Grenze International Journal of Engineering and Technology, June Issue

Machine Learning Approach for Tuberculosis Classification

Suguna G C[1], Sheela S N[2], Veerabhadrappa S T[3] and Ravikumar K P[4]
[1-4]Department of Electronics and Communication Engineering, JSS Academy of Technical Education, Bengaluru Affiliated to Visvesvaraya Technological University, Belagavi, Karnataka, Bengaluru, India
Email: sugunagc@jssateb.ac.in or, sheelasn@jssateb.ac.in, veerabhadrappast@jssateb.ac.in ravikumarkp@jssateb.ac.in

***Abstract*— The prevalence of tuberculosis, a severe sickness, is widespread in contemporary times. Given the severity of tuberculosis as a disease, it is imperative to adopt a proactive and expeditious approach to its treatment. Mycobacterium tuberculosis is a bacterial species responsible for the pathogenesis of tuberculosis. In order to overcome problem, a machine-learning methodology was proposed for diagnosis tuberculosis. Chest X-ray (CXR) images have served as a vital diagnostic tool for the respective medical condition. This study utilizes a three-step categorization procedure in order to ascertain the infection status of TB cases. The three steps of learning encompass histogram segmentation, GLCM feature extraction, and random forest classification. Histogram-based image segmentation is employed to partition the chest X-ray image into distinct segments according to the pixel values. The GLCM technique has demonstrated efficacy in segmenting CXR pictures for illness diagnosis due to its ability to capture the grayscale connection between neighboring pixels through distance and angle measurements. The Modified Random Forest algorithm is a computational approach utilized to distinguish between tuberculosis-infected and normal CXR image data. The aforementioned approach provides the most precise methodology for diagnosing a patient's illness, with the random forest model achieving a 95.23% an accuracy rate.**

***Keywords*— Tuberculosis, Histogram segmentation, GLCM features, Random-forest classifier.**

I. INTRODUCTION

The medical condition referred to as "pulmonary tuberculosis" primarily impacts the respiratory system, particularly the lungs, and possesses a high degree of transmissibility. The term "extra pulmonary tuberculosis" is used to describe the manifestation of tuberculosis in organs other than the lungs. The rapid global dissemination of this phenomenon has led to its present classification as one of the most significant threats to the human population. According to the World Health Organization (WHO), tuberculosis was responsible for over 1.8 million deaths worldwide in 2015 [1]. Tuberculosis is a clinical condition commonly caused by the bacterium Mycobacterium tuberculosis. This condition can be classified as a multisystem infectious disease due to its impact on multiple vital organ systems [1]. Furthermore, tuberculosis can be classified into two categories: "latent" or "active." Based on a 2018 estimate provided by the World Health Organization, it is indicated that approximately 25% of the global population is affected by latent tuberculosis [1-2]. Tuberculosis can be contracted by individuals across all age groups, ranging from infants to the elderly.

Grenze ID: 01.GIJET.12.2.93_21

Nevertheless, certain age cohorts may exhibit a higher susceptibility to contracting tuberculosis or a greater propensity to manifest severe symptoms. Individuals who are in the early stages of life, the elderly population, individuals with HIV or other chronic medical conditions, and those with weakened immune systems are at a higher risk of contracting active tuberculosis disease. In numerous nations, there is a heightened vulnerability to tuberculosis among children who are under the age of five. In general, the risk of tuberculosis increases as individual's age, with the highest incidence rates observed among individuals aged 65 years and older. It is imperative for individuals across all age groups to possess knowledge on the indicators and manifestations of tuberculosis, and to promptly pursue medical assistance in the event of potential exposure to the causative agents. Tuberculosis can affect individuals of all genders. Based on empirical study, it has been observed that there may be a somewhat higher incidence of tuberculosis among men compared to women. The observed heterogeneity in this context could potentially be influenced by various factors, including but not limited to variations in exposure to risk factors and differences in behavioral patterns. Mycobacterium tuberculosis is the bacterial pathogen responsible for the development of tuberculosis. The transmission of Mycobacterium tuberculosis, the causative agent of active tuberculosis disease, occurs when an infected individual expels small droplets containing the bacteria into the surrounding air through coughing or sneezing. These airborne particles can subsequently be inhaled by others, facilitating person-to-person transmission.

Not all individuals who get the pathogenic microorganisms will subsequently develop symptomatic tuberculosis. The bacterium has the potential to enter a dormant state within the host's body, wherein it does not manifest any discernible symptoms. The condition is commonly referred to as latent tuberculosis infection. Several variables contribute to the predisposition of active tuberculosis disease, such as compromised immune system function, proximity to an infected individual, and suboptimal living situations.

Tuberculosis is a grave ailment that requires prompt medical intervention and appropriate treatment. The treatment regimen for certain conditions often involves the concurrent administration of multiple antibiotics over an extended period of several months. In conjunction with pharmaceutical interventions, many home remedies can be employed to bolster the immune system and alleviate symptoms associated with tuberculosis therapy. These encompass the adoption of nutritionally balanced dietary practices, the maintenance of proper hygiene, and the attainment of sufficient sleep.

Machine learning, a subfield of artificial intelligence, facilitates the autonomous acquisition of knowledge by computers through the analysis of training data, hence enhancing their performance without the need for explicit programming. The identification of patterns within data and subsequent acquisition of knowledge from these patterns enables machine learning algorithms to autonomously generate predictions. In brief, algorithms and models utilized in machine learning acquire information by means of experiential learning.

The integration of machine learning techniques into the healthcare sector has the potential to facilitate the advancement of diagnostic tools for the analysis of medical pictures. Machine learning has been shown to reveal distinct patterns particular to certain diseases. To facilitate the early identification of diseases, it is plausible to enhance low-level data by extracting relevant information from the database [3-5]. Histogram-based segmentation is a widely employed and straightforward technique for picture segmentation. The histogram distribution of the CXR picture is utilized for the purposes of segmentation and thresholding. The histogram curves of all the photos are exhibited, resulting in the process of segmentation. The CXR images are effectively analyzed by the utilization of histogram segmentation, and relevant characteristics are retrieved as required [5-7]. The TB detection investigation utilized the Grey level co-occurrence matrix (GLCM), also referred to as the GLCM, yielding promising results. The combination of GLCM features and manual CXR picture segmentation has been found to yield effective results in the detection of tuberculosis [8-10]. The random forest algorithm is widely favored in the field of supervised machine learning for its effectiveness in addressing classification and regression tasks. The approach is grounded in the principle of majority voting, wherein many classifiers are combined to address a complex problem and enhance the model's performance [11-12].

II. Related Work

Olfa et al. introduced a machine learning model that utilizes hyper parameters for classification purposes and extracts the most impactful patterns from images associated with tuberculosis (TB). Various features, including frequency domain, spatial domain, angular second moment, correlation, contrast variance, inverse difference moment, sum variance, sum average, difference variance, entropy, sum entropy, difference entropy, etc., were extracted. A genetic algorithm (GA) technique was employed to select the most optimal features, which were then utilized as input for a support vector machine (SVM) classifier. The results illustrate the increased utility of the Support Vector Machine (SVM) classifier compared to alternative classifiers. The SVM algorithm demonstrated an accuracy rate of approximately 84% [1].

Zotin et al. proposed a methodology that incorporates lung margin detection and CXR categorization for the analysis of internal chest radiographs. The authors employed segmentation techniques commonly utilized in radiographic imaging to exclude lung area, big area, and irregular pictures from their analysis. In their study, the researchers utilized a gray-level co-occurrence matrix to extract a total of 18 features. These features encompassed various aspects such as contrast, correlation, energy, entropy, homogeneity, dissimilarity, autocorrelation, maximum probability, cluster prominence, cluster shade, information measure of correlation, sum of square, inverse difference moment, sum average, sum variance, sum entropy, difference variance, and sum entropy. Additionally, they employed a probabilistic neural network classifier to effectively differentiate between normal and pathological chest X-ray images. The employed technique demonstrates exceptional results with minimal training duration, achieving a commendable accuracy rate of 96% on average [2].

In their study, Sergii et al. presented a method for lung segmentation utilizing deep convolutional neural networks, incorporating lossless and lossy data augmentation techniques. The CADx system possesses the ability to extract small and disparate data by utilizing enhanced separation strategies, data augmentation methods, data set stratification procedures, and eliminating conspicuous outliers [4].

Tej et al. propose a methodology aimed at identifying various anomalies observed in chest X-ray pictures. These anomalies encompass pleural effusions, infiltrates, fibrosis, hilar swelling, dense masses attributed to tuberculosis, and other related illnesses. The strategy employed in this study is founded upon the hierarchical feature gathering technique, wherein attributes are employed to differentiate between healthy and sick groups at two distinct levels of hierarchy [5].

Regina et al. examine the durability of machine learning models utilized in X-ray machinery intended for assisting in the detection of tuberculosis. Two LML algorithms, namely the Efficient Lifelong Learning Algorithm and Learning Unforgotten (LwF), were employed for the job of classifying tuberculosis. The findings indicate that the LML paradigm is effective in sustaining performance in both tasks. An accuracy rate of around 84.6% was achieved [6].

Mohamed et al. discuss the presence of a set of graphical metrics and heuristics in the field of mathematics. They propose a straightforward spatial characteristic and a technique for distinguishing between various vertical continuous lines using the histogram method. This approach considers both the vertical and horizontal distances between symbols. The two entities exhibit a degree of intersection; however they are associated with distinct numerical categories. The researchers achieved the highest level of accuracy, which was 96.3% [7].

Alphonse et al. conducted a study on the identification of pulmonary disease in chest X-ray pictures. The study involved three main steps: segmentation, feature extraction, and classification. The CXR photos were analyzed using a decision tree classifier known as the layered iterative decision tree learner in order to extract relevant characteristics. The researchers achieved an accuracy ranging from 98% to 99% depending on the dataset [8].

III. Dataset Description

The collection comprises of chest X-ray (CXR) images, consisting of 3500 photos of individuals without tuberculosis (TB) and 700 images of patients diagnosed with TB. Additionally, there are 2800 TB images available for download through the NIAID TB portal, which can be accessed by consenting to an agreement [24]. The dataset comprises 8-bit images with size ranging from 512x512 to 1024x1024.

IV. Methodology

The present study employs a three-stage methodology to categories lung chest X-ray pictures from the specified dataset into two distinct groups: tuberculosis-infected and non-infected. The process encompasses three main components: segmentation, feature extraction, and classification. The X-ray images underwent pre-processing procedures, including noise reduction and the use of the Gabor filter.

Segmentation was performed on all chest X-ray images in the designated dataset using a histogram-based method. Subsequently, features were extracted from the dark-level co-occurrence matrix. After being stored, the collected features were inputted into the random forest classifier. The random forest classifier was employed to train and evaluate the model, effectively discerning between photographs that exhibited signs of tuberculosis infection and those that did not. The data used for training and testing purposes were divided into proportions of 80% for training and 20% for testing.

A. Histogram-based Segmentation and Thresholding

During the initial stage of this review, segmentation based on histograms was performed. Segmentation is a widely employed method for dividing an image into its constituent regions. The calculation of segmentation relies on factors such as intensity values, irregularity, and similarity. The split of histograms in this review is contingent upon the intensity values. The histogram curve displays pixel intensities along the horizontal axis and the number of image segments with corresponding pixel intensity values along the vertical axis. The histogram curve in this study is constructed throughout a range of 0 to 255-pixel intensities.

Histogram thresholding was performed on both tuberculosis-infected and non-infected X-ray images from the dataset. The results of this process are depicted in Figure 2 and Figure 3.

Figure 2 illustrates the application of thresholding in the context of histogram partitioning for images depicting tuberculosis-infected and uninfected conditions. The compilation of the histogram-based partitioning was completed for the dataset, which consisted of a range of 50 to 60 images. The study observed that images contaminated with tuberculosis had a higher degree of fixation on white pixels with values ranging from 160 to 200 on the horizontal axis. The photographs that were not affected primarily emphasized areas with low values ranging from 0 to 10. The removal of highlights from the Gray-Level Co-occurrence Matrix is facilitated through the process of histogram analysis.

The study involves a three-step process for tuberculosis categorization, with feature extraction being the second stage. The extraction of features is performed on the grey-level co-occurrence matrix. Following the segmentation process using histogram analysis, the subsequent step involves the extraction of GLCM features. Therefore, this particular stage can be referred to as GLCM feature extraction.

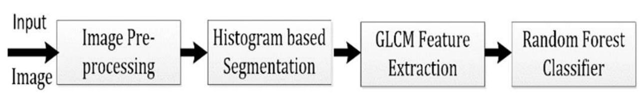


Figure 1: Methodology for categorizing Tuberculosis

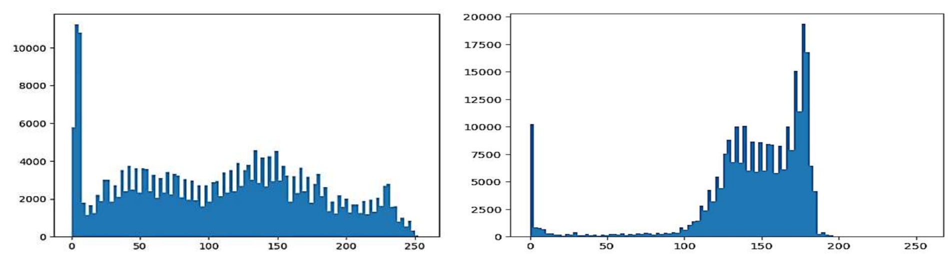


Figure 2.: Histogram for tuberculosis (a) infected and (b) non-infected CXR image

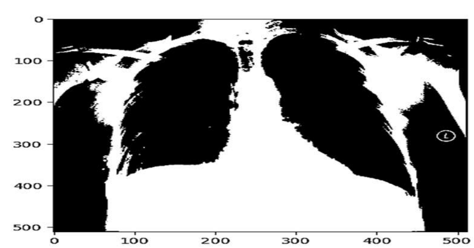


Figure 3(a) Thresholding of normal X-ray image of a lung

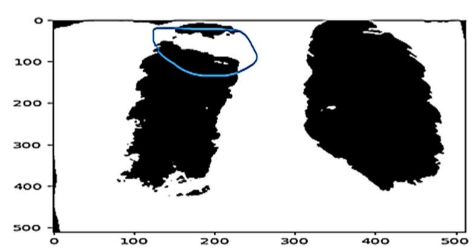


Figure 3 (b) Thresholding of tuberculosis-infected X-ray image of a lung with marking of segmenting region

B. GLCM features

GLCM features, including dissimilarity, correlation, homogeneity, contrast, angular second moment (ASM), and energy, were computed for each picture within the designated dataset. These computations were performed at a specific displacement and over many angles.

The displacement of the Gray-Level Co-occurrence Matrix (GLCM) refers to the distance between two pixels that must be taken into account during feature computation. When producing GLCM features, it is essential to consider not only the displacement but also the orientation of adjacent pixels. The calculation of the specified features [27-28] involves the inclusion of angles at 0, 45, 90, 135, and 180 degrees. The GLCM features calculated are described as

Dissimilarity

Dissimilarity can be defined as the variation of grey-level pairs in an image. The equation for dissimilarity is given by

$$Dissimilarity = \sum_{i,j=0}^{M-1} Pij|i-j| \qquad (1)$$

Where 'i' represents horizontal cell coordinates and 'j' represents vertical cell coordinates [0, 1] provides the range of dissimilarity. When the reference and neighboring pixels' grey levels are at the furthest edges of the range that image pixels might have for grey levels, it is maximum.

Correlation

It is a measurement of the linear relationships between the image's grey tones. The equation provides the correlation.

$$Correlation = \sum_{i,j=0}^{M-1} Pij\left[\frac{(i-\mu_i)(i-\mu_j)}{(\sigma_i^2)(\sigma_j^2)}\right] \qquad (2)$$

Homogeneity

The homogeneity of the GLCM's non-zero entries is measured. The equation gives homogeneity.

$$Homogeneity = \sum_{i,j=0}^{M-1} \frac{Pij}{1+(i-j)^2} \qquad (3)$$

Any image has a high GLCM homogeneity if it has a large number of pixels with the same or nearly identical values for their grey levels. Lower homogeneity results from bigger pixel changes. It varies inversely when the grey level values of adjacent pixels vary.

Contrast

Contrast is the linear dependency of grey levels of the neighboring pixels. The equation to represent contrast is given by,

$$Contrast = \sum_{i,j=0}^{M-1} Pij(i-j)^2 \qquad (4)$$

Contrast is inversely proportional to the similarity of the neighboring pixels. If the contrast is 0 then the neighboring pixels are identical. The range of contrast is given by [0, 1].

Angular Second Moment (ASM)

ASM is used to quantify repetitions or uniformity in pixel pair texturing. It identifies anomalies in the picture textures. The representational equation is given by,

$$\text{ASM}=\sum P_{ij^2} \qquad (5)$$

ASM has a range of [0,1]. The grey level distribution takes on a constant periodic pattern when the ASM, which has a maximum value of 1, is present.

Energy

It can be defined as the measure of local homogeneity [27].

$$\text{Energy}=\sum (Pij)^2 \qquad (6)$$

Energy is directly proportional to the homogeneity and is highest i,e; 1 for a constant image. The range of energy is from 0 to 1.

C. Random Forest Classification

The concluding phase of our research entails the utilization of a random forest classifier to classify photographs as either tuberculosis-infected or non-infected. The random forest classifier is provided with the GLCM features that were extracted in step 4.2 as input. The random forest algorithm is widely acclaimed in the field of machine learning for its effectiveness in addressing classification and regression tasks. The supervision is conducted collaboratively. The random forest is composed of several decision trees with varying choices. The methodology does not hinge solely on a single tree; rather, it leverages predictions from multiple trees and determines the projected outcome by aggregating the majority votes from these predictions. The operational mechanism of a random forest algorithm for addressing classification problems is as follows. The process of calculating the mean value from the 'n' outcomes of a decision tree is referred to as averaging . Averaging is commonly employed in the context of regression problems, but majority voting is typically preferred for classification tasks. The programmer has the ability to adjust the hyper parameter "estimators" in order to specify the quantity of decision trees within the model. The random forest classifier is formed by combining the bagging and decision tree approaches. The aforementioned averaging strategy is employed by random forest algorithms in order to enhance accuracy and mitigate the issue of over-fitting. Random forest has been found to be highly effective in situations involving huge dimensionality and varied feature types. The parameters of the random forest approach, namely n_estimators, max_depth, and min_samples_split, hold significant importance. The utilizations of n-estimators, which refers to the quantity of decision trees employed in the Gini index-based random forest methodology, has been documented in [12].

The Mathematical Formula

$$Gini=1-n\sum i=1(pi)^2 \qquad (7)$$

yields the Gini index, where "pi" denotes the likelihood that an item will be classified into a specific class.
The assessment of information impurity is conducted by the utilizations of the Gini index, whereby a lower Gini index signifies a reduced level of impurity. The decision tree will be generated based on m different attributes. According to the documented predictions of each tree, individuals will thereafter select the optimal conclusion or ultimate result.
The operational procedure of the random forest algorithm can be succinctly summarized as follows:

- Initially, subsets of the original dataset are generated, followed by the selection of replacement rows and columns to form subsets of the training dataset.
- Unique decision trees will be generated from each of the groups utilized.
- Each decision tree will generate a result.

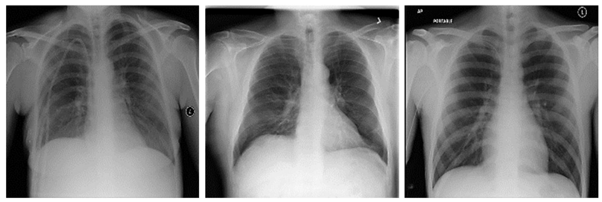
In the context of the classification problem, the final output is determined by the process of majority voting. Conversely, in the regression problem, the final output is obtained by averaging the relevant data [31-35].

IV. EXPERIMENT AND ANALYSIS

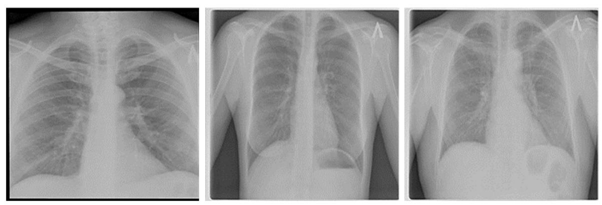
This section provides a more detailed description of the hierarchical feature extraction approach that is recommended for classifying tuberculosis, normal and infected chest X-ray images. Patients who were included in the dataset were categorized based on whether they had tuberculosis or did not have TB, using chest X-ray images.
This study employed a random forest classifier, a machine learning technique, to classify patients as either infected or non-infected with tuberculosis. In contrast to the asymmetrical distribution and long-tailed pattern observed in the distribution of image areas, the distribution of lung masks exhibited a far closer resemblance to the typical distribution. The training and testing data were partitioned into two distinct groups, with a distribution ratio of 80% for the training data and 20% for the testing data.
This study aims to demonstrate the efficacy of the proposed lung segmentation strategy in combination with lossless and lossy data augmentation techniques, in order to generate statistically reliable predictions of tuberculosis even while working with very small datasets including less than 10^3 pictures. The segmentation of tuberculosis-infected and non-infected images was performed using histogram analysis. The intensity values of the images were measured on a scale ranging from 0 to 255. It was observed that tuberculosis-infected images exhibited a higher concentration of white patches in the histogram curve, particularly in the intensity range of 160 to 200 on the x-axis. In contrast, non-infected images showed a greater concentration of dark regions, specifically in the intensity range of 0 to 10. Histogram segmentation was performed on approximately 50 to 60 pictures. In order to facilitate segmentation and feature extraction, all images within the dataset were converted

to grayscale and shrunk to dimensions of 256x256.The extraction of features is performed by utilizing the grey-level co-occurrence matrix in conjunction with the histogram.

Figure 4 displays photos depicting individuals with tuberculosis infection and individuals without tuberculosis infection. The characteristics of the Gray-Level Co-occurrence Matrix exhibit variations between images with and without Tuberculosis illness, hence offering highly correlated features that can be utilized for effective categorization. An analysis of variance was conducted to determine the significance of qualities that were and were not impacted by tuberculosis. Table 1 presents the mean and standard deviation of the features of the Gray-Level Co-occurrence Matrix.



(a) Tuberculosis infected X-Ray image



(b) Non-infected X-Ray image

Figure 4: Sample of tuberculosis-infected and non-infected images from the Kaggle database.

TABLE I: MEAN AND STANDARD DEVIATION OF TUBERCULOSIS AND NON- TUBERCULOSIS OF GLCM FEATURES

GLCM	TUBERCULOSIS INFECTED Mean±SD	TUBERCULOSIS NON-INFECTED Mean±SD
Dissimilarity	14.5±3.8	15.8±3.2
Correlation	0.87±0.06	0.91±0.03
Homogeneity	0.18±0.09	0.14±0.05
Contrast	665.6±350.2	642.5±249.6
Angular Second Moment	0.02±0.06	0.01±0.01
Energy	0.054±0.09	0.034±0.04

Table 2 presents the results of the random forest classification technique applied on GLCM features for distinguishing tuberculosis-infected data from non-tuberculosis data.

The training data is fitted to the random forest classifier after the data is randomly divided into training and testing samples. Now the test data are being predicted by the random forest classifier. The accuracy is measured by comparing the output of the random forest classifier with the benchmark result. The tiny size of the dataset, which may be reduced by the well-known data augmentation methods, can be used to explain such overfitting. The role of the lossless and lossy forms of data augmentation was examined and employed for this goal.

TABLE II: PERFORMANCE METRICS

Validation	Sensitivity	Specificity	Accuracy	Precision	F1 score
5-Fold	80.28%	98.28%	95.23%	90.47%	85.1%

VI. CONCLUSION

The article proposes a machine learning approach for the detection of tuberculosis. The proposed methodology utilizes a three-phase approach, which includes histogram segmentation, GLCM feature extraction, and random forest classification, in order to accurately detect tuberculosis .In the initial part of the investigation, histogram-based segmentation was conducted on the chest X-ray images of patients with tuberculosis and those without TB. Subsequently, the GLCM properties, encompassing energy, homogeneity, contrast, and ASM, were computed and documented in a CSV file. The random forest classifier was trained and tested using the input data supplied from the CSV file.The experimental findings demonstrated that tuberculosis might be effectively predicted by the utilization of diverse datasets. The aforementioned methodology has several advantages, including full automation, efficient handling of large datasets, and improved accuracy.

REFERENCES

[1] Hrizi, O., Gasmi, K., Ben Ltaifa, I., Alshammari, H., Karamti, H., Krichen, M., Ben Ammar, L. and Mahmood, M.A., 2022. Tuberculosis disease diagnosis based on an optimized machine learning model. Journal of Healthcare Engineering, 2022.

[2] Ignatius, J.L.P., Selvakumar, S., Paul, K.G.J.L., Kailash, A.B., Keertivaas, S. and Prajan, S.A.R., 2023. Histogram Matched Chest X-Rays Based Tuberculosis Detection Using CNN. COMPUTER SYSTEMS SCIENCE AND ENGINEERING, 44(1), pp.81-97.

[3] Singh, M., Pujar, G.V., Kumar, S.A., Bhagyalalitha, M., Akshatha, H.S., Abuhaija, B., Alsoud, A.R., Abualigah, L., Beeraka, N.M. and Gandomi, A.H., 2022. Evolution of machine learning in tuberculosis diagnosis: a review of deep learning-based medical applications. Electronics, 11(17), p.2634

[4] Faruk, O., Ahmed, E., Ahmed, S., Tabassum, A., Tazin, T., Bourouis, S. and Monirujjaman Khan, M., 2021. A novel and robust approach to detect tuberculosis using transfer learning. Journal of healthcare engineering, 2021.

[5] Ignatius, J.L.P., Selvakumar, S., Paul, K.G.J.L., Kailash, A.B., Keertivaas, S. and Prajan, S.A.R., 2023. Histogram Matched Chest X-Rays Based Tuberculosis Detection Using CNN. COMPUTER SYSTEMS SCIENCE AND ENGINEERING, 44(1), pp.81-97.

[6] Zotin, A., Hamad, Y., Simonov, K. and Kurako, M., 2019. Lung boundary detection for chest X-ray image classification based on GLCM and probabilistic neural networks. Procedia Computer Science, 159, pp.1439-1448.

[7] Ayas, S. and Ekinci, M., 2014. Random forest-based tuberculosis bacteria classification in images of ZN-stained sputum smear samples. Signal, Image and Video Processing, 8, pp.49-61.

[8] Dinnes, J., Deeks, J., Kunst, H., Gibson, A., Cummins, E., Waugh, N., Drobniewski, F. and Lalvani, A., 2007. A systematic review of rapid diagnostic tests for the detection of tuberculosis infection. Health Technology Assessment (Winchester, England), 11(3), pp.1-196.

[9] Stirenko, S., Kochura, Y., Alienin, O., Rokovyi, O., Gordienko, Y., Gang, P. and Zeng, W., 2018, April. Chest X-ray analysis of tuberculosis by deep learning with segmentation and augmentation. In 2018 IEEE 38th International Conference on Electronics and Nanotechnology (ELNANO) (pp. 422-428). IEEE.

[10] Rachna, H.B. and Swamy, M.M., 2013. Detection of Tuberculosis bacilli using image processing techniques. International Journal of Soft Computing and Engineering (IJSCE), 3(4).

[11] Raju, P.D.R. and Neelima, G., 2012. Image segmentation by using histogram thresholding. International Journal of Computer Science Engineering and Technology, 2(1), pp.776-779.

[12] Suguna, G.C., Veerabhadrappa, S.T., Tejas, A., Vaishnavi, P., Gowda, R., Udupa, P., Reddy, S. and Sudarshan, E., 2022. A Machine learning Classification approach for detection of Covid 19 using CT images. EMITTER International Journal of Engineering Technology, pp.183-194.

[13] Mall, P.K., Singh, P.K. and Yadav, D., 2019, December. Glcm-based feature extraction and medical x-ray image classification using machine learning techniques. In 2019 IEEE Conference on Information and Communication Technology (pp. 1-6). IEEE.

[14] Chandra, T.B., Verma, K., Singh, B.K., Jain, D. and Netam, S.S., 2020. Automatic detection of tuberculosis related abnormalities in Chest X-ray images using hierarchical feature extraction scheme. Expert Systems with Applications, 158, p.113514.

[15] Pal, M., 2005. Random forest classifier for remote sensing classification. International journal of remote sensing, 26(1), pp.217-222.

[16] Hussainy, S.F., Zaffar, F., Zaffar, M.A., Khaliq, A., Khan, I.H. and Ahmad, R., 2017. Decision-tree-inspired classification algorithm to detect Tuberculosis (TB).

[17] Shabut, A.M., Tania, M.H., Lwin, K.T., Evans, B.A., Yusof, N.A., Abu-Hassan, K.J. and Hossain, M.A., 2018. An intelligent mobile-enabled expert system for tuberculosis disease diagnosis in real-time. Expert Systems with Applications, 114, pp.65-77.

[18] Boehme, C.C., Nabeta, P., Hillemann, D., Nicol, M.P., Shenai, S., Krapp, F., Allen, J., Tahirli, R., Blakemore, R., Rustomjee, R. and Milovic, A., 2010. Rapid molecular detection of tuberculosis and rifampin resistance. New England Journal of Medicine, 363(11), pp.1005-1015.

[19] Rahman, T., Khandakar, A., Kadir, M.A., Islam, K.R., Islam, K.F., Mazhar, R., Hamid, T., Islam, M.T., Kashem, S., Mahbub, Z.B. and Ayari, M.A., 2020. Reliable tuberculosis detection using chest X-ray with deep learning, segmentation, and visualization. IEEE Access, 8, pp.191586-191601.

[20] Alsaffar, M., Alshammari, G., Alshammari, A., Aljaloud, S., Almurayziq, T.S., Hamad, A.A., Kumar, V. and Belay, A., 2021. Detection of tuberculosis disease using image processing technique. Mobile Information Systems, 2021, pp.1-7.

[21] Lopes, U.K. and Valiati, J.F., 2017. Pre-trained convolutional neural networks as feature extractors for tuberculosis detection. Computers in biology and medicine, 89, pp.135-143.

[22] da Costa Alves, R.R., de Assis Tavares, F.C.J., de Seixas, J.M. and Trajman, A., Introducing lifelong machine learning in the active tuberculosis classification through chest radiographs.

[23] Delon, J., Desolneux, A., Lisani, J.L. and Petro, A.B., 2006. A nonparametric approach for histogram segmentation. IEEE Transactions on image processing, 16(1), pp.253-261.
[24] The database is collected from this source: https://www.kaggle.com/datasets/tawsifurrahman/tuberculosis-tb-chest-xray-dataset
[25] Xavier, A.I., Villavicencio, C., Macrohon, J.J., Jeng, J.H. and Hsieh, J.G., 2022. Object Detection via Gradient-Based Mask R-CNN Using Machine Learning Algorithms. Machines, 10(5), p.340.Lopes, U.K. and Valiati, J.F., 2017. Pre-trained convolutional neural networks as feature extractors for tuberculosis detection. Computers in biology and medicine, 89, pp.135-143.
[26] Sahlol, A.T., Abd Elaziz, M., Tariq Jamal, A., Damaševičius, R. and Farouk Hassan, O., 2020. A novel method for detection of tuberculosis in chest radiographs using artificial ecosystem-based optimization of deep neural network features. Symmetry, 12(7), p.1146.
[27] https://support.echoview.com/WebHelp/Windows_And_Dialog_Boxes/Dialog_Boxes/Variable_Properties_Dialog_Box/Operator_Pages/GLCM_Texture_Features.htm#Energy to understand GLCM features.
[28] Iqbal, N., Mumtaz, R., Shafi, U. and Zaidi, S.M.H., 2021. Gray level co-occurrence matrix (GLCM) texture based crop classification using low altitude remote sensing platforms. PeerJ Computer Science, 7, p.e536.
[29] https://in.mathworks.com/help/images/texture-analysis-using-the-gray-level-co-occurrence-matrix-glcm.html
[30] Mohanaiah, P., Sathyanarayana, P. and GuruKumar, L., 2013. Image texture feature extraction using the GLCM approach. International journal of scientific and research publications, 3(5), pp.1-5.
[31] Gebejes, A. and Huertas, R., 2013. Texture characterization based on grey-level co-occurrence matrix. Databases, 9(10), pp.375-378.
[32] Hall-Beyer, M., 2017. GLCM texture: a tutorial v. 3.0 March 2017.
[33] Alkalai, M. and Sorge, V., 2013. A Histogram-Based Approach to Mathematical Line Segmentation. In Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications: 18th Iberoamerican Congress, CIARP 2013, Havana, Cuba, November 20-23, 2013, Proceedings, Part I 18 (pp. 447-455). Springer Berlin Heidelberg.
[34] PS, S.K. and Vs, D., 2016. Extraction of texture features using GLCM and shape features using connected regions. International journal of engineering and technology, 8(6), pp.2926-2930.
[35] Xavier Alphonse Inbaraj, Charlyn Villavicencio, Julio Jerison Macrohon, Jyh-Horng Jeng, JerGuang Hsieh. "A Novel Machine Learning Approach for Tuberculosis Segmentation and Prediction Using Chest-X-Ray (CXR) Images", Applied Sciences, 2021