar neighbors in the "Eigenfaces" space, using the Euclidean distance. In order to reduce the computational cost, the calculations are performed in a 64x64 resolution.

Once the nearest neighbors have been identified, a regression is performed using the coordinates of the ROIs of these neighbors with the aim of predicting the coordinates of the lungs in the test image. For this, the regression equations (1 and 2) are used, which are applied to each coordinate, either x or y, of each

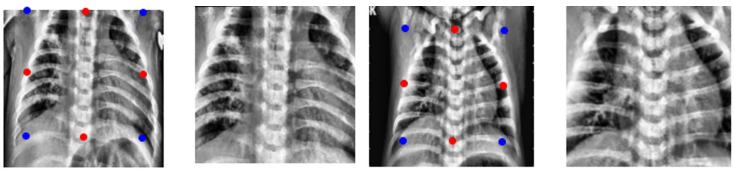

Fig. 5. Two examples of new images with their estimated ROI coordinates used to warp the inside region towards a fixed and normalized template.

Q landmark, until completing the entire set of landmarks (Q1, Q2, Q3 and Q4). The regression equations are detailed below:

$$x_i = \frac{1}{k} \sum_{i=1}^{k} x_{ni}, \tag{1}$$

$$y_i = \frac{1}{k} \sum_{i=1}^{k} y_{ni}. \tag{2}$$

2.4 Image Warping

Once the coordinates are obtained through regression, a Warping operation [36] is used to extract the region of interest. In figure 5, examples of different test radiographs from the data set are presented along with their automatically estimated ROI coordinates of provisional landmarks (red dots). The calculated coordinates are geometrically transformed to obtain the corners of the ROI (final landmarks depicted as blue dots) that are used in the Warping operation towards a standard fixed size template. On the right side of each image the normalized image resulting from the LFA is shown.

3 Feature Reduction and Selection

After using the LFA on all radiographs in the data set to extract all regions of interest, these new images undergo additional preprocessing before being processed by a classifying algorithm. For our work, we propose the use of [39,18] Eigenfaces as a feature reduction method. In addition, we incorporated a statistical analysis of these features using Fisher's linear discriminant in order to preserve only the most discriminating features and weighing each of them according to their power of discrimination between classes. Together these two methods ensure obtaining a reduced number of discriminant features suitable for efficient classification using traditional classifiers.

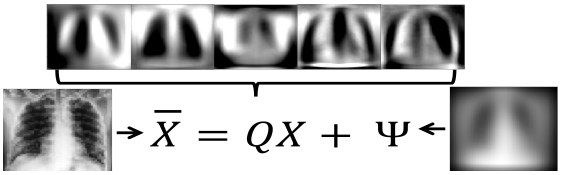

Fig. 6. Reconstructed image (left) is computed as a linear combination of the columns of matrix Q (in the middle) plus the mean image (right).

3.1 Eigenfaces for Dimensionality Reduction

Eigenfaces [39,18] is based on principal component analysis (PCA) and its objective is to reduce the dimensionality of the images in the [16] dataset. Because each pixel becomes a dimension or feature to be analyzed, processing 256x256 images can be time consuming. On the other hand, a large number of features, in comparison to a smaller number of training examples, could produce missclassification when euclidean distance based approaches as k-NN are used.

The resulting eigenfaces are sorted according to the greater variances of the training set, and can be used to reconstruct every image in the training set as a linear combination of them. Because the greatest amount of variance is concentrated in the first eigenfaces, we can use only a few number of them to efficiently represent all the training images and even novel ones. Thus, every normalized image from the training set can be represented with this compact set of features. Figure 6 shows the Eigenfaces equation, and the matrix Q which columns are the Eigenfaces. The $eigenfaces$ method works better and is capable of concentrating more variance in a less number of eigenfaces when training images are more similar. In our case, the normalized images are more similar to each other than the original images from the dataset. For this reason, the number of useful PCA features is necessarily reduced when using the proposed LFA.

3.2 Using the Fisher Discriminant to Reduce the Number of Useful Features

Fisher discriminant criterion also known as Fisher ratio FR has been used in Linear Discriminant Analysis for finding a linear projection of features that maximizes the separation between classes. Typically, only one important feature survives this process in two classes problems. However, since the PCA features are to some degree independent, we can use, in a naive fashion, the fisher ratio as a measure of separation between classes for a each feature.

This process is done by evaluating each feature individually, and making sure that the means of the observations in each class are as far apart as possible, while the variances within each class are as small as possible. Using this analysis, it is possible to select a number greater than 2 of those features obtained by the Eigenfaces method that best discriminate the classes in the data set [35].

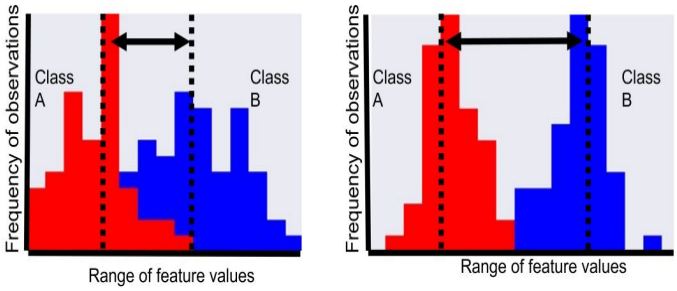

Fig. 7. Example of frequency distributions for each class. The discriminative capability of a feature can be visually assessed by the separation between the means of the histograms. The pair of histograms on the right shows a greater separation, indicating higher discrimination between classes. Conversely, the pair of histograms on the left exhibits lower discrimination.

The FR has been used in works such as the one mentioned in [14], and we denoted it as J. The FR formula is found in equation 3:

$$J_i = \frac{(\mu_{ic_0} - \mu_{ic_1})^2}{\sigma^2_{ic_0} + \sigma^2_{ic_1}}. \tag{3}$$

3.3 The Fisher Ratio as a Weight for each Feature

We propose to use the FR value as a weigh for each feature, in such a way that those features that possess a greater capacity for discrimination are amplified.

As a first step we standardize all selected features in order to give them a uniform relevance. Then, we calculate $\rho_K = \sqrt{J_K}$ for each feature k. Next, we normalize ρ_K as shown in equation 4:

$$\varrho_k = \frac{\rho_k}{\sum_{i=1}^{k} \rho_i}. \tag{4}$$

Finally, each ϱ_k is used to weigh all the standardized observations for the feature k.

4 Experiments Setup

In this work, the weighted K-NN and MLP algorithms were used for classification. Several experiments were conducted to compare the impact of different image preprocessing and feature enhancement algorithms on classification accuracy. The algorithms used in the training and testing stages included LFA for image normalization and preprocessing, Eigenfaces for dimensionality reduction, FR for selection of the best features, and W for

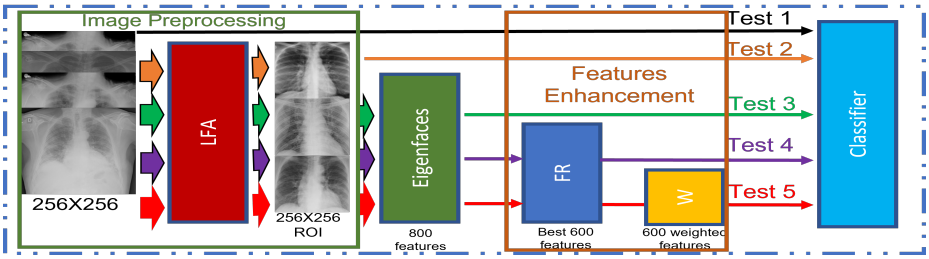

Fig. 8. Graphical representation of the different experiments conducted in image preprocessing. Each arrow represents a sequence of algorithms that may include image preprocessing or feature enhancement. A classification accuracy value is calculated for each arrow.

Table 2. Results of the Weighted K-NN and the MLP for the experiments using different preprocessing methods.

Classifier	E1	E2	E3	E4	E5
Weighted K-NN	82%	88%	88.3%	92.3%	95.6%
MLP	86%	90.8%	91%	93%	97.3%

weighting the features based on their discriminative capacity between classes. These two algorithms together aim to improve the discriminative ability of the features across classes. A total of five experiments were conducted for each classifier, which are described in Figure 8.

A total of 1300 COVID-19 images and 1300 normal images, all of size 256x256 pixels, were used. The region of interest was extracted from these images using the LFA algorithm, forming a bank of normalized images. The images were divided into 2000 training images, with 1000 from each class. For the testing phase, 600 images were selected, with 300 from each class. In experiments 1 and 2, 65,536 pixels, which constitute all the pixels of the images, were used. For experiments 3, 4, and 5, 600 features were employed.

For the MLP topology, 4 hidden layers with 120 neurons each and a single neuron in the output layer were utilized. The training was conducted for 100 epochs.

5 Experimental Results

Various values were tested for the parameter K in the weighted K-NN, and it was determined that the optimal value is 11. Conversely, experiments were conducted with various topologies and number of epochs in the MLP, yet no notable enhancements in classification precision were detected. The classification accuracy results for all experiments of each classifier are displayed in Table 2.

Additional tests were conducted in Experiment 5, varying the number of features for both classifiers. However, it was found that 600 is the optimal

Table 3. Results of Weighted K-NN and MLP for cross-validation.

Classifier	Test 1	Test 2	Test 3	Test 4	Test 5	Mean	Std
Weighted K-NN	95.6%	94.8%	95%	94%	95.2%	94.92%	0.593
MLP	97%	97.3%	96.4%	97%	96.8%	96.9%	0.331

number of features for both classifiers. Furthermore, Experiment 5 underwent cross-validation to demonstrate the consistency of the proposed set of algorithms in this work. Table 3 displays the results of the 5 tests, along with the average and standard deviation for each classifier.

6 Discussion of Results

For both classifiers, the following statements can be made regarding the experiments conducted in image preprocessing:

-Experiment one, where images undergo no preprocessing, generally displays the worst results.

-Experiment two illustrates that image normalization improves results compared to experiment one.

-In experiment three, where an image representation is projected onto the Eigenfaces space, no noteworthy enhancement is discernible.

-Experiment four highlights the importance of feature selection that effectively separates classes using FR, resulting in improved accuracy.

-Experiment five showcases the effectiveness of our algorithm sequence, which includes image normalization, feature selection, and weighting, yielding the best results.

Furthermore, the results exhibit robust consistency with minimal variability during cross-validation. Finally, the MLP achieved accuracy results that can compete with other state-of-the-art algorithms for classifying chest X-ray images.

7 Conclusions

In this paper, we have introduced a technique for the automatic detection and normalization of the Region of Interest (ROI) in chest radiographs. This approach is complemented by a feature selection method grounded in Fisher's criterion (FR) and utilizes PCA for automated COVID-19 detection. Through this approach, a reduced set of highly discriminative features is extracted. The outcomes underscore that the combination of both ROI alignment and feature selection processes leads to a significant improvement in classification accuracy. Notably, this improvement is evident when utilizing conventional classifiers such as weighted K-NN and MLP. These enhanced features demonstrate a notably superior classification capacity in comparison to the original pixel values. The reliability of the reported results is further solidified by the incorporation of cross-validation techniques in their acquisition.

The contributions of this study encompass a method for normalizing the ROI in lung images and a technique for selecting highly discriminative features using FR. Our approach achieves accuracy values that compete with other state-of-the-art works employing CNN-based techniques.

For future work, the ROI normalization technique can be applied to other databases and for the detection of other lung diseases. Additionally, the feature selection and weighting approaches can be tested to enhance the accuracy of other classification algorithms. Finally, we provide the link to download and use the LFA code (https://github.com/picazo07/LFA.git).

References

1. Alzahrani, S. A., Al-Salamah, M. A., Al-Madani, W. H., Elbarbary, M. A.: Systematic review and meta-analysis for the use of ultrasound versus radiology in diagnosing of pneumonia. Crit Ultrasound Journal, vol. 9, no. 1, pp. 1–11 (2017)
2. Amatya, Y., Rupp, J., Russell, F. M., Saunders, J., Bales, B., House, D. R.: Diagnostic use of lung ultrasound compared to chest radiograph for suspected pneumonia in a resource-limited setting. Int J Emerg Med, vol. 11, no. 8, pp. 1–5 (2018)
3. Changawala, V., Sharma, K., Paunwala, M.: Averting from convolutional neural networks for chest x-ray image classification. In: 2021 IEEE International Conference on Signal Processing, Information, Communication and Systems (SPICSCON). pp. 14–17 (2021)
4. Chowdhury, M. E. H., Rahman, T., Khandakar, A., Mazhar, R., Kadir, M. A., Mahbub, Z. B., Islam, K. R., Khan, M. S., Iqbal, A., Emadi, N. A., Reaz, M. B. I., Islam, M. T.: Can AI help in screening viral and COVID-19 pneumonia? IEEE Access, vol. 8, pp. 132665–132676 (2020)
5. Cleophas, T., Zwinderman, A.: Machine Learning in Medicine: Part Two. Machine Learning in Medicine, Springer Netherlands (2013)
6. in Data, O. W.: Covid-19 data explorer. https://ourworldindata.org/explorers/coronavirus-data-explorer (2022), accessed: February 2022
7. Do, T.-N., Le, V.-T., Doan, T.-H.: Svm on top of deep networks for covid-19 detection from chest x-ray images. Journal of information and communication convergence engineering, vol. 20, pp. 219–225 (2022)
8. El-kenawy, E.-S., Mirjalili, S., Ibrahim, A., Alrahmawy, M., Elsaid, M., Mounir, R., Eid, M.: Advanced meta-heuristics, convolutional neural networks, and feature selectors for efficient covid-19 x-ray chest image classification. IEEE Access, vol. 9, pp. 36019 – 36037 (2021)
9. Ertel, W., Black, N.: Introduction to Artificial Intelligence. Undergraduate Topics in Computer Science, Springer International Publishing (2018)
10. Gadermayr, M., Cooper, S. S., Klinkhammer, B., Boor, P., Merhof, D.: A quantitative assessment of image normalization for classifying histopathological tissue of the kidney. In: Pattern Recognition: 39th German Conference, GCPR 2017. pp. 3–13. Springer (2017)
11. Gazda, M., Plavka, J., Gazda, J., Drotar, P.: Self-supervised deep convolutional neural network for chest x-ray classification. IEEE Access, vol. 9, pp. 151972–151982 (2021)
12. González, R., Woods, R.: Digital Image Processing, Global Edition. Pearson Education (2018)

13. Hamza, A., Attique Khan, M., Wang, S.-H., Alhaisoni, M., Alharbi, M., Hussein, H. S., Alshazly, H., Kim, Y. J., Cha, J.: Covid-19 classification using chest x-ray images based on fusion-assisted deep bayesian optimization and grad-cam visualization. Frontiers in Public Health, vol. 10, pp. 1–17 (2022)
14. Ibis, E.: Sistema de aprendizaje automático para la detección de neumonía. Master's thesis, Benemérita Universidad Autónoma de Puebla, Puebla, México (2022)
15. Islam, N., Ebrahimzadeh, S., Salameh, J.-P., Kazi, S., Fabiano, N., Treanor, L., Absi, M., Hallgrimson, Z., Leeflang, M. M., Hooft, L., van der Pol, C. B., Prager, R., Hare, S. S., Dennie, C., Spijker, R., Deeks, J. J., Dinnes, J., Jenniskens, K., Korevaar, D. A., Cohen, J. F., Van den Bruel, A., Takwoingi, Y., van de Wijgert, J., Damen, J. A., Wang, J., McInnes, M. D.: Thoracic imaging tests for the diagnosis of covid-19. Cochrane Database Syst Rev, vol. 3, no. 3, pp. 1–145 (2021)
16. Jolliffe, I.: Principal Component Analysis. Springer Series in Statistics, Springer (2002)
17. Khan, A., Khan, S., Saif, M., Batool, A., Sohail, A., Khan, M.: A survey of deep learning techniques for the analysis of covid-19 and their usability for detecting omicron. Journal of Experimental and Theoretical Artificial Intelligence, vol. 2023, pp. 1–43 (2023)
18. Kirby, M., Sirovich, L.: Application of the karhunen-loeve procedure for the characterization of human faces. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 12, no. 1, pp. 103–108 (1990)
19. Kociolek, M., Strzelecki, M., Obuchowicz, R.: Does image normalization and intensity resolution impact texture classification?. Computerized Medical Imaging and Graphics, vol. 81, pp. 1–17 (2020)
20. Liu, W., Zheng, Y., Zhou, X., Chen, Q.: Axis orbit recognition of the hydropower unit based on feature combination and feature selection. Sensors, vol. 23, no. 6, pp. 1–18 (2023)
21. Lv, C., Lu, Y., Lu, M., Feng, X., Fan, H., Xu, C., Xu, L.: A classification feature optimization method for remote sensing imagery based on fisher score and mrmr. Applied Sciences, vol. 12, pp. 1–19 (2022)
22. Mikołajczyk-Bareła, A., Grochowski, M.: Data augmentation for improving deep learning in image classification problem. In: 2018 international interdisciplinary PhD workshop (IIPhDW). pp. 117–122. IEEE (2018)
23. Moberg, A., Taléus, U., Garvin, P., Fransson, S.-G., Falk, M.: Community-acquired pneumonia in primary care: Clinical assessment and the usability of chest radiography. Scandinavian journal of primary health care, vol. 34, pp. 1–7 (2016)
24. Moeslund, T. B.: Introduction to Video and Image Processing: Building Real Systems and Applications. Undergraduate Topics in Computer Science, Springer London (2012)
25. Muljo, H. H., Pardamean, B., Purwandari, K., Cenggoro, T. W.: Improving lung disease detection by joint learning with covid-19 radiography database. Communications in Mathematical Biology and Neuroscience, vol. 2022, no. 1, pp. 1–24 (2022)
26. Mustafa Ghaderzadeh, M. A., Asadi, F.: X-ray equipped with artificial intelligence: Changing the COVID-19 diagnostic paradigm during the pandemic. BioMed research international, vol. 2021, pp. 1–16 (2021)
27. Niederman, M. S.: Community-acquired pneumonia. Annals of Internal Medicine, vol. 163, no. 7, pp. 1–16 (2015)

28. Nillmani, Sharma, N., Saba, L., Khanna, N. N., Kalra, M. K., Fouda, M. M., Suri, J. S.: Segmentation-based classification deep learning model embedded with explainable ai for covid-19 detection in chest x-ray scans. Diagnostics, vol. 12, no. 9, pp. 1–32 (2022)
29. Park, S.-H., Lee, S.-G.: A method of feature extraction on motor imagery eeg using fld and pca based on sub-band csp. Journal of KIISE, vol. 42, pp. 1535–1543 (2015)
30. Qin, C., Yao, D., Shi, Y., Song, Z.: Computer-aided detection in chest radiography based on artificial intelligence: a survey. BioMedical Engineering OnLine, vol. 17, no. 1, pp. 1–23 (2018)
31. Rahman, T., Khandakar, A., Qiblawey, Y., Tahir, A., Kiranyaz, S., Abul Kashem, S. B., Islam, M. T., Al Maadeed, S., Zughaier, S. M., Khan, M. S., Chowdhury, M. E.: Exploring the effect of image enhancement techniques on covid-19 detection using chest x-ray images. Computers in Biology and Medicine, vol. 132, pp. 1–16 (2021)
32. Ridzuan, M., Bawazir, A. A., Navarette, I. G., Almakky, I., Yaqub, M.: Self-supervision and multi-task learning: Challenges in fine-grained covid-19 multi-class classification from chest x-rays. In: Annual Conference on Medical Image Understanding and Analysis. pp. 234–250. Springer (2022)
33. Salvatore, C., Interlenghi, M., Monti, C. B., Ippolito, D., Capra, D., Cozzi, A., Schiaffino, S., Polidori, A., Gandola, D., Alì, M., Castiglioni, I., Messa, C., Sardanelli, F.: Artificial intelligence applied to chest x-ray for differential diagnosis of covid-19 pneumonia. Diagnostics, vol. 11, no. 3, pp. 1–12 (2021)
34. Shorten, C., Khoshgoftaar, T. M.: A survey on image data augmentation for deep learning. Journal of Big Data, vol. 6, no. 1, pp. 1–48 (2019)
35. Silva, T. S.: An illustrative introduction to fisher's linear discriminant. https://sthalles.github.io/fisher-linear-discriminant/ (2019)
36. Szeliski, R.: Computer Vision: Algorithms and Applications. Springer-Verlag (2010)
37. Talaat, A., Yousri, D., Ewees, A., Al-qaness, M. A. A., Damaševičius, R., Elsayed Abd Elaziz, M.: Covid-19 image classification using deep features and fractional-order marine predators algorithm. Scientific reports, vol. 10, pp. 15364 (2020)
38. Ticinesi, A., Lauretani, F., Nouvenne, A., Mori, G., Chiussi, G., Maggio, M., Meschi, T.: Lung ultrasound and chest x-ray for detecting pneumonia in an acute geriatric ward. Medicine, vol. 95, no. 27, pp. 1–7 (2016)
39. Turk, M., Pentland, A.: Eigenfaces for Recognition. Journal of Cognitive Neuroscience, vol. 3, no. 1, pp. 71–86 (1991)