



Research Article

Chest Disease Image Classification Based on Spectral Clustering Algorithm

Song Jiang^{1*}, Yuan Gu², Ela Kumar³

¹Department of Biochemistry, Huzhou Institute of Biological Products Co., Ltd. China

²Department of Statistics, The George Washington University, Washington, DC, USA

³Department of Computer Science and Engineering, K L College of Engineering, Vaddeswaram, India

E-mail: songjiang@hzbio.net

Received: 10 April 2023; **Revised:** 31 May 2023; **Accepted:** 7 June 2023

Abstract: Nowadays, the emergence of new technologies gives rise to a huge amount of data in different fields such as public transportation, community services, scientific research, etc. Due to the aging population, healthcare is becoming more important in our daily life to reduce public burdens. For example, manually archiving massive electronic medical files, such as X-ray images, is impossible. However, precise classification is essential for further work, such as diagnosis. In this report, we applied a spectral clustering algorithm to classify chest disease X-ray images. We also employed the “pure” K-means algorithm for comparison. Three types of indexes are used to quantify the performances of both algorithms. Our analysis result shows that spectral clustering can successfully classify chest X-ray images based on the presence of disease spots on the lungs and the performance is superior to “pure” K-means clustering.

Keywords: spectral clustering, K-means, chest disease, image, classification

1. Introduction

Nowadays, chest diseases have become a significant threat to human health [1]. These diseases manifest in various types, each with its own set of symptoms and severity levels. To accurately diagnose chest diseases, medical professionals rely heavily on diagnostic imaging techniques, with X-ray being the most commonly used due to its speed and affordability [2]. X-ray has become a primary tool for screening and identifying various chest ailments such as pneumonia, pneumothorax, and masses [3].

Among these diseases, about half a billion persons are suffered from pneumonia and about 4 million people die from it per year [4]. For example, lung cancer is one of the malignant tumors, which has the highest incidence and mortality in the world, since the 5-year survival rate of the patient is around 16% [5]. Early diagnosis and treatment can significantly reduce mortality resulting from chest diseases [6]. However, understanding chest X-ray needs a lot of professional knowledge. So far, X-ray images are typically explained by radiologists [7]. But misdiagnosis always happens, because of the diverse chest pathological features and the potential fatigue or lack of experience of radiologists [8]. Therefore, there is a crucial need to develop chest X-ray-assisted diagnostic algorithms that can aid radiologists in providing timely and effective treatment to patients with chest diseases [9].

Recently, the emergence of artificial intelligence techniques has attracted popularity worldwide because they

Copyright ©2023 Song Jiang, et al.

DOI: <https://doi.org/10.37256/rrcs.2120232742>

This is an open-access article distributed under a CC BY license
(Creative Commons Attribution 4.0 International License)

<https://creativecommons.org/licenses/by/4.0/>

can be applied in the biomedical fields, including skin cancer diagnosis, standard plane detection and localization in fetal ultrasound, lung nodule detection, etc. [10]. At present, this technology has made some progress in chest disease diagnosis [11]. However, artificial intelligence-based diagnosis technique still has many obvious drawbacks. For example, these include low-level feature recognition, small proportion of disease spots, and changeable disease location between the chest disease and the remaining normal area, compared with those from real images, due to the coarse-grained recognition [12]. As a consequence, it is hard to make connections between features, and finding the subtle traits that fully characterize the object is not straightforward, so the fine-grained classification of chest diseases is very challenging due to the difficulty of finding discriminative features. It is also difficult to find the most representative features for fine-grained classification of chest diseases [13, 14].

On the other hand, the imbalance between normal and disease X-ray images makes it challenging to design accurate diagnostic algorithms. This can lead to misdiagnosis, as the gradient direction tends to favor normal X-ray images [15, 16]. Current algorithms can only detect whether an image is diseased or not, but have difficulty identifying the specific type of disease and differences between them [17]. To conquer those problems, binary cross-entropy is used in this study to balance the number of disease and normal X-ray images and compensate for the loss function [18].

Among graph mining techniques for imaging classification, clustering algorithms are extraordinarily critical, since they can categorize objects into different groups according to their similarities. Traditional clustering algorithms, such as K-means, are limited by their spherical assumptions, which can lead to local optima. Spectral clustering has rapidly developed in recent years to compensate for these limitations by leveraging the eigenvalues of the similarity matrix [19-21].

Fiedler initialized spectral clustering algorithm, unlike traditional clustering algorithms, spectral clustering does not rely on shape assumptions and can find the global optimum [22]. It has been widely used in very large-scale integration (VLSI) design, load balancing, parallel computing, and sparse matrix partitioning. Spectral clustering is advantageous in both theoretical and practical applications, and it can be applied to any distribution pattern without considering dependency assumptions [23, 24]. In the biomedical field, spectral clustering can be used for image segmentation, including classifying normal and abnormal medical images such as pneumonia X-ray images. If successful, the algorithm could be applied to other chest diseases and other medical imaging modalities like magnetic resonance imaging (MRI) and ultrasound images.

Previous articles have shown that researchers are focusing on image classification. We plan to use the spectral clustering algorithm to classify normal and abnormal medical images, starting with pneumonia X-ray images and expanding to other chest diseases such as infiltration, effusion, mass, and nodule [25, 26]. The algorithm will also be applied to other medical imaging modalities like MRI and ultrasound images. These automatic classification models can be used as a pre-screening step to save time and facilitate doctors in making diagnoses.

In this paper, we will present an overview of the relevant research related to our topic of interest, followed by a detailed description of our data and the methodology we will use. Then, we will discuss the data extraction and preprocessing steps, the application of various models supported by theoretical underpinnings, and a comparative analysis of the performance of these models. In the final section, we will provide a comprehensive evaluation of the results obtained and highlight potential areas for future research.

2. Related research

Image classification is an important task in the field of medical imaging, particularly for X-ray images. Over the years, various techniques and methods have been developed to classify X-ray images accurately. One of the most popular approaches for X-ray image classification is deep learning. Among those deep learning models, numerous studies have investigated the performance of convolutional neural networks (CNNs). For example, Rajpurkar et al. proposed a deep learning algorithm that achieved state-of-the-art performance in classifying 14 different thoracic diseases using X-ray images [27].

Transfer learning has also been investigated for X-ray image classification. It involves utilizing pre-trained models on large data sets to extract features. For instance, Zhang et al. used a transfer learning approach by utilizing a pre-trained VGG-16 model for classifying chest X-ray images into normal and abnormal categories. Their study showed that the approach achieved high accuracy rates of over 90% [28].

In addition to utilizing the aforementioned deep learning techniques, traditional machine learning methods have also been studied for X-ray image classification. For instance, Hamed et al. developed a hybrid approach that combined histogram equalization with machine learning algorithms, such as K-Nearest Neighbor (KNN), Decision Tree, Support Vector Machine (SVM), etc. [29].

Overall, there have been significant developments in X-ray image classification using various techniques, ranging from deep learning to traditional machine learning methods. These studies have demonstrated high accuracy rates, which can be beneficial for early diagnosis and treatment of various diseases. However, there is still scope for improvement in terms of the different methods and their accuracy, as well as expanding the range of diseases that can be classified using X-ray images.

3. Methodology

3.1 Dataset

The dataset in our study includes chest X-ray images that have been categorized as normal or abnormal, with the abnormal category being caused by pneumonia (as shown in Figure 1). To ensure the accuracy of the analysis of chest X-ray images, a quality control process was implemented where all radiographs were initially screened to remove any scans that were deemed unreadable or of low quality. The images were evaluated and diagnosed by expert physicians before being cleared for use in the artificial intelligence (AI) system [30]. The dataset is quite large, with a total size of 1.15 GB and thousands of images. In order to improve efficiency and save time, a subset of 200 normal and 200 pneumonia images were randomly selected for further processing. The data can be downloaded from Kaggle, <https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia> [31].

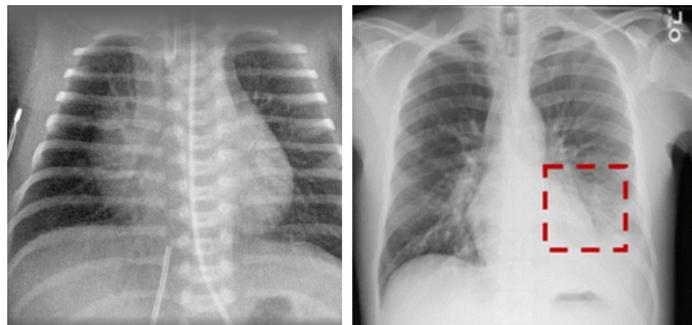


Figure 1. Normal (left) and abnormal (right, pneumonia) chest X-ray images

3.2 Classification

3.2.1 Spectral clustering classification

The primary approach of the whole spectral clustering classification procedure is depicted in the flowchart as shown in Figure 2. First, all randomly selected images are imported and then subjected to grayscale conversion and resizing as spectral clustering has specific requirements for the image name, grayscale, and size of images. These three steps are necessary for preparing the images for spectral clustering. Next, after applying spectral clustering classification, the classified images are restored into two different folders.

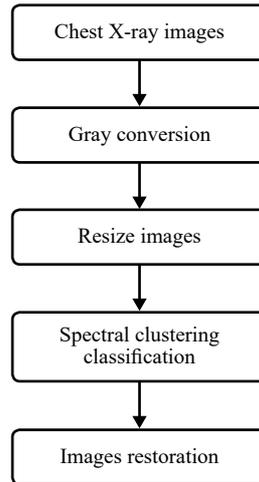


Figure 2. The flowchart of spectral clustering procedure

In spectral clustering, images are converted into nodes in a high-dimensional space. The distance between nodes determines their edge weights.

Once the nodes are created, a distance matrix is computed and used to generate the diagonal degree matrix. This degree matrix is then used to compute the Laplacian matrix, which is a key step in the spectral clustering algorithm. The Laplacian matrix helps to identify clusters of images based on their similarities and differences, and ultimately leads to the accurate classification of the images.

To find the distance matrix, usually V is used to represent a set of nodes, and use E to represent a set of edges. And a graph is G , so $G(V, E)$. Among them, V is the set of all nodes (v_1, \dots, v_n) . w_{ij} is the weight between v_i and v_j , $w_{ij} = w_{ji}$. If there is an edge between two nodes, exists $w_{ij} > 0$, otherwise, $w_{ij} = 0$. For any node in the graph, the corresponding degree W [Equation (1)] is the sum of all its adjacent edges [32].

In spectral clustering, a distance matrix is obtained by representing nodes as a set V and edges as a set E in a graph $G(V, E)$. V represents all nodes (v_1, \dots, v_n) , and w_{ij} is the weight between v_i and v_j , where $w_{ij} = w_{ji}$. If an edge exists between two nodes, $w_{ij} > 0$; otherwise, $w_{ij} = 0$. The degree d (Equation (1)) of a node in the graph is the sum of its adjacent edge weights, and is used to generate the Laplacian matrix for spectral clustering [32].

$$d_i = \sum_{j=1}^n w_{ij} \quad (1)$$

To describe the degrees of all nodes, we can get a $n \times n$ matrix of degrees D , which is a diagonal matrix [Equation (2)] [33].

$$D = \begin{pmatrix} d_1 & \cdots & \cdots \\ \vdots & \ddots & \vdots \\ \cdots & \cdots & d_n \end{pmatrix} \quad (2)$$

The adjacent matrix is named W [Equation (3)]. It's also an $n \times n$ matrix, the i th row and j th column is w_{ij} . Usually, we can use the weight among all nodes to construct a $m \times m$ similarity matrix, in which i th row and j th column have a corresponding weight W_{ij} [34].

$$W = \begin{pmatrix} w_{11} & w_{12} & \cdots & w_{1m} \\ w_{21} & w_{22} & \cdots & w_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ w_{m1} & w_{m2} & \cdots & w_{mm} \end{pmatrix} \quad (3)$$

There are three common methods to construct a similarity matrix: epsilon nearest neighbor, KNN, and fully connected graph.

In the epsilon nearest neighbor method, a distance threshold epsilon is set, and the Euclidean distance S_{ij} [Equation (4)] is used to evaluate the distance between any two nodes x_i and x_j [35].

$$S_{ij} = \|x_i - x_j\|_2^2 \quad (4)$$

In the KNN method, all nodes are traversed using KNN to find their KNNs, and the edge weights w_{ij} for these neighbors are computed. In a fully connected graph, all nodes have weights greater than zero and are connected to each other. Before performing clustering, the Laplacian matrix computation is computed as $L = D - W$ where D is the diagonal degree matrix and W is the adjacent matrix described earlier. The Laplacian matrix offers advantages over traditional clustering with complex shapes and varying densities in high-dimensional spaces due to the following reasons:

- (a) It is a symmetric matrix, this can be easily found because both D and W are symmetric matrices.
- (b) For any vector f , we have

$$f^T L f = \frac{1}{2} \sum_{i,j=1}^n w_{ij} (f_i - f_j)^2$$

- (c) The matrix is positive semi-definite and n is all nonnegative real number eigenvalues

$$0 = \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$$

In order to prevent bad results of clustering, it is necessary to limit the size of all sub-graphs. The most common way to do this is normalized cut, the formula [36] is

$$NCut(A_1, A_2, \dots, A_k) = \frac{1}{2} \sum_{i=1}^k \frac{W(A_i, \bar{A}_i)}{vol(A_i)}$$

The weight $\frac{1}{\sqrt{vol(A_j)}}$ is used to represent vector h for $h_i^T L h_i$

$$\begin{aligned} h_i^T L h_i &= \frac{1}{2} \sum_{m=1} \sum_{n=1} w_{mn} (h_{im} - h_{in})^2 \\ &= \frac{1}{2} \left(cut(A_i, \bar{A}_i) \frac{1}{vol(A_i)} + cut(\bar{A}_i, A_i) \frac{1}{vol(A_i)} \right) \\ &= \frac{cut(A_i, \bar{A}_i)}{vol(A_i)} \end{aligned}$$

Based on the basic knowledge of the clustering algorithm, it is clear that clustering can be used on various kinds of sample data and converges at global optimal. If the input $D = (x_1, \dots, x_n)$ contains k_1 dimensions after dimension reduction and k_2 dimensions after clustering, the output is a set of k clusters $C(c_1, \dots, c_k)$.

3.2.2 “Pure” K-means clustering classification

To compare the results, “pure” K-means is used to classify the images. The procedure is depicted in the flowchart as shown in Figure 3. The procedure is similar to the spectral clustering algorithm, except for the last two steps. Unlike spectral clustering, the “pure” K-means algorithm does not require complex computations such as distance matrix and Laplacian matrix and it can be imported from sklearn by setting the number of clusters to be 2. The last step involved attaching classified labels to the corresponding images, rather than creating two folders to store the classified images separately.

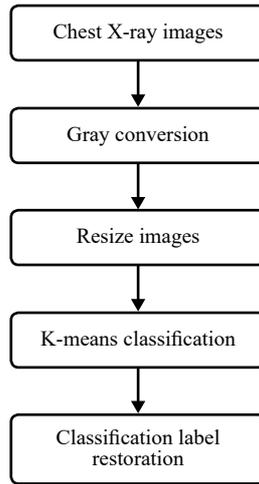


Figure 3. The flowchart of “pure” K-means clustering procedure

3.3 Measurement

Since clustering is an unsupervised learning, the common supervised learning measurement, such as precision, recall, accuracy, etc. based on the confusion matrix, cannot be used in this project. To qualify the performances of spectral clustering and “pure” K-means clustering, three types of indexes: silhouette coefficient, Davies-Bouldin index, and Calinski-Harabasz index are applied.

3.3.1 Silhouette coefficient

The silhouette coefficient, as shown in Equation (5), is calculated using the mean intra-cluster distance (a) and the mean nearest-cluster distance (b). The coefficient is from -1 to 1 with a higher coefficient indicating better cluster classification. A coefficient of 0 indicates that the clusters are close to each other, and the classification is poor. A negative coefficient indicates some data are assigned to the wrong clusters [37].

$$\begin{aligned}
 s(i) &= \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \\
 s(i) &= \begin{cases} 1 - a(i) / b(i), & \text{if } a(i) < b(i) \\ 0, & \text{if } a(i) = b(i) \\ b(i) / a(i) - 1, & \text{if } a(i) > b(i) \end{cases} \\
 -1 \leq s(i) \leq 1 & \qquad \qquad \qquad (5)
 \end{aligned}$$

3.3.2 Davies-Bouldin index

The Davies-Bouldin index, as shown in Equation (6), is used to measure the average similarity of each cluster with its most similar cluster. The index ranges from 0 to infinity, with a lower value indicating better cluster classification. A value of 0 indicates perfect clustering, while a higher value indicates worse clustering [38].

$$DB = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} R_{ij} \quad (6)$$

3.3.3 Calinski-Harabasz index

The Calinski-Harabasz index, as shown in Equation (7), is also known as the variance ratio criterion. It measures the value of the ratio between the within-cluster dispersion and the between-cluster dispersion. A higher index value indicates better cluster separation and classification, while a lower value indicates less distinct clusters [39].

$$s = \frac{\text{tr}(B_k)}{\text{tr}(W_k)} \times \frac{n_E - k}{k - 1} \quad (7)$$

where

$$W_k = \sum_{q=1}^k \sum_{x \in C_q} (x - c_q)(x - c_q)^T$$

$$B_k = \sum_{q=1}^k n_q (c_q - c_E)(c_q - c_E)^T$$

4. Experiments and results

4.1 Spectral clustering procedure

4.1.1 Renaming

To begin the spectral clustering process, 200 normal images and 200 pneumonia images were randomly selected and imported. The images were not named sequentially, so a renaming process was carried out. For instance, the original names of the first 3 images were “normal 1,” “normal 10,” and “normal 20.” After renaming, they became “NORMAL-1,” “NORMAL-2,” and “NORMAL-3,” respectively. Figure 4 shows the first 20 renamed normal images.

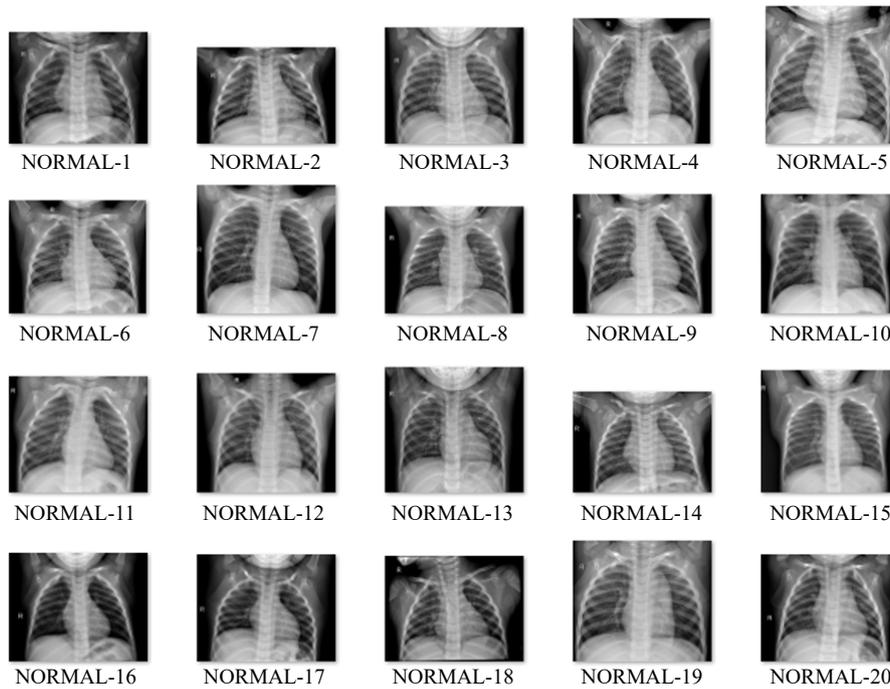


Figure 4. Sample of renamed normal images

4.1.1 Renaming

After renaming, the images were converted to grayscale using the Open Source Computer Vision Library (OpenCV), a powerful image processing tool [40]. The cv2.COLOR_BGR2GRAY function from OpenCV was applied to convert the images to grayscale. Figure 5 displays the first 20 normal images after grayscale conversion..

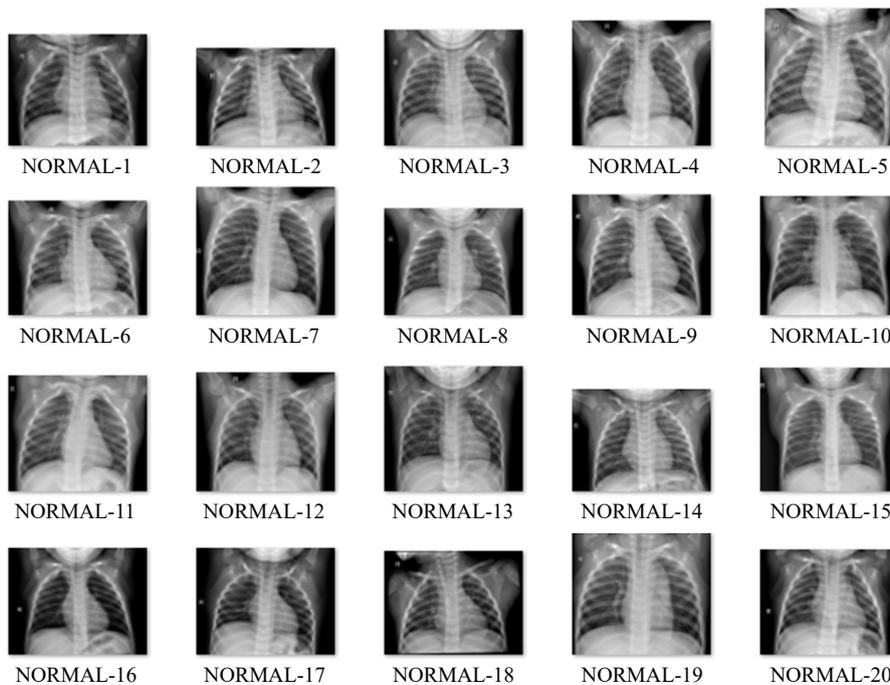


Figure 5. Sample of renamed normal images after grayscale conversion

4.1.3 Renaming

After grayscale conversion, the images needed to be resized to a smaller size than their original size. This was accomplished using the Python Imaging Library (PIL), another powerful tool for image processing [41]. Figure 6 displays the first 20 resized normal images. The images are presented as large icons in their original renamed, grayscale-converted, and resized forms. The resized images are significantly smaller in size compared to the remaining images.



Figure 6. Sample of grayscale conversion images after resizing

4.2 Classification

The created spectral clustering algorithm was then applied to process the images. First, the images were loaded and Principal Component Analysis (PCA) was performed. Next, the distance matrix and Laplacian matrix were computed. After getting eigenvectors, a feature vector was created from k first eigenvectors by stacking them as columns, followed by applying K-means clustering. Here, k was set to 2. The classified images were saved into two different folders. Figure 7 shows the classified images by spectral clustering, with one folder containing a sample of one type classified first 30 images and another folder containing a sample of another type classified first 30 images.

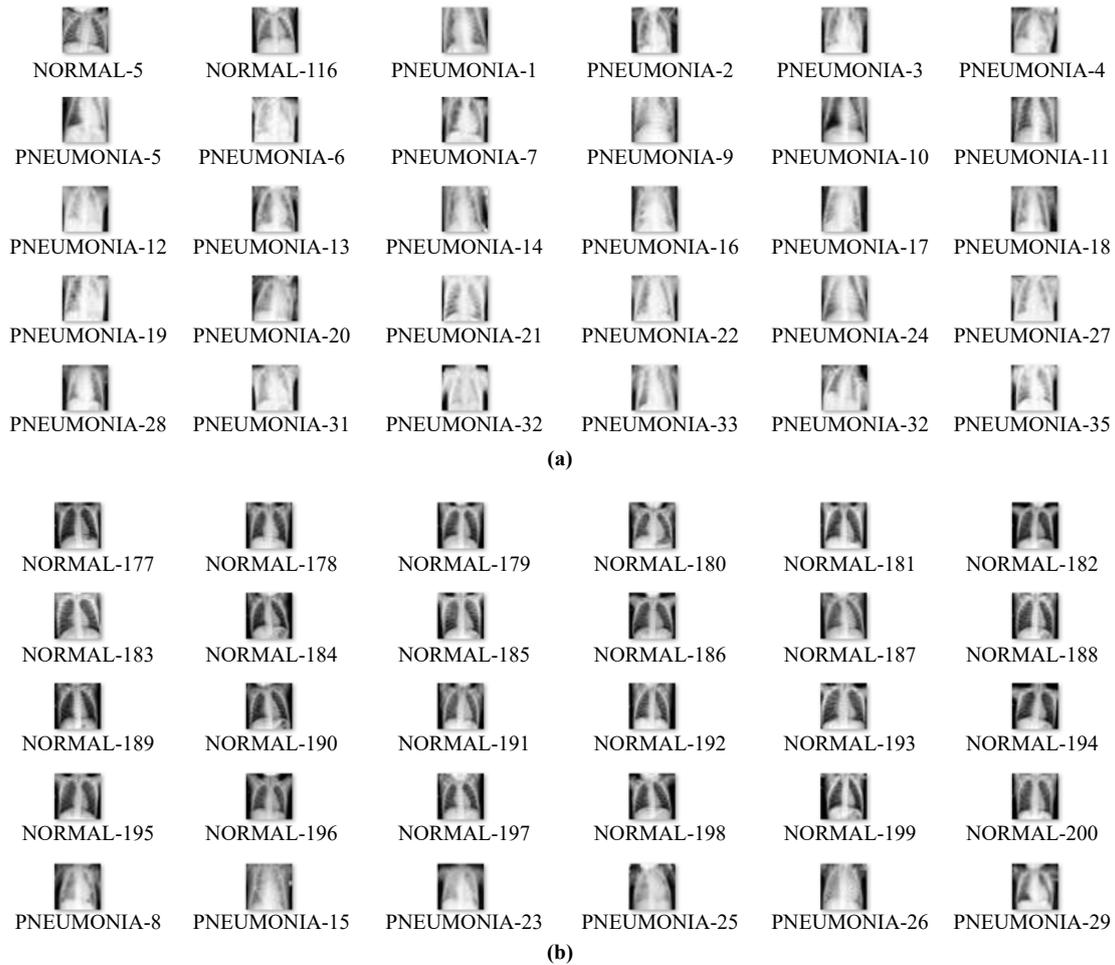


Figure 7. The classified images by spectral clustering for each folder: (a) sample of one type classified first 30 images in one folder and (b) sample of another type classified first 30 images in the second folder

Figure 8 and Figure 9 display the results of spectral clustering and “pure” K-means clustering on 400 images, with the expectation of 200 normal images and 200 pneumonia images being divided into two separate folders. In practice, after applying spectral clustering, 214 images are in one folder and 186 images are in another folder. In the 214 images folder, there are 198 normal images and 16 pneumonia images. In the 186 images folder, there are 2 normal images and 184 pneumonia images. This indicates that the classification was mostly successful. However, the results of “pure” K-means clustering are very different. One folder contains 59 images, and another folder contains 341 images, indicating a severe imbalance. The folder with 59 images has 22 normal images and 37 pneumonia images while the folder with 341 images has 178 normal images and 163 pneumonia images. It is obvious that in each folder, no one category dominates, indicating an unsuccessful classification.

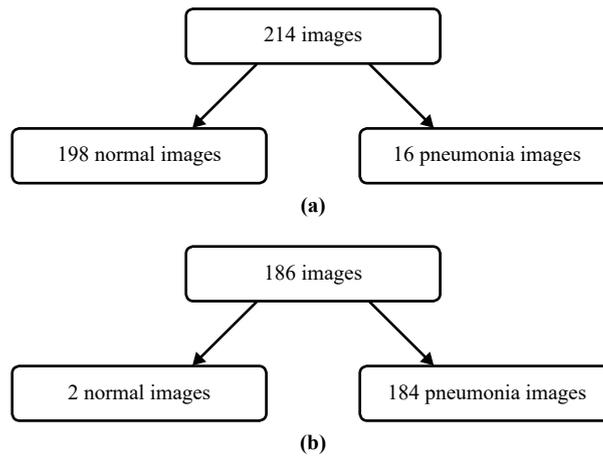


Figure 8. The details of classified images by spectral clustering

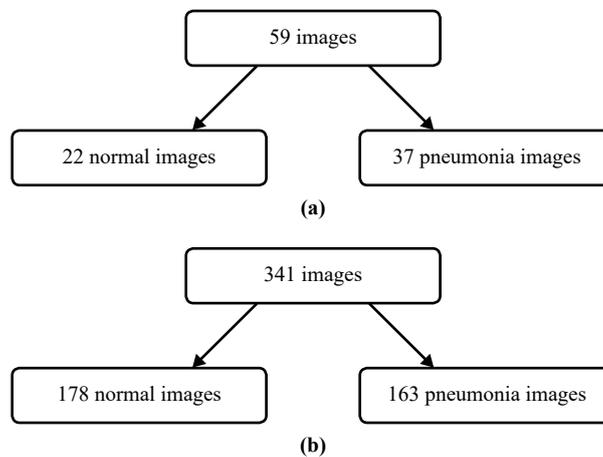


Figure 9. The details of classified images by “pure” K-means clustering

4.3 Performance

Table 1 shows the performance results of spectral clustering and “pure” K-means clustering. For the silhouette coefficient, the spectral clustering is 0.443, which is significantly higher than the “pure” K-means coefficient of 0. A silhouette coefficient of 0 indicates overlapping clusters [42]. While 0.443 is not a “perfect” value, it is an improvement over the alternative. Spectral clustering also outperformed “pure” K-means in terms of the Davies-Bouldin index, achieving a score of 1 compared to the “pure” K-means score of 8. Scores closer to zero indicate better partitioning [43]. Additionally, the Calinski-Harabasz index was higher for spectral clustering at 311 compared to the “pure” K-means score of 2.8, demonstrating that spectral clustering outperformed “pure” K-means clustering [44].

Table 1. The results of quantified performances made by spectral clustering and “pure” K-means clustering

	Spectral clustering	“Pure” K-means clustering
Silhouette coefficient	0.443	0.00
Davies-Bouldin index	1.025	8.076
Calinski-Harabasz index	311.582	2.852

5. Conclusion

Based on the above results and analysis, it can be concluded that: (1) spectral clustering is able to successfully classify chest X-ray images based on the presence of disease spots on the lungs and (2) the performance of spectral clustering is superior to “pure” K-means clustering as shown by the three unsupervised indexes used in the evaluation.

6. In the Future

Despite achieving satisfactory results, there are still some limitations and areas for improvement in this study. Firstly, the computational efficiency of spectral clustering can be optimized, especially for larger datasets. This can be achieved by implementing parallel computing techniques or using more efficient algorithms for computing the distance and Laplacian matrices.

Secondly, to make the classification more meaningful and realistic, it is necessary to include more types of chest diseases in the dataset and to increase the number of clusters accordingly. This will provide a more comprehensive analysis of the chest X-ray images and improve the accuracy of the classification results.

Lastly, the “pure” K-means clustering method used for comparison performed poorly in this study, and further improvements are needed to make it more effective. This can include using more advanced initialization methods, such as K-means++, or applying feature engineering techniques to improve the quality of the input data. Overall, these improvements can enhance the performance and applicability of the spectral clustering algorithm in the field of medical image analysis.

Conflict of interest

The authors declare that there is no conflict of interest.

References

- [1] Coultas DB, Samet JM. Chest diseases: Passive smoking and health. *Western Journal of Medicine*. 1986; 144(3): 350.
- [2] Graham RA, Wang S, Catalano PJ, Haller DG. Postsurgical surveillance of colon cancer: preliminary cost analysis of physician examination, carcinoembryonic antigen testing, chest X-ray, and colonoscopy. *Annals of Surgery*. 1998; 228(1): 59-63. <https://doi.org/10.1097/00000658-199807000-00009>
- [3] Seder DB, Christman RA, Quinn MO, Knauft ME. A 45-year-old man with a lung mass and history of charcoal aspiration. *Respiratory Care*. 2006; 51(11): 1251-1254.
- [4] Whizar-Lugo V, Saucedo-Gastelum A, Hernández-Armas A, Garzón-Garnica F, Granados-Gómez M. Chest trauma: an overview. *Journal Anesthesia & Critical Care: Open Access*. 2015; 3(1): 00082. <https://doi.org/10.15406/jaccoa.2015.03.00082>
- [5] Benediktsson K, Perbeck L. Capsular contracture around saline-filled and textured subcutaneously-placed implants in irradiated and non-irradiated breast cancer patients: five years of monitoring of a prospective trial. *Journal of Plastic, Reconstructive & Aesthetic Surgery*. 2006; 59(1): 27-34. <https://doi.org/10.1016/j.bjps.2005.08.005>
- [6] National Lung Screening Trial Research Team. Reduced lung-cancer mortality with low-dose computed tomographic screening. *New England Journal of Medicine*. 2011; 365(5): 395-409. <https://doi.org/10.1056/NEJMoa1102873>
- [7] Spampinato C, Palazzo S, Giordano D, Aldinucci M, Leonardi R. Deep learning for automated skeletal bone age assessment in X-ray images. *Medical Image Analysis*. 2017; 36: 41-51. <https://doi.org/10.1016/j.media.2016.10.010>
- [8] Liang G, Zheng L. A transfer learning method with deep residual network for pediatric pneumonia diagnosis. *Computer Methods and Programs in Biomedicine*. 2020; 187: 104964. <https://doi.org/10.1016/j.cmpb.2019.06.023>
- [9] Waite S, Kolla S, Jeudy J, Legasto A, Macknik SL, Martinez-Conde S, et al. Tired in the reading room: the influence of fatigue in radiology. *Journal of the American College of Radiology*. 2017; 14(2): 191-197. <https://doi.org/10.1016/j.jacr.2016.11.001>

org/10.1016/j.jacr.2016.10.009

- [10] Schlemper J, Oktay O, Chen L, Matthew J, Knight C, Kainz B, et al. Attention-gated networks for improving ultrasound scan plane detection. *ArXiv [Preprint]* 2018. Version 1. <https://doi.org/10.48550/arXiv.1804.05338>
- [11] Er O, Yumusak N, Temurtas F. Chest diseases diagnosis using artificial neural networks. *Expert Systems with Applications*. 2010; 37(12): 7648-7655. <https://doi.org/10.1016/j.eswa.2010.04.078>
- [12] Alemzadeh M, Boylan C, Kamath M. Review of texture quantification of CT images for classification of lung diseases. *Critical Reviews™ in Biomedical Engineering*. 2015; 43(2-3): 183-200. <https://doi.org/10.1615/CritRevBiomedEng.2015011026>
- [13] Dunnmon JA, Yi D, Langlotz CP, Ré C, Rubin DL, Lungren MP. Assessment of convolutional neural networks for automated classification of chest radiographs. *Radiology*. 2019; 290(2): 537-544. <https://doi.org/10.1148/radiol.2018181422>
- [14] Qin R, Wang Z, Jiang L, Qiao K, Hai J, Chen J, et al. Fine-grained lung cancer classification from PET and CT images based on multidimensional attention mechanism. *Complexity*. 2020; 2020: 6153657. <https://doi.org/10.1155/2020/6153657>
- [15] Tonpho T, Leelasantham A, Kiattisin S. Investigation of chest x-ray images based on medical knowledge and balanced histograms. In: *2010 International Symposium on Intelligent Signal Processing and Communication Systems*. Chengdu, China: IEEE; 2010. p.1-4. <https://doi.org/10.1109/ISPACS.2010.5704636>
- [16] Butnor KJ. Avoiding underdiagnosis, overdiagnosis, and misdiagnosis of lung carcinoma. *Archives of Pathology & Laboratory Medicine*. 2008; 132(7): 1118-1132. <https://doi.org/10.5858/2008-132-1118-AUOAMO>
- [17] Singh V, Misra AK. Detection of plant leaf diseases using image segmentation and soft computing techniques. *Information Processing in Agriculture*. 2017; 4(1): 41-49. <https://doi.org/10.1016/j.inpa.2016.10.005>
- [18] Zhang J, Xie Y, Li Y, Shen C, Xia Y. Covid-19 screening on chest x-ray images using deep learning based anomaly detection. *ArXiv [Preprint]* 2020. Version 4. <https://doi.org/10.48550/arXiv.2003.12338>
- [19] Hornik K, Feinerer I, Kober M, Buchta C. Spherical k-means clustering. *Journal of Statistical Software*. 2012; 50(10): 1-22. <https://doi.org/10.18637/jss.v050.i10>
- [20] Yan D, Huang L, Jordan MI. Fast approximate spectral clustering. In: *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, United States: Association for Computing Machinery; 2009. p.907-916. <https://doi.org/10.1145/1557019.1557118>
- [21] Sachdev V, Tian X, Gu Y, Nichols J, Sidenko S, Li W, et al. A phenotypic risk score for predicting mortality in sickle cell disease. *British Journal of Haematology*. 2021; 192(5): 932-941. <https://doi.org/10.1111/bjh.17342>
- [22] Ning H, Xu W, Chi Y, Gong Y, Huang TS. Incremental spectral clustering by efficiently updating the eigen-system. *Pattern Recognition*. 2010; 43(1): 113-127. <https://doi.org/10.1016/j.patcog.2009.06.001>
- [23] Li Y, Sze SM, Chao TS. A practical implementation of parallel dynamic load balancing for adaptive computing in VLSI device simulation. *Engineering with Computers*. 2002; 18(2): 124-137. <https://doi.org/10.1007/s003660200011>
- [24] Zhu X, Zhang S, He W, Hu R, Lei C, Zhu P. One-step multi-view spectral clustering. *IEEE Transactions on Knowledge and Data Engineering*. 2018; 31(10): 2022-2034. <https://doi.org/10.1109/TKDE.2018.2873378>
- [25] Wielpütz M, Kauczor HU. MRI of the lung: state of the art. *Diagnostic and Interventional Radiology*. 2012; 18(4): 344-353. <https://doi.org/10.4261/1305-3825.dir.5365-11.0>
- [26] Kaplan I, Oldenburg NE, Meskell P, Blake M, Church P, Holupka EJ. Real time MRI-ultrasound image guided stereotactic prostate biopsy. *Magnetic Resonance Imaging*. 2002; 20(3): 295-299. [https://doi.org/10.1016/S0730-725X\(02\)00490-3](https://doi.org/10.1016/S0730-725X(02)00490-3)
- [27] Rajpurkar P, Irvin J, Zhu K, Yang B, Mehta H, Duan T, et al. Chexnet: Radiologist-level pneumonia detection on chest X-rays with deep learning. *ArXiv [Preprint]* 2017. Version 3. <https://doi.org/10.48550/arXiv.1711.05225>
- [28] Zhang X, Xue Y, Su X, Chen S, Liu K, Chen W et al. A transfer learning approach to correct the temporal performance drift of clinical prediction models: Retrospective cohort study. *JMIR Medical Informatics*. 2022; 10(11): e38053. <https://doi.org/10.2196/38053>
- [29] Hamed Y, Alzahrani AI, Mustaffa Z, Ismail MC, Eng KK. Two steps hybrid calibration algorithm of support vector regression and K-nearest neighbors. *Alexandria Engineering Journal*. 2020; 59(3): 1181-1190. <https://doi.org/10.1016/j.aej.2020.01.033>
- [30] Kermany DS, Goldbaum M, Cai W, Valentim CC, Liang H, Baxter SL, et al. Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell*. 2018; 172(5): 1122-1131. <https://doi.org/10.1016/j.cell.2018.02.010>
- [31] Kermany D, Zhang K, Goldbaum M. (2018). Labeled optical coherence tomography (OCT) and chest X-ray images for classification. *Mendeley Data*. <https://doi.org/10.17632/rscbjbr9sj.2>

- [32] Bettstetter C. On the minimum node degree and connectivity of a wireless multihop network. In: *Proceedings of the 3rd ACM international symposium on Mobile ad hoc networking & computing*. New York, United States: Association for Computing Machinery; 2002. p.80-91. <https://doi.org/10.1145/513800.513811>
- [33] Ballantine CS. Stabilization by a diagonal matrix. *Proceedings of the American Mathematical Society*. 1970; 25(4): 728-734. <https://doi.org/10.1090/S0002-9939-1970-0260765-X>
- [34] Ding C, He X. Linearized cluster assignment via spectral ordering. In: *Proceedings of the twenty-first international conference on Machine learning*. New York, United States: Association for Computing Machinery; 2004. p.30. <https://doi.org/10.1145/1015330.1015407>
- [35] Deng S, Du L, Li C, Ding J, Liu H. SAR automatic target recognition based on Euclidean distance restricted autoencoder. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. 2017; 10(7): 3323-3333. <https://doi.org/10.1109/JSTARS.2017.2670083>
- [36] Shi J, Malik J. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2000; 22(8): 888-905. <https://doi.org/10.1109/34.868688>
- [37] Aranganayagi S, Thangavel K. Clustering categorical data using silhouette coefficient as a relocating measure. In: *International Conference on Computational Intelligence and Multimedia Applications (ICCIMA 2007)*. Sivakasi, India: IEEE; 2007. p.13-17. <https://doi.org/10.1109/ICCIMA.2007.328>
- [38] Thomas JC, Peñas MS, Mora M. New version of Davies-Bouldin index for clustering validation based on cylindrical distance. In: *2013 32nd International Conference of the Chilean Computer Science Society (SCCC)*. Temuco, Chile: IEEE; 2013. p.49-53. <https://doi.org/10.1109/SCCC.2013.29>
- [39] Łukasik S, Kowalski PA, Charytanowicz M, Kulczycki P. Clustering using flower pollination algorithm and Calinski-Harabasz index. In: *2016 IEEE Congress on Evolutionary Computation (CEC)*. Vancouver, Canada: IEEE; 2016. p.2724-2728. <https://doi.org/10.1109/CEC.2016.7744132>
- [40] Bradski G. The OpenCV library. *Dr. Dobb's Journal of Software Tools*. 2000; 25(11): 120-125.
- [41] Umesh P. Image processing in python. *CSI Communications*. 2012; 36(9): 23-24.
- [42] Dembele D, Kastner P. Fuzzy C-means method for clustering microarray data. *Bioinformatics*. 2003; 19(8): 973-980. <https://doi.org/10.1093/bioinformatics/btg119>
- [43] Wang G, Wang Z, Chen W, Zhuang J. Classification of surface EMG signals using optimal wavelet packet method based on Davies-Bouldin criterion. *Medical and Biological Engineering and Computing*. 2006; 44: 865-872. <https://doi.org/10.1007/s11517-006-0100-y>
- [44] Kermani S, Samadzadehaghdam N, EtehadTavakol M. Automatic color segmentation of breast infrared images using a Gaussian mixture model. *Optik*. 2015; 126(21): 3288-3294. <https://doi.org/10.1016/j.ijleo.2015.08.007>