TEM Journal. Volume 11, Issue 3, pages 1331-1340, ISSN 2217-8309, DOI: 10.18421/TEM113-43, August 2022.

Convolutional Neural Network as an Image Processing Technique for Classification of Bacilli Tuberculosis Extra Pulmonary (TBEP) Disease

Bob Subhan Riza [1,2], Jufriadif Na'am [1], S Sumijan [2]

[1] *Computer Science Faculty, Universitas Putra Indonesia, Padang, Indonesia*
[2] *Faculty of Engineering and Computer Science, Universitas Potensi Utama, Medan, Indonesia*

Abstract **– Tuberculosis Extra Pulmonary (TBEP) is an infectious disease caused by the bacterium Mycobacterium tuberculosis and can cause death. Patients suffering from this disease must be treated quickly without waiting for a long time. Biopsy is one of the techniques used to take the patient's lung fluid and given Ziehl Neelsen chemical dye and then observed using a microscope to determine this TBEP disease. In this research, the TBEP detection process was developed using a classification method, namely CNN with feature extraction and feature selection. The feature uses 5 features where these features are a combination of shape features and texture features with the highest information gain value. From the results of research conducted through the training and testing stages of the classification method using feature selection, the accuracy rate is higher than not using feature selection with a comparison of the feature selection stage increasing 0.6536% for the training process, and 0.8942% for the testing process.**

DOI: 10.18421/TEM113-43
https://doi.org/10.18421/TEM113-43

Corresponding author: Bob Subhan Riza,
Computer Science Faculty, Universitas Putra Indonesia, Padang, Indonesia.
Email: bob.potensi@gmail.com

Received: 24 April 2022.
Revised: 08 August 2022.
Accepted: 15 August 2022.
Published: 29 August 2022.

Keywords **– Tuberculosis Extra Pulmonary, Otsu Thresholding, Hue Saturation Value (HSV), feature extraction, feature selection, Convolutional Neural Network**

1. Introduction

Tuberculosis is a life-threatening infectious disease worldwide caused by the bacterium Mycobacterium tuberculosis. These bacteria are in the form of Acid-Fast Bacilli (AFB). These bacilli are 1-4 m long and 0.3-0.56 m wide as shown in Figure 1., are non-spore-forming, non-motile, and facultative. Bacterial cell walls contain long chain glycolipids that are mycolic, rich in acids and phosphopoglycans [1], [2].

Tuberculosis (TB) is a chronic and infectious disease that affects the world's human population and requires complex treatment. It is a public health problem with more than 9 million estimated new cases and 1.5 million deaths annually worldwide [3]. Of the estimated 9 million people who contracted TB in 2013, more than 80% were in Southeast Asia, the Western Pacific and Africa. Most of the infected population comes from poor and marginalized communities with weak health services infrastructure [4].

Tuberculosis can affect every human being regardless of region, usually the ones who develop this disease are adults, with 30 countries affected by tuberculosis where almost 90% of those affected fall ill. Those of productive age are susceptible to TB between 15 to 50 years and children. TB usually comes out through phlegm and coughing, if the saliva is scattered at low temperatures, the possibility for germs to survive will be long enough to allow the transmission process to occur. There are 2 types of Tuberculosis, namely Pulmonary Tuberculosis (TBP) and Extra Pulmonary Tuberculosis (TBEP). TBP affects the lungs, whereas TBEP can affect any organ of the body except the spine, heart, pancreas, skeletal muscle, and thyroid.

So far, to detect Extra Pulmonary Tuberculosis (TBEP) through a biopsy, namely by taking fluid from a person's lymph nodes which will be detected by a doctor or health analyst then placed on the preparation and viewed through a microscope for readings on the preparation to see the presence of germs or Bacilli tuberculosis. Detecting what is happening at this time will take a long time because the liquid preparations are viewed under the microscope one by one carefully and the liquid preparations contain 150 fields of vision [5]. Extrapulmonary tuberculosis can cause complications. Based on this, we need a system for good reporting and recording for TB control, see **Figure 1**. [6].

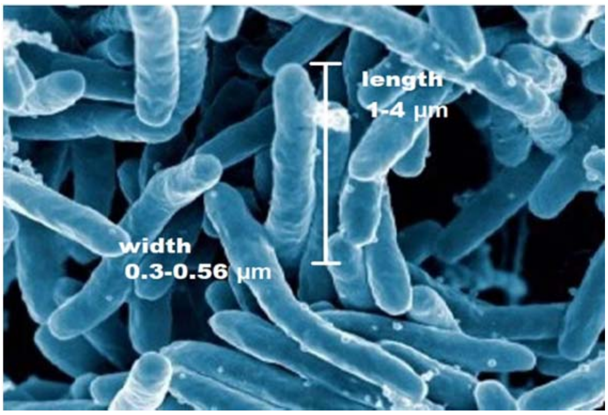


Figure 1. Mycobacterium Tuberculosis [6]

The WHO explained the impact of the Covid-19 Pandemic, the number of deaths due to Tuberculosis globally could increase by around 0.2 -0.4 million in 2020. The estimated impact of the Covid-19 Pandemic on the number of deaths due to Tuberculosis is shown in **Figure 2**.

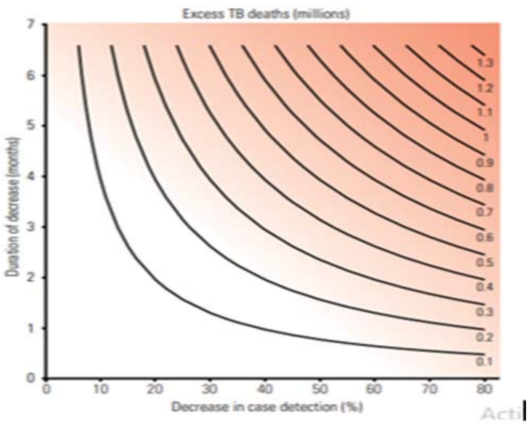


Figure 2. Estimated impact of the COVID-19 pandemic on the number of TB deaths globally in 2020, for various combinations of decreased case detection and duration of decline

Tuberculosis can damage the human lungs or other parts of the body and cause severe illness. Tuberculosis is spread through the air when a person who has TB in the lungs coughs, sneezes or talks and transmits it through the air [7]. Further spread is if the germ is inhaled by a person then it can be infected. Tuberculosis is mostly spread from frequent and long associations, for example with family members or friends.

Symptoms of Tuberculosis are cough that lasts more than 3 weeks, sudden weight loss, tiredness, night sweats, decreased appetite, blood-stained sputum. Symptoms of Extra Pulmonary Tuberculosis are pain and swelling in the affected area [8], [9]. Several image processing methods have been reported for the detection of Extra Pulmonary Tuberculosis (TBEP).

Chetan detects tuberculosis images from screening results using thresholding technique [10], Jorge researched Extra Pulmonary Tuberculosis (TBEP) from the results of screening with a segmentation process [11].

AlSaffar et al., in their study detect Tuberculosis bacteria with vector machines, logistic regression, and nearest neighbors and there are several steps that are automatically carried out; for the segmentation stage is to identify possible Tuberculosis objects by eliminating artifacts where the resulting outcome is "definite", "probably" or "non Tuberculosis". After testing for all data collected from the TB Hospital laboratory, the accuracy rate is very good with an accuracy value of 99% [12].

Hamed et al.'s research in developing tuberculosis detection using various methods to classify individual and overlapping bacilli from the rest of the images based on the eigenvalues of the shape and color models. By using statistical shape model and statistical color model and KNN classifier produces an average accuracy value of 82.7% for the detection of single bacilli and overlapping bacilli, to identify only individual bacilli from overlapping bacilli and other objects, and the accuracy value is 99, 1% [13].

Aeri et al. created an expert system in the identification of mycobacterium tuberculosis using feature extraction method by comparing colors. The image data in this study were 1266 image data from sputum. The method used in image improvement is a median filter with a color histogram to extract color features. The extracted colors are HSV colors, where a collection of hue values forms a histogram cluster. In the classification method with AdaBoost and an expert system (random forest), the results of the Tuberculosis classification get an accuracy value of 85% [14].

Aeri et al. conducted a study to detect tuberculosis by classifying mycobacterium tuberculosis based on color feature extraction using HSV (Hue Saturation Value) with the adaptive boosting method. Before the classification stage is carried out, first the filter process is carried out using the median filter method. Extraction of color histogram features is carried out using several quantization measures. The color histogram is formed from the combined results of the hue values of each image pixel. The hue value itself ranges from 0-360 which represents the color of each image pixel. The steps to be taken are 8, 16, 32 and 64 using a combination of Adaboost from decision-making to learning methods [15]. The results of this study obtained the best Tuberculosis bacteria classification accuracy value in the testing process using the Adaboost method, namely 81.7% [16].

2. Research Methodology

A. Block Diagram

In the process of making a system to diagnose Extra Pulmonary Tuberculosis (TBEP) for the segmentation stage, there are several processes that are carried out, and can be seen in **Figure 3**.

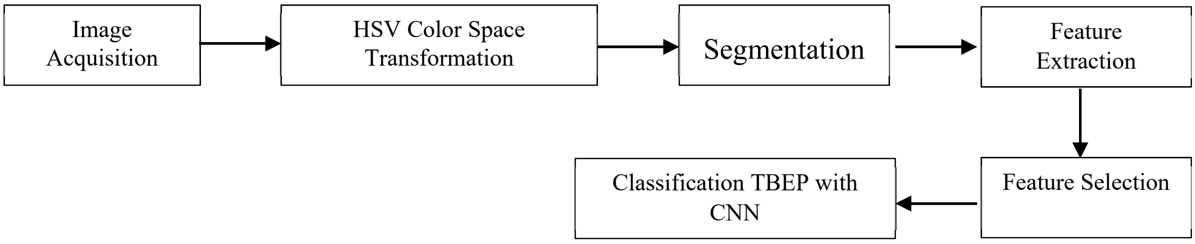


Figure 3. Block Diagram Classification

B. Image Acquisition

At the stage of image acquisition of Extra Pulmonary Tuberculosis (TBEP) by means of a biopsy, namely taking fluid from patients with Extra Pulmonary Tuberculosis, then the fluid is placed into a preparation, so that the object to be seen through a microscope can be seen clearly, the fluid that will be seen in the microscope is treated with Ziehl Neelsen chemical dye. The equipment used to analyze the preparations was an Olympus BX53 digital microscope with a 1000x magnification lens, because to be able to see the structure of the BTA (mycobacterium) 1000x magnification was used from the microscope as shown in **Figure 4**. The number of images captured was 78 images consisting of 51 TBEP images and 27 non-TBEP images with dimensions of 1920x1440 pixels and taken from the Central General Hospital H. Adam Malik, Medan and the sample preparations were examined and labeled and validated by a specialist in Clinical Microbiology by the name of dr. Rina Yunita, SpMK (K).

Figure 4. Digital microscope

Extrapulmonary Tuberculosis images seen under a microscope can be shown in **Figure 5**.

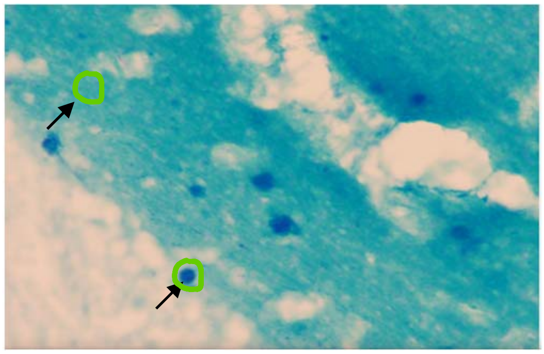


Figure 5. Extrapulmonary Tuberculosis (TBEP) images

Figure 5. shows Extra Pulmonary Tuberculosis, where can be seen the presence of Acid-Fast Bacilli (AFB). This BTA shows that the sample seen in the Mycobacterium Tuberculosis image is in the form of a red rod (bacillus) inside a green circle. Mycobacterium Tuberculosis is red due to previous staining with Ziehl Neelsen. This staining makes it easier to differentiate between bacilli and non-bacilli.

3. Result and Discussion

A. HSV Color Space Transformation

There are several methods used in the transformation of the RGB color space, including CMYK, HSV and L*a*b. HSV (Hue Saturation Value) color is a color that is derived from the RGB (Red Green Blue) color model, so to produce HSV color you must go through the process of converting colors from RGB to HSV [17]. In this study, we have tried various color spaces including RGB, the better result is HSV to get to the segmentation stage.

After taking Tuberculosis and Non Tuberculosis images, the next step is to transform the HSV (Hue, Saturation, Value) color space. Hue is the actual color, Saturation is the purity of the color and Value is the brightness of the color. The advantage of using the HSV color space is that there are colors that are the same as those captured by the human senses. The results of the HSV color can be seen in **Figure 6**.

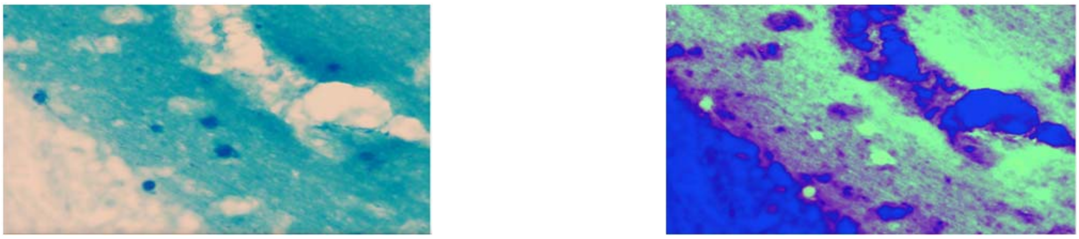
The image resulting from the HSV transformation will be used for the segmentation stage, following equation 1. for HSV color [18]:

$$H = \begin{cases} \begin{cases} \theta & if\ B \le G \\ 360\ \theta & if\ B > G \end{cases} \\ cos^{-1}\left\{ \dfrac{\frac{1}{2}[(R-G)+(R-B)]}{[(R-G)^2+(R-B)\,(G-B)]^{1/2}} \right\} \end{cases} \quad \text{............(1)}$$

$$S = \begin{cases} 0 & if\ R = G + B \\ 1 - \dfrac{3}{(R+G+B)}[\min(R,G,B)] & other \end{cases}$$

$$V = \frac{1}{3}\,(R+G+B)$$

HSV Algorithm
1. Start
2. Declaration
3. var Img,H,S,V
4. Call out image Img
5. Image transformation Img into HSV (RGB Color space to HSV) :
6. Calculating H,S and V
7. Extracting Hue component
   H = HSV(:,:,1)
8. Saving result component Hue
   handles.H = H;
   guidata(hObject, handles)
9. Finish



(a) (b)

Figure 6. (a) RGB Image, (b) Image converted from RGB to HSV

Figure 6. (a) shows the original image, namely RGB, then the RGB image is converted to HSV (b). The transformation of the RGB color space into HSV can be assumed that the coordinates R, G, B (0 or 1) are a sequence of red, green, blue in the RGB color space, max is the maximum value of the values (red, green, blue), min is the minimum value of value (red, green, blue). After converting the color space to HSV, the Hue component is used, because of the three HSV components that can visually distinguish between objects and the background, the Hue component is the Hue component where the Hue component is obtained from the HSV image. The image of the Hue component can be seen in **Figure 7**.

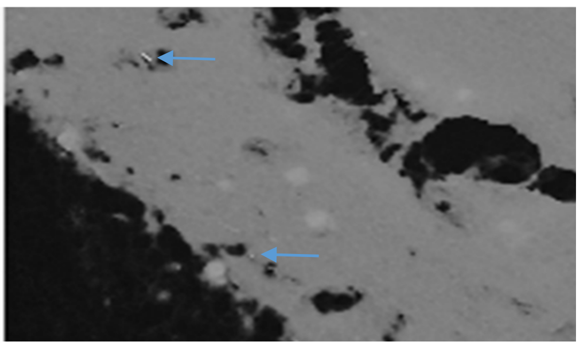


Figure 7. Hue component of RGB to HSV. Color Space Transformation

Figure 7. is the Hue component of the result of the RGB to HSV color space transformation. In Figure 7., the Bacilli object can already be different from the color of other objects. The color of the Bacilli object as indicated by the arrow is white.

B. Segmentation

Image segmentation is the identification and isolation of an image into certain areas with the aim of conforming to the structural unit. Segmentation is an important operation in biomedical image processing because it is used to isolate physiological and biological structures [19]. The general approach to segmentation can be grouped into three classes: pixels, regions and edges. In this study, the purpose of the segmentation process is to divide the network image into two regions, TBEP and non-TBEP regions. The TBEP region refers to objects that characterize Bacilli TBEP while the non-TBEP region refers to the background and objects in the image. In the research it was conducted using several algorithms for the segmentation process.

Otsu thresholding is used because it can calculate the threshold value automatically and maximize the variance between classes from classes separated by a threshold, which aims to separate BTA objects from the background and can detect Bacilli in the image or image used. The Otsu thresholding equation can be seen in equation 2 [20]:

$$\sigma^2 = P_{nw}(M_{nw} - M)^2 + P_w(M_w - M)^2$$
$$M = P_{nw}.M_{nw} + P_w.M_w \quad (2)$$
$$P_{nw} + P_w = 1$$

$$t^* = \underset{a \le t \le b}{\text{Arg Max}}\{P_{nw}(M_{nw} - M)^2 + P_w.(M_w - M)^2$$

Where:
σ^2 : variance within the TBEP and non-TBEP classes
P_{nw} : pixel value probability for non TBEP class
P_w : pixel value probability for TBEP class
M_{nw} : average non-TBEP class pixel value
M_w : TBEP class average pixel value
M : average pixel value of the image
t^* : threshold value

Otsu Thresholding Algorithm and Morphological Operation
1. Start
2. Var H Declaration ;
3. Call out image result Hue : H
4. Calculate otsu equation
5. Perform morphological operations (closing and opening)
6. Calculating MSE, RMSE and PSNR value
7. Take the bounding box of each segmented subject
8. Save the results of closing and opening
9. Finish

The results of segmentation using Otsu thresholding with the HSV (Hue, Saturation, Value) coloring technique can be shown in **Figure 8.** and **Figure 9**.

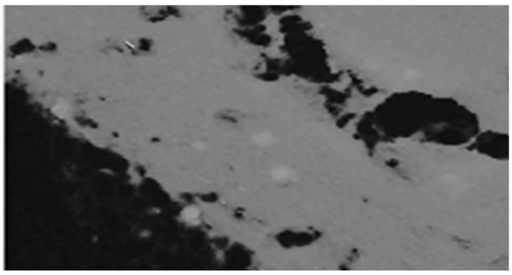


Figure 8. Image result of HSV staining technique

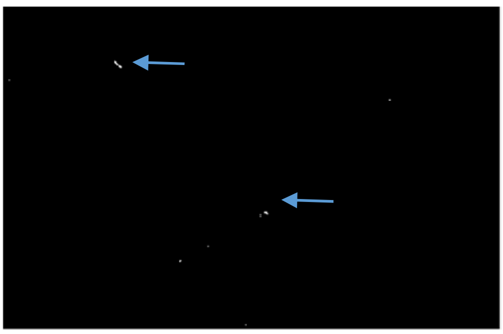


Figure 9. Image segmentation results from Otsu thresholding

Figure 8. results of segmentation using HSV staining technique. Figure 9. is the result of thresholding using Otsu thresholding. The results show that the Bacilli object can be separated from other objects where in the image objects other than Bacilli have a black background while the Bacilli object is in the foreground with white color, the blue arrow shows an image of Bacilli which can be detected by the Otsu method.

C. *Feature Extraction*

The resulting image from the segmentation process is then processed to the feature extraction stage to get good features, based on the shape and texture features of the AFB object. The features used in this study are shape features with the following criteria: perimeter, eccentricity, metric and area ratio and texture features with the criteria of contrast, energy, homogeneity, and entropy. The results of the morphological analysis using closing and opening can be seen in **Figure 10.** where the morphological operation aims to change the shape of the object from the initial image, which is to remove elements around the AFB object.

Figure 10. Results of morphological operations (closing and opening)

Figure 10. is an example of the results of morphological analysis for AFB positive class from Figure 11. (b) with a value of MSE = 0.69838 and PSNR = 1.5591, the results of this morphological analysis process are used to obtain shape features in the form of perimeter, eccentricity, metric, and area ratio.

The results of the average PSNR, MSE and RMSE values using 78 TBEP and non TBEP images with Otsu thresholding can be seen in **Table 1**.

Table 1. PSNR, MSE, RMSE value results using Otsu Thresholding

Otsu Thresholding			
No	**PSNR**	**MSE**	**RMSE**
1	2,2102	0,6011	0,7753
2	2,4660	0,5668	0,7528
3	2,6746	0,5402	0,7350
4	2,7486	0,5311	0,7287
5	3,4080	0,4562	0,6755
6	4,8394	0,3281	0,5728
7	5,1470	0,3057	0,5529
8	5,5008	0,2818	0,5308
9	6,3957	0,2293	0,4789
10	5,9604	0,2535	0,5035
11	5,0977	0,3092	0,5561
12	4,9120	0,3227	0,5681
13	5,0553	0,3122	0,5588
14	5,0506	0,3126	0,5591
15	3,8207	0,4149	0,6441
16	2,9244	0,5100	0,7141
17	3,1688	0,4821	0,6943
18	2,7541	0,5304	0,7283
19	2,6516	0,5430	0,7369
20	2,8176	0,5227	0,7230
21	3,7067	0,4259	0,6526
22	6,2006	0,2398	0,4897
23	3,9510	0,4026	0,6345
24	5,3972	0,2886	0,5372
25	4,1861	0,3814	0,6176
26	4,5030	0,3546	0,5955
27	6,4186	0,2281	0,4776
28	1,5591	0,6984	0,8357
29	4,0421	0,3943	0,6279
30	1,8040	0,6601	0,8125
31	0,6529	0,8604	0,9276
32	1,7338	0,6708	0,8190
33	2,3160	0,5867	0,7659
34	3,1939	0,4793	0,6923
35	1,0734	0,7810	0,8837
36	1,5833	0,6945	0,8334
37	1,2177	0,7555	0,8692
38	5,1746	0,3038	0,5512
39	4,4428	0,3595	0,5996
40	2,5706	0,5533	0,7438
41	3,3871	0,4584	0,6771
42	3,7963	0,4172	0,6459
43	3,9608	0,4017	0,6338
44	3,4631	0,4505	0,6712
45	3,1121	0,4884	0,6989
46	2,7322	0,5331	0,7301
47	3,3512	0,4623	0,6799
48	5,6254	0,2738	0,5233
49	5,5596	0,2780	0,5273
50	7,5990	0,1738	0,4169
51	5,8512	0,2599	0,5098
52	6,1711	0,2415	0,4914
53	4,7589	0,3343	0,5782
54	3,4361	0,4533	0,6733
55	4,7371	0,3360	0,5796
56	1,2544	0,7491	0,8655
57	1,8035	0,6602	0,8125
58	2,5426	0,5568	0,7462
59	3,9482	0,4029	0,6347
60	4,5514	0,3506	0,5921
61	6,2642	0,2364	0,4862
62	6,6402	0,2168	0,4656
63	1,9515	0,6380	0,7988
64	2,2019	0,6023	0,7761
65	2,0694	0,6209	0,7880
66	1,4433	0,7173	0,8469
67	2,0084	0,6297	0,7936
68	1,6311	0,6869	0,8288
69	3,6895	0,4276	0,6539
70	2,0753	0,6201	0,7875
71	1,3375	0,7349	0,8573
72	1,8096	0,6592	0,8119
73	1,6061	0,6909	0,8312
74	1,3623	0,7308	0,8548
75	1,2864	0,7436	0,8623
76	1,0887	0,7783	0,8822
77	2,9500	0,5070	0,7120
78	1,0544	0,7844	0,8857
Total	269,4423	37,7792	53,3661
Average	3,4544	0,4843	0,6842

Table 1. shows the PSNR, MSE and RMSE values with Otsu Thresholding, using 78 TBEP and non TBEP images producing an average PSNR value = 3.4544 decibels, MSE = 0.4843 and RMSE = 0.6842. The development of the model which is a novelty in this research, namely the extraction of shape and texture features is to combine several shape and texture features with a total of 8 features, then feature selection is carried out, 5 features are taken in the order of the best gain values. Furthermore, calculations for the perimeter, eccentricity, metric and area ratio are carried out, namely:

Perimeter is the length of the TBEP and non TBEP Bacilli objects. As for the similarities

$$P_i = \sum_{x=0}^{M_i-1} \sum_{y=0}^{N_i-1} O_i(x,y), \forall O_i\,(\mathrm{x,y}) \in 4 - \mathrm{neighbour}(O_i(\mathrm{x,y})) = 0 \quad (3)$$

Where $M_i, and\ N_i$ each based on the number of rows and columns to – i. x,y limit for Bacilli TBEP object images or non TBEP images from closing and opening results.

Eccentricity is the ratio of the minor elliptical foci to the major elliptical foci in Bacilli TBEP. Equation of eccentricity

$$E_i = \sqrt{1 - \frac{b_i}{a_i}} \quad (4)$$

Where $a_i, and\ b_i$ is the size of the major and minor axes of the object O_i(x,y). Metric is a comparison between the area and perimeter of an object's area. Metric equation

$$M = \frac{4\mu x A}{Cf^2} \text{...}(5)$$

where M=*metric*, A=area, Cf = *circumference*

Area ratio: the ratio of the number of pixels of the object, is the size of the object. Equation of area ratio

$$A_i = \sum_{x=0}^{M_i-1} \sum_{y=0}^{N_i-1} O_i(x, y) \text{.............}(6)$$

Where $M_i, and\ N_i$ each based on the number of rows and columns to – i.

Texture feature extraction used with GLCM (Gray Level Co-Occurrence Matrix) is by calculating the probability of neighboring relationships within a certain distance. Distance selection is one of the most important parameters in the GLCM.

The development of texture feature extraction begins with calling the grayscale image and forming a framework with a size of 256x256. Furthermore, the co-occurrence matrix is compiled by filling in the number of spatial relationships that exist in the matrix. The matrix is then transposed in order to obtain a symmetrical angle of 180o. Next, add the co-occurrence matrices so that they are symmetrical and then normalize where by adding up all the symmetric matrices, it is used as a divisor for all pixels in the symmetric matrix. Finally, the calculation of Contrast, Energy, Homogeneity and Entropy is done. The calculation of texture features for 1 data is:

1) Contrast Equation:

$$\mathrm{C} = \sum_{a=1}^{m} \sum_{b=1}^{n} P_{a,b}(a-b)^2 \text{..........} (7)$$

2) Energy Equation:

$$E_1 = \sum_{a}^{m} \sum_{b}^{n} \frac{p\,(a,b)}{1+(a-b)^2} \text{......................} (8)$$

3) Homogeneity Equation:

$$H = \sum_{a}^{m} \sum_{b}^{n} P(a,b)^2 \text{.....................} (9)$$

4) Entropy Equation:

$$E_2 = -\sum_{a}^{m} \sum_{b}^{n} p(a,b) \log\{p(i,j)\}(10)$$

D. *Feature Selection*

Feature selection is one of the feature selection processes that affect the accuracy of TBEP detection. The purpose of feature selection is to speed up the training process. The feature selection used is Information gain. Information gain can measure how influential a feature is on the measurement results. This algorithm can reduce feature dimensions by measuring the entropy reduction before and after separation, the result as can see in **Table 2**.

Table 2. Feature selection results using Information Gain

Feature	*Gain Value*
Perimeter	0.0684
Eccentricity	0.0180
Metric	0.0153
Area *Ratio*	0.0249
Contrast	0.4036
Energy	0.3393
Homogeneity	0.3360
Entropy	0.3831

Based on the calculation of the information gain algorithm, the results of the order of features from the highest to the lowest are contrast, entropy, energy, homogeneity, perimeter, eccentricity, metric, and area ratio. This study uses five (5) highest features, namely contrast, entropy, energy, homogeneity, perimeter, see **Table 3**.

Table 3. CNN parameters training and testing process for TBEP detection without feature selection

Number of features	*epoch*	Training			Test		
		Accuracy (%)	Sensitivity (%)	Specification (%)	Accuracy (%)	Sensitivity (%)	Specification (%)
8	100	81.3725	82.6667	80.1282	68.5544	62.3894	81.2785
8	200	81.0458	84	78.2051	71.237	67.6991	78.5388
8	300	81.3725	84.6667	78.2051	71.9821	68.8053	78.5388
8	400	81.3725	84.6667	78.2051	72.4292	69.2478	78.9954
8	500	81.6993	85.3333	78.2051	72.4292	69.2478	78.9954

The features used in this training process are 8 features. At the beginning of the training, the training data were trained using epoch=100 with a batch size of 128, learning rate 0.1, and momentum 0.9 resulting in an accuracy of 81.3725%, sensitivity 82.6667% and specificity 80.1282%, epoch=200 yielding an accuracy of 81.0458%, sensitivity 84% and specificity 78.2051%, epoch=300 produces 81.3725% accuracy, sensitivity 84.6667% and specificity 78.2051%, epoch=400 produces 81.3725% accuracy, sensitivity 84.6667% and specificity 78.2051%, epoch=500 produces 81.6993% accuracy, sensitivity 85.3333% and specificity 78.2051%.

In the test, the test data using epoch=100 with batch size 128, learning rate 0.1, and momentum 0.9 resulted in an accuracy of 68.5544%, sensitivity 62.3894% and specificity 81.2785%, epoch=200 yielding an accuracy of 71.9821%, sensitivity 68.8053 % and specificity 78,5388%, epoch=300 yielded accuracy 71,6841%, sensitivity 68,3628% and specificity 78,5388%, epoch=400 yielded accuracy 72,4292%, sensitivity 69,2478% and specificity 78,9954%, epoch=500 yielded accuracy 72.4292%, sensitivity 69.2478% and specificity 78.9954%, epoch=500 yielded accuracy of 72.4292%, sensitivity 69.2478% and specificity 78.9954%.

The results show that the accuracy value is better at the epoch value = 500 where the training process with an accuracy value of 81.6993%. The results show that the accuracy value is better at the epoch value = 500 where the testing process with an accuracy value of 72.4292%, see **Table 4**.

Table 4. CNN parameter training and testing process for TBEP detection with feature selection

Number of features	***epoch***	**Training**			**Test**		
		Accuracy (%)	Sensitivity (%)	Specification (%)	Accuracy(%)	Sensitivity (%)	Specification (%)
5	100	82.3529	78.6667	85.8974	69.2996	63.0531	82.1918
5	200	80.719	81.3333	80.1282	69.4486	63.9381	80.8219
5	300	82.0261	81.3333	82.6923	72.2802	67.9204	81.2785
5	400	81.3725	82.6667	80.1282	72.5782	69.2478	79.4521
5	500	81.6993	84.6667	78.8462	73.3234	70.5752	78.9954

There are 5 features used in this training process. At the beginning of the training, the training data were trained using epoch=100 with batch size 128, learning rate 0.1, and momentum 0.9 producing an accuracy of 82.3529%, sensitivity 78.6667% and specificity 85.8974%, epoch=200 yielding an accuracy of 80.719%, sensitivity 81.3333% and specificity 80.1282%, epoch=300 produces 82.0261% accuracy, sensitivity 81.3333% and specificity 82.6923%, epoch=400 produces 81.3725% accuracy, sensitivity 82.6667% and specificity 80.1282%, epoch=500 produces 81.6993% accuracy, sensitivity 84.6667% and specificity 78,8462%.

In the test, the test data using epoch=100 with batch size 128, learning rate 0.1, and momentum 0.9 resulted in 69.2996% accuracy, 63.0531% sensitivity and 82.1918% specificity, epoch=200 yielded 69.4486% accuracy, sensitivity 63.9381 % and specificity 80.8219%, epoch=300 resulted in 72.2802% accuracy, sensitivity 67.9204% and specificity 81.2785%, epoch=400 yielded an accuracy of 72.5782%, sensitivity 69.2478% and specificity 79.4521%, epoch=500 yielded accuracy of 73.3234%, sensitivity of 70.5752% and specificity 78.9954%.

Better accuracy results with the selection of training process features for Tuberculosis detection based on parameters with batch size 128, learning rate 0.01, and momentum 0.9, namely at epoch 100 it produces 82.3529% accuracy, 78.6667% sensitivity, 85.8974% specificity. %. The results show that the accuracy value is better at the epoch value = 500 where the testing process with an accuracy value of 73.3234%, sensitivity 70,5752%, specificity 78,9954%. **Table 5.** as result test parameters and **Table 6.** for CNN Training and testing result.

Table 5. Test parameters using CNN

Parameter	Value with feature selection	Value without feature selection
learning rate value	0,1	0,1
Batch size	128	128
momentum	0,9	0,9

Table 6. Results of training and testing using CNN

Method	Data	Training Accuracy (%)	Test Accuracy (%)
CNN	AFB features with feature selection	82.3529	73.3234
CNN	AFB features without feature selection	81.6993	72.4292

Based on the results from Table 6. on the training and testing process using the Convolutional Neural Network for classification, the better results obtained are using feature selection than without using feature selection, where using the information gain algorithm for the feature selection stage increases 0.6536% for the training process and 0.8942% for the testing process, see **Figure 11**.

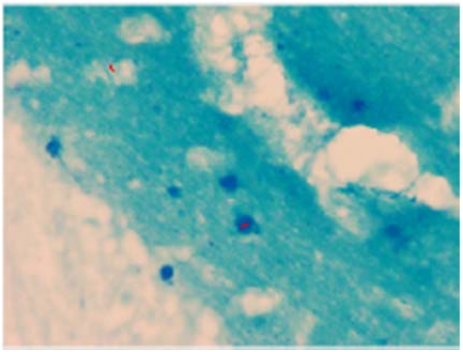


Figure 11. TBEP detection results using CNN

In Figure 11, it can be explained that the CNN method can detect Bacilli TBEP well and the red color in the image above indicates that Bacilli TBEP.

4. Conclusion

From the results of the research conducted, explaining the CNN method and using the feature selection method it can detect Bacilli TBEP with higher accuracy than without using feature selection.

Acknowledgment

The research team would like to thank those who have facilitated and provided full support and trust as the winner of the Internal Grant "Pengembangan Algoritma CNN pada Pendeteksian Penyakit Tubercolosis" funded by the Universitas Potensi Utama in accordance with register number 0369/UPU/PJJ/PPM/IV/2020 from April 2020 until April 2022 so that this research can be completed properly.

References

[1]. Riza, B. S., Mashor, M. Y., Osman, M. K., & Jaafar, H. (2017, August). Automated segmentation procedure for Ziehl-Neelsen stained tissue slide images. In *2017 5th International Conference on Cyber and IT Service Management (CITSM)* (pp. 1-5). IEEE. doi:10.1109/CITSM.2017.8089292

[2]. Zayet, S., Berriche, A., Ammari, L., Kanoun, F., Kilani, B., & Benaissa, H. T. (2022). Pseudotumoural pulmonary tuberculosis: a case series. *IJID Regions*, *2*, 158-161. doi:10.1016/J.IJREGI.2022.01.009

[3]. Ferkol, T., & Schraufnagel, D. (2014). The global burden of respiratory disease. *Annals of the American Thoracic Society*, *11*(3), 404-406. doi:10.1513/ANNALSATS.201311-405PS/SUPPL_FILE/DISCLOSURES.PDF

[4]. Cao, Y., Liu, C., Liu, B., Brunette, M. J., Zhang, N., Sun, T., ... & Curioso, W. H. (2016, June). Improving tuberculosis diagnostics using deep learning and mobile health technologies among resource-poor and marginalized communities. In *2016 IEEE first international conference on connected health: applications, systems and engineering technologies (CHASE)* (pp. 274-281). IEEE. doi:10.1109/CHASE.2016.18

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

[6]. Raviglione, M. C., & Rieder, H. L. (2021). Synergy between government and non-governmental organizations in health: WHO and the Union collaboration in tuberculosis control. *Journal of clinical tuberculosis and other mycobacterial diseases*, *24*, 100251. doi:10.1016/J.JCTUBE.2021.100251

[7]. Setiawan, B. U. D. I., Zarqya, I. C. H. W. A., Putro, S. A. P. T. O. N. O., & Khasanah, F. U. R. A. I. D. A. (2019). The effect of red betel leaf's essential oil (Piper Crocatum Ruiz & Pav.) against Third Instar Aedes aegypti Larvae. *Pakistan J Med Heal Sci*, *13*, 1162-1164.

[8]. Pangaribuan, L., Kristina, K., Perwitasari, D., Tejayanti, T., & Lolong, D. B. (2020). Faktor-Faktor yang mempengaruhi kejadian tuberkulosis pada umur 15 tahun ke atas di Indonesia. *Buletin Penelitian Sistem Kesehatan*, *23*(1), 10-17. doi:10.22435/HSR.V23I1.2594

[9]. Hartatik, H. (2016). Diagnosa Penyakit Pulmonary Tuberculosis Dan Extrapulmonary Tuberculosis Menggunakan Algoritma Certainty Factor (CF). *CSRID (Computer Science Research and Its Development Journal)*, *8*(1), 11-24. doi:10.22303/CSRID.8.1.2016.11-24

[10]. Pawar, C. C., & Ganorkar, S. R. (2016). Tuberculosis screening using digital image processing techniques. *Int Res J Eng Technol*, *3*, 623-7.

[11]. Díaz-Huerta, J. L., Téllez-Anguiano, A. D. C., Fraga-Aguilar, M., Gutierrez-Gnecchi, J. A., & Arellano-Calderón, S. (2019). Image processing for AFB segmentation in bacilloscopies of pulmonary tuberculosis diagnosis. *Plos one*, *14*(7), e0218861. doi:10.1371/JOURNAL.PONE.0218861

[12]. Alsaffar, M., Alshammari, G., Alshammari, A., Aljaloud, S., Almurayziq, T. S., Hamad, A. A., ... & Belay, A. (2021). Detection of tuberculosis disease using image processing technique. *Mobile Information Systems*, *2021*. doi:10.1155/2021/7424836

[13]. Yousefi, H., Mohammadi, F., Mirian, N., & Amini, N. (2020, December). Tuberculosis bacilli identification: a novel feature extraction approach via statistical shape and color models. In *2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA)* (pp. 366-371). IEEE. doi:10.1109/ICMLA51294.2020.00065

[14]. Rachmad, A. E. R. I., Chamidah, N., & Rulaningtyas, R. I. R. I. E. S. (2020). Mycobacterium tuberculosis identification based on colour feature extraction using expert system. *Ann. Biol*, *36*, 196-202.

[15]. Syafar, F.; Husain, H.; Ridwansyah; Harun, S.; Sokku, S. (2017). Key data and information quality requirements for asset management in higher education: A case study, *Proceedings of the 30th International Business Information Management Association Conference, IBIMA 2017 - Vision 2020: Sustainable Economic Development, Innovation Management, and Global Growth*, Vols 2017-January, 1670–1677.

[16]. Rachmad, A., Chamidah, N., & Rulaningtyas, R. (2021, February). Classification of mycobacterium tuberculosis based on color feature extraction using adaptive boosting method. In *AIP Conference Proceedings* (Vol. 2329, No. 1, p. 050005). AIP Publishing LLC. doi:10.1063/5.0042283

[17]. Haryanto, E. V., Lubis, E. L., Saleh, A., & Lubis, N. I. (2019, November). Implementation of Augmented Reality of Android Based Animal Recognition using Marker Based Tracking Methods. In *Journal of Physics: Conference Series* (Vol. 1361, No. 1, p. 012019). IOP Publishing. doi:10.1088/1742-6596/1361/1/012019

[18]. Aris, T. A., Nasir, A. S. A., Chin, L. C., Jaafar, H., & Mohamed, Z. (2020, November). Fast k-Means Clustering Algorithm for Malaria Detection in Thick Blood Smear. In *2020 IEEE 10th International Conference on System Engineering and Technology (ICSET)* (pp. 267-272). IEEE. doi:10.1109/ICSET51301.2020.9265380

[19]. Mashor, M. Y., Nasir, A. A., & Mohamed, Z. (2018, August). Identification of giemsa staind of malaria using k-means clustering segmentation technique. In *2018 6th International Conference on Cyber and IT Service Management (CITSM)* (pp. 1-4). IEEE. doi:10.1109/CITSM.2018.8674254

[20]. Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics*, *9*(1), 62-66. doi:10.1109/TSMC.1979.4310076