Identification of The Tuberculosis (TB) Disease Based on X-Ray Images Using Probabilistic Neural Network (PNN)

U Andayani, R F Rahmat, Nani Sylviana Pasi, B Siregar, M F Syahputra, M A Muchtar

1Department of Information Technology, Faculty of Computer Science, Universitas Sumatera Utara, Indonesia

ulfi.andayani@usu.ac.id

Abstract. Tuberculosis is an airborne contagious disease caused by the Bacterium Mycobacterium Tuberculosis which means major threat to global health. If Tuberculosis is not treated promptly and quickly, it will cause the disease to get worse. For that, the treatment should not be done carelessly. Therefore, a system is needed to identify the Tuberculosis diseases to help the doctors as an early stage in diagnosing the disease. In this research, Probabilistic Neural Network (PNN) method was used to identify the presence of Tuberculosis by system to help the doctor to make the decision. The stages before doing an identification such as the image acquisition, pre-processing, and feature extractions using Invariant Moment. This study was conducted using train data of 105 normal X-Ray images and 105 images of pulmonary Tuberculosis X-Rays. For the testing data used of 50 X-Ray images. The results showed that the proposed method was able to identify the Tuberculosis disease with an accuracy of 96%.

1. Introduction
Tuberculosis is an airborne disease caused by Mycobacterium Tuberculosis and result in impaired breathing. On the data WHO (World Health Organization) region the number of tuberculosis cases of the disease in Southeast Asia by the year 2015 reached 2,656,560 cases. Whereas in Indonesia reached 330,729 cases. And in the case of tuberculosis (TB) is still growing and reaching 8.7 million people. Tuberculosis (TB) is a major threat to global health. The emergence of new techniques hardware and software triggers an attempt to develop a diagnosis system for detecting the disease tuberculosis (TB). But the real progress has been hampered by less availability of radiography.

Research on the identification of the disease tuberculosis (TB) is a growing research up to this point. Previous research done in the year 2012, Thoma, et al., took a series of features to form grooves and textures of the extensive pulmonary organs using three different Mask for stages of segmentation i.e. Mask Model Statistical Intensity, Mask, and logs Gog Gabor Masks with the Support Vector Machine method for identification of X-Ray images of normal and abnormal [1].

Previous research done in the year 2013, Rohmah, et al., conducted a study of diagnosis of tuberculosis or not tuberculosis based on Lung X-Ray image using Histogram and feature extraction using Mean, Standard Deviation, Kewnes, Kurtosis, Entropy and classification using Principal Component Analysis (PCA) and the accuracy of 95.4. [2].

Other research in the year 2014, Xue, et al., conducted a study using the same dataset writer with classifying image X-Ray into two categories namely the Frontal View and Lateral View using Image features Profile, body Size Ratio
and features PHOG (Pyramid of the Histogram the orientation of the Gradient method with Support Vector Machine (SVM) with an accuracy of 99% [3].

Other research in the year 2014, Jaeger, et al., conducted a study to detect manifestations of tuberculosis with early stages of lung region segmentation using extract Graph Cut, to describe the pattern of normal and abnormal in the pulmonary organs with use the two features, namely Object Detection Features and Inspired CBRI-Based Image Feature with the method of Support Vector Machine (SVM) for classification [4].

In the year 2016, Poornimadevi, et al., conducted a study with the same dataset to detect the lung limits using the Radon Transform and Shape Similarity Bhattacharyya, created a model of the patient's specific anatomical shape of early pulmonary organs using Pythagoreans flow nonrigid registration, and extraction of pulmonary organs limits using Graph Cut Segmentation. Then count pulmonary organs limits the study achieve the accuracy of 94,1% [5].

2. Identification
The experts identified the tuberculosis based on X-Ray image still manually so that add to the waiting time of patients to get the results of the inspection that was done the doctor. Problems also occur due to a lack of beginner doctor against the dataset available as material for practice or learning about these diseases. So, it takes an approach for identification of tuberculosis pulmonary organs infected with the help of a computer system in an effort to help advanced tuberculosis examination will be undertaken by expert doctors. Therefore, this study made for the identification of the disease tuberculosis (TB) using a feature extraction method using Invariant Moment Probabilistic Neural Network (PNN).

3. Methodology
The method to be used for the identification of tuberculosis disease in the X-Ray image, there are several stages. The beginning stages of resizing image aims to reduce the resolution at a later stage, the image of pre-processing such as Contrast aims to clarify an image so that the higher the brightness in the X-Ray image, after image brighter then resumed Close Morphological stage aims to add pixels on the holes that are there on an image, then Gaussian stage aims to remove the noise in the image, the next stage of Thresholding aiming to separate between the object and the background at this stage of changing the image into binary, next Contour for extracting pulmonary pattern on these studies after lung pattern is detected then the Canny stage aims to detect the edges on the lung. After the finished pre-processing stage will be resumed at the stage of feature extraction using methods of Invariant Moment and Probabilistic Neural Network (PNN) as tuberculosis identification or X-Ray image normally. As for the above stages can be seen in the form of public architecture in Figure 1.
3.1. Pre-processing

a. **Resizing**
   
   The process of resizing is the process of changing the size of the image, either zoom in or zoom out the image resolution, resizing is used in research to minimize the size of lung X-Ray image into 1000x1000 pixels.

b. **Contrast**
   
   The process of contrast is the process of improving the contrast of the image of X-Ray's to look clearer colour grayscale on X-Ray images. The density of values drawn far apart by expanding rarely larger. The effect is to increase visual contrast between the two regions, it allows the analysis to more easily distinguish between those areas before.

c. **Morphological Close**
   
   This process is the fill pixels on the holes present on the image. This process is a process of dilation followed stages of erosion. On the value of each operator's image to be included based on a comparison between the corresponding pixels of the image of its neighbouring pixels with the value entered. By choosing the size and shape of the image, then it can build a morphological operator to process the image inserted in order to be more specific.
d. **Gaussian**

   Gaussian blur is often called smoothing which is the value of neighbouring pixels which removes fine detail of an image, the main purpose of the consummation of illumination to change brightness and contrast image with reduced noise in the image and reduce the detail.

3.2. **Segmentation**

   Segmentation is the process of partitioning a digital image into multiple areas of the image with the aim of simplifying or changing the representation of the image becomes more easily analyzed for the next stage.

   a. **Thresholding**

      On the binary-shaped image Thresholding all pixels of the object have a level of grey and pixel threshold, generally that black is white, and the background is the best threshold. Thresholding can be defined as a grey scale into binary [0,1].

   b. **Active Contour**

      In this process to detect pulmonary extract pulmonary pattern on an image. Consists of a set of connecting control points by straight lines. Active contour is associated with multiple control points to each other. A simple segmentation in transform by specifying projected contour is then transferred to the style of the image pushed to the limits of the desired object.

   c. **Canny Edge Detection**

      Canny edge detection techniques used to detect the various sides of the image. Edge detection of canny operator best in detection of edge. Edge detection methods which have a wider range of detection. This method can detect the edge as much as possible, to determine the edge as close as possible with the original and does not detect the edge repeatedly to reduce the errors at the time of identify.

3.3. **Feature Extraction**

   **Invariant Moment**

   Invariant Moment has been widely applied to pattern recognition image in a variety of applications because it is Invariant features for the translation image, scale and rotation. The moment known as enter geometric Moments, Zernike Moments, Rotational Moment and Complex Moments.

3.4. **Identification**

   - **Introduction Probabilistic Neural Network (PNN)**

      Probabilistic Neural Network (PNN) was developed by Donald Specht. These networks provide a general solution to re-purposed pattern by following the approach called Bayesian statistical classifier. PNN has advantages i.e. can overcome the existing problems on the Back-Propagation method by overcoming a long training time and a great need dataset not in the stage of learning.

      Probabilistic Neural Network classifier mostly. PNN is a Neural Network algorithm using probability distribution functions and the implementation of statistical algorithms known as kernel discriminatory analysis, where these operations are organized into four layers with feedforward networks i.e. the input layer, layer, layer pattern summation and output layers [6].

      Excellence that distinguish the Probabilistic Neural Network (PNN) is training a fast, parallel structures that cannot be separated, in the discovery of optimal classification data with increased training. PNN can learn faster than a lot of neural network model and have had success in some applications [7].

      PNN use activation layer second IE hidden layer which aims to make the local decision function based on a sample of the input layer. After the training, the function is added in summation
layer. The result of the function is a probability. So the probability that the maximum most belonging to a specific class.

- **Input Layer**
  The input Layer is an input that is used for the calculation of PNN.

- **Pattern Layer**
  On the Pattern layer approach distance between the vector of weights with input is the value of the training data, whereas the extraction vectors are the value of the extraction attribute will be on the test. In this process using equation 1:

  \[ W_{ij}(x) = \frac{1}{2\pi d^2 \sigma^d} \exp \left( -\frac{\|x-x_{ij}\|^2}{2\sigma^2} \right) \]

  Where:
  - \( W_{ij} \) = gaussian kernel
  - \( D \) = the dimension of vector \( x \)
  - \( \sigma \) = spread/smoothing parameter
  - \( x \) = vector testing
  - \( x_{ij} \) = vector training to \( j \) from class \( i \)

  to determine the value of the smoothing parameter by using the techniques of trial and error.

- **Summation Layer**
  In this process the maximum number of possible counts from each neuron layer in layers with the same class with an average amount of test data with a class. The process of using the equation 2:

  \[ g_i(x) = \frac{1}{2\pi d^2 \sigma^d} \frac{1}{N_i} \sum_{j=1}^{N_i} \exp \left( -\frac{\|x-x_{ij}\|^2}{2\sigma^2} \right) \]

  Where:
  - \( g_i(x) \) = function of density of probability
  - \( D \) = the dimension of vector \( x \)
  - \( \sigma \) = spread/smoothing parameter
  - \( N_i \) = the amount of training data class \( i \)
  - \( x \) = vector testing
  - \( x_{ij} \) = training vector to \( j \) from class \( i \)

- **Output Layer**
  On the process of this compared value between results from the two classes. Probabilistic value is high will be classified a class. The process of using the equation 3:

  \[ G_i(x) = \arg\max\{g_i(x)\}, \quad i = 1, 2, ..., m \]

4. Result and discussions

Following are the results of experiments conducted for the identification of the disease tuberculosis. The results obtained in the form of tuberculosis disease identification based on normal or tuberculosis.

4.1. Pre-processing

The result of this stage in the form of images that have been through the process of Resizing, Contrast, Morphological Close, Gaussian, Thresholding, Contour and Canny Edge Detection. An example of a pre-processing stage can be seen in Figure 2.
4.2. Feature Extraction

The next stage is a feature extraction. At this stage, the image of the results of the previous process, namely the process of Canny Edge Detection. After segmentation stage next stage canny edge detection, namely the extraction of features. Feature extraction is a process to obtain the characteristic of an object or characteristic. This process requires a method to generate the value of the image, one of the methods used in this study are Invariant Moment.

Invariant Moment shape characteristics extraction methods is in the field of image processing. In this method there are seven values in the image of each object, the values are independent against the translation, rotation, and per Scala. After the image of the objects, then the seventh value calculated Invariant Moment. Process calculation of Invariant Moment image file with examples of the results of the Canny Edge Detection with example files as shown in Figure 3.

4.3. Identification of the disease Tuberculosis Using Probabilistic Neural Network (PNN)

The next stage after stage of the obtained value extraction characteristics i.e. identification method using Probabilistic Neural Network. Feature extraction of characteristics obtained from the previous stage are used at this stage. These stages are distinguished into two stages, namely training and testing. In the stage of training will be used is the hallmark of some of the data extraction training. While in testing phase value extraction characteristics of test data would be entered the pattern layer and then to summation layer and the value of the highest probabilities will be grouped into these classes.

4.4. The accuracy

Data based on test results that have been made on the application of the identification of the disease tuberculosis, the accuracy value can be obtained with an average of 96%.

\[
\text{%Accurate} = \frac{\text{The amount of true testing data images}}{\text{The amount of all true testing data images}} \times 100\%
\]
\[
\frac{48}{50} \times 100\% = 96\%
\]

From the above calculation can be showed that the level of accuracy of the method of Probabilistic Neural Network (PNN) in the identification of the disease tuberculosis is 96%.

5. Conclusion and future works

5.1. Conclusion

The results of pulmonary tuberculosis disease detection using neural network Probabilistic Neural Network (PNN) are as follows:

- Detection of tuberculosis disease types can be done by using the neural network Adaptive Probabilistic Neural Network (PNN) as a method for detecting tuberculosis disease with data training as much as 105 normal X-Ray image and X-Ray image of 105 tuberculosis, and on the test data by as much as 50 normal X-Ray image and tuberculosis with 96% accuracy.
- In the process of image processing by using the Contrast, Morphological Close, Gaussian, Thresholding, Contour, Edge Canny Detection and feature extraction Invariant Moment for lung extraction of X-Ray images so that it displays the results from the normal lung or the form of pulmonary tuberculosis.

5.2. Future works

The future development of research are as follows:

- On the research of the next expected image pre-processing can identify tuberculosis based on texture patches found on the X-Ray image.
- Research the next developing identification based on a diagnosis such as phlegm etc.

References

[1] Thoma G, Jaeger, S, Karargyris A and Antani S 2012 *Detecting Tuberculosis in Radiographs Using Combined Lung Masks* Annual International Conference of the IEEE EMBS (San Diego: California USA)

[2] Rohmah R N, Susanto A and Soesanti I 2013 *Lung Tuberculosis Identification Based on Statistical Feature of Thoracic X-Ray* IEEE Quality in Research pp. 19-20 (Indonesia)

[3] Xue Z, You D, Candemir S, Jaeger S, Antani S, Long L R and et al 2014 Chest X-Ray Image View Classification *Lister Hill National Center for Biomedical Communications National Library of Medicine* Bethesda (USA: National Institutes of Health)

[4] Jaeger S, Karargyris A, Candemir S, Folio L, Siegelman J, Callaghan F and et al 2014 *Automatic Tuberculosis Screening Using Chest radiographs* IEEE Transaction On Medical Imaging (Shenzhen)

[5] Poornimadevi C, and C H S 2016 *Automatic Detection Of Pulmonary Tuberculosis IUsing Processing Techniques* (India)

[6] Bakos M 2007 *Active Contours and their Utilization at Image Segmentation* 5th Slovakian-Hungarian Joint Symposium on Applied Machine Intelligence and Informatics

[7] Karargyris A, Jaeger S, Antani S and Thoma G 2012 *Detection Tuberculosis in Radiographs Using Combined Lung Masks* Annual International Conference (San Diego: California USA)