Segmentation Algorithm to Determine Group for Hand Gesture Recognition

Fifin A. Mufarroha, Fitri Utaminingrum, Wayan F. Mahmudy
Faculty of Computer Science, Brawijaya University, Indonesia
fayucandra@gmail.com

***Abstract*—The main principle of hand gesture is recognizing any forms of gesture in the form of alphabet letters. The goal is to help the disabled to communicate with each other. Our system runs in real time without the help of sensors, gloves, etc. With such lighting conditions, different conditions of human hand and background of shooting become a problem in the completion of the process. This research proposed a segmentation method to resolve these problems. The method begins with capturing a picture using a webcam, which is followed by the segmentation process. We also proposed several conditions of skin detection. In this research, the segmented image undergoes the extraction process, which adopts three forms of feature extraction, namely slimness, roundness, and rectangularity. The final step of the method is measuring the resemblance of the images data features using adaptive neuro fuzzy inference system.**

***Index Terms*—Hand Gesture; Segmentation; Adaptive Neuro-Fuzzy Inference System (ANFIS).**

I. Introduction

American sign language is a sign language used as a communication for people with special needs. In this case, special needs refer to those who have difficulties in communicating. Communication is the exchange of information between two parties. It is commonly done using your voice and accepted by a sense of hearing. The problem of sign language is the difficulty of normal people to understand the meaning of each form of the sign; hence, communication between normal people and the disabilities becomes a problem. This research identifies and distinguishes the hand shape in groups. We can use the form of sign as parameters to distinguish between the gesture from attitudes and things, such as changing the orientation [1]. Some human activities were conducted indoors in the translation process [2].

One of the important steps in the process of recognizing hand gestures is segmentation. Firstly, we must take a hand as an object of the image. In real time, it is a challenge to get the object of the hand in the condition of the room. We can use skin detection techniques and add another method to ease the process of obtaining the object [3]. However, adding a method is problematic as the computation process becomes longer. Another method can be implemented for the segmentation process. Hand detection has been done using contrast adjustment on each frame in a video clip. The accuracy of this method becomes a disorder when the background is converted to adaptive. The reason of this problem is the contrast adjustment method conducted by the intensity values of the object and background [4], and another field has been conducted [5] [6].

We solved these problems by proposing a segmentation method. We created conditions where we took advantage of HSV and YCbCr color channels. Considering that the segmentation process can run quickly and precisely, the process of determining the group will achieve maximum results. Adaptive neuro-fuzzy inference system (ANFIS) is a classification method for the introduction of extracted features. Few studies have been implemented ANFIS as a method of classification. We used this method to determine the group, for example, ANFIS has been proven as a sophisticated framework for the classification of multi-object drawn from the findings of people with a brain tumor [7].

II. Design System

In this study, we created a system that classifies 26 signs gestures from the American Sign Language into groups. The 26 signs of hand gesture were used as samples in the testing. Figure 1 shows the 26 American Sign Language gestures. In this research, there were several processes such as segmentation, feature extraction, and grouping.

The process began by taking pictures using a webcam. We did not use other tools such as gloves or sensor to facilitate the process of identifying the object of the hand. We resized the image to 300x400 pixels. The captured image went through a segmentation process to separate the object (hand) from the background. Next, we recommended the group of the images: An overview of the algorithm step is shown in Figure 2.

Based on the observation of the American Sign Language (ASL) as shown in Figure 1, we created three types of hand gestures, namely the finger grip, finger upwards, and finger sideways. From this analysis, we derived three groups of determination as shown in Table 1.

Table 1
Determination of Groups

Group	Alphabet
Group 1	A, E, M, N, O, S, T
Group 2	B, C, D, F, I, K, L, R, U, V, W, X, Y
Group 3	G, H, J, P, Q, Z

The aim of the grouping is to facilitate the classification process. From all these reports, the classification process was carried out within each group rather than the whole class so that the system can make the process of classification within smaller coverage.

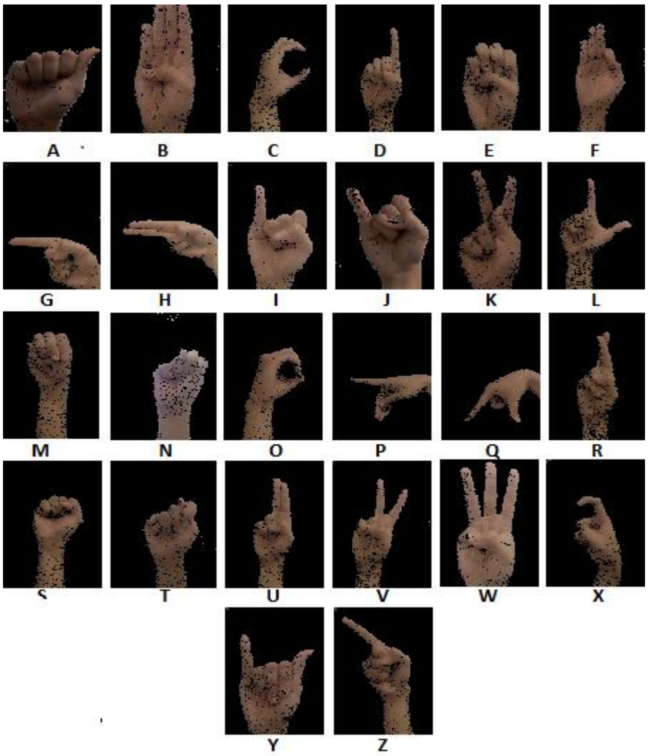


Figure 1: American sign language

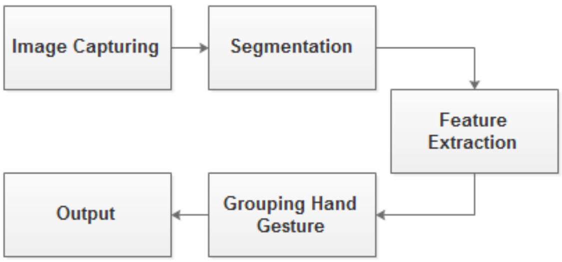


Figure 2: Overview system

III. SEGMENTATION

Segmentation is the process of separating the objects to be processed from the background. Since the object is a hand, we decided to use the techniques of skin detection. Skin detection is the technique of separating the skin with no skin object. This technique uses HSV [8] and YCbCr [9] color space. After conducting several trials using both of the color channels, we found that the channel Cb, Cr, and Hue matches the lighting conditions at the time the image has been captured. The transformation of the HSV color space can be drawn from equation (1-4).

$$Hue = \begin{cases} 0 & if\ max = min \\ \left(60^0 \times \dfrac{G-B}{max-min}\right) mod\ 360^0 & if\ max = R \\ 60^0 \times \dfrac{B-R}{max-min} + 120^0 & if\ max = G \\ 60^0 \times \dfrac{R-G}{max-min} + 240^0 & if\ max = B \end{cases} \tag{1}$$

$$Saturation = 1 - \frac{3}{R+G+B}[min] \tag{2}$$

$$value = \frac{1}{2}[(R+G+B)] \tag{3}$$

Where:

$$max = \max(R,G,B) \text{ and } min = \min(R,G,B) \tag{4}$$

YCbCr space becomes the most attractive space for skin detection, which is calculated from the RGB space as shown in Equation (5).

$$\begin{aligned} Y &= 0.299\,R + 0.587G + 0.114B \\ Cr &= R - Y \\ Cb &= B - Y \end{aligned} \tag{5}$$

The use of both color space is getting the only skin object. We created some conditions, as follows:

- If the pixels in the image are in accordance with the conditions of the channel, then it is considered to be a skin; hence, the pixels in the image will be set as 1.
- Otherwise, if it does not match with the existing values, the image pixels will be set as 0.

From these conditions, the RGB images, which have been entered will be multiplied with the changed pixels values, so the image will be transformed into a binary image consisting of 1 and 0. The details of the segmentation process can be seen in Figure 3.

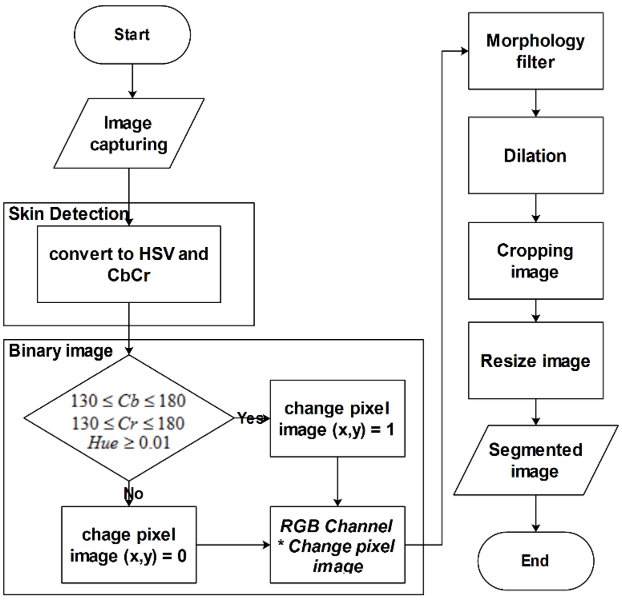


Figure 3: Proposed method.

The object derived from this step was not perfect. Therefore, we applied a morphology filter to fill the holes of the object and dilation to thicken the object on the images. If there is a noise, we can use a filtering process to remove the noise in the images [10]. By applying all of these processes, we get objects that were detected in the perfect form. Next, we cut the image according to the size of the detected object and resized the images to 300 x 400 pixels according to the size of the original image.

IV. FEATURE EXTRACTION

We used three types of shape features extraction, namely the slimness, roundness, and rectangularity used to get the feature value.

- Slimness: This feature is defined as the ratio of the length and width of the shape gesture [11].

$$slimness = \frac{L_p}{W_p} \tag{6}$$

L_p is the length image and W_p is the width image.

- Roundness is defined as :

$$roundness = \frac{4\pi A}{P^2} \quad (7)$$

  A is the area of hand image and P is the circumference of hand image.
- Rectangularity: This feature was illustrated as a technique of hand that resembles [12], which was defined:

$$rectangularity = \frac{L_p W_p}{A} \quad (8)$$

V. Group Based on Similarity of Hand Gesture

Although fuzzy logic is a not new method, the combination of fuzzy logic and the new method was just found several years ago. A pair of input-output data set used to perform the learning procedure with optimizing parameters given by the fuzzy inference system is known as ANFIS [13]. In order to derive an easy task, x and y are two inputs from fuzzy interference system and the output is f [14]. Figure 4 shows that Takagi-Sugeno-Kang (TSK) reasoning fuzzy model.

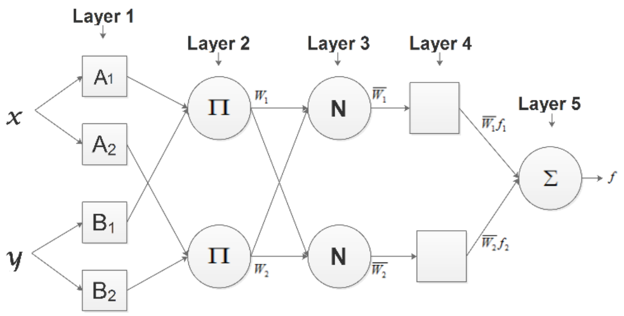


Figure 4: Takagi-Sugeno-Kang model type-3 ANFIS (Equivalent ANFIS)

This architecture describes the five layers, as follows:

Layer 1: Every node i in this layer is an adaptive node.

$$Out_i^1 = \mu A_i(x) \quad (9)$$

This node function i is Gaussian function to measure A_i membership of the activation function node as shown in equation (10):

$$\mu A_i(x) = \frac{1}{1 + \left| \frac{x - c_i}{a_i} \right|^{b_i}} \quad (10)$$

a_i, b_i and c_i are the premise parameters and they conform the shape and the location of the membership function.

Layer 2: Activation degree (firing strength) for every fuzzy rule implemented in every node output. The output is the multiplication of all inputs entered in this layer.

$$Out_i^2 = w_i = \mu A_i(x) \times \mu B_i(y) \quad (11)$$

Layer 3: Normalized firing strength. The i-th node calculates the ratio of the i-th rule's firing strength to the sum of all rules' firing strengths as shown in equation (12).

$$Out_i^3 = \overline{w}_i = \frac{w_i}{w_1 + w_2} \quad (12)$$

Layer 4: Each node in this layer is an adaptive node. The output of the layer 3 is $\overline{w_i}$ and the parameters set are $\{p_i, q_i, r_i\}$. The parameters in this layer will be referred as parameters consequent as shown in Equation (13).

$$Out_i^4 = \overline{w}_i f_i = \overline{w}_i (p_i x + q_i y + r_i) \quad (13)$$

Layer 5: The summary of all the entries that calculate the output of the fuzzy system.

$$Out_i^5 = f = \sum_{i=1}^{4} \overline{w}_i f_i = \frac{\sum_{i=1}^{4} w_i f_i}{\sum_{i=1}^{4} w_i} \quad (14)$$

In this research, the segmented image was derived from the feature values in the previous process. Based on the three types of features, they were then grouped by ANFIS.

VI. Experiment And Results

The segmented images are the result of applying a segmentation algorithm that we proposed, as shown in Table 2. Based on the image segmentation, the image can be segmented perfectly without any noise and hole in the object of the hand. The next phase is cropping and resizing the image from the segmented image into 300x400 pixel.

Research conducted in [4] created a proposed method using contrast adjustment for segmentation process. Consider some background conditions in the image, the proposed method has a risk that the results obtained were not reliable. The background and the object in the image determine the accuracy of the results, in which all of them were located from the value of its intensity. The comparison between the proposed method and the segmentation method [4] is shown in Table 3 and Table 4.

Testing for comparison method was conducted in two sessions, applying both methods with different data. The first session applied data in [4] and the second session used data in this research. The results of the first session showed the method from research [4] and the proposed method can solve the segmentation process with great results, as seen in Table 3.

Table 2
Implementation Proposed Algorithm of Segmentation Process

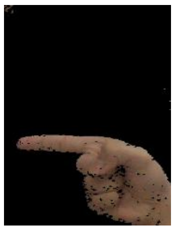
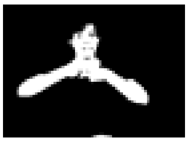
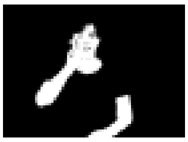
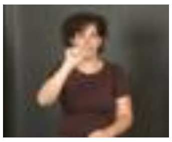
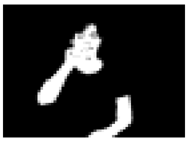
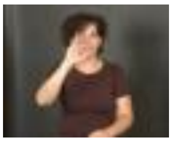
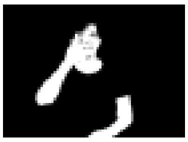
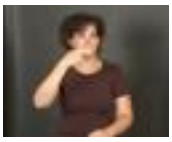
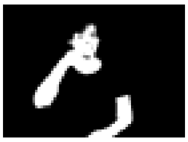
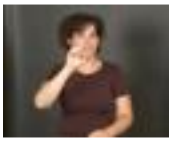
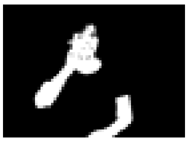
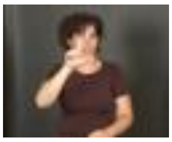
Image captured	Proposed algorithm	Segmented image
		
		

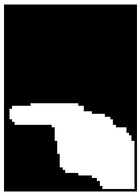
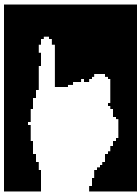
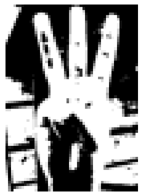
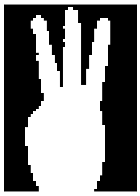
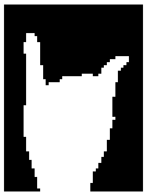
Table 3
Comparisons Using Data on Research [4]

ASL	Original Image	Segmentation method [4]	Proposed method
My			
Name			
R			
A			
C			
H			
E			
L			

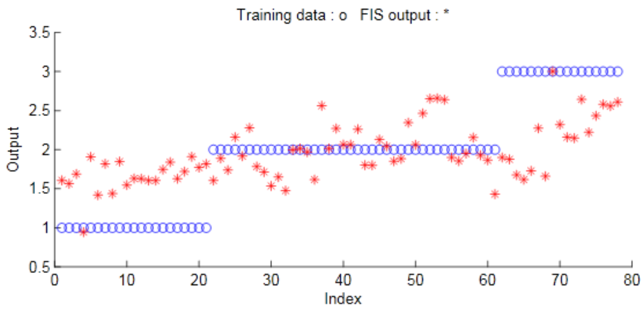
As shown in Table 4, the results of the testing conducted in the second session show dissimilarity when using the data in this research. The result of the contrast adjustment method depends on the background and the object in the image, resulting in the object that is less than perfect. Otherwise, the method that we proposed is powerful on adaptive background because we applied the skin detection techniques. Our method searches for the value that is near to the value of the skin.

Another experiment performed the classification phase from the result of the segmentation process, involving 78 training data and 78 testing data. The training data include three images with every form from the hand sign ANFIS. The results from the simulation result of the hand form category input that was based on three feature categories. This study performed membership function as Gaussian with epoch numbers, which are 5, 10, and 1000, and one membership function set as 3.

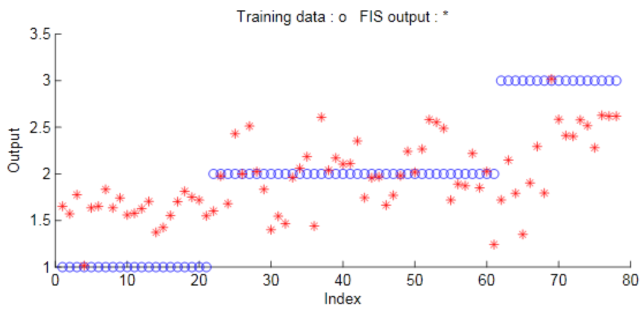
Table 4
Comparisons Using Data on This Research

ASL	Original Image	Segmentation method [4]	Proposed method
G			
I			
W			
Y			

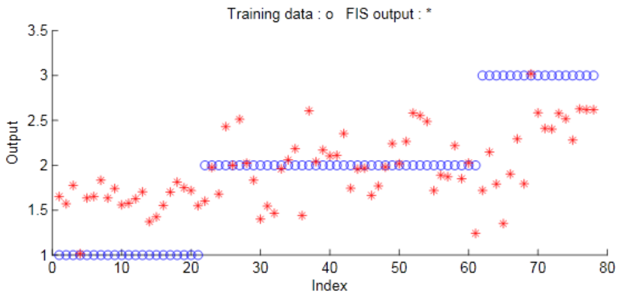
Learning data distribution with epoch number can be seen in Figure 5, whereas the output of training is divided into three classes grouping, as shown in blue color and red color that represents the distribution of data. 78 data were distributed based on its value. These values were located at the target group.



(a)



(b)



(c)

Figure 5: Output of Training ANFIS with (a) number of epochs 5, (b) number of epochs 10, (c) number of epochs 1000

The performance of both the training and testing sets were evaluated by the following measure of mean square error (MSE) and model efficiency (EFF) [15]. The mean square error (MSE) measures mistake of the means quadrate. Smaller values from MSE ensure a better performance. MSE was calculated as shown in Equation (15) [16]:

$$MSE = \frac{1}{k}\sum_{n=1}^{k}(p_n - q_n)^2 \tag{15}$$

Table 5 shows the result of the use of ANFIS on the learning data with the 78 samples image. The testing detail on every group is described in Table 6. The results of experiments conducted on learning data showed that the number of epoch 5 and 10 have the same MSE. However, when we raise drastically the number of digits epoch until 1000, the results showed that there was a decrease of MSE. The result of testing, as shown in Table 7 indicates that the best accuracy in the number epoch 1000 is 87.18%

Table 5
Accuracy Rate of Training

Scenario	Number of epoch	MSE	Group samples	EFF
1	5	1.28	68	87.18%
2	10	1.28	68	87.18%
3	1000	1.04	69	88.46%

Table 6
The result of grouping using ANFIS

Group	Alphabet	#of samples	Number of epoch		
			5	10	1000
1	a, e, m, n, o, s, t	21	16	16	19
2	b, c, d, f, i, k, l, r, u, v, w, x, y	39	36	36	36
3	g, h, j, p, q, z	18	11	11	13
	Aggregate		63	63	68

Table 7
Accuracy rate on testing

Scenario	Number of epoch	MSE	Group samples	EFF
1	5	2.88	63	80.77%
2	10	2.88	63	80.77%
3	1000	1.28	68	87.18%

VII. Conclusion

The implementation of the proposed algorithm by utilizing HSV and YCbCr color channels shows that the proposed algorithm has good results and efficient computation time where the hand image can be segmented properly. Considering several issues such as lighting, background and without tools, the algorithm that we proposed obtains a good result.

The experiments were performed using 78 training images and 78 testing images. We chose a simple feature extraction method by considering the rapid computation time but the method influenced the accuracy result on the hand gesture identification. This result is caused by the application of extraction feature that used three different method types, which were still very weak in the implementation, even when using adaptive neuro-fuzzy inference system as the classification method. Future research is to continue this research by conducting classification process and the result is the alphabet letters.

References

[1] M. Panwar, "Hand gesture recognition based on shape parameters," *Int. Conf. Comput. Commun. Appl.*, pp. 1–6, 2012.

[2] A. Jalal, M. Uddin, and T. S. Kim, "Depth video-based human activity recognition system using translation and scaling invariant features for life logging at smart home," *IEEE Trans. Consum. Electron.*, vol. 58, no. 3, pp. 863–871, 2012.

[3] A. A. Randive and S. D. Lokhande, "Hand Gesture Segmentation," *Int. J. Comput. Technol. Electron. Eng.*, vol. 2, no. 3, pp. 125–129, 2012.

[4] A. Y. Dawod, M. J. Nordin, and J. Abdullah, "Gesture Segmentation : Automatic Continuous Sign Language Technique Based on Adaptive Contrast Stretching Approach Pattern Recognition Research Group , Centre for Artificial Intelligence Technology (CAIT), Assistive Technology SIG , Faculty of Computing," *Middle-East J. Sci. Res.*, vol. 24, no. 2, pp. 347–352, 2016.

[5] H. S. Abdulbaqi, M. Z. M. Jafri, A. F. Omar, K. N. Mutter, L. K. Abood, and I. S. Bin Mustafa, "Segmentation and estimation of brain tumor volume in computed tomography scan images using hidden Markov random field Expectation Maximization algorithm," *2015 IEEE Student Conf. Res. Dev.*, vol. 8, no. 3, pp. 55–60, 2015.

[6] Z. Zainal Abidin *et al.*, "Brain Lesion Segmentation from Diffusion-weighted MRI based on Adaptive Thresholding and Gray Level Co-occurrence Matrix Faculty of Electrical Engineering , Universiti Teknikal Malaysia Melaka ," *2015 IEEE Student Conf. Res. Dev.*, vol. 8, no. 2, pp. 41–48, 2011.

[7] M. Sharma and S. Mukharjee, "Artificial Neural Network Fuzzy Inference System (ANFIS) For Brain Tumor Detection," *Int. J. Comput. Appl. Technol. Res.*, vol. 3, no. 3, pp. 150–154, 2014.

[8] S. Manjare and S. . Chougule, "Skin Detection for Face Recognition Based on HSV Color Space," *Int. J. Eng. Sci. Res. Technol.*, vol. 2, no. 7, pp. 3–7, 2013.

[9] A. N. Ghomseh, "Pixel-based Skin Detection Based on Statistical Models," *J. Telecommun. Electron. Comput. Eng.*, vol. 8, no. 5, pp. 7–14, 2016.

[10] F. Utaminingrum, K. Uchimura, and G. Koutaki, "Mixed gaussian and impulse noise removal based on kernel observation and edge direction," *Int. J. Innov. Comput. Inf. Control*, vol. 11, no. 5, pp. 1509–1523, 2015.

[11] Q. Wu, C. Zhou, and C. Wang, "Feature Extraction and Automatic Recognition of Plant Leaf Using Artificial Neural Network," *Av. en Ciencias la Comput.*, pp. 5–12, 2006.

[12] K. Singh, I. Gupta, and S. Gupta, "SVM-BDT PNN and Fourier Moment Technique for Classification of Leaf Shape," *Int. J. Signal Process. Image Process. Pattern Recognit.*, vol. 3, no. 4, pp. 67–78, 2010.

[13] G. D. Santika, W. F. Mahmudy, and A. Naba, "Electrical Load Forecasting using Adaptive Neuro-Fuzzy Inference System," *Accept. IJASCA*, pp. 1–20, 2016.

[14] J. S. R. Jang, "ANFIS: Adaptive-Network-Based Fuzzy Inference System," *IEEE Trans. Syst. Man Cybern.*, vol. 23, no. 3, pp. 665–685, 1993.

[15] J. E. Nash and J. V. Sutcliffe, "River flow forecasting through conceptual models part I - A discussion of principles," *J. Hydrol.*, vol. 10, no. 3, pp. 282–290, 1970.

[16] A. A. M. Ahmed and S. M. A. Shah, "Application of adaptive neuro-fuzzy inference system (ANFIS) to estimate the biochemical oxygen demand (BOD) of Surma River," *J. King Saud Univ. - Eng. Sci.*, p. , 2015.