

Comprehensive Classification Model for Diagnosing Multiple Disease Condition from Chest X-Ray

Savitha S K¹

Research Scholar,
Department of CSE
Visvesvaraya Technological University
Bangalore, India

N.C. Naveen²

Professor
Department of CS & E, JSSATE
Bangalore, India

Abstract—Classification plays a significant role in the diagnosis of any form of radiological images in the healthcare sector. After reviewing existing classification approaches carried out over chest radiographs, it was explored that existing techniques are highly restricted to perform binary classification that is not comprehensive for assisting in an effective diagnosis process of chest disease condition. This paper presents a novel approach to classifying chest x-rays on the basis of the practical disease condition. Harnessing the potential features of content-based image retrieval, the proposed system introduces a novel concept of attribute map that not only performs comprehensive classification but also makes the complete computational model extremely lightweight. The study outcome proved to offer better accuracy with the proposed non-iterative process in contrast to existing classifier design.

Keywords—Chest x-ray; classification; supervised learning; radiographs; accuracy

I. INTRODUCTION

There has been a significant improvement in the area of medical image processing with an evolution of various devices and forms of medical images [1][2]. Essential body parts that are found to be investigated most frequently in the majority of the hospital using medical image processing is the chest region. At present, there is a various form of advanced radiological techniques used for assessing the condition of the chest, e.g., Magnetic Resonance Imaging (MRI), Computed Tomography (CT), Positron Emission Tomography (PET), etc. [3]. Medical assessment using these processes are quite less frequent as they are used either to assess the advance critical condition of the chest. In any cases of diagnosis, a chest x-ray is the most frequently adopted diagnosis method [4] recommended by physicians. Various research papers have discussed challenges associated with diagnosing medical images, but very few are found to be associated with chest x-ray [5][6]. Over the period of time, certain researchers have dedicated themselves to discuss the existing challenges as well as problems associated with image processing techniques over chest radiographs [7][8][9]; however, there is no effective disclosure about effectivity in the approaches presented by researchers. The automated diagnosis processes (on the basis of research) are categorized into two problems, i.e., identification process and classifying the disease. There are various significant techniques for assisting in detection operation for chest radiographs [10]-[15]. The detection process is mainly dependent on

segmentation techniques [16] and feature-based aspects [17] while the classification techniques are more into involving machine learning approaches [18]-[20]. However, there are various associated problems of existing classification techniques that are found unspoken in any research papers. The *first* problem is associated with the classification objective itself. Almost all the researchers have performed binary classification problems that are quite essential but only in preliminary level. For example, if the queried image is found to be malignant in binary classification process than it should further show the stage of cancer or exhibit more comprehensive information. Disease conditions of lungs are quite lethal, and it requires precise diagnosis, whereas all the existing researchers have only presented the solution to the just preliminary stage. The *second* problem associated with any of the existing approaches is related to the cardinality of the outcomes. 99% of the research work uses a trained dataset to be compared with the test image with the aid of features-based method. The outcome is the only one trained image to show the similarity match with the test image. This reduces the scope of inferring the outcome of classification and this lead the researchers to highly depend on the flow of the techniques used for database building as well as for classification. Adoption of such a process is evaluated only with respect to the empirical approach adopted, or the feature-based approached utilized and possibly it does so by overlooking the actual clinical context of it. For example, a chest x-ray of pleural effusion, as well as a chest x-ray for any pulmonary edema within the lungs, is visibly the same. There are good chances that if existing classification approaches are applied that it could not be able to identify the difference between them. It should be noted that pleural effusion, as well as pulmonary edema, are two different disease condition of the chest which has to be diagnosed discretely.

Hence, the proposed paper discusses a novel technique of classification that is capable of addressing the problem discussed above. The organization of the present manuscript is as follow: Section II presents a discussion of existing research approaches towards classification, Section III highlights the identified research problems addressed in this paper, Section IV briefs about research methodology undertaken, Section V discusses algorithms implemented for accomplishing the goal of this paper. Finally, result analysis is discussed in Section VI, and summary of contribution is briefed in Section VII in the form of a conclusion.

II. RELATED WORK

There are various literatures contributing towards unique techniques of classification in medical images. This section updates more research work as an extension to our prior review work [21].

Classification performance can be optimized if a closer connection is set with the identification technique. One such attempt was found to be seen in the work of Katre and Thakare [22]. However, an in-depth study shows that the study is more focused on implementing *the watershed algorithm* for segmentation as well as it addresses the basic noise removal process. The overall claim of detection and classification of chest x-ray is not seen in this work, and hence the work just offers elementary theoretical guidelines. The work of Mukherjee et al. [23] has also claimed combinably addressing detection as well as classification problems in lung images. The authors have mechanized a unique bilateral filtering approach for eliminating the noise followed by segmentation and thresholding process. Finally, a binary classification technique was applied over the extracted features (e.g., eccentricity, circularity, and aspect ratio). The study offers good detection implementation; however, the binary classification process is not a novel or unique approach. Apart from this, the applicability of approach towards multiple disease condition of the chest is very unlikely. Research towards identification problems has been carried out by Bar et al. [24] where *convolution neural network* has been used along with binary supervised classification technique, which is quite a common approach adopted by frequent researchers. Unfortunately, binary classification is only meant for critical disease condition of the chest and doesn't relate to other potential disease condition.

Early research work has also focused on using fuzzy logic for the purpose of classification of the lung nodules. Roy et al. [25] have used contour-based model for accomplishing segmentation. Hence, the study was more inclined towards detection and less toward classification. Similar research flow was also adopted by Zhang et al. [26]. Zhang et al. [27] have presented a classification process that uses rank for classifying. Researchers have also presented a mechanism for unique binary classification process that is meant for assists in further classification process but not directly for disease diagnosis (Xue et al.[28]). A simplified *rule-based method* was reported to assist in classification as reported in work of Taher et al. [29]. It was also seen that the usage of the Bayesian classifier as reported in the work of Dhaware and Pise [30] is found to improve identification of lung cancer. *Fuzzy-c means* clustering was applied for improving the detection as well as for addressing directionality reduction. However, the study lacks extensive evaluation to prove its robustness against different disease condition of lungs. One of the early initiatives for classification process has been initiated by Kulkarni and Panditrao [31] where the authors have signified the essential contribution of medical image processing for diagnosis of lung cancer and importance of classification. The neural network is another frequently used approach that offers the simple classification approach along with feature extraction. However, the outcomes are not subjected to comparative evaluation for assessing the outcomes practicality. A block-based approach

has been used by Ahmad et al. [32] where the classification has been carried out by *the Bayes classifier*. The target of classification was mainly to identify the pleural effusion in chest x-ray to show 100% accuracy with Gabor filter. However, the study can only classify any case that comes under the pleural effusion condition of the chest x-ray. Learning approach plays a significant role in the classification of chest x-rays. A study conducted by Dong et al. [33] have shown that it is feasible to train the multi-class model for assisting in further classification using a *convolution neural network*. The study has carried out analysis on the dataset of different disease condition using standard accuracy parameters; however, the technique adopted was quite iterative and less challenging datasets. Similar usage of convolution neural network was also proven to assists in classifying lung nodule as claimed by Liu et al. [34]. The study outcome shows the comprehensive benefits of performing classification operation; however, the classification rate is slightly diminished compared to another existing system.

Investigation towards classifying lung nodule was also reported in the work of Mao and Deng [35] where a *pattern-based approach* was used for performing representation of effective features. The study uses a linear classifier, e.g., supports vector machine with the impressive rate of classification outcomes. However, there is no evidence if it can be used for classifying other disease condition of a chest. Usage of deep learning algorithm can also improve the representation performance of features. Literature also claimed of using an encoding technique of local binary pattern for classification. [36] The work of Mao et al. [37] has used deep learning approach to perform classification of the lung nodule. The outcome shows better performance in contrast to other conventional classifiers. Work carried out by Kumar et al. [38] has presented the discussion of usage of *deep features* for assisting in solving classification problems of chest radiographs. Harnessing in the potential of deep learning, the author discussed that it is an effective classifier for binary classification of lung cancer stage. Unfortunately, deep learning is the quite advanced version of the machine learning approach using a neural network, and it is capable of doing more than binary classification, which was not found to be attempted. Deep learning and convolution neural network were also reported to improve the classification performance. This fact was discussed by Li et al. [39] where *region-of-interest* was used for performing the classification process. However, this complexity of analysis by the computational model was not checked owing to the usage of the region of interest based approach with almost information defined for critical classification. Wang et al. [40] have discussed the impact of database towards benchmarking problems in the classification of chest x-rays. Apart from this, there are also reported usages of Principal Component Analysis (PCA) for solving classification problems in chest x-ray [41]-[45]. The next section briefs the research problems associated with the existing research techniques of classification.

III. RESEARCH PROBLEM

From the prior section, it can be seen that there are various existing techniques of classification in medical images especially considering chest x-ray. However, are various

significant problems have been found unaddressed. This section briefs about such forms of unaddressed research problems as follows:

- *The gap between detection and classification problems:* It has been observed that detection, as well as classification, has always been treated as separate problems in an existing system. The integrated study of connection of detection and classification with respect to accuracy has been less emphasized.
- *Existing Classification Uncertainty:* There are various research works towards classification where the outcomes show significant accuracy. However, it should be known that such accuracy is claimed on the basis of the single instance of a match with the database or trained images. The classifications are numerically proven correct, but there was no explicit discussion of accuracy with respect to disease-based classification.
- *Unprepared Dataset:* The existing publically used dataset doesn't have any well-structured categorization of the chest x-ray on the basis of any disease. However, there are disease-specific dataset too [46]-[49].
- *Untapped Potential of Features:* Features are one of the significant information set to represent the complete dataset. Various empirical attributes within the features have been less explored for hence granularity of the classification process has never been testified in this regards.

Therefore, the proposed study targets to address the entire above problem in order to evolve up with a novel and robust classification process of the chest x-ray.

IV. RESEARCH METHODOLOGY

The proposed study is a continuation of our prior research work [50]. By adopting an analytical research methodology, the proposed study presents a novel classification process of chest x-ray very differently than existing approaches. In order to address the *first* research problem, the proposed study introduces a novel attribute map-based method for assisting in the classification process. This mechanism is used in both detection and classification problems. The *second* research problem is addressed by applying content-based image retrieval technique where classified images have multiple options to be double checked for its accurate prediction. The *third* research problem is addressed by synthetically developing a well-structured dataset of trained images. The training is carried out using attribute maps and a supervised learning technique. The *fourth* problem is addressed by completely developing the model using attribute and not features. The difference between an attribute and features is that features are a subset of information for attribute and hence working with attribute offer more granularities in system design as well as assist in assessing the classification performance. Various empirical variables associated with the image-related pixel and numerical information forms an attribute. Fig.1 highlights that complete model consists of i) training phase and ii) classification phase. The training phase consists of applying enhanced supervised learning algorithm over the disease-specific dataset. The classification phase consists of applying a series of algorithms in order to obtain classified images. For better study accuracy, the proposed model contributes to offering better inference in its outcome by incorporating ranks with the outcome of the similar set of images and thereby redefining the conventional process of classification of a chest x-ray. The next section illustrates about the algorithm implemented for the proposed system.

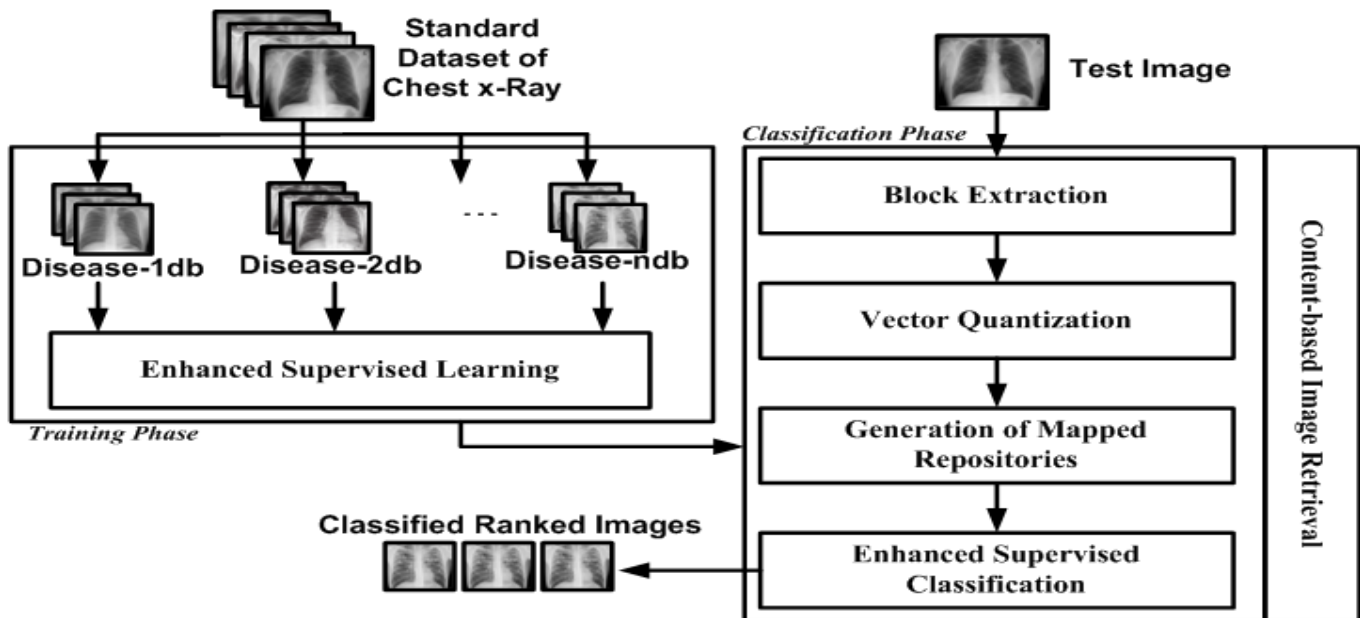


Fig. 1. Schematic Architecture of Proposed System.

V. ALGORITHM IMPLEMENTATION

This section discusses the series of algorithms that have been constructed in order to perform classification of the chest radiographs. There are various inclusions of enhanced machine learning approach as well as content-based image retrieval approach implemented in order to perform classification of the chest x-ray. The discussions of the implemented algorithms are as follows:

A. Enhanced Supervised Learning

This is the first algorithm that is responsible for generating a trained dataset of a chest x-ray. The complete training is carried out using enhanced supervised learning approach (as Shown in Fig.2). From the application viewpoint, this algorithm can be considered to be configured on various terminals of the healthcare facility that takes the feed of different forms of chest x-rays. All the input images are not directly stored in physical/cloud storage units, but they are trained, categorized, and arranged in a highly structured manner.

According to the above scenario, all the input chest x-rays are forwarded to the terminals by all respective radiologists. However, they are all mixed and will be needed to be well categories. The initial categorization is carried out manually as only these images will be later treated as a reference image. Hence, it is considered that such images are error-free and involved all forms of information that are required to be used by the physician to perform an effective classification process during the query. It is also considered that each category represents a specific form of cancer or certain abnormalities of the chest x-ray. The prime target is that while performing a query, it is essential that the queried image be subjected to classification with a higher degree of similarity with a maximum number of images, unlike conventional approaches of disease classification. Therefore, the process of generating trained data is essentially important to be effectively maintained. The algorithmic steps of the supervised learning are as followed:

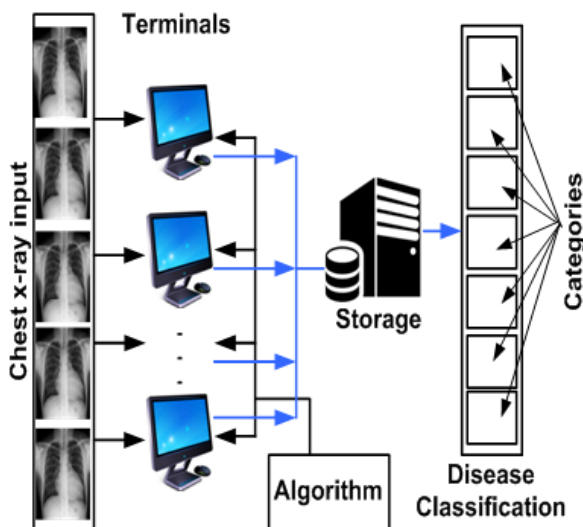


Fig. 2. Considered Scenario of Algorithm Implementation.

Algorithm for Enhanced Supervised Learning

Input: I, d, n

Output: t_d

Start

1. $I = \{d(I_{n1}), d(I_{n2}), \dots, d(I_{nn})\}$

2. **For** $i=1: n$

3. $a_{\text{map}} \rightarrow f_1(d(I_i))$

4. $c_{\text{class}} \rightarrow f_2(I, a_{\text{map}})$

5. $c_{\text{mat}} \rightarrow f_3(c_{\text{class}}, I)$

6. **End**

7. $t_d = \{a_{\text{map}}, c_{\text{class}}, c_{\text{mat}}\}$

End

The illustration of the lines of the above algorithms is as follow: The algorithm takes the input I (chest x-ray), d (dataset), and n (number of the dataset) that after processing yields an outcome of t_d (trained data). The algorithm doesn't take a single image as input, but it takes the complete folder with category mentioned for the pre-defined classified disease as the input. This design principle also offers the advantage of investigating by considering various types of dataset d towards the classification process. The complete dataset of an image I is formed by considering each individual dataset d with different types, formats, and the number of images within it. Consider that dataset $d(I_{n1})$ consists of n_1 number of images, dataset $d(I_{n2})$ consists of n_2 number of images, and so on (Line-1). Considering all the n number of images of chest x-ray to be trained (Line-2), the first processing carried out by the proposed system is to extract an attribute map a_{map} (Line-3). This extraction process is carried out by an explicit function $f_1(x)$ that returns a matrix of all the visual-based attributes. As it is a matrix-based operation, therefore, the input $d(I_i)$ is required to be in an array form of the chest x-ray image. Owing to the matrix-based form of an input data set, it is feasible to extract the potential attributes from all the digitized chest x-ray irrespective of any shape and size. The algorithm specifies the location of an image (I_{loc}) by defining the aggregated images of chest x-rays. The study represents I_{loc} in the form of the string that represents an array of input images. The algorithm does this operation recursively for all the input chest x-ray images. It then generates a set of the map against all the significant visual attributes of an input image. The process of execution of $f_1(x)$ is carried out in three different steps, i.e., identification of attributes, describing the attributes, and generation of the final matrix of reference attribute. The proposed study considers high-level extraction of attributes in its first stage followed by abstraction of all the images by individual blocks as a step to describe the attributes. The mechanism of representing attribute is all about the mechanism of representing this block as the numerical vector. The final process leads to a generation of referential attribute matrix that maps the information about the similar form of blocks. The next step in algorithm implementation is to performing training over the targeted classifier in order to obtain better results in the trained dataset. For this purpose, a different function $f_2(x)$ is constructed (Line-

4) that takes the input of the input images as well as generated attribute map (Line-4). The prominent potential of this function $f_2(x)$ is that it constructs a specific category for the different set of chest x-ray images by harnessing the potential of machine learning and statistical approach. All the computed elements of I will now be represented by a specific set of categories. Further, this process is optimized using a supervised learning approach of Support Vector Machine in order to perform the more enhanced form of classification and further results in category class matrix c_{class} (Line-4). The final step of this algorithm is to carry out an assessment using third explicit function $f_3(x)$ considering input arguments of category class c_{class} and input images I (Line-5). This also leads to a generation of an error matrix that can be used for further analysis. All the outcomes, i.e., attribute map a_{map} , category classifier c_{class} , and category matrix c_{mat} are stored in the form of trained data t_d (Line-7). This completes the algorithm operation, and now the system is ready to perform classification with the aid of a test image.

B. Block Extraction

This algorithm is responsible for generating an image block from the input of the test image, where each block is the direct representation of a localized form of the chest x-ray. The blocks also represent some of the essential information pertaining to image maps.

According to the Fig.3, the outcome of the block extraction process is basically an attribute matrix, which contributes towards the classification process. Each block represents the overall test image in multiple form and dimension and is obtained by specifying a particular size of cells in the form of a number of rows nr and columns nc . A statistical operation is further applied in this algorithm to ensure that the outcome is least affected by any form of variability factors (e.g., illumination, brightness, contrast, etc.). The steps of the algorithm are discussed as follows:

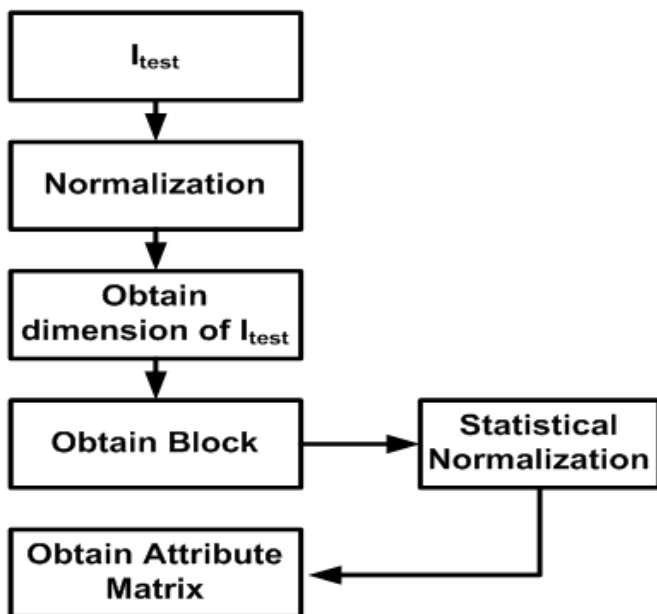


Fig. 3. Proposed Block Extraction.

Algorithm for block extraction

Input: I, b

Output: B

Start

1. init b
2. $I = g(I)$
3. **For** $i=1:b:nr_b$
4. **For** $j=1:b:nc_b$
5. $B \rightarrow I(i:i_b, j:j_b)$
6. $B = \text{apply norm}(B)$
7. $A \rightarrow [A, B]$
8. **End**

9. **End**

End

The algorithm takes the input of I (queried image) and b (block size) in order to generate the blocks. After initializing with the number of block size b (Line-1), the algorithm performs slight fine-tuning of an image using a function $g(x)$ in order to offer better precise during the classification process. The fine-tuning is carried out by resizing the image followed by increasing the image to double precision (Line-2). The size of the test image is obtained to generate a matrix of dimension $nr \times nc$; however, for blocking operation, this size will be required to be changed. Hence, a modified size of all rows nr_b and columns nc_b is obtained as $nr-(b-1)$ and $nc-(b-1)$ respectively. The abstract representation of this is nr_b and nc_b (Line-3 and 4) is used in order to construct a new block B (Line-5). Although the steps till now assist in constructing a new block, its outcome will be affected if the original test image is subjected to a different form of events that results in significant variance to the constructed blocks. This could also reduce the identification as well as classification performance. This problem is mitigated by applying a standard deviation to the obtained block B followed by further normalization where B is subtracted from the mean value of the blocks. The resultant is further divided by the obtained value of standard deviation that leads to the generation of final block B (Line-6). These extracted blocks also act as the specific attribute of high level for assisting in the classification process.

C. Vector Quantization

After the blocks have been extracted in the prior algorithm, now the system is nearly ready to perform classification. The Fig.4 represents the flow of vector quantization. However, the information obtained from the attribute matrix is so massive that it requires performing the certain form of clustering operation. The proposed system applies vector quantization in order to carry out clustering operation. It was seen in the prior algorithm implementation step that vector is transformed into blocks after the process of generation of attribute map that finally results in forming a map repository (refer next algorithm). The proposed system applies vector quantization as a process of clustering that segregates all the vectors into

different groups on the basis of their attributes. The center of cluster obtained from this assists in forming the elements of mapped repositories.

According to the above flow of the vector quantization technique in clustering, the proposed system takes the input in the form of cluster definition in terms of a number followed by computing the center of cluster and computing spatial distance of all objects to the center of clusters. Using the criteria of the reduced distance scale, the system performs clustering. This results in obtaining a faster clustering operation. The steps of the algorithm are discussed as follow:

Algorithm for Vector Quantization

Input: A (attribute)

Output: C_{id} (center of cluster)

Start

1. C_{id} → vect_quan (A, M)
 2. Function vect_quan
 3. (s₁, s₂, ..., s_m) ← arb((x₁, x₂, ..., x_N)M)
 4. **For** m ← 1 to M
 5. **Do** μ_m ← s_m
 6. **While** the stopping criterion has not been met
 7. **Do for** m ← 1 to M
 8. **Do** ω_m ← {}
 9. **For** n ← 1 to N
 10. **Do** j ← arg_{min} ||μ_j - x_n||
 11. ω_j ← ω_j ∪ {x_n} //reassignment of vectors
 12. **For** m ← 1 to M
 13. **do** μ_m = $\frac{1}{|\omega_m|} \sum_{x \in \omega_m} x$
 14. return {μ₁, ..., μ_m}
 15. **End**
 16. **End**
 17. **End**
- End**

Owing to the generation of a massive set of information, the Euclidean distance among the data points will actually increase and will be highly uncertain. Therefore, the prime motive of the presented vector quantization technique is to reduce the mean of the squared distance of all the data points with respect to the center point of the cluster. The algorithm computes the center of cluster μ of all images corresponding to ω cluster. The computation of the center of the cluster is carried out considering a function called as vector quantization (Line-1). The input to the algorithm is an attribute matrix A and limit the size of M for clustering. The algorithm formulates seeds by capturing different data points of order M and selects them randomly (Line-3). For all the specified number of clusters M (Line-4), the algorithm assigns respective seeds to the center of clusters μ (Line-5). In case the criteria of terminating the iteration have not been met than the algorithm attempts for minimization of the Euclidean space (Line-9-11) by performing reassignment of the vectors as well as re-computation of the center-of-clusters (Line-13). The outcome of the algorithm is finally a center of cluster C_{id} that will be used as an input to the next algorithm for generating mapped repositories.

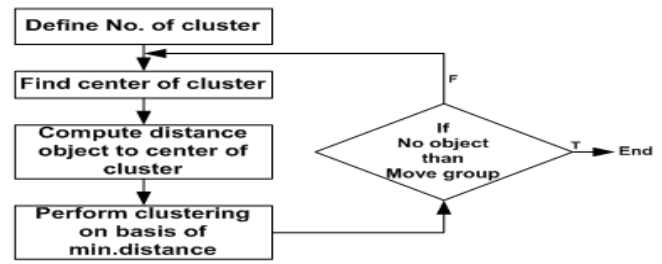


Fig. 4. Vector Quantization.

D. Generating Mapped Repositories

This algorithm is responsible for generating mapped repositories, which is a collection of all significant attributes from the images. The operations are two-sided. The first side of the operation is associated by identifying significant attributes followed by storing all this mapped information in one matrix called as map repository. This repository generates a perfect representation of an image. This repository will also be used by the test image in order to perform classification of the disease categories corresponding to the unique dataset of chest x-ray (refer Fig.5).

Algorithm for Generating Mapped Repositories

Input: C_{id}

Output: MR

Start

1. MR → mat (axb)
2. **For** i=1:p: M
3. **For** j=1:p:M-1
4. a_{vec} → C_{id}(q)
5. q=q+1
6. a_{vec} → δ(a_{vec})
7. MR ← a_{vec}
8. **End**
9. **End**

End

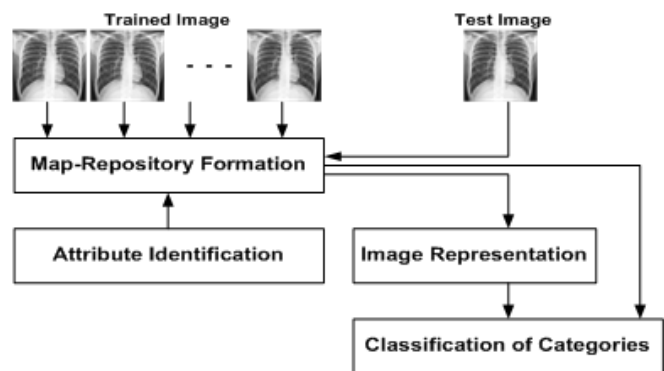


Fig. 5. Process of Map Repositories.

The algorithm begins by developing a matrix of size $a \times b$ and it takes the center of cluster C_{id} as an input (Line-1). An effective center of cluster C_{id} is considered by following empirical expression,

$$C_{id} \rightarrow C_{id-\text{argmin}}(C_{id}) \quad (1)$$

The above empirical expression (1) represents the displacement of the scope of C_{id} from 0 to a positive value. The complete process of generation of the map repositories MR is carried out for M number of clusters. For all the maximum value of the clusters (Line-2 and 3), the attribute vector is obtained for the center of clusters corresponding to its specific index value (index q) (Line-4). The index is incremented (Line-5) further followed by reshaping δ of the attribute vector (Line-6). Finally, the obtained result of attribute vector a_{vec} is stored back to a matrix MR (Line-7).

E. Enhanced Supervised Classification

The prime purpose of this algorithm is to perform disease classification for the test image of the chest x-ray. The operation is meant for forecasting the relevance of the input image of a chest x-ray (test image) with the most relevant trained dataset.

As the proposed concept is designed on the basis of content-based image retrieval process, therefore, the outcome of the classification will be more comprehensive, unlike the existing system. The proposed system uses a support vector machine for this purpose in order to perform classification operation. One of the significant enhancements of supervised technique applied in this algorithm is that it offers a significant control of iteration towards performing forecasting for a given set of data. The forecasting process carried out for the proposed system is done only once for one image, and that makes its non-recursive learning algorithm, unlike the existing applications of using a supervised technique that is highly recursive in order. Another significant contribution of the proposed algorithm is that it offers a ranking mechanism to exhibit the classified images as an outcome. The complete process of enhanced supervised classification is exhibited in Fig.6.

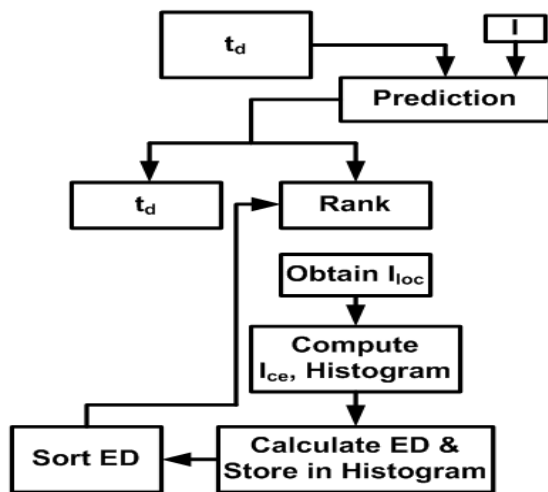


Fig. 6. Process of Enhanced Supervised Classification.

Fig.6 showcases the complete process of applying the predictive-based approach in order to obtain highly precise classification process. A similar illustration is given in the form of algorithmic steps as following:

Algorithm for Enhanced Supervised Classification

Input: t_d, I

Output: I_{type}, I_{ranked}

Start

1. load t_d
2. $[I_q, \text{rank}] \leftarrow \rho(c_{\text{class}}, I)$
3. $I_{type} \leftarrow c_{\text{class}}(I_q)$
4. $\text{svm}_{op} \leftarrow I_q$
5. flag I_{type} as an identified category of trained data
6. $I_{loc} \rightarrow f(\text{svm}_{op})$
7. $[H_o \text{ bins}] \rightarrow \tau(I_1)$
8. **For** $i=1:\text{size}(I_{loc})$
9. $I_t \rightarrow [I=(i)]$
10. Compute $E_{\text{dist}}(H, H_o)$
11. $ED_{\text{vec}} \rightarrow [ED]$
12. $I_{ranked} \rightarrow \text{show ranked images} (\text{sort}(ED_{\text{vec}}))$
13. **End**

End

The algorithm takes the input of t_d (trained data) and I (queried image) that after processing leads to the generation of I_{type} (Classified image) and I_{ranked} (sorted images). This algorithm has dependency over the trained data t_d (Line-1) that will be used mainly for checking the relevancy between the queried image and trained images. A function $\rho(x)$ is constructed to carry out a predictive operation using input arguments of category class c_{class} and test image I (Line-2). The process results in label I_q and rank (Line-2) where the labels stored back in an explicit matrix svm_{op} (Line-4). Upon this step of implementation, the algorithm starts exhibiting the nearest and similar dataset that corresponds to the queried image. The next process involved in the algorithm is to obtain the information related to the location of an image I_{loc} (Line-6). A new function τ is applied over resized image I_1 (Line-7) in order to obtain histograms H_o as well as a bin. The algorithm also obtains another form of histogram H that is applied over the trained dataset. An interesting point to observe is that this step of histogram calculation is not carried out in database construction stage (first algorithm). Therefore, for all the sizes of image locations (Line-9), the algorithm applies Euclidean distance formula considering histogram from queried image H_o and that from trained image H . Finally, correlation is obtained between the queried image I_1 and trained image I_t and the value is stored back in matrix E_{dist} (Line-10). After obtained vectors of distance (Line-11), the sorting is carried out using Euclidean distance ED_{vec} to complement the ranking process. The accomplishment of this final step of algorithm results in the

display of relevant images that are ranked based on their relevancy level with the test image.

VI. RESULT ANALYSIS

As the proposed system targets to offer a classification of the chest x-ray; therefore, the performance parameters selected in this process has to be closely related to the accuracy of the classification process. However, unlike any existing classification process where there is the only a single outcome of the classified image, the outcome of the proposed system will result in exhibiting multiple relevant images from the dataset. The outcome also comes with an identification of a relevant dataset that represents multiple disease categories. This section discusses the result analysis with respect to the assessment and implementation strategy.

A. Database Considered

All the images for assessing the algorithm implementation is carried out considering the Japanese Society Radiological Technology (JSRT) dataset which hosts various forms and modalities of chest x-rays [51]. The average size of all the grey-scaled chest x-rays is approximately 8192 kilobytes, and it bears 12 bit of color. However, one of the biggest hurdles in using this database is that it has combined collection of all images together whereas the proposed system demands a highly structured categorized database as per the disease condition. Therefore, after obtaining the complete database, it was subjected to further manual categorization as per critical disease condition of the lungs. The proposed system considers 7 disease conditions of the lungs e.g. i) Chronic obstructive pulmonary disease [52], ii) Emphysema [53], iii) Bronchitis [53], iv) Cystic Fibrosis [51], v) Pleural Effusion [51], vi) Tuberculosis [54], and vii) Adenocarcinoma [55]. All the above mentioned disease conditions are most frequently encountered and the present database has possession of chest x-ray corresponding to this disease condition.

B. Visual Outcomes

Implemented on MATLAB, the visual outcomes of the proposed study was feasible to be exhibited properly. Fig.7 highlights the input image that is subjected to blocking operation (Algorithm-B) to obtain the blocked image as shown in Fig.8. The blocks are then subjected to a consecutive algorithm for obtaining the attribute space (Fig.9). Finally, the center of the cluster is obtained (Fig.10) while the map repository is constructed (Fig.11).



Fig. 7. Input Image.

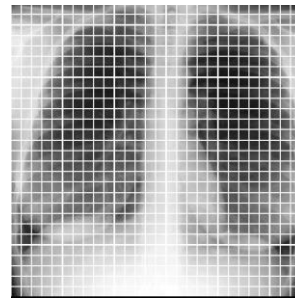


Fig. 8. Image with Blocks.



Fig. 9. Extracted Attribute Space.

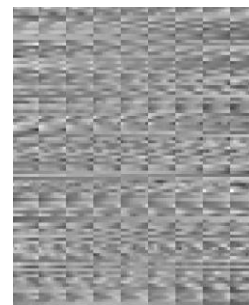


Fig. 10. Center of Cluster.

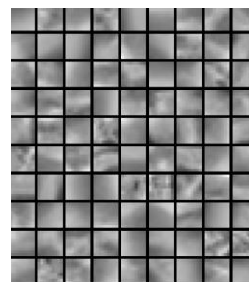


Fig. 11. Map Repository.

For final assessment of proposed classification using content-based image retrieval technique, the proposed system subjects multiple forms of test images to find the visual match of the most relevant images from the disease-specific dataset. One interesting part to observe is that the proposed system encapsulates complete information about the prominent attributes within a map repository, thereby making the complete process more lightweight when the classification is performed. The significant contribution are mainly two fold viz. i) the massive information of the large number of images in the dataset can be now stored in the form of map repository, which is comparatively very much smaller in size in contrast to total size of all images considered for training, Although, the map-repository shown in Fig.12 belongs to a single test image, but the same process can also be applied over the trained attributes of the images. This process makes the classification quite faster and highly percussive. Fig.13 highlights some of the sample outcomes of the classification process.

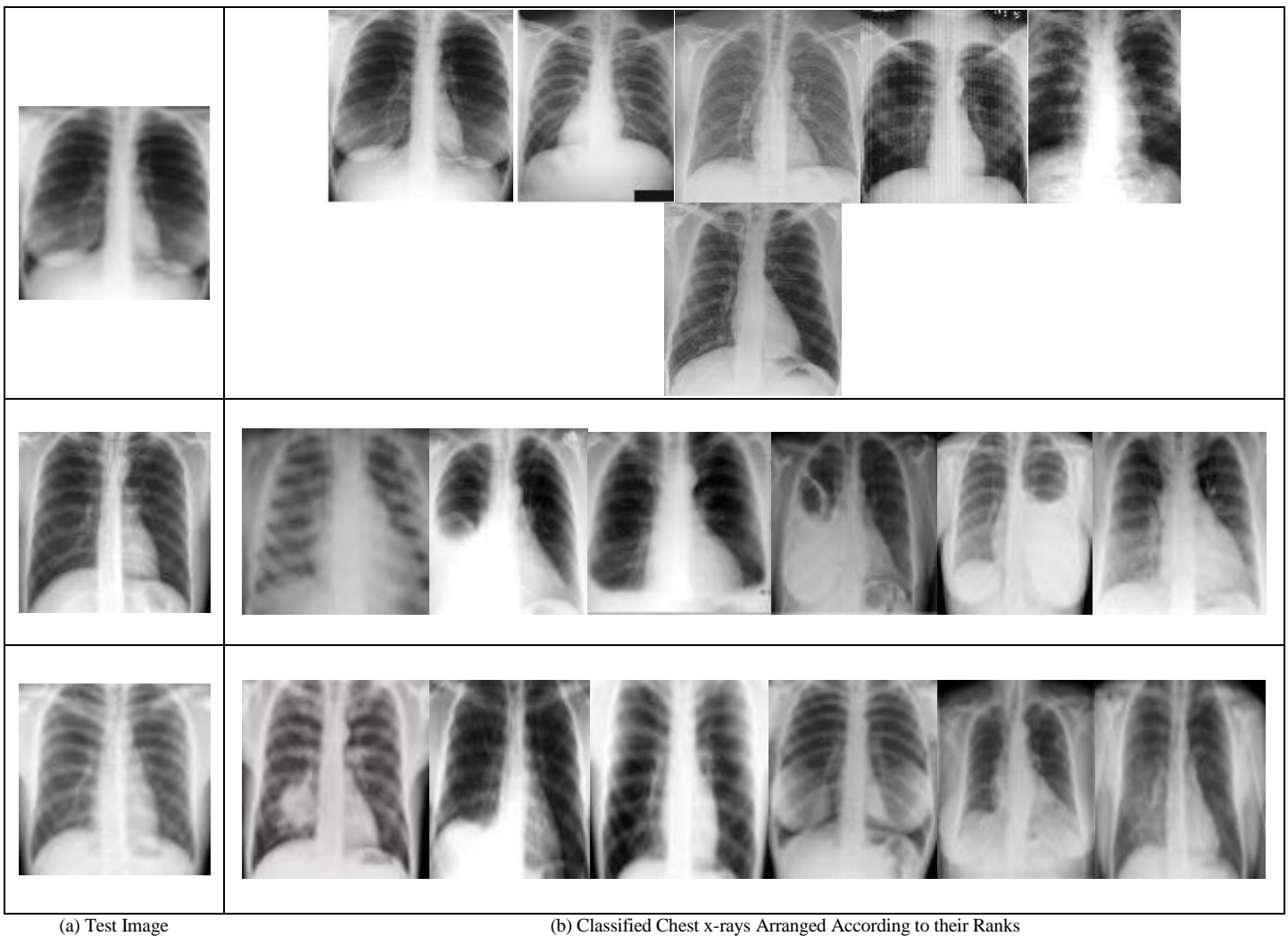


Fig. 12. Outcomes of Classification.

Exhibition of all visual outcomes is out of the scope of this paper. The novelty/contribution of this visual outcomes are again multi-fold viz. i) existing algorithms discussed in literature mainly classify one medical image (test image) with multiple trained images, but ultimately only one trained images are considered as the most matched outcome. Although such outcomes of the existing system may be technically correct from an image processing viewpoint, they could be clinically wrong. This can be illustrated in following-usually, there are good possibilities that chest x-ray of a subject suffering from two different diseases bears similar visual appearance, but clinically they are wrong. The root cause of this problem is the existing database system which is not arranged or organized with respect to multiple disease condition of a chest x-ray. Hence, the proposed system performs the classification of chest x-rays corresponding to some critical disease condition of the chest and performs the visual analysis on that basis. The outcome shown in Fig.13 only highlights that irrespective of any form of test images, the proposed system could successfully perform an effective classification of respective disease, ii) another significant benefit of this visual outcome is that a physician is offered with multiple referential images from the trained dataset that matches with the queried image.

This phenomenon allows the physician to correlate the disease condition with more helping referential / clinical files associated with the extracted ranked classified images. Hence chances of true and reliable diagnosis of a disease condition are highly made possible using content-based image retrieval system implied on the classification technique. Another significant advantage is iii) proposed system offers highly precise display of the classified image irrespective of different forms, shape, orientation, and format. Hence, the proposed system offers true test-bed to perform diagnosis of critical chest disease.

C. Numerical Outcomes

The numerical assessment of outcomes obtained from the proposed system has been assessed with sensitivity and precision as core performance parameters. Although, there are many numbers of images within the JSRT dataset, the discussion of the outcomes is carried out using a small number of images in order to practically understand the numerical parameters, e.g., attributes are represented in Table.1. The assessment has been carried out considering 12 images each in a different dataset.

TABLE I. NUMERICAL INFORMATION OBTAINED FROM A DATASET

A dataset with disease Name	Number of images	Number of extracted attributes
COPD	12	858336
Emphysema	12	228108
Bronchitis	12	434380
Cystic Fibrosis	12	1072636
Pleural Effusion	12	617056
Tuberculosis	12	342112
Adenocarcinoma	12	273416

A simple table like this one above with only 12 images shows that it could have a massive number of attributes. This is important as the proposed system performs enhancement of the supervised learning approach using an attribute map. The analysis has exhibited that there are approximately 80% of the potential attributes that can be utilized with each image of the chest x-ray. The analysis also performs balancing of the attributes across all the chest x-rays in order to enhance the performance of clustering using vector quantization. For an effective analysis of the proposed system, the outcomes have been compared with another frequently used classifier, i.e., neural network and principal component analysis.

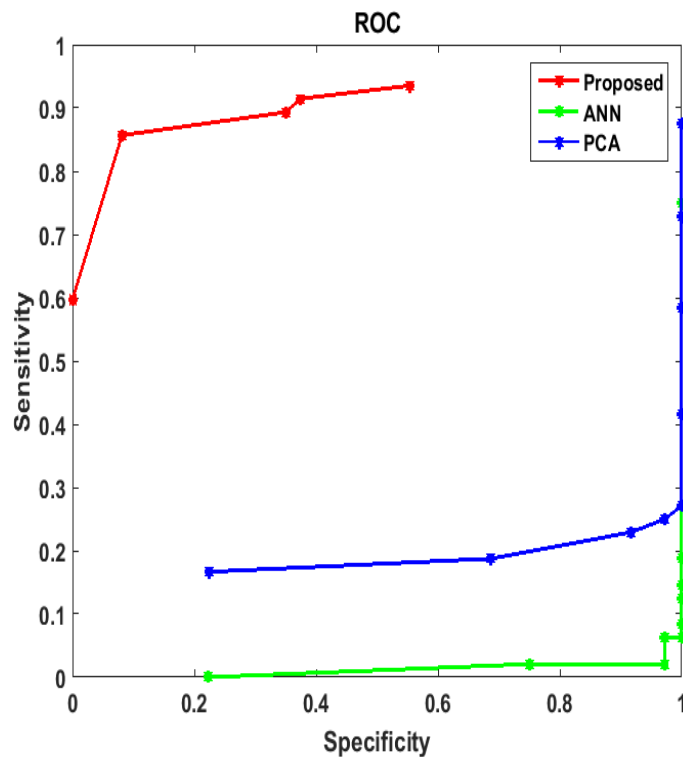


Fig. 13. Comparative Analysis of Sensitivity Vs. Specificity.

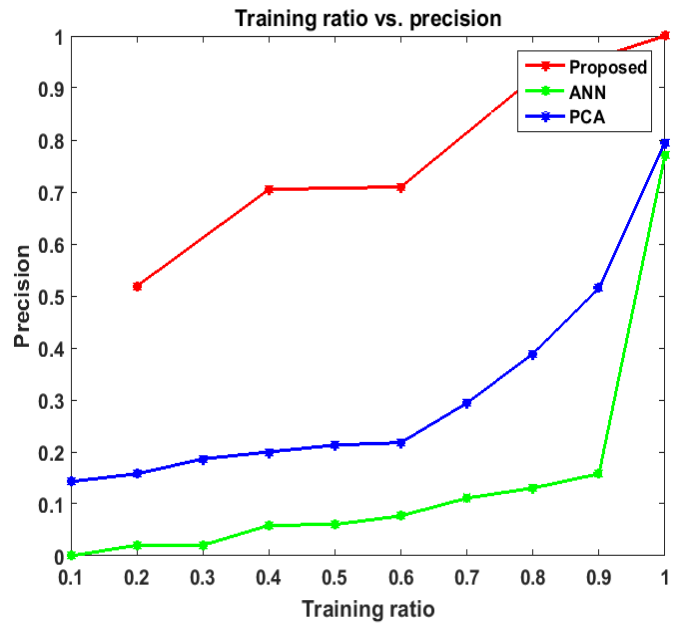


Fig. 14. Comparative Analysis of Precision Vs. Training Ratio.

A closer look at both the outcomes of Fig.13 and Fig.14 shows that the proposed system offers good accuracy performance as compared to existing classified design using machine learning (ANN) and statistical approach (PCA). The prime reasons behind this are multifold-the filtration process of ANN to obtain the elite outcome is highly accurate, but on the process of training, it cancels out the prior outcomes that it believes will adversely affect the faster convergence performance. Apart from this, the training process of ANN is quite iterative that could offer significant computational complexity while performing training. Similarly, usage of PCA is more emphasized on exploring orthogonal projections associated with the images of chest x-rays. It also looks for images with a higher degree of variance values. The good part of using PCA is to understand the correlation of the data within the same diseases as the data points are linearly correlated. However, a problem starts to surface when the PCA algorithm needs to read another dataset that corresponds to the different disease. Hence, while building / training dataset, PCA just filters out the data point with less variance value thereby leading to dropping of some good cases of similarity with the test data. Hence, it doesn't perform better than the proposed system. However, it performs better than ANN only for the reason that it is less iterative and it retains the potential amount of information associated with attributes for each dataset, which is not carried out by the neural network. The prime reason behind the outcomes of the proposed system is the mechanism of constructing attribute map that offers an efficient collection of all potential attributes specific to each disease for which reason classification can be carried out effectively. Analysis of computational complexity has been carried out with respect to algorithm processing time. Following are the outcomes.

The analysis has been carried out on the core i3 processor with all the trained images in a dataset where the algorithm processing time has been programmatically executed. A closer

look at the outcome (Table 2) shows that the processing time of the proposed system is slightly lower than a neural network based approach and slightly higher than principal component analysis. Although, with an increase of more images will definitely increase the processing time, but they are found to be within the tolerable limits. However, memory and processor is widely available in existing healthcare industry but what is required is a device with the capability of precise diagnosis of disease condition that can be effectively done by the proposed system in contrast to the existing system.

TABLE II. COMPARATIVE ANALYSIS OF ALGORITHM PROCESSING TIME

Algorithms	Training Time	Classification time
Neural Network	6.29388 sec	0.14 sec
Principal Component Analysis	7.89224 sec	0.07 sec
Proposed	9.29331 sec	0.11 sec

VII. CONCLUSION

This paper has presented a unique approach to classifying the disease condition of chest x-ray using non-conventional and out of mainstream research methodologies. Review of existing approaches shows that binary classifier design assists in foretelling the preliminary stage of disease condition of the lung, but it is not sufficient for diagnosis. For an effective diagnosis, it is essential to extract more clinical information from the chest x-ray. Therefore, this paper presents one such solution to address this problem and introduce a new avenue of classification algorithms concerning chest x-rays. The significant contributions of the proposed study are as follows viz. i) unlike existing classifiers, the proposed classifier can perform comprehensive classification to offer maximum information associated with the test image. When the test image is found to be matched with multiple numbers of images from the trained dataset, more information from the trained images can be used to relate the actual disease condition. ii) one significant contribution of the proposed technique is that it doesn't use any of the conventional classifier found in literature e.g. neural network, deep learning, convolution neural network, fuzzy logic, etc. Unlike these existing algorithms, the proposed algorithm offers non-iterative operations, and hence they are likely to be more practically compatible with existing mobile healthcare devices. iii) The novel introduction to attribute map and map repository offers a highly compact mechanism of converging the information associated with attributes, which is a very good alternative of dimensional reduction concept (e.g., PCA used in existing times) with better accuracy performance. Hence, the applicability of the proposed study is quite more and comprehensive as it addresses some of the practical requirement of classification of a chest x-ray. The limitation of the paper is that it only deals with the classification time while other parameters like accuracy, PSNR and MSE were not considered.

The paper can be considered in future for other medical imaging like brain MRI, Mammogram classification for disease analysis. Further, the paper can be implemented as base paper to analyze other performance parameters like accuracy, PSNR, MSE, etc.

REFERENCES

- [1] R. Suganya, S. Rajaram, A. Sheik Abdullah, Big Data in Medical Image Processing, CRC Press, 2018
- [2] Nilanjan Dey, Amira Ashour, Fuquian Shi, Valentina E. Balas, Soft Computing Based Medical Image Analysis, Academic Press, Technology and Engineering, 2018
- [3] Terry Des Jardins, George G. Burton, Clinical Manifestations & Assessment of Respiratory Disease - E-Book, Elsevier Health Sciences, 2015
- [4] Kaplan, Kaplan Medical, USMLE Step 2 CK Lecture Notes 2018: Internal Medicine, Simon, and Schuster, 2017
- [5] U. E. Peter, U. Paramalli, C. I. Uchechi and O. Nnaemeka, "Challenges and prospects of blind spread spectrum medical image watermarking," 2017 IEEE 3rd International Conference on Electro-Technology for National Development (NIGERCON), Owerri, 2017, pp. 10-18.
- [6] C. Cavaro-Menard, L. Zhang and P. Le Callet, "Diagnostic quality assessment of medical images: Challenges and trends," 2010 2nd European Workshop on Visual Information Processing (EUVIP), Paris, 2010, pp. 277-284.
- [7] C. Cavaro-Menard, L. Zhang and P. Le Callet, "Diagnostic quality assessment of medical images: Challenges and trends," 2010 2nd European Workshop on Visual Information Processing (EUVIP), Paris, 2010, pp. 277-284.
- [8] G. Niranjana and M. Ponnaivaikko, "A Review on Image Processing Methods in Detecting Lung Cancer Using CT Images," 2017 International Conference on Technical Advancements in Computers and Communications (ICTACC), Melmaurvathur, 2017, pp. 18-25.
- [9] K. Devaki and V. MuraliBhaskaran, "Study of computed tomography images of the lungs: A survey," 2011 International Conference on Recent Trends in Information Technology (ICRTIT), Chennai, Tamil Nadu, 2011, pp. 837-842.
- [10] K. Devaki and V. MuraliBhaskaran, "Study of computed tomography images of the lungs: A survey," 2011 International Conference on Recent Trends in Information Technology (ICRTIT), Chennai, Tamil Nadu, 2011, pp. 837-842.
- [11] R. Agarwal, A. Shankhadhar and R. K. Sagar, "Detection of Lung Cancer Using Content Based Medical Image Retrieval," 2015 Fifth International Conference on Advanced Computing & Communication Technologies, Haryana, 2015, pp. 48-52.
- [12] X. Li, L. Shen and S. Luo, "A Solitary Feature-Based Lung Nodule Detection Approach for Chest X-Ray Radiographs," in IEEE Journal of Biomedical and Health Informatics, vol. 22, no. 2, pp. 516-524, March 2018.
- [13] H. Chung, H. Ko, S. J. Jeon, K. H. Yoon and J. Lee, "Automatic Lung Segmentation With Juxta-Pleural Nodule Identification Using Active Contour Model and Bayesian Approach," in IEEE Journal of Translational Engineering in Health and Medicine, vol. 6, pp. 1-13, 2018.
- [14] R. Sarkar and S. T. Acton, "SDL: Saliency-Based Dictionary Learning Framework for Image Similarity," in IEEE Transactions on Image Processing, vol. 27, no. 2, pp. 749-763, Feb. 2018.
- [15] A. Zamani, S. A. Rezaeieh and A. M. Abbosh, "Lung cancer detection using frequency-domain microwave imaging," in Electronics Letters, vol. 51, no. 10, pp. 740-741, 5 14 2015.
- [16] Awais Mansoor, Ulas Bagci, Brent Foster, "Segmentation and Image Analysis of Abnormal Lungs at CT: Current Approaches, Challenges, and Future Trends", PMC-Radiographics. July-August 2015; 35(4): 1056-1076.
- [17] X. Li, L. Shen and S. Luo, "A Solitary Feature-Based Lung Nodule Detection Approach for Chest X-Ray Radiographs," in IEEE Journal of Biomedical and Health Informatics, vol. 22, no. 2, pp. 516-524, March 2018.
- [18] QingZeng Song, Lei Zhao, * XingKe Luo, and XueChen Dou, "Using Deep Learning for Classification of Lung Nodules on Computed Tomography Images", PMC-J Healthc Eng. 2017; 2017: 8314740.
- [19] Yu.Gordienko, Yu.Kochura, O.Alienin, "Dimensionality Reduction in Deep Learning for Chest X-Ray Analysis of Lung Cancer", arXiv.org, Cornell University Library, 2018

- [20] R. Hooda, S. Sofat, S. Kaur, A. Mittal and F. Meriaudeau, "Deep-learning: A potential method for tuberculosis detection using chest radiography," 2017 IEEE International Conference on Signal and Image Processing Applications (ICSIPA), Kuching, 2017, pp. 497-502.
- [21] Savitha S.K. and N C Naveen. Article: Study for Assessing the Advancement of Imaging Techniques in Chest Radiographic Images. Communications on Applied Electronics 4(5):22-34, February 2016. Published by Foundation of Computer Science (FCS), NY, USA.
- [22] P. R. Katre and A. Thakare, "Detection of lung cancer stages using image processing and data classification techniques," 2017 2nd International Conference for Convergence in Technology (I2CT), Mumbai, 2017, pp. 402-404.
- [23] J. Mukherjee, A. Chakrabarti, S. H. Shaikh and M. Kar, "Automatic Detection and Classification of Solitary Pulmonary Nodules from Lung CT Images," 2014 Fourth International Conference of Emerging Applications of Information Technology, Kolkata, 2014, pp. 294-299.
- [24] Y. Bar, I. Diamant, L. Wolf, H. Greenspan, "Deep learning with non-medical training used for chest pathology identification", Research Gate, 2015
- [25] T. S. Roy, N. Sirohi and A. Patle, "Classification of lung image and nodule detection using fuzzy inference system," International Conference on Computing, Communication & Automation, Noida, 2015, pp. 1204-1207.
- [26] F. Zhang, Y. Song, W. Cai, Y. Zhou, S. Shan and D. Feng, "Context Curves for Classification of Lung Nodule Images," 2013 International Conference on Digital Image Computing: Techniques and Applications (DICTA), Hobart, TAS, 2013, pp. 1-7.
- [27] F. Zhang et al., "A ranking-based lung nodule image classification method using unlabeled image knowledge," 2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI), Beijing, 2014, pp. 1356-1359.
- [28] Z. Xue et al., "Chest X-ray Image View Classification," 2015 IEEE 28th International Symposium on Computer-Based Medical Systems, Sao Carlos, 2015, pp. 66-71.
- [29] F. Taher, N. Werghi and H. Al-Ahmad, "Rule based classification of sputum images for early lung cancer detection," 2015 IEEE International Conference on Electronics, Circuits, and Systems (ICECS), Cairo, 2015, pp. 29-32.
- [30] B. U. Dhaware and A. C. Pise, "Lung cancer detection using Bayasein classifier and FCM segmentation," 2016 International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT), Pune, 2016, pp. 170-174.
- [31] A. Kulkarni and A. Panditrao, "Classification of lung cancer stages on CT scan images using image processing," 2014 IEEE International Conference on Advanced Communications, Control and Computing Technologies, Ramanathapuram, 2014, pp. 1384-1388.
- [32] W. S. H. M. Wan Ahmad, W. M. D. Wan Zaki, M. F. Ahmad Fauzi and W. H. Tan, "Classification of Infection and Fluid Regions in Chest X-Ray Images," 2016 International Conference on Digital Image Computing: Techniques and Applications (DICTA), Gold Coast, QLD, 2016, pp. 1-5.
- [33] Y. Dong, Y. Pan, J. Zhang and W. Xu, "Learning to Read Chest X-Ray Images from 16000+ Examples Using CNN," 2017 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE), Philadelphia, PA, 2017, pp. 51-57.
- [34] X. Liu, F. Hou, H. Qin, A. Hao, "Multi-view multi-scale CNNs for lung nodule type classification from CT images", Elsevier, Vol. 77, pp. 262-275, 2018
- [35] K. Mao and Z. Deng, "Research Article Lung Nodule Image Classification Based on Local Difference Pattern and Combined Classifier", Hindawi Publishing Corporation, pp. 7, 2016
- [36] W. Wong, A. A. Abu-Shareha, M. F. Pasha and R. Mandava, "Enhanced local binary pattern for chest X-ray classification," 2013 IEEE Second International Conference on Image Information Processing (ICIIP-2013), Shimla, 2013, pp. 271-275.
- [37] K. Mao, R. Tang, X. Wang, W. Zhang, and H. Wu, "Research Article Feature Representation Using Deep Autoencoder for Lung Nodule Image Classification", Hindawi, pp. 11, 2018
- [38] D. Kumar, A. Wong and D. A. Clausi, "Lung Nodule Classification Using Deep Features in CT Images," 2015 12th Conference on Computer and Robot Vision, Halifax, NS, 2015, pp. 133-138.
- [39] W. Li, P. Cao, D. Zhao, and J. Wang, "Research Article Pulmonary Nodule Classification with Deep Convolutional Neural Networks on Computed Tomography Images", Hindawi Publishing Corporation, pp. 7, 2016
- [40] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri and R. M. Summers, "ChestX-Ray8: Hospital-Scale Chest X-Ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 2017, pp. 3462-3471.
- [41] R. N. R. A. Susanto, I. Soesanti and Maesadj, "Thoracic X-ray features extraction using thresholding-based ROI template and PCA-based features selection for lung TB classification purposes," 2013 3rd International Conference on Instrumentation, Communications, Information Technology and Biomedical Engineering (ICICI-BME), Bandung, 2013, pp. 65-69.
- [42] S. Widodo, R. N. Rohmah and B. Handaga, "Classification of lung nodules and arteries in computed tomography scan image using principle component analysis," 2017 2nd International conferences on Information Technology, Information Systems and Electrical Engineering (ICITISEE), Yogyakarta, 2017, pp. 153-158.
- [43] S. Widodo, R. N. Rohmah and B. Handaga, "Classification of lung nodules and arteries in computed tomography scan image using principle component analysis," 2017 2nd International conferences on Information Technology, Information Systems and Electrical Engineering (ICITISEE), Yogyakarta, 2017, pp. 153-158.
- [44] X. Li, S. Luo, Q. Hu, J. Li and D. Wang, "Rib suppression in chest radiographs for lung nodule enhancement," 2015 IEEE International Conference on Information and Automation, Lijiang, 2015, pp. 50-55.
- [45] J. Bai, X. Huang, S. Liu, Q. Song and R. Bhagalia, "Learning orientation invariant contextual features for nodule detection in lung CT scans," 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI), New York, NY, 2015, pp. 1135-1138.
- [46] "ELCAP Public Lung Image Database", <http://www.via.cornell.edu/lungdb.html>, Retrived on 24-September-2018
- [47] "CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning", <https://stanfordmlgroup.github.io/projects/chexnet/>, Retrived on 24-September-2018
- [48] "gregwchase/nih-chest-xray", <https://github.com/gregwchase/nih-chest-xray>, Retrived on 24-September-2018
- [49] "COPD Machine Learning Datasets", <http://bigr.nl/research/projects/copd>, Retrived on 24-September-2018
- [50] Savitha S. K and N. C. Naveen, "Algorithm for pre-processing chest-x-ray using multi-level enhancement operation," 2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), Chennai, 2016, pp. 2182-2186
- [51] "Digital Image Database", Digital Image Database, Retrived on 24-September-2018
- [52] "Chronic obstructive pulmonary disease (COPD) Dataset", <https://healthcaregateway.co.uk/chronic-obstructive-pulmonary-disease-copd-dataset/>, Retrived on 24-September-2018
- [53] "Random Sample of NIH Chest X-ray Dataset", <https://www.kaggle.com/nih-chest-xrays/sample>, Retrived on 24-September-2018
- [54] "Tuberculosis Chest X-Ray Image Data Sets", <https://ceb.nlm.nih.gov/repositories/tuberculosis-chest-x-ray-image-data-sets/>, Retrived on 24-September-2018
- [55] "Cancer Data Access System", <https://biometry.nci.nih.gov/cdas/datasets/plco/21/>, Retrived on 24-September-2018