Machine learning model based on Gary-level co-occurrence matrix for chest Sarcoidosis diagnosis

Attar, Hani; Solyman, Ahmed; Deif, Mohanad A.; Hafez, Mohamed; Kasem, Hager M.; Mohamed, Abd-Elnaser Fawzy

Published in:
2023 2nd International Engineering Conference on Electrical, Energy, and Artificial Intelligence (EICEEAI)

DOI:
10.1109/EICEEAI60672.2023.10590168

Publication date:
2024

Document Version
Author accepted manuscript

Machine Learning Model Based on Gary-Level Co-occurrence Matrix for Chest Sarcoidosis Diagnosis

Hani Attar
Faculty of Engineering,
Zarqa University, Zarqa, Jordan
College of Engineering
University of Business and Technology
Jeddah, Saudi Arabia
Hattar@zu.edu.jo

Ahmed Solyman
School of Computing, Engineering and Built Environment
Glasgow Caledonian University
Glasgow, UK
ahmed.solyman@gcu.ac.uk

Mohanad A. Deif
Department of Artificial Intelligence, College of Information Technology
Misr University for Science & Technology (MUST)
October 6 City 12566, Egypt
Mohanad.Deif@must.edu.eg

Mohamed Hafez
Faculty of Engineering FEQS
INTI-IU-University
Nilai, Malaysia
mohdahmed.hafez@newinti.edu.my

Hager M. Kasem
Department of Bioelectronics
Modern University of Technology and Information (MTI) University
Cairo, Egypt
HAJAR.89898@eng.mti.edu.eg

Abd-Elnaser Fawzy Mohamed
Communication Department
Bilbies Higher Institute for Engineering
Sharqia, Egypt
dr_naser62@yahoo.com

***Abstract*— Sarcoidosis is often misdiagnosed and mistreated due to the limitations of radiological presentations. With the recent emergence of COVID-19, doctors face challenges distinguishing between the symptoms of these two diseases. As a result, people are adapting to new practices such as working from home, wearing masks, and using disinfectants. The similarity in symptoms between sarcoidosis and COVID-19 has made it difficult to differentiate between the two conditions, potentially impacting patient outcomes. The diagnostic process for distinguishing between them is time-consuming, labor-intensive, and costly. Researchers and medical practitioners have gained significant attention to computer-aided detection (CAD) systems for sarcoidosis using radiological images to address this issue. This study uses machine learning classifiers, ensembles, and features such as Gray-Level Co-occurrence Matrix (GLCM) and histogram analysis to identify lung sarcoidosis infection from chest X-ray images. The proposed method extracts statistical texture features from X-ray images by calculating a GLCM for each image using various stride combinations. These GLCM features are then used to train the machine learning classifiers and ensembles. The research focuses on multi-class classification, categorizing X-ray images into three classes: sarcoidosis-affected, COVID-19-affected, and regular lungs, as well as binary classification, distinguishing sarcoid-affected cases from others. The proposed method, known for its simplicity and computational efficiency, demonstrates significant accuracy in identifying sarcoidosis and COVID-19 from chest X-ray images.**

Keywords—component, formatting, style, styling, insert (keywords)

I. INTRODUCTION

Sarcoidosis is a disease that causes the body to make small lumps of inflammatory cells, called granulomas, in the lungs, skin, or lymph nodes. It is a systemic disease that can affect any part of the body. The exact cause of sarcoidosis remains unknown, but experts speculate it stems from an immune system response to an unidentified substance. Research suggests that factors such as infectious agents, chemicals, dust, and abnormal reactions to the body's proteins (auto-proteins) may contribute to the development of granulomas, especially in individuals with a genetic predisposition [1],[2].

The National Heart, Lung, and Blood Institute conducted a workshop to understand this disease better and enhance outcomes for sarcoidosis patients. This workshop reviewed available data concerning the healthcare burden and outcomes associated with sarcoidosis in the United States. Inequalities in outcomes were identified across various factors, including ethnicity, sex, race, and socioeconomic status. Notably, African Americans were found to experience more severe disease manifestations disproportionately, and they had the highest mortality rates. Interestingly, mortality rates among white individuals may also be on the rise [3].

The impact of sarcoidosis on healthcare is not limited to its physical symptoms but also includes economic, psychosocial, and other medical conditions associated with the disease. Sarcoidosis and COVID-19 are two diseases with significant similarities. COVID-19 first appeared in Wuhan, China, in December 2019 and quickly spread to nearly every country globally [4]. On January 30, 2020, the World Health Organization (WHO) declared the COVID-19 outbreak a public health emergency of international concern (PHEIC). In March 2020, the WHO officially classified the outbreak as a pandemic [5]. COVID-19 has caused a significant number of deaths worldwide.

Medical professionals continue to face formidable challenges in combating this virus. These challenges include its recent emergence, delayed detection, limited testing knowledge, inexperience among healthcare practitioners, and its striking resemblance to well-known pneumonia and the unfamiliar sarcoidosis. The symptoms observed in affected individuals, such as cough, shortness of breath, fatigue, fever, and sweating, closely mirror those seen in persons afflicted with sarcoidosis and COVID-19. This resemblance further complicates differentiation, especially in regions with both prevalent conditions [6], [7].

Numerous serum markers associated with sarcoidosis, including neopterin, soluble IL-2 receptor (sIL-2R), and angiotensin-converting enzyme (ACE), serve as staging parameters for the disease. Although there is a strong association between these markers and the etiology of the disease, the available evidence is inadequate to establish their effectiveness in predicting clinical worsening. The

quantification of sIL-2R, in particular, suffers from low sensitivity, leading to a high rate of false-negative results [8]. Consequently, confidently diagnosing sarcoid infection becomes challenging, affecting the timely treatment and management of infected patients.

Conversely, radiological imaging, specifically chest X-ray (CXR), plays a pivotal role in assessing the clinical development of sarcoidosis and guiding appropriate treatment decisions for infected individuals. Chest X-ray patient's images with sarcoidosis exhibit characteristic patterns that closely resemble those found in X-ray images of individuals infected with COVID-19. Extensive research has been conducted to identify and classify sarcoidosis using chest X-ray images [9]. Although Reverse Transcription Polymerase Chain Reaction (RT-PCR) is widely used to test for COVID-19, it may not be enough to stop the virus from spreading globally. Chest X-rays can be valuable for triaging COVID-19 patients because they can be taken quickly and easily. The virus's rapid spread has pressured medical and health systems worldwide [10][11]. Amidst this immense healthcare challenge, there is an immediate necessity to develop a method capable of effectively distinguishing patients suspected of having sarcoidosis from those who are average or infected with COVID-19, using chest X-ray images. Researchers worldwide are actively exploring texture-based cognitive methods to identify and differentiate between these diseases, given the similarities in their symptoms.

The proposed framework extracts a succinct collection of texture characteristics from the Gray-Level Co-occurrence Matrix (GLCM). This approach aims to effectively capture the texture information inherent in chest X-ray (CXR) pictures [12]. The aforementioned condensed set of textural data is subsequently fed into a discriminative K-Nearest Neighbors (KNN) model, facilitating accurate classification for the identification of sarcoidosis and distinguishing it from COVID-19 in chest X-ray (CXR) images. The findings suggest that utilizing a smaller amount of features in the proposed framework not only simplifies the jobs of the discriminative KNN classifier but also has the potential to improve the classification accuracy and scalability of the technique. Furthermore, reducing the number of extracted features significantly decreases the system's time complexity.

II. RELATED WORK

In light of the prevalence of the sarcoid epidemic, there has been a notable surge in scholars' interest in the domains of artificial intelligence (AI) and medical image analysis. Specifically, there is a rising focus on developing advanced Computer-Aided Detection (CAD) systems that exhibit exceptional efficiency in detecting sarcoidosis by analyzing chest radiography images. Researchers' commitment has led to several studies in artificial intelligence (AI) that employ chest radiography images to aid clinicians in managing sarcoid patients [13].

Chest radiography, a cost-effective imaging modality that provides ample data for training machine learning algorithms, has led to the rising popularity of deep learning techniques in identifying lung conditions based on chest radiographs. Previous research efforts have been directed towards understanding sarcoidosis. For instance, Prof. Mohanad A. A. Deif et al. [2] researched the detection of sarcoidosis in the lungs through chest X-ray images. Young Jae Kim [14] presented research on utilizing a machine-learning model to identify COVID-19 in chest X-rays.

Based on prior research studies, various methods have been proposed to differentiate between individuals with sarcoidosis and those with COVID-19. In Reference [15], researchers devised a patch-based deep learning Convolutional Neural Network (CNN) strategy to diagnose COVID-19 from chest X-ray (CXR) images, utilizing a compact set of trainable parameters. The CXR image dataset employed in this study encompassed samples from healthy subjects and individuals diagnosed with tuberculosis, bacterial pneumonia, and viral pneumonia resulting from COVID-19 infection. The diagnostic model demonstrated excellent performance, achieving an overall accuracy of 88.9% and an F1 score of 84.40%.

Similarly, a recent study by Eriko Koda and colleagues [16][17] showcased the efficacy of employing complementary texture analysis techniques for identifying mediastinal lymphadenopathy caused by Sarcoid on unenhanced CT images. The integration of these methods with the experimental semi-variogram feature resulted in an area under the receiver operating characteristic curve (AUROC) of 0.89. Furthermore, it demonstrated sensitivity, specificity, and accuracy of 75%, 70%, and 90%.

Various research groups [18] have employed conventional machine-learning algorithms to differentiate between regular and Sarcoid patients using CXR images. Likewise, in Reference [19], Fan and colleagues introduced Inf-Net, a deep-learning network designed to segment COVID-19 lung infections and identify areas in chest CT images that exhibit suspicious signs of COVID-19. Their methodology incorporates a parallel partial decoder to generate comprehensive representations of segmented maps. Additionally, they employ implicit reverse attention and explicit edge attention mechanisms to delineate boundaries and enhance representations. This model yielded remarkably favorable segmentation results, achieving a Dice score of 0.74 and an alignment index of 0.89.

In Reference [20], Albert Salazar and his team conducted a study to identify factors that predict the persistence of disease activity at the time of diagnosis. Over 14 years (from 1974 to 1987), 209 patients received a diagnosis of sarcoidosis at Belloite Hospital, a teaching institution in Barcelona, Spain, with a capacity of 1,000 beds. The study included 193 patients who were subsequently monitored. The persistence of disease activity was revealed to be independently influenced by numerous variables using a Cox proportional-hazards regression model. The parameters examined in this study encompassed the lack of erythema nodosum (risk ratio = 2.37), the identification of pulmonary infiltrates on chest X-rays (risk ratio = 1.89), the existence of splenomegaly (risk ratio = 3.67), and the nonexistence of lymphadenopathy in chest X-rays (risk ratio = 2.26).

Moreover, in Reference [21], an innovative deep-learning model known as COVID-19 was introduced to distinguish between patients diagnosed with COVID-19 and individuals without the condition. This model was developed using a dataset comprising 50 chest X-ray images. Seven established deep-learning networks were utilized as feature extractors to achieve efficient feature extraction. A thorough examination of the COVIDXNet model about two alternative deep learning models, specifically VGG-19 and DensNet201, demonstrated

that the COVIDXNet model exhibited superior performance, potentially attaining a diagnostic accuracy rate of up to 90%.

In our study, a machine learning model was adopted that utilized GLCM (Gray-Level Co-occurrence Matrix) features and histogram-based training data to differentiate between cases of sarcoidosis and COVID-19. Although there is no definitive treatment for sarcoidosis, many individuals with the condition lead healthy lives with minimal or no medical intervention. Sarcoidosis can, in some instances, resolve spontaneously, while in others, it may persist over extended periods and lead to organ damage. To ensure high accuracy in our system, image preprocessing tools were employed to enhance the quality of medical images, even when computational resources were limited. GLCM, which captures second-order statistical information about gray levels between adjacent pixels in an image, played a crucial role in this process. Additionally, advanced histogram-based techniques were employed to improve the accuracy of our classification model further [22].

The proposed approach identifies and differentiates suspected sarcoid cases early by classifying X-ray images into three distinct categories: sarcoid, regular, and COVID-19 infected. Furthermore, it also includes a binary classification task that categorizes images into sarcoid or normal. The key contributions of this research and the proposed method encompass:

- Determination of the most effective GLCM (Gray-Level Co-occurrence Matrix) and histogram feature stride combinations for enhancing the accuracy of sarcoid classification.
- A comprehensive performance evaluation of Machine Learning classifiers, particularly K-Nearest Neighbors (KNN), to achieve improved identification of sarcoidosis and COVID-19 cases.

III. MATERIALS AND METHODS

A. MATERIALS

To assess the performance of the proposed CAD (Computer-Aided Detection) system, a diverse dataset was compiled by amalgamating multiple publicly available datasets. 92 chest X-ray (CXR) images was randomly selected, encompassing confirmed cases of sarcoidosis, COVID-19, and normal cases obtained from various sources. This dataset comprised labeled CXR images categorized into three classes: sarcoid class (31 images), COVID-19 class (30 images), and Normal class (31 images). These images varied significantly, ranging from 912 × 456 pixels to 2721× 2438 pixels. In the current study, the chest X-ray images belonging to the Sarcoid, normal, and COVID-19 classes were merged, and subsequently, the dataset was curated for three-class classification, categorizing images based on whether the patient had sarcoidosis, was normal, or had COVID-19. Figure 1 shows the visual challenge of distinguishing sarcoidosis from COVID-19, although comprehensive metadata for all patients in this dataset was not provided [23]. A selection of sample images from the dataset is depicted in Figure 1.

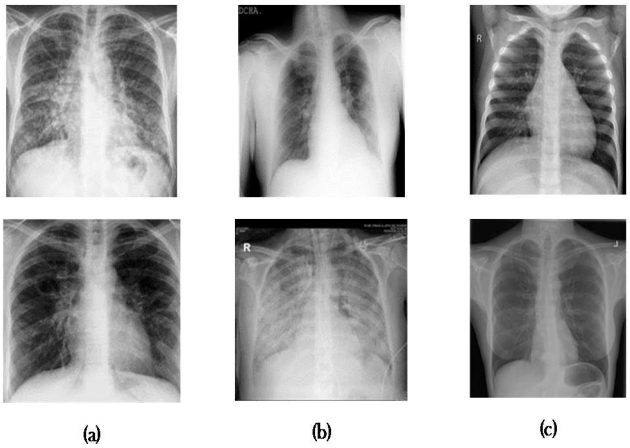

Figure 1. Different classes of chest X-ray images: (a) sarcoidosis-infected, (b) covid-19-infected, (c)normal.

In this section, the focus lies on the identification of sarcoidosis and COVID-19. This is achieved by presenting a method that involves the extraction of texture features from 8-bit JPEG input chest X-ray images [10]. This approach's foundation lies in utilizing the Gray Level Co-occurrence Matrix (GLCM). Specifically, the co-occurrence matrix is generated for each chest X-ray image provided as input. five different combinations of strides were considered, all at the same angle of (45,1), (45, 2), (45, 8), (45,32), and (45,128), for computing the GLCM. The six statistical features, namely contrast, homogeneity, dissimilarity, correlation, angular second moment, and energy, are extracted from the GLCM [24]. The features mentioned above are subsequently integrated into a unified feature vector. It is essential to acknowledge that the dimensions of the co-occurrence matrix are contingent upon the maximum count of unique or distinct intensity values exhibited by pixels in a picture. An increased quantity of unique intensity values leads to enhanced accuracy in extracting textural information.

Nevertheless, a higher quantity of unique gray levels results in a larger co-occurrence matrix, increasing the computational cost and time required. Furthermore, histogram-based features was integrated into the methodology, obtaining five distinct features from this approach. A total of 11 characteristics are collected by employing a mix of GLCM (Gray-Level Co-occurrence Matrix) and histogram analysis. These characteristics and different combinations of stride are utilized to train machine learning classifiers, specifically emphasizing K-Nearest Neighbors (KNN), to identify sarcoidosis from chest X-ray pictures.

➢ METHODS

A. Overview

Within this part, the proposed methodology for the automated identification of sarcoidosis is presented. Figure 2 displays the comprehensive conceptual block diagram that outlines the fundamental stages of our Computer-Aided Detection (CAD) system. This method categorizes X-ray pictures into Sarcoid, COVID-19, and normal classifications.

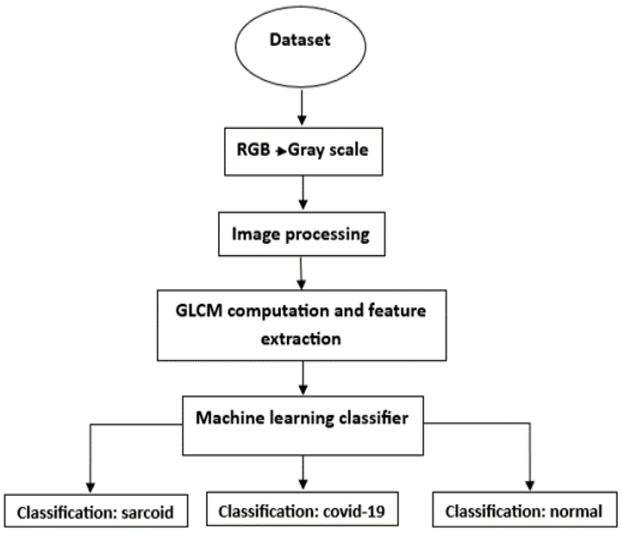

Figure 2. Functional block diagram of classifying X-ray images using a machine learning strategy.

The overall structure of the proposed machine learning framework functions as follows: in the initial step, the chest X-ray is segmented into regions suspected of being infected with the disease. A refined collection of second-order statistical texture characteristics is derived from the gray level co-occurrence matrix (GLCM) of sarcoidosis, COVID-19, and normal healthy chest X-ray (CXR) image samples. The GLCM-based texture features that have been collected are transformed into a one-dimensional vector representation. This vector representation is subsequently utilized as input for the K-nearest neighbors (KNN) model, which categorizes the data into one of three categories: sarcoidosis, COVID-19, or normal. In the subsequent sections, the intricate procedural stages of the proposed technique will be expounded upon.

B. Image Preprocessing

The initial stage of image preprocessing is primarily tasked with identifying and removing artifacts present in the image [25]. Including this phase is crucial in the context of CXR images, as many exhibit many manifestations of noise and undesired artifacts, such as patient clothing and wires, which necessitate their elimination to ensure precise sarcoidosis diagnosis. The subsequent section delineates the procedures executed during the preprocessing stage, which encompasses three primary operations: (i) scaling the image, (ii) mitigating noise, and (iii) augmenting image contrast using pixel correction and histogram equalization. The main goal of lung segmentation is to delineate and separate the regions corresponding to the lungs in chest X-ray (CXR) images [26]. Automated lung segmentation typically serves as a preprocessing step in lung CXR image analysis and holds significant importance in diagnosing lung diseases.

- Image resizing: The X-ray image datasets used in this study were sourced from multiple origins, resulting in images of varying sizes. To accommodate the GLCM program's specific image size requirement, each image was resized to the necessary dimensions while preserving its essential features [27].
- Noise removal: This step focuses on eliminating noise from the images, enhancing them for more accurate feature extraction. After inputting the images into the GLCM program and saving all the extracted features, a "diagnose" column was added to the data, allowing for the removal of unnecessary data dimensions to ensure the quality of data for machine learning.

C. features extraction

Feature extraction is pivotal in pattern recognition and image classification applications 28]. The main aim of this task is to extract distinctive characteristics from image data, which are then utilized as input vectors for machine learning models. This process allows machine learning models to automatically provide visual descriptions, interpretations, or understandings of the image contents. Within the framework of the proposed computer-aided diagnosis (CAD) system for Sarcoid, a specific method for feature extraction based on the Gray-Level Co-occurrence Matrix (GLCM) is utilized. This method aims to acquire a succinct collection of discerning textural properties from chest X-ray pictures, including contrast, energy, homogeneity, and several others.

Python is the programming language to extract features from the input chest X-ray images[25]. The Python Scikit-image package provides a range of valuable functions for extracting diverse characteristics from an image's Gray-Level Co-occurrence Matrix (GLCM). This study employs a second-order methodology to extract second-order statistical texture features from an input chest X-ray image[30]. The strategy considers the connectedness of clustered pixels within the image. To enhance the efficiency of GLCM computation, the gray levels of pixels in each preprocessed CXR image are quantized to a lower number of gray levels. The dimensions of the Co-occurrence matrix are determined by the most significant count of unique or distinct intensity values found within the image. The accuracy of the extracted textural information increases with a more significant number of distinct intensity values. However, to manage computational costs, the distance for GLCM computation is limited to no more than 128, as computing GLCM for all 256 gray levels would be highly computationally intensive.

An image's GLCM (Gray Level Co-occurrence Matrix) is calculated for a specified offset, as defined in Equation 1.

$$Offset = (Stride, Angle) = (\delta, \theta). \quad (1)$$

$$Where, \delta = [1], [2], [8], [32], [128]$$

$$\theta = [0, \Pi/4, \Pi/2, 3\Pi/4]$$

The derivation of Statistical Haralick features involves the utilization of a co-occurrence matrix, which is generated from an image. The following sections provide a detailed explanation of the features and the mathematical equations employed for their calculation.

Contrast–measurement of the local variations: This feature assesses the local intensity variations within an image. In further explication, the contrast measure serves to quantify the disparity in luminance values between a given pixel and its adjacent pixel within the entirety of the image. Therefore, a low-contrast image is proposed to demonstrate a steady progression of grayscale values, whereas a high-contrast image showcases more pronounced values at the spectrum's extremes (black and white). This implies that the lack of contrast in an image is mainly associated with reduced spatial frequencies rather than diminished gray levels. Therefore,

calculating GLCM contrast may be efficiently performed using Equation 2, which strongly correlates with spatial frequencies.

$$contrast = \sqrt{\sum_{i,j=0}^{n-1} M_{ij}(i-j)^2} \quad (2)$$

Here, 'n' denotes the total number of gray levels employed, corresponding to the GLCM dimension.

Energy, sometimes called uniformity, characterizes the level of homogeneity in grayscale distribution and the texture's overall consistency. It serves as a measure of grayscale pattern stability. Consequently, a higher Energy value indicates a more consistent and stable pattern. In mathematical terms, energy represents the quadratic sum of GLCM elements and quantifies the concentration of gray levels in the GLCM. It can be calculated from the GLCM using Equation 3.

$$\text{energy} = \sqrt{\sum_{i,j=0}^{n-1} M_{ij}^2} \quad (3)$$

The Angular Second Moment can be defined as the mathematical square of energy. The calculation can be performed by utilizing the formula provided in Equation 4.

$$ASM = \sum_{i,j=0}^{n-1} M_{ij}^2 \quad (4)$$

The Inverse Difference Moment (IDM) is a metric used to evaluate the level of local homogeneity in an image, offering valuable information about the intensity variance present within the image. The presence of increased homogeneity within an image is indicative of a decrease in the level of intensity variation. Moreover, it may be stated that IDM has an inverse relationship with contrast, whereby an increase in contrast results in a decrease in homogeneity, and conversely, a decrease in contrast leads to an increase in homogeneity. To enhance precision, homogeneity is a measure that evaluates the degree to which the distribution of GLCM (Gray-Level Co-occurrence Matrix) elements aligns with its diagonal. This measure essentially serves to characterize the uniformity of texture. It is derived from the GLCM using Equation 5:

$$homogenity = \sum_{i,j=0}^{n-1} \frac{1}{1+(i-j)^2} M_{ij} \quad (5)$$

Correlation: The discriminative capacity of the Correlation feature extracted from the Gray-Level Co-occurrence Matrix (GLCM) is comparable to that of the Contrast feature. The quantitative assessment of the level of association or correlation between a pixel and its neighboring pixels throughout the entire image is performed. In essence, the concept of correlation involves evaluating the linear relationship between the gray levels of pixels situated in specified relative positions to one another. The correlation property is constrained to the interval [-1, 1]. The correlation score serves as a measure of the magnitude of the relationship between images. A correlation value close to +1 or -1 signifies a strong positive or negative association between two images, respectively. The correlation of the Gray-Level Co-occurrence Matrix (GLCM) can be computed by utilizing Equation 6.

$$correlation = \sum_{i,j=0}^{n-1} \frac{(i-\mu i)(j-\mu j)}{(\sqrt{\sigma i^2 \sigma j^2})} \quad (6)$$

In this context, μi and σi represent the mean and standard deviation values for the normalized GLCM elements calculated in the horizontal direction, while μj and σj refer to the corresponding values computed in the vertical direction.

Dissimilarity, also referred to as Difference Average, calculates the mean difference between the distributions of gray levels inside an image, as denoted by Equation 7.

$$dissimilarity = \sum_{i,j=0}^{n-1} M_{ij} |i-j| \quad (7)$$

The Histogram classifier provides several features, including mean, max, min, standard deviation, and variance, which aid in classifying chest X-rays into disease categories [31].

The mean, also known as the average, is calculated by summing all the values in a column and then dividing that sum by the total number of values, as described in Equation 8.

$$Mean = \Sigma mini / N \quad (8)$$

The standard deviation, typically denoted as "s," is merely the square root of the variance, as shown in Equation 9.

$$standard\ deviation = \sqrt{\sum n_i (m_i - \mu)^2/(n-1)} \quad (9)$$

Variance is a metric that provides insight into the dispersion or spread of a specific data distribution. The sample variance is typically represented as s^2, as seen in Equation 10.

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 \quad (10)$$

D. Classification and Evaluation Stage

This section focuses on the feature classification module used in our CAD system, which aims to differentiate between cases of Sarcoid and COVID-19. The primary purpose of the classification module inside our computer-aided design (CAD) system is to assign each chest X-ray (CXR) image from the gathered dataset to one of three diagnostic categories: Sarcoid, COVID-19, or normal. This classification is based on the utilization of extracted textural information. The efficacy of the classification module is contingent upon the availability of a dataset comprising a collection of clinical diagnostic cases. In the context of COVID-19 classification, numerous classification strategies have been documented in the extant literature. These techniques include Artificial Neural Networks (ANN), Support Vector Machines (SVMs)[32], Bayesian Networks (BN), and k-nearest Neighbor (k-NN) [32]. Conditional Random Fields (CRFs), among others. In the present study, the k-nearest Neighbor (k-NN) model has been selected for feature categorization [33]. To enhance computational efficiency, the images within the dataset undergo an initial transformation from color to 8-bit grayscale format. Subsequently, these images are down-sampled to a standardized resolution of 128 × 128 pixels. The preprocessing stage precedes the feature extraction phase. In the absence of an independent dataset, the dataset images are randomly partitioned into training and test sets, with 80% allocated for training and 20% for testing. Various assessment criteria are employed to quantitatively evaluate the performance of our computer-aided design (CAD) system. These metrics encompass accuracy, precision, and recall, which may be mathematically represented as follows.

The level of precision or correctness in the information or data being presented. Accuracy is widely recognized as a

fundamental and frequently utilized statistic for evaluating performance. The concept of accuracy pertains to the probability of correctly identifying a randomly chosen instance, regardless of its positive or negative nature. To enhance precision, accuracy refers to the likelihood that the diagnostic test yields a correct diagnosis, as denoted by equation 11[34].

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (11)$$

where TN refers to true negatives, which signify the accurate categorization of chest X-ray pictures into normal and sick classifications. Conversely, FN denotes false negatives, which pertain to cases in which chest X-ray images are erroneously classified as belonging to the normal or infected categories.

Precision, often known as precision, is a metric that emphasizes the correctness of identifying positive instances within the set of all projected positive instances. The mathematical representation of this concept is denoted as the quotient between the number of accurately predicted positive classes and the overall number of expected positive classes, as seen in equation 12 [35].

$$\text{Precision} = \frac{TP}{TP+FP} \quad (12)$$

Equation 12 demonstrates a clear correlation between high precision and a consistently low false positive rate.

Recall, often referred to as sensitivity, hit rate, or true positive rate, can be conceptualized as the ability of a model to identify all instances of positive situations. The mathematical representation of this concept is denoted by equation 13 [34]:

$$\text{Recall} = \frac{TP}{TP+FN} \quad (13)$$

It's essential to observe that equation 13 implies that a high recall is undeniably linked to a low false negative rate.

IV. Results and Discussion

This part explores the comprehensive outcomes derived from several rigorous experiments. These tests aimed to assess the efficacy of our proposed computer-aided design (CAD) model in automating the detection of sarcoidosis and differentiating between sarcoidosis, COVID-19, and normal cases in chest X-ray pictures. The objective was to confirm the model's exceptional performance characteristics and determine the optimal distance and angle parameters for GLCM to yield the best results.

The X-ray images outlined in section 2.2.2 were initially processed and standardized to achieve this objective. Subsequently, these images were categorized into three distinct classes: Sarcoid, COVID-19, and normal, as described in section 2.1. Six features were then extracted from these images using GLCM. Furthermore, the distance parameter was systematically varied each time the data was processed with GLCM and its features were extracted. These variable stride values for GLCM (1, 2, 8, 32, 128) were selected based on the superior performance of individual machine learning classifiers.

In the subsequent step, the optimal GLCM distance was assessed using KNN classification. The accuracy gradually improved as the transition was made from stride combination 1 to (1, 2, 8, 32, 128). As illustrated in Figure 3, our CAD model exhibits significant potential, attaining competitive levels of average accuracy, precision, and recall. The model demonstrated a level of performance characterized by an accuracy rate of 93.50%, precision rate of 98%, and recall rate of 93.50% when evaluated under the conditions of a distance of 128 units and an angle of 45 degrees.

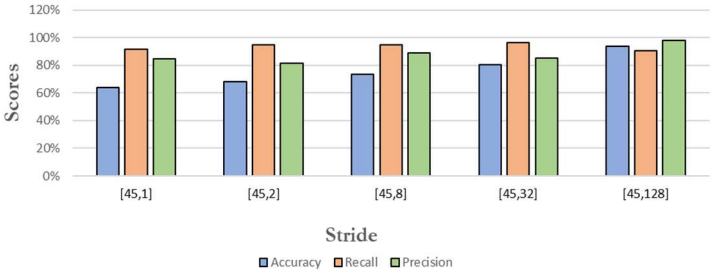

Figure 3. The performance of the three sets in the GLCM was studied using the KNN approach concerning accuracy, recall, and precision.

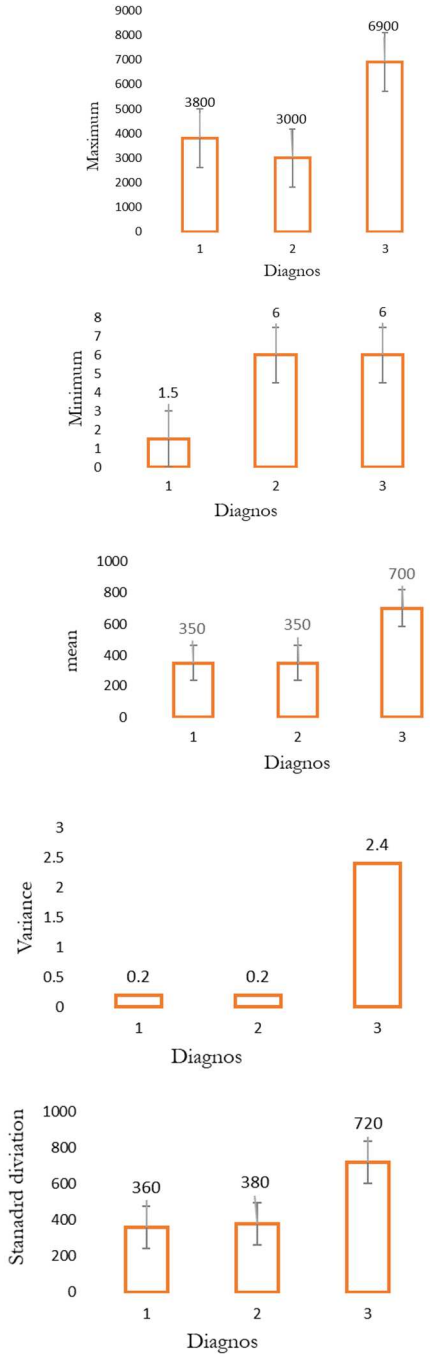

Figure 4. (3) classifiers give (mean, sta, min, max, var)

Consequently, it becomes evident that the GLCM stride mixture 128 outperforms the other four stride combinations attempted for GLCM feature extraction. The histogram was employed for classifying X-ray images and detecting sarcoidosis. Each image underwent analysis using five equations, yielding respective results. Based on these outcomes, the image was categorized as sarcoid or non-sarcoid. This process is visually represented in Figure 4, where the features contribute to the diagnosis, with values assigned as follows: normal = 0, COVID-19 = 1, and sarcoid = 2.

CONCLUSION

This paper has introduced a Computer-Aided Detection (CAD) method for distinguishing sarcoid cases from CXR images and distinguishing between sarcoidosis, COVID-19, and normal cases. The proposed methodology utilizes a refined collection of GLCM-based texture features to represent the segmented lung tissue accurately in every CXR image. The collected features undergo normalization and are subsequently employed in a discriminative K-nearest neighbors (KNN) model for the ultimate classification of Sarcoid. The chest X-rays are categorized into sarcoidosis, COVID-19, and normal.

A comprehensive examination and verification process was carried out on a substantial dataset of frontal chest X-ray (CXR) pictures that are publicly accessible. The approach demonstrated a noteworthy mean accuracy of 93.50%, accompanied by precision and recall metrics of 98% and 90.50%, respectively. The findings mentioned above suggest that the suggested computer-aided detection (CAD) approach has the potential to be a valuable asset for radiologists and medical physicists. It can facilitate the reliable differentiation between sarcoid and COVID-19 instances and expedite and enhance the accuracy of infection detection procedures.

Additionally, the paper explored a histogram method, but it yielded weaker results than GLCM. The optimal angle was 45 degrees, while the best distance was 128, resulting in the highest resolution of 93.50%. Smaller distances produced lower resolutions.

In the context of future endeavors, it is imperative to delineate two primary aims. To begin with, it is proposed to optimize the process of extracting lung tissue features by using a combination of several texture features, including Local Binary Patterns (LBP), Local Directional Contrast Patterns (LDCRF), Histogram of Oriented Gradients (HOG), among others. This integration aims to develop a hybrid feature descriptor that can effectively capture the relevant characteristics of lung tissue. Furthermore, it is recommended to broaden the scope of their investigations by incorporating supplementary public chest X-ray (CXR) datasets that encompass individuals diagnosed with sarcoidosis, COVID-19, or other viral and bacterial pneumonia, whether confirmed or suspected instances.

REFERENCES

[1] M. A. Judson, "Sarcoidosis: clinical presentation, diagnosis, and approach to treatment," Am. J. Med. Sci., vol. 335, no. 1, pp. 26–33, 2008.

[2] N. Baghdadi, A. S. Maklad, A. Malki, and M. A. Deif, "Reliable sarcoidosis detection using chest x-rays with efficientnets and stain-normalization techniques," *Sensors*, vol. 22, no. 10, p. 3846, 2022.

[3] A. K. Gerke, M. A. Judson, Y. C. Cozier, D. A. Culver, and L. L. Koth, "Disease burden and variability in sarcoidosis," Ann. Am. Thorac. Soc., vol. 14, no. Supplement 6, pp. S421–S428, 2017.

[4] N. Mustafa, "Research and statistics: coronavirus disease (COVID-19)," Int. J. Syst. Dyn. Appl., vol. 10, no. 3, pp. 67–86, 2021.

[5] P. Ji, A. Song, P. Xiong, P. Yi, X. Xu, and H. Li, "Egocentric-vision based hand posture control system for reconnaissance robots," J. Intell. Robot. Syst., vol. 87, pp. 583–599, 2017.

[6] P. Rai, B. K. Kumar, V. K. Deekshit, I. Karunasagar, and I. Karunasagar, "Detection technologies and recent developments in the diagnosis of COVID-19 infection," Appl. Microbiol. Biotechnol., vol. 105, pp. 441–455, 2021.

[7] C. Huang et al., "Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China," Lancet, vol. 395, no. 10223, pp. 497–506, 2020.

[8] M. W. Ziegenhagen, U. K. Benner, G. Zissel, P. Zabel, M. A. X. Schlaak, and J. Muller-Quernheim, "Sarcoidosis: TNF-α release from alveolar macrophages and serum level of sIL-2R are prognostic markers," Am. J. Respir. Crit. Care Med., vol. 156, no. 5, pp. 1586–1592, 1997.

[9] C. Tana et al., "Sarcoidosis and COVID-19: at the cross-road between immunopathology and clinical manifestation," Biomedicines, vol. 10, no. 10, p. 2525, 2022.

[10] R. M. Pereira, D. Bertolini, L. O. Teixeira, C. N. Silla Jr, and Y. M. G. Costa, "COVID-19 identification in chest X-ray images on flat and hierarchical classification scenarios," Comput. Methods Programs Biomed., vol. 194, p. 105532, 2020.

[11] T. Ozturk, M. Talo, E. A. Yildirim, U. B. Baloglu, O. Yildirim, and U. R. Acharya, "Automated detection of COVID-19 cases using deep neural networks with X-ray images," Comput. Biol. Med., vol. 121, p. 103792, 2020.

[12] F. L. Utino et al., "Second-harmonic generation imaging analysis can help distinguish sarcoidosis from tuberculoid leprosy," J. Biomed. Opt., vol. 23, no. 12, p. 126001, 2018.

[13] D. Ganeshan, C. O. Menias, M. G. Lubner, P. J. Pickhardt, K. Sandrasegaran, and S. Bhalla, "Sarcoidosis from head to toe: what the radiologist needs to know," Radiographics, vol. 38, no. 4, pp. 1180–1200, 2018.

[14] Y. J. Kim, "Machine learning model based on radiomic features for differentiation between covid-19 and pneumonia on chest x-ray," Sensors, vol. 22, no. 17, p. 6709, 2022.

[15] Y. Oh, S. Park, and J. C. Ye, "Deep learning COVID-19 features on CXR using limited training data sets," IEEE Trans. Med. Imaging, vol. 39, no. 8, pp. 2688–2700, 2020.

[16] E. Koda et al., "CT texture analysis of mediastinal lymphadenopathy: Combining with US-based elastographic parameter and discrimination between sarcoidosis and lymph node metastasis from small cell lung cancer," PLoS One, vol. 15, no. 12, p. e0243181, 2020.

[17] T. D. Pham, Y. Watanabe, M. Higuchi, and H. Suzuki, "Texture analysis and synthesis of malignant and benign mediastinal lymph nodes in patients with lung cancer on computed tomography," Sci. Rep., vol. 7, no. 1, p. 43209, 2017.

[18] S. Katsushika et al., "Deep learning algorithm to detect cardiac sarcoidosis from echocardiographic movies," Circ. J., vol. 86, no. 1, pp. 87–95, 2021.

[19] D.-P. Fan et al., "Inf-net: Automatic covid-19 lung infection segmentation from ct images," IEEE Trans. Med. Imaging, vol. 39, no. 8, pp. 2626–2637, 2020.

[20] J. Mañá, A. Salazar, and F. Manresa, "Clinical factors predicting persistence of activity in sarcoidosis: a multivariate analysis of 193 cases," Respiration, vol. 61, no. 4, pp. 219–225, 1994.

[21] E. E.-D. Hemdan, M. A. Shouman, and M. E. Karar, "Covidx-net: A framework of deep learning classifiers to diagnose covid-19 in x-ray images," arXiv Prepr. arXiv2003.11055, 2020.

[22] P. Dayal et al., "Performance limitations of flat-histogram methods," Phys. Rev. Lett., vol. 92, no. 9, p. 97201, 2004.

[23] M. Manansala et al., "COVID-19 and sarcoidosis, readiness for vaccination: challenges and opportunities," Front. Med., vol. 8, p. 672028, 2021.

[24] A. Baraldi and F. Panniggiani, "An investigation of the textural characteristics associated with gray level co-occurrence matrix statistical parameters," IEEE Trans. Geosci. Remote Sens., vol. 33, no. 2, pp. 293–304, 1995.

[25] R. E. Hammam, A. A. A. Solyman, M. H. Alsharif, P. Uthansakul, and M. A. Deif, "Design of Biodegradable Mg Alloy for Abdominal Aortic

Aneurysm Repair (AAAR) Using ANFIS Regression Model," *IEEE Access*, vol. 10, pp. 28579–28589, 2022.

[26] L. Cadena, A. Zotin, and F. Cadena, "Enhancement of medical image using spatial optimized filters and OpenMP technology," Lect. Notes Eng. Comput. Sci., 2018.

[27] P. K. Mall, P. K. Singh, and D. Yadav, "GLCM based feature extraction and medical X-RAY image classification using machine learning techniques," in 2019 IEEE Conference on Information and Communication Technology, 2019, pp. 1–6.

[28] R. E. Hammam *et al.*, "Prediction of wear rates of UHMWPE bearing in hip joint prosthesis with support vector model and grey wolf optimization," *Wirel Commun Mob Comput*, vol. 2022, pp. 1–16, 2022.

[29] R. E. Hammam, A. A. A. Solyman, M. H. Alsharif, P. Uthansakul, and M. A. Deif, "Design of Biodegradable Mg Alloy for Abdominal Aortic Aneurysm Repair (AAAR) Using ANFIS Regression Model," *IEEE Access*, vol. 10, pp. 28579–28589, 2022.

[30] R. M. Haralick, K. Shanmugam, and I. H. Dinstein, "Textural features for image classification," IEEE Trans. Syst. Man. Cybern., no. 6, pp. 610–621, 1973.

[31] H. Zhang, W. Gao, X. Chen, and D. Zhao, "Object detection using spatial histogram features," Image Vis. Comput., vol. 24, no. 4, pp. 327–341, 2006.

[32] Q. I. Ahmed, H. Attar, A. Amer, M. A. Deif, and A. A. A. Solyman, "Development of a Hybrid Support Vector Machine with Grey Wolf Optimization Algorithm for Detection of the Solar Power Plants Anomalies," *Systems*, vol. 11, no. 5, p. 237, 2023.

[33] S. Zhang, X. Li, M. Zong, X. Zhu, and R. Wang, "Efficient kNN classification with different numbers of nearest neighbors," IEEE Trans. neural networks Learn. Syst., vol. 29, no. 5, pp. 1774–1785, 2017.

[34] M. A. Deif, M. A. A. Eldosoky, H. W. Gomma, A. M. El-Garhy, and A. S. Ell-Azab, "Adaptive neuro-fuzzy inference system controller technique for lower urinary tract system disorders," *J Clin Eng*, vol. 40, no. 3, pp. 135–143, 2015.

[35] E. M. O. Mokhtar and M. A. Deif, "Towards a Self-sustained House: Development of an Analytical Hierarchy Process System for Evaluating the Performance of Self-sustained Houses," *ENGINEERING JOURNAL*, vol. 2, no. 2, 2023.