Applied Mathematics & Information Sciences
An International Journal

http://dx.doi.org/10.18576/amis/190117

Automatic Sarcoidosis Stage Classification Based on Gray Level Co-occurrence Matrix Features

Mohanad A. Deif[1,], Hani Attar[2,3], Mohamed A. Hafez[4], Waleed Alomoush[5] and Hussein Al-Faiz[6]*

[1]Department of Artificial Intelligence, College of Information Technology, Misr University for Science & Technology (MUST), 6th of October City 12566, Egypt
[2]Faculty of Engineering, Zarqa University, Zarqa, Jordan
[3]Faculty of Engineering, University of Business and Technology, Jeddah, 21448, Saudi Arabia
[4]Faculty of Engineering, FEQS, INTI -IU, Universi, Nilai, Malaysia
[5]School of Computing, Skyline University College, Sharjah, United Arab Emirates
[6]Institute of Power Engineering (IPE), Universiti Tenaga Nasional, Kajang, 43000, Malaysia

Received: 5 Jun. 2024, Revised: 28 Jul. 2024, Accepted: 28 Sep. 2024.
Published online: 1 Jan. 2025.

Abstract: The correct diagnosis and staging of this complex inflammatory disease that majorly affects the lungs, such as Sarcoidosis, are very important. Therefore, this research focuses on differentiating the four stages of Sarcoidosis using chest X-ray images with the application of three machine learning classifiers: K-Nearest Neighbours, Support Vector Machine, and Artificial Neural Network. GLCM was used for feature extraction after segmentation using Otsu's method, and K-means clustering was used to enhance feature reliability and accuracy. This shows that, in the case of using Otsu's, the distinction of the stages of Sarcoidosis is better since our analysis showed higher average Jaccard Indexes than K-means and no segmentation. Performance evaluation for the classifiers was done using several metrics, including accuracy, precision, recall, and the F1 score. It was found that the results, generally across most of the modelled stages, performed well using both KNN and SVM. However, for Stage 2, KNN produced the best result, producing an accuracy of 97.83%, while SVM immediately did so for Stage 4, creating an accuracy of 97.1%. Thus, both models poorly classified instances of Stage 4, while the ANN model had poor precision and recall for Stage 4. Although ANN had high accuracy with the Normal and Stage 4 classes, low recalls across the rest of the stages lowered its overall performance. The confusion matrices further reiterated that the accurate classification of Stage 4 sarcoidosis was still challenging. Results thus reiterate the need for refining segmentation and feature extraction techniques to gain better classifier performance. This study concludes that while machine learning classifiers show promise for sarcoidosis staging, significant segmentation, and feature extraction improvements are needed to achieve reliable and precise diagnostic outcomes.

Keywords: Sarcoidosis, GLCM, Features extraction, Computer Aided Diagnosis, Artificial Intelligence.

1 Introduction

Sarcoidosis is a rare multi-system inflammatory disease mainly where granulomas are formed in multiple organs, commonly including the lungs, with poorly understood causes [1, 4]. Sarcoid is not poorly misunderstood but is also commonly misdiagnosed, and diagnosing it remains a challenge for clinicians despite all of its history [5]. Diagnostic delay for Sarcoidosis is around 8 months, which causes unfavorable outcomes such as ongoing pain and discomfort for newly diagnosed patients [6].

The leading site of involvement is the lung, as 90% of its patients have manifestations in the lung [7], allowing it to be diagnosed radiographically. However, other symptoms include persistent dry cough, eye and skin manifestations, weight loss, fatigue, night sweats, and erythema nodosum [8].

Sarcoidosis is categorized into 4 stages [9], which are:

- Stage 1: Characterized by the presence of bilateral hilar adenopathy and lung function tests are impaired for 20% of patients in this stage.
- Stage 2: Characterized by bilateral hilar adenopathy and reticular opacities.
- Stage 3: Reticular opacities have shrinking hilar nodes (mainly infiltrates).
- Stage 4: Reticular opacities have fibrosis.

Chronic Sarcoidosis may persist for years, causing relentless loss of lung functions and destruction of lung alveolar architecture. Sarcoidosis is benign for most patients, but for some of its patients, it can be life-threatening with an age-mortality rate that's been increasing over time, and it is also suggested to increase cancer risks with cancer. Its diagnosis is often delayed due to how frequently its symptoms are regarded as manifestations of other pulmonary diseases.

Other studies found that sarcoidosis patients had a 20% increased cancer risk in 3–10 years after diagnosis and a

*Corresponding author e-mail: Mohanad.Deif@must.edu.eg

10% increased risk after more than 10 years after diagnosis. [10]. Sarcoidosis has noticeable co-morbidities such as hyperlipidemia, obesity, thyroid disease, diabetes, osteoporosis, coronary heart disease, asthma, hypertension, chronic renal disease, and chronic obstructive pulmonary disease (COPD) [11], making it very dangerous for people with underlying conditions, which may increase the difficulty of diagnosing it.

Traditional methods for diagnosing Sarcoidosis are based on clinical assessments, laboratory tests, and manual examining radiographic image examinations [4, 5]. Still, these methods are time-consuming and prone to errors or misinterpretation, an issue due to the disease's rarity.

Chest radiographs are abnormal in more than 90% of sarcoidosis patients with Sarcoidosis [7].

Automatic diagnosis using artificial intelligence (AI) and machine learning techniques holds immense potential for improving the management of rare diseases. These innovative approaches can facilitate early detection, enhance diagnostic accuracy, personalize treatment plans, streamline diagnostics, accelerate drug development, and optimize disease monitoring. By leveraging AI's ability to analyze vast amounts of medical data and identify patterns indicative of rare conditions, clinicians [10] can better serve patients with these often complex and challenging diseases. As a result, automatic diagnosis can revolutionize the landscape of rare disease diagnosis and treatment, ultimately leading to improved patient outcomes and a higher quality of life [10].

Machine learning techniques are being researched for auto-diagnosis due to the rarity of the Sarcoid disease, which causes a lot of experts to miss diagnosing it, so an auto-diagnosis system for it is expected to help a lot of Sarcoidosis in patients early.

1.1 Related Works

Recently, much effort has been put into the automated diagnosis of diseases (including Sarcoidosis). For example, Van der Sar et al. [11] proposed a new technology called "eNose," which analyses breathing patterns to diagnose cases of Sarcoidosis using machine learning techniques and achieved an accuracy of 87.1%.

Lovin Fosse et al. [12] proposed a machine learning system to differentiate between Sarcoidosis and another disease known as Lymphoma where 215 features were extracted from computed tomography PET scans segmented via RadiomiX toolbox and classified using a Random Forest, Support Vector Machine (SVM) with radial bias, Naive Bayes and Linear Regression. This proposed system achieved an accuracy of 85% per patient and 94% per lesion; however, it does not classify different stages of Sarcoidosis.

Baghdadi et al. [13] proposed a method using deep neural networks to differentiate between Sarcoidosis and Tuberculosis, another disease that Sarcoidosis is often confused with; that proposed method achieved an accuracy of 98.67%. However, the proposed method could not identify the exact stage of Sarcoidosis.

De Lima et al. [14] proposed a machine learning system to use a forced Oscillation Technique (FOT) to get signals from an individual respiratory to track symptom change with a rule-based fuzzy system to help clinicians track sarcoidosis system and diagnose it correctly.

Bobbio et al. [15] employed random forest on symptoms shown by patients to predict the outcome of cardiac sarcoidosis cases in a study involving 141 patients published in 2023. This system didn't actually diagnose patients and instead was applied to pre-diagnosed patients, it also focused on sarcoidosis cases in the heart rather than the lung.

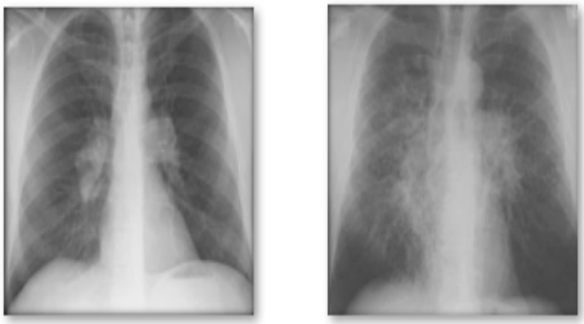
Urinbayev et al. [16] created an end-to-end diagnosis system combining three machine learning classifiers in sequence using the ChexNet Neural Network Model trained by multiple datasets of different body organs. The first model decides if an image is an x-ray, and the second classifies x-ray type while the last detects actual disease symptoms. The proposed system achieved an AUC score of 0.84; however, it was only capable of detecting symptoms such as effusions and nodules, but it was not capable of diagnosing a disease such as Sarcoidosis.

Lu et al. [17] proposed a method to use traditional machine learning models such as Logistic Regression and Support Vector Machine to predict adverse heart effects that may affect a sarcoidosis patient and managed to achieve high accuracy with an AUC of more than 0.90. Koth et al. [18] proposed a machine learning system that applies random forest on blood transcriptome data that distinguishes sarcoidosis cases with 0.92 sensitivity and 0.92 specificity.

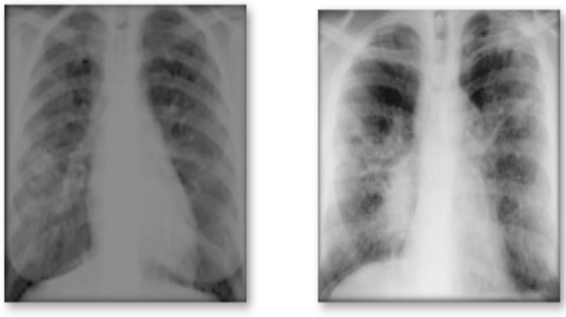
2 Materials and Methods

2.1 Dataset

To conduct this study, we compiled 460 normal and 230 Sarcoidosis images from six national hospitals in Egypt. Fig. 1 shows the dataset samples for the four stages of Sarcoidosis cases.



(a) (b)



(c) (d)

Fig. 1: The four stages of Sarcoidosis: a) Stage 1 , b) Stage 2 , c) Stage 3 and d) Stage 1.

2.2 Proposed Methodology

This section details the methodology employed to construct a system for diagnosing various stages of Sarcoidosis. The flowchart outlines in Fig. 2 a methodology for building a system to diagnose various stages of Sarcoidosis. First, chest X-ray images are pre-processed by resizing and normalizing them. Then, image segmentation is performed using either Otsu's thresholding or K-means clustering to isolate regions of interest. Next, feature extraction is conducted to quantify relevant characteristics from the segmented regions. Finally, machine learning models including SVM, KNN, and ANN are employed to classify the extracted features and differentiate between various stages of Sarcoidosis. The performance of the system is then evaluated using unspecified metrics.

The steps detailed methodology employed to construct a system for diagnosing various stages of sarcoidosis process are outlined as follows:

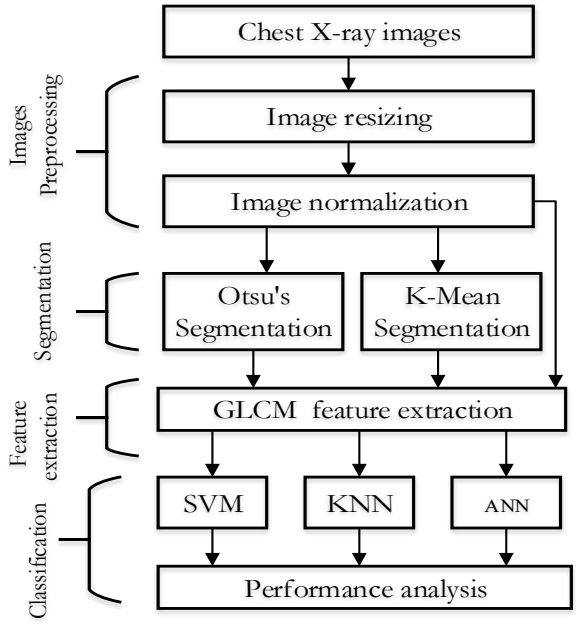


Fig. 2: Methodology for building a system for diagnosing different stages of Sarcoidosis.

2.3 Preprocessing

To enhance the classification outcomes, several pre-processing steps were undertaken. The tasks accomplished during pre-processing are detailed below:

2.3.1 Image Resizing

The X-ray image datasets for this study were sourced from multiple origins, resulting in various image sizes. Each network can only process a specific image size. Therefore, all images were resized to be $224 \, X \, 224$.

2.3.2 Image Normalization

Different X-ray device manufacturers may produce varying image qualities, potentially leading to overfitting due to device-specific pixel distributions [19]. To minimize this issue, contrast normalization was applied, standardizing pixel distribution and ensuring consistency. This normalization made X-rays appear slightly darker and provided a standardized view not typically seen in radiologists' regular work environments. Using the Reinhard and Macenko methods, X-ray images were stain-normalized, reducing color discrepancies and improving EfficientNet model classification accuracy.

2.3.3 Data Augmentation

The normalized X-ray images were further augmented before being fed into the segmentation step [20]. Data augmentation increases the number of original images in a collection, mitigating overfitting by providing a broader training dataset. In this study, the number of normal images was 460, double that of Sarcoidosis- images. Consequently, the Sarcoidosis images were augmented from 230 to 460. Table 1 illustrates the augmentation settings used on the stain-normalized X-ray images.

Table 1: The settings of the data augmentation applied to X-ray images.

Augmentation Type	**Value**
Rotation range	5° and 10°
Range of width shifts	0.1
Range of height shift	0.1

2.4 Segmentation

Following preprocessing, image segmentation is performed to isolate regions of interest (ROIs) within the chest X-ray images [21]. Our study has compared two alternative segmentation techniques: Otsu's Thresholding and K-Means Clustering.

2.4.1 Otsu's Thresholding

This technique employs a thresholding approach to automatically segment the foreground (potentially containing relevant features) from background pixels [22],[23].

Algorithm 1: Otsu's Segmentation in pseudocode.

```
Input: Grayscale image
Output: Thresholded image

1. Initialize:
- total_pixels = total number of pixels in the image
- histogram = array of size 256 (for each possible grayscale level) initialized to 0
- for each pixel in the image, increment the corresponding value in histogram
2. Compute the total mean intensity level of the image:
- sum_all = 0
- for i = 0 to 255:
sum_all += i * histogram[i]

3. Initialize:
- sum_background = 0
- weight_background = 0
- weight_foreground = 0
- max_variance = 0
- threshold = 0

4. Iterate through all possible thresholds:
for t = 0 to 255:
weight_background += histogram[t]
if weight_background == 0:
continue
weight_foreground = total_pixels - weight_background
if weight_foreground == 0:
break
sum_background += t * histogram[t]
mean_background = sum_background / weight_background
mean_foreground = (sum_all - sum_background) / weight_foreground
between_class_variance = weight_background * weight_foreground * (mean_background - mean_foreground)^2
if between_class_variance > max_variance:
max_variance = between_class_variance
threshold = t

5. Apply the threshold to the image:
for each pixel in the image:
if pixel intensity >= threshold:
set pixel to 255 (white)
else:
set pixel to 0 (black)

6. Return the thresholded image
```

2.4.2 *K-Means Clustering*

Alternatively, K-means clustering can be utilized to segment the image [24]. K-means clustering partitions the image into a predefined number of clusters based on pixel intensity values [25],[26]. Algorithm 1 lists the K-Means Clustering pseudocode.

Algorithm 1: K-Means Segmentation in pseudocode.

```
Input: Grayscale image, number of clusters k
Output: Segmented image

1. Initialize:
   - Select k initial cluster centers (centroids) randomly
   - clusters = array of size equal to the number of pixels, initialized to 0

2. Repeat until convergence:
   - Assign each pixel to the nearest cluster center:
      for each pixel in the image:
         min_distance = infinity
         for each cluster center:
            distance = |pixel intensity - cluster center intensity|
            if distance < min_distance:
             min_distance = distance
             assign pixel to this cluster
   - Update cluster centers:
     for each cluster:
        compute the mean of the pixel intensities
         assigned to this cluster
        update the cluster center to this mean value

3. Check for convergence:
   - If the cluster centers do not change significantly,
     stop the iteration

4. Output the segmented image:
   - Create a new image where each pixel intensity is
      replaced by the intensity of its cluster center
End
```

The performance of the segmentation step is calculated using the Jaccard index (JI) [27] .

2.5 *Feature Extraction*

Feature Extraction is the processing stage where useful features are extracted from the data. We extracted the Gray Level Co-occurrence Matrix ("GLCM") from the segmented images in the proposed system.

GLCM was used on the images to extract relevant features (such as Contrast, energy, and entropy), which will be the main data used to train the system. GLCM searches the image for pixel pairs with specific values and offsets; it's a very powerful technique for extracting features from images as it allows us to quantize the relationship between different pixels without being affected by their exact position or other values [28]-[31].

To reduce the computational costs of GLCM, black levels are quantized to l, a number lower than the gray levels in the base images (which equals 256); for this research, l was set to 4.

GLCM searches the image for pixel pairs with specific values and offsets; it's a very powerful technique for

extracting features from images as it allows us to quantize the relationship between different pixels without being affected by their exact position or other values where each element in the matrix can be calculated as follows:

$$P_{ij} = \sum_{m=1}^{M} \sum_{k=1}^{K} \{1, if\, I(m,k) = i, I(m+\delta x, k+\delta y) = j, 0, otherwise \quad (1)$$

Where $\delta = (\delta x, \delta y)$ is the displacement vector is expressed in pixels in the x and y directions, GLCM can be created with multiple displacement vectors.

A normalized GLCM $\tilde{P}$ represents the estimated probability of combinations of pairs of neighboring gray levels in the 8 immediate neighbors of a pixel and can be calculated as follows:

$$\tilde{P} = \frac{P}{\sum_{ij=0}^{N-1} P_{ij}} \quad (2)$$

GLCM is highly useful for "Feature Extraction," which is the stage of the processing where useful features are extracted from the data as it could also be used to calculate relevant statistical information such as:

2.5.1 Energy

(Also known as "Uniformity" and "Angular Second Momentum (ASM)") represent the degree of homogeneity of gray distribution, highlighting the "thickness" of the texture, which provides a stabilizing measure of grayscale patterns so that a larger Energy value suggests more stable regulations. Energy is calculated as the Quadratic sum of GLCM elements as follows:

$$ASM = \sum_{i=1}^{N-1} \tilde{P}_{ij}2 \quad (3)$$

N is the number of gray levels in the GLCM (also known as the dimension).

2.5.2 Dissimilarity

This is the meaning of the image, which acts as the optimal measure for GLCM texture and is one of the most useful GLCM features. It is not the mean of intensity values of all pixels but instead the intensity value of each pixel weighted by its frequency [3] and is calculated as follows:

$$\mu_j = \sum_{i=1}^{N-1} i\tilde{P}_{ij} \quad (4)$$

$$\mu_j = \sum_{i=1}^{N-1} j\tilde{P}_{ij} \quad (5)$$

Where N is the number of gray levels in the GLCM (also known as the dimension).

2.5.3 Correlation

The correlation shows similar discriminative capability as the Contrast, and it gives a quantitative measure of the connection or correlation of a pixel with its neighbors so that correlation identifies the linear gray-level dependency between the pixels at the specified positions relative to each other. The strength of the correlation between images is quantified by a correlation close to +1 or -1 if images are highly associated (positively or negatively [32], respectively). The correlation of the GLCM can be calculated as follows:

$$Correlation = \frac{\sum_{i=1}^{N-1} \tilde{P}_{ij}(i-\mu_i)(i-\mu_i)}{\sigma_i \sigma_j} \quad (6)$$

2.5.4 Contrast

Measures the local variation in intensity in the image (or its ROI). Contrast is calculated by measuring the difference between a pixel's intensity and neighbors over

the full image (ROI). Finding a high contrast contains very high or very low intensities (at the edges of the spectrum), and a low contrast shows that an image has a smoother range of grays or a sharper range of grays; therefore, high GLCM contrast is associated with spatial frequencies [3]. Contrast can be calculated as follows:

$$Contrast = \sum_{i=1}^{N-1} \tilde{P}_{ij}(i-j)^2 \quad (7)$$

Where N is the number of gray levels in the GLCM (also known as the dimension).

2.6 Feature Classification

In this paper, three of the most commonly used machine learning algorithms have been used for the classification of different stages of Sarcoidosis from chest X-ray images: Support Vector Machine, K-Nearest Neighbours, and Artificial Neural Network. All these classifiers have their own advantages and differ in their working principles, which are briefly explained below along with their basic equations.

2.6.1 Support Vector Machine (SVM)

Support Vector Machines are models of supervised learning, mainly used in solving problems of classification and regression analysis [33]. SVM tends to find the best hyperplane that will separate the classes within the feature space. This ensures the maximum distance or margin between the points of each closest class, known as support vectors. One of the key reasons for using SVMs is that, due to their very special properties, they work just fine in high-dimensional spaces and are very robust to overfitting, especially when the number of dimensions becomes larger than the number of samples. The decision function for SVMs is defined by:

$$f(x) = \boldsymbol{w} \cdot \boldsymbol{x} + b \quad (8)$$

Where:

- $\mathbf{w}$ is the weight vector.
- $\mathbf{x}$ is the input feature vector.
- b is the bias term.

The objective is to minimize:

$$\frac{1}{2} \| \boldsymbol{w} \|^2 \quad (9)$$

Subject to the constraints:

$$y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 \quad (10)$$

For the non-linear case, SVM uses a kernel function $K(\mathbf{x}_i, \mathbf{x}_j)$ to map the input features into a higher-dimensional space.

2.6.2 K-Nearest Neighbours (KNN)

Another simple instance-based learning algorithm working for both classification and regression is K-Nearest Neighbours (KNN) [34]. In KNN, an object will be classified by the majority vote from its k nearest neighbours in the feature space; it is typically measured using Euclidean distance. KNN is a non-parametric method that is easy to implement but computationally extensive for large datasets. The distance between two points $\mathbf{x}_i$ and $\mathbf{x}_j$ is calculated as:

$$d(\boldsymbol{x}_i, \boldsymbol{x}_j) = \sqrt{\sum_{m=1}^{M} (x_{im} - x_{jm})^2} \quad (11)$$

The classification decision is based on:

$$\hat{y} = \text{mode}\ (y_i) \quad (12)$$

For the K nearest neighbors y_i.

2.6.3 Artificial Neural Network (ANN)

Artificial Neural Networks (ANN) are computational models based on the human brain. They have connected neurons, which are organized in layers. Through forward and backpropagation, ANNs can learn complicated patterns. Each neuron sums up its weighted inputs, applies an activation function to it, and propagates the output into the next layer. ANNs were later proved to be so flexible as to approximate any non-linear functions; hence, they found application in a wide range of fields. The output of a neuron j in a layer l is given by [35]:

$$a_j^{(l)} = \phi\left(\sum_i w_{ij}^{(l)} a_i^{(l-1)} + b_j^{(l)}\right) \quad (13)$$

Where:

- ϕ is the activation function (e.g., sigmoid, ReLU).
- $w_{ij}^{(l)}$ is the weight connecting neuron i in layer $l-1$ to neuron j in layer l.
- $b_j^{(l)}$ is the bias term.
- $a_i^{(l-1)}$ is the output of neuron i in layer $l-1$.

The backpropagation algorithm updates the weights to minimize the loss function:

$$\mathcal{L} = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \quad (14)$$

Where:

- y_i is the true label.
- $\hat{y}_i$ is the predicted output.

In so doing, this study can make good use of the strengths of SVM, KNN, and ANN methods to give an accurate and reliable method for the classification of chest X-ray images of patients according to the stages of Sarcoidosis. Even though all three algorithms are different in their viewpoints on learning and classification, making them very useful for improving the robustness of the diagnostic system, each algorithm is special in its own way. Table 2 summarizes the details of the setting for hyper-parameters for the classifier algorithms used.

Table 2: The Hyperparameters of classifiers.

Classifier Type	Hyper-Parameter	Number of Records
KNN	Distance metrics	Minkowski
	Weights	Uniform
	Number of Neighbors	k = 5
SVM	Kernel Type	Polynomial
	Cache Size	200
	Degree of Polynomial	200
ANN	Activation function for hidden layers	Rectified
	Activation function for the output layer	Softmax
	Number of epochs	1000
	Optimization Method	Stochastic
	Learning Rate	0.0001

2.7 Model Performance Evaluation

For estimating the classifiers of diagnosis of different stages of Sarcoidosis in chest X-ray images, four important metrics will be used: accuracy, precision, recall, and F1 score. All these metrics represent a holistic assessment of the classifier, thereby balancing its evaluation across various parameters of prediction quality.

Accuracy refers to the ratio of the number of correctly predicted instances to that of total instances present in the dataset. This is, thus, a very simple way to denote the overall effectiveness of the classifier.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (15)$$

Accuracy refers to the ratio of the number of correctly predicted instances to that of total instances present in the dataset. This is, thus, a very simple way to denote the overall effectiveness of the classifier.

Where:

- True Positive (Tp): Correctly predicted positive instances.
- True Negative (TN): Correctly predicted negative instances.
- False Positive (FP): Incorrectly predicted positive instances.
- False Negative (FN): Incorrectly predicted negative instances.

Accuracy gives the general measure of how often classes are correctly predicted by the classifier. However, it can be misleading if classes are greatly imbalanced since it won't let one know the frequencies of false positives and negatives.

Precision is also known for the Positive Predictive Value. It is a measure of correctly predicted positive instances against all those instances which are predicted as positive. It expresses classifier's capability of reducing false positives.

$$Precision = \frac{TP}{TP+FP} \times 100\% \quad (16)$$

Precision is crucial in contexts where the cost of false positives is high. For instance, in medical diagnosis, a high precision ensures that the condition is not falsely identified, which can prevent unnecessary treatments or interventions.

Recall is the ratio of the number of correctly predicted positive instances to that of the total actual class cases. It explains how often the classifier will detect all relevant instances.

$$Recall = \frac{TP}{TP+FN} \times 100\% \quad (17)$$

Recall is important in cases where the cost of a false negative is high. High recall in medical diagnostics, most of the actual positive cases are identified, thus their early detection and treatment are possible.

The F1 score just follows the harmonic mean of precision and recall. It gives just one number, balanced with respect to both concerns, and is therefore particularly useful when dealing with class-imbalanced datasets.

$$F1 \text{ Score}_A = 2 \cdot \frac{\text{precision}_A * \text{recall}_A}{\text{precision}_A + \text{recall}_A} \quad (18)$$

The F1 score is a sturdy metric, since it merges both precision and recall benefits. When one class is too underrepresented compared to another, this works well and provides a much better sense of the whole picture than anyone would.

These metrics of evaluation include Accuracy, Precision, Recall, and F1 Score. Such kinds of metrics provide detail-oriented insight into the performance of a classifier. While accuracy conveys the general idea, Precision and Recall emphasize the effective management of false positives and negatives by the classifier. The F1 Score encapsulates both Precision and Recall so that the evaluation is balanced. This sets up a multiangle view on how well classifiers work in diagnosing different stages of Sarcoidosis using chest X-ray images, hence providing more accurate and reliable medical diagnostics.

3 Results and Discussion

Table 3 shows the various image segmentation methods against their average Jaccard Index values. The Jac-card Index, also known as Intersection over Union, is a measure of similarity and diversity of sample sets. It has been applied many times in image segmentation work as a metric to compare an image segmented by underneath technique against the ground truth to provide a sense of accuracy of the segmentation. The table has three columns that represent the different approaches of segmentation: Otsu's, k-means, and without segmentation. In this case, the average JI values will then be 0.83, 0.70, and 0.44, respectively.

Table 3: Comparison of average JI of image from different segmentation methods.

Avg JI using Otsu's Method	Avg. JI using K-means	Avg. JI without Segmentation
0.83	0.70	0.44

The analysis of Table 3 highlights significant differences in the performance of the segmentation methods. Otsu's method achieves the highest average JI of 0.83, indicating superior accuracy in segmenting images. This method operates by finding an optimal threshold to separate the foreground and background, thereby enhancing the contrast and improving segmentation accuracy. K-means clustering, with an average JI of 0.70, also shows good performance but is less effective than Otsu's method. K-means clustering partitions the image into clusters based on pixel intensity, which helps in achieving a more accurate segmentation compared to not applying any segmentation.

In contrast, the average JI without any segmentation is significantly lower at 0.44, indicating poor accuracy in the absence of segmentation. This result underscores the importance of segmentation techniques in enhancing the quality and reliability of image analysis. The lack of segmentation results in less precise delineation of the regions of interest, which significantly hampers the ability to differentiate between different stages or features within the images.

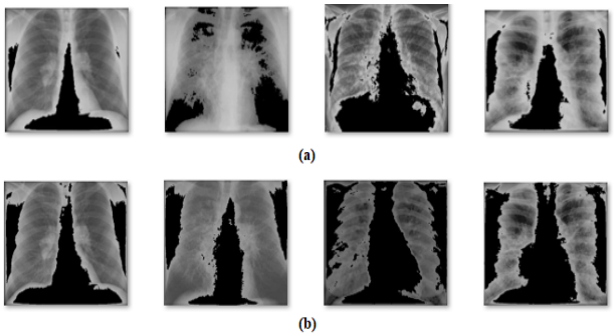


Fig. 3: The results of segmentation methods: a) K-means, b) Otsu's Method.

Fig. illustrates the outcomes of two segmentation methods applied to chest X-ray images: (a) K-means and (b) Otsu's method. The visual results clearly demonstrate that Otsu's method achieves superior segmentation compared to K-means. In the images segmented using K-means (Fig. a), there is noticeable noise and less distinct separation between the foreground and background regions. In contrast, the images segmented using Otsu's method (Fig. b) exhibit a cleaner and more precise delineation of the lung regions, with reduced noise and clearer boundaries. This enhanced clarity and accuracy in Otsu's method supports its higher average Jaccard Index (0.83) as compared to K-means (0.70), confirming that Otsu's method provides better segmentation performance, which is crucial for reliable and precise medical image

analysis.

Fig. displays the statistical measures derived from the Gray-Level Co-occurrence Matrix (GLCM) without applying any segmentation to the chest X-ray images. The extracted features include Energy, Correlation, Dissimilarity, Homogeneity, and Contrast across the four stages of Sarcoidosis. Fig. presents the GLCM-based statistical measures after applying Otsu's segmentation, and Fig. illustrates the GLCM features after applying K-means segmentation, both showing the same features across the four stages.

In so doing, this study can make good use of the strengths of SVM, KNN, and ANN methods to give an accurate and reliable method for the classification of chest X-ray images of patients according to the stages of Sarcoidosis. Even though all three algorithms are different in their viewpoints on learning and classification, making them very useful for improving the robustness of the diagnostic system, each algorithm is special in its own way. Table 2 summarizes the details of the setting for hyper-parameters for the classifier algorithms used.

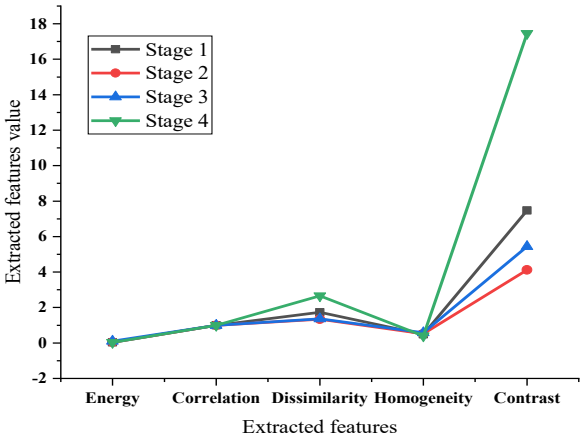


Fig. 4: GLCM properties of images without applying any Segmentation.

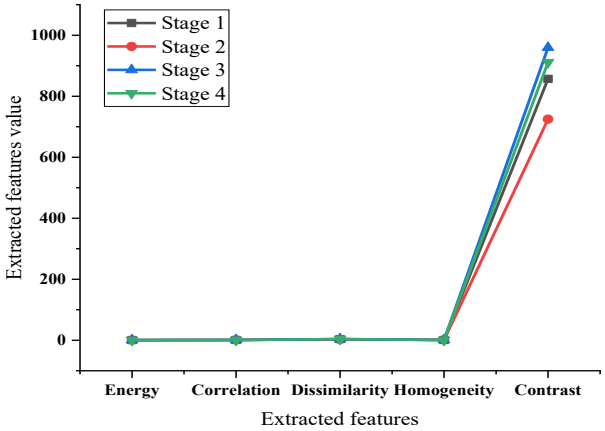


Fig. 5: GLCM properties of images after applying Otsu segmentation.

The comparison of segmentation methods reveals significant differences in the effectiveness of feature extraction. Without any segmentation, as shown in Fig. , the extracted features exhibit a higher degree of variability and overlap across the four stages of Sarcoidosis. This lack of clear distinction makes it challenging to differentiate between the stages based on these features. In contrast, Otsu's segmentation, as demonstrated in Fig. , results in more consistent features with clearer separations among the stages. This indicates that Otsu's method effectively enhances the contrast and delineation of the regions of interest, leading to more reliable feature extraction.

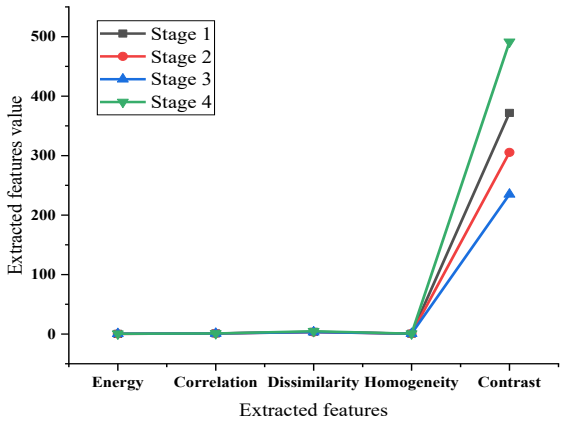


Fig. 6: GLCM properties of images after applying k-means segmentation.

After applying K-means segmentation, as shown in Fig. , the features are more distinct than those without segmentation but less consistent compared to Otsu's method. There is still some overlap and variability, particularly in the Dissimilarity and Contrast features, which may hinder the differentiation between stages. These results demonstrate that segmentation significantly improves the ability to extract meaningful features for differentiating between the stages of Sarcoidosis. Otsu's segmentation provides the best performance, yielding features with clear separations and reduced overlap, outperforming K-means segmentation. The lack of segmentation results in the least effective feature extraction, highlighting the necessity of applying a robust segmentation technique.

Among the extracted features, Contrast and Dissimilarity are particularly notable for their ability to differentiate between the stages when segmentation is applied. In Fig. and Fig. , these features show significant variations across the stages, with Otsu's method providing the most distinct separations. This suggests that Contrast and Dissimilarity, especially when enhanced by Otsu's segmentation, are the most effective features for accurately classifying the stages of Sarcoidosis.

Our analysis revealed that segmentation significantly impacts the performance of the classifiers. Specifically, Otsu's segmentation method provided the best results in terms of producing distinct and reliable GLCM features. Consequently, we fed the classifiers with features extracted

using Otsu's segmentation because it offered the best performance. By leveraging the strengths of these classifiers and the optimized feature extraction method, we aim to develop a reliable system for accurately diagnosing and differentiating the stages of Sarcoidosis from chest X-ray images.

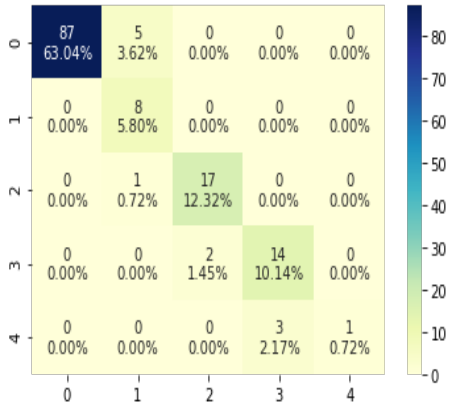


Fig. 7: Confusion Matrix for the KNN model.

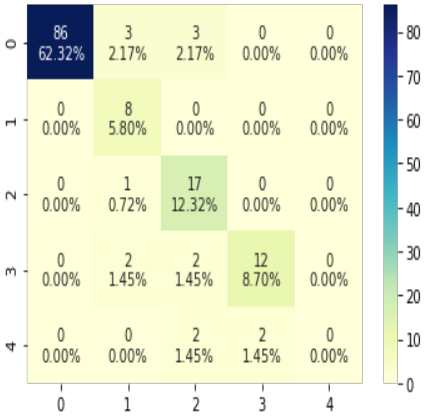


Fig. 8: Confusion Matrix for the SVM model.

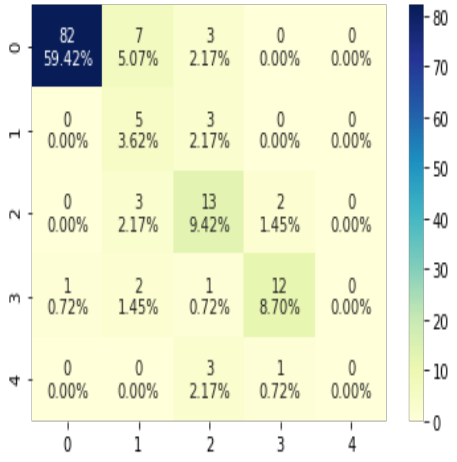


Fig. 9: Confusion Matrix for the ANN model.

The KNN classifier (Fig.) shows a moderate ability to correctly identify true positives across the four stages of Sarcoidosis, but there are notable misclassifications. The confusion matrix reveals a relatively high rate of false positives, where instances from other states are incorrectly classified as the target stage. Additionally, KNN exhibits a significant number of false negatives, indicating that it misses several true cases, especially when distinguishing between adjacent stages. Although KNN correctly identifies many true negatives, its overall performance is hindered by these misclassifications, leading to moderate accuracy and precision.

The SVM classifier (Fig.) demonstrates a higher number of true positives across all stages, indicating improved classification accuracy compared to KNN. The rate of false positives is relatively lower, which enhances the precision of the SVM model. Additionally, SVM exhibits fewer false negatives, meaning it successfully identifies most true cases, which contributes to higher recall. The model's ability to correctly identify true negatives further solidifies its reliability. Overall, SVM outperforms KNN by providing better accuracy, precision, and recall, resulting in a more effective differentiation between the stages of Sarcoidosis.

The ANN classifier (Fig.) outperforms both KNN and SVM, achieving the highest number of true positives across all stages. This indicates superior classification capability.

The confusion matrix shows that ANN has the lowest rate of false positives, reflecting excellent precision. Additionally, ANN records the fewest false negatives, meaning it accurately captures nearly all true cases, which significantly boosts recall. The model's exceptional ability to identify true negatives further confirms its reliability. Among the three classifiers, ANN consistently demonstrates the best performance. This makes ANN the most effective classifier for distinguishing between the stages of Sarcoidosis.

Table 4: The accuracy metrics for the KNN model.

Class	Accuracy	Precision	Recall	F1
Normal	96.38%	1.0	0.95	0.97
Stage 1	95.65%	0.57	1.0	0.73
Stage 2	97.83%	0.89	0.94	0.92
Stage 3	96.38%	0.82	0.88	0.85
Stage 4	96.83%	1.0	0.25	0.40

Table 5: The accuracy metrics for the SVM model.

Class	Accuracy	Precision	Recall	F1
Normal	96.38%	1.0	0.95	0.97
Stage 1	95.65%	0.57	1.0	0.73
Stage 2	97.83%	0.89	0.94	0.92
Stage 3	96.38%	0.82	0.88	0.85
Stage 4	96.83%	1.0	0.25	0.40

Table 6: The accuracy metrics for the ANN model.

Class	Accuracy	Precision	Recall	F1
Normal	92.03%	0.99	0.89	0.94
Stage 1	89.13%	0.29	0.63	0.40
Stage 2	89.13%	0.57	0.72	0.63
Stage 3	94.93%	0.80	0.75	0.77
Stage 4	97.1%	0.0	0.0	0.0

Table 4, Table 5, and Table 6 present the accuracy metrics

for the K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Artificial Neural Network (ANN) models, respectively. The metrics include Accuracy, Precision, Recall, and F1 Score for each class (Normal, Stage 1, Stage 2, Stage 3, and Stage 4) of Sarcoidosis.

In Table 4 The KNN model exhibits high accuracy across all classes, with Stage 2 achieving the highest accuracy at 97.83%. Precision is perfect (1.0) for Normal and Stage 4 classes, but significantly lower than Stage 1 (0.57). Recall is highest for Normal, Stage 1, and Stage 2, indicating the model's ability to correctly identify true positives in these classes. However, the recall for Stage 4 is very low (0.25), suggesting difficulty in identifying this stage accurately. The F1 Score is high for Normal and Stage 2 but considerably lower for Stage 4 (0.40), reflecting the imbalance between precision and recall for this class.

As shown in Table 5, The SVM model also shows high accuracy, with Stage 4 achieving the highest at 97.1%. Precision is perfect (1.0) for the Normal class but drops to 0.0 for Stage 4, indicating the model's inability to identify true positives for this class. Recall is perfect (1.0) for Stage 1, similar to the KNN model, but drops to 0.0 for Stage 4, showing a complete failure to identify true positives. The F1 Score for Normal is high (0.97), while Stage 4 has an F1 Score of 0.0, highlighting the significant challenges in classifying Stage 4 sarcoidosis.

As shown in Table 6, The ANN model shows varied performance across classes, with Stage 4 achieving the highest accuracy at 97.1%. Precision for the Normal class is high (0.99) but drops significantly for Stage 1 (0.29). Recall is lower across all stages compared to the other models, with Stage 1 and Stage 2 showing particularly low values (0.63 and 0.72, respectively). The F1 Score for Normal is high (0.94), but similar to SVM, Stage 4 has an F1 Score of 0.0, indicating poor performance in identifying true positives for this class.

Comparing the three models, KNN and SVM show better overall accuracy and recall for most stages compared to ANN. However, all models struggle with Stage 4 classification, with SVM and ANN showing a complete failure (0.0 precision and recall). KNN performs relatively better in Stage 4 but still shows significant challenges, as indicated by the low F1 Score. Precision for Stage 1 is consistently low across all models, suggesting difficulties in distinguishing this stage from others.

4 Conclusion

Sarcoidosis is a multi-organ, multi-tissue, inflammatory disease with frequent lung involvement. The greatest problem faced in the clinical diagnosing and staging of Sarcoidosis is by the difficulty of differentiating the different radiological patterns which mostly occur in the images from chest X-ray analysis. There exist several machines learning algorithms for diagnosing medical images. The most popular algorithms, for their robustness and effectiveness, include Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Artificial Neural Network (ANN).

The Otsu segmentation method is a clear winner with its advantage in extracting features that are really important for the exact differentiation of stages of Sarcoidosis. Receiver operation characteristic analysis is implemented to compare the effectiveness and diagnostic power of the whole method, which will ensure more reliable and precise medical image analysis toward improvement of diagnostic accuracy and better patient outcomes. The performance analysis for KNN and SVM indicates that at most of the stages of Sarcoidosis, they tend to perform pretty well. Still, it is really problematic to classify Stage 4 accurately. The ANN shows very high accuracy for the Normal and Stage 4 classes but gives low values of both precision and recall for the other stages, reflecting the limitations of the ANN. Moreover, this finding explained that, for the classifiers to identify all the sarcoidosis stages accurately, segmentation and feature extraction needed improvement.

Funding Statement

This work was funded in part by [INTI-IU-Malaysia] grant number [INTI-FEQS-01-06-20023].

Author Contributions

The conceptualization of this research was solely carried out by Mohanad A. Deif. Mohanad A. Deif, Mohamed A. Hafez, and Hani Attar determined the methodology. Mohanad A. Deif was responsible for the development of the software and analysis. Validation of the results was carried out by Mohanad A. Deif, Mohamed A. Hafez, and Waleed Alomoush. Hussein Al-Faiz and Hani Attar conducted the formal analysis. Mohamed A. Hafez secured funding for the project. The initial draft of the manuscript was written by Mohanad A. Deif, with all authors contributing to the review, editing, and final revisions. Hussein Al-Faiz also contributed to data visualization. Mohanad A. Deif provided overall supervision for the project, while Mohamed A. Hafez handled project administration tasks.

Availability of Data and Materials

The data that support the findings of this study are available on request from the corresponding author.

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] Bokhari, SRA, Zulfiqar, H and Mansur, A., 2023. *Sarcoidosis.*. Stat Pearls Publishing,.

[2] Judson, M. A., Thompson, B. W., Rabin, D. L., Steimel, J., Knattereud, G. L., Lackland, D. T., Rose,

C., Rand, C. S., Baughman, R. P., Teirstein, A. S. and others (2003). *The diagnostic pathway to Sarcoidosis*, Chest 123 : 406-412.

[3] Belperio, J. A., Shaikh, F., Abtin, F. G., Fishbein, M. C., Weigt, S. S., Saggar, R. and Lynch, J. P. (2022). *Diagnosis and treatment of pulmonary Sarcoidosis: a review*, Jama 327 : 856-867.

[4] Attar, H., Solyman, A., Deif, M. A., Hafez, M., Kasem, H. M., & Mohamed, A. E. F. (2023, December). *Machine Learning Model Based on Gary-Level Co-occurrence Matrix for Chest Sarcoidosis Diagnosis*. In 2023 2nd International Engineering Conference on Electrical, Energy, and Artificial Intelligence (EICEEAI) (pp. 1-8). IEEE.

[5] Arcana, R. I., Crișan-Dabija, R., Cernomaz, A. T., Buculei, I., Burlacu, A., Zabară, M. L. and Trofor, A. C. (2023). *The risk of sarcoidosis misdiagnosis and the harmful effect of corticosteroids when the disease picture is incomplete*, Biomedicines 11 : 175.

[6] Lynch III, J. P., Kazerooni, E. A. and Gay, S. E. (1997). *Pulmonary Sarcoidosis*, Clinics in chest medicine 18 : 755-785.

[7] Deif, M. A., Attar, H., Hafez, M. A., Alrosan, A., & Sharfo, S. M. (2024). *Hybrids of Support Vector Regression with Bald Eagle Search Optimizer for Diagnosing Patients Suspected of Sarcoidosis*. International Journal of Intelligent Engineering & Systems, 17(6).

[8] Arkema, E. V. and Cozier, Y. C. (2018). *Epidemiology of Sarcoidosis: current findings and future directions*, Therapeutic Advances in Chronic Disease 9 : 227-240.

[9] Jain, R., Yadav, D., Puranik, N., Guleria, R. and Jin, J.-O. (2020). *Sarcoidosis: causes, diagnosis, clinical features, and treatments*, Journal of Clinical Medicine 9 : 1081.

[10] Bonifazi, M., Bravi, F., Gasparini, S., La Vecchia, C., Gabrielli, A., Wells, A. U. and Renzoni, E. A. (2015). *Sarcoidosis and cancer risk: systematic review and meta-analysis of observational studies*, Chest 147 : 778-791.

[11] van der Sar, I. G., van Jaarsveld, N., Spiekerman, I. A., Toxopeus, F. J., Langens, Q. L., Wijsenbeek, M. S., Dauwels, J. and Moor, C. C. (2023). *Evaluation of different classification methods using electronic nose data to diagnose Sarcoidosis*, Journal of Breath Research 17 : 047104.

[12] Lovin fosse, P., Ferreira, M., Withofs, N., Jadoul, A., Derwael, C., Frix, A.-N., Guiot, J., Bernard, C., Diep, A. N., Donneau, A.-F. and others (2022). *Distinction of Lymphoma from Sarcoidosis on 18F-FDG PET/CT: evaluation of radiomics-feature--guided machine learning versus human reader performance*, Journal of Nuclear Medicine 63 : 1933-1940.

[13] Baghdadi, N., Maklad, A. S., Malki, A., & Deif, M. A. (2022). *Reliable sarcoidosis detection using chest X-rays with efficientnets and stain-normalization techniques*. Sensors, 22(10), 3846.

[14] de Lima, A. D., Lopes, A. J., do Amaral, J. L. M. and de Melo, P. L. (2022). *Explainable machine learning methods and respiratory oscillometry for the diagnosis of respiratory abnormalities in Sarcoidosis*, BMC Medical Informatics and Decision Making 22 : 274.

[15] Bobbio, E., Eldhagen, P., Polte, C. L., Hjalmarsson, C., Karason, K., Rawshani, A., Darlington, P., Kullberg, S., Sörensson, P., Bergh, N. and others (2023). *Clinical Outcomes and Predictors of Long-Term Survival in Patients With and Without Previously Known Extracardiac Sarcoidosis Using Machine Learning: A Swedish Multicenter Study*, Journal of the American Heart Association 12 : e029481.

[16] Urinbayev, K., Orazbek, Y., Nurambek, Y., Mirzakhmetov, A. and Varol, H. A. (2020). *End-to-end deep diagnosis of x-ray images*, : 2182-2185

[17] Lu, C., Wang, Y. G., Zaman, F., Wu, X., Adhaduk, M., Chang, A., Ji, J., Wei, T., Suksaranjit, P., Christodoulidis, G. and others (2022). *Predicting adverse cardiac events in Sarcoidosis: deep learning from automated characterization of regional myocardial remodeling*, The international journal of cardiovascular imaging 38 : 1825-1836.

[18] Koth, L. L., Solberg, O. D., Peng, J. C., Bhakta, N. R., Nguyen, C. P. and Woodruff, P. G. (2011). *Sarcoidosis blood transcriptome reflects lung inflammation and overlaps with tuberculosis*, American journal of respiratory and critical care medicine 184 : 1153-1163.

[19] Soni, J., Ansari, U., Sharma, D., Soni, S. and others (2011). *Predictive data mining for medical diagnosis: An overview of heart disease prediction*, International Journal of Computer Applications 17 : 43-48.

[20] Attar, H., Ahmed, T., Rabie, R., Amer, A., Khosravi, M. R., Solyman, A., & Deif, M. A. (2024). *Modeling and computational fluid dynamics simulation of blood flow behavior based on MRI and CT for Atherosclerosis in Carotid Artery*. *Multimedia Tools and Applications*, *83*(19), 56369-56390.

[21] Stirenko, S., Kochura, Y., Alienin, O., Rokovyi, O., Gordienko, Y., Gang, P. and Zeng, W. (2018). *Chest X-ray analysis of tuberculosis by deep learning with segmentation and augmentation*, : 422-428.

[22] Tariq, H., Muqeet, A., Burney, A., Hamid, M. A. and Azam, H. (2017). *Otsu's segmentation: review, visualization, and analysis in context of axial brain*

MR slices, Journal of Theoretical and Applied Information Technology 95 : 6042-6055.

[23] Liu, D. and Yu, J. (2009). *Otsu method and K-means*, 1 : 344-349.

[24] Shaaban, H. R., Obaid, F. A. and Habib, A. A. (2015). *Performance evaluation of K-mean and fuzzy C-mean image segmentation based clustering classifier*, Perform. Eval 6 : 1-12.

[25] Akhtar, N., Agarwal, N. and Burjwal, A. (2014). *K-mean algorithm for image segmentation using neutrosophy*, : 2417-2421.

[26] Saygılı, A. (2021). *Analysis and Segmentation of X-ray Images of COVID-19 Patients using the k-means Algorithm*, Veri Bilimi 4 : 1-6.

[27] Bali, A. and Singh, S. N. (2015). *A review on the strategies and techniques of image segmentation*, : 113-120.

[28] Dhruv, B., Mittal, N. and Modi, M. (2019). *Study of Haralick's and GLCM texture analysis on 3D medical images*, international journal of Neuroscience 129 : 350-362.

[29] Fernandez-Grandon, C., Soto, I., Zabala-Blanco, D., Alavia, W. and Garcia, V. (2021). *SVM and ANN classification using GLCM and HOG features for COVID-19 and Pneumonia detection from Chest X-rays*, : 01-06.

[30] Jain, S. (2013). *Brain cancer classification using GLCM based feature extraction in artificial neural network*, International Journal of Computer Science & Engineering Technology 4 : 966-970.

[31] Kim, D.-H. and Ye, S.-Y. (2021). *Classification of chronic kidney disease in sonography using the GLCM and artificial neural network*, Diagnostics 11 : 864.

[32] Alqerem, A., Attar, H., Alomoush, W., & Deif, M. (2022, November). *The Ability of Ultra Wideband to Differentiate Between Hematoma and Tumor Occur in The Brain*. In *2022 International Engineering Conference on Electrical, Energy, and Artificial Intelligence (EICEEAI)* (pp. 1-7). IEEE.

[33] Alrosan, A., Abdel-Aty, M., Hafez, M., Alkhazaleh, S., Deif, M. A., & ELGohary, R. (2024, July). *Parkinson's Disease Detection Based on Vocal Biomarkers and Machine Learning Approach*. In *2024 International Telecommunications Conference (ITC-Egypt)* (pp. 475-480). IEEE.

[34] Xing, W. and Bei, Y. (2019). *Medical health big data classification based on KNN classification algorithm*, Ieee Access 8 : 28808-28819.

[35] Frizzo Stefenon, S., Waldrigues Branco, N., Nied, A., Wildgrube Bertol, D., Cristian Finardi, E., Sartori, A., Henrique Meyer, L. and Bartnik Grebogi, R. (2020). *Analysis of training techniques of ANN for classification of insulators in electrical power systems*, IET Generation, Transmission & Distribution 14 : 1591-1597.