Dynamic Indonesian Sign Language Recognition by Using Weighted K-Nearest Neighbor

***Abstract*— People with disability in Indonesia is 2.45% of the total population. And people with hearing disability are the second largest. Usually, they have a problem with communication. A sign language is created to solve the problem. However, it is less popular so only limited people understand it. Therefore, a system which able to recognise sign language is required. This research introduces dynamic feature extraction to recognise sign language. Dynamic features were used to capture trajectory movement of hand skeleton. Because sign languages are characterized not only by hand movement but also by hand position, besides dynamic features this research also used hand position feature. For classification, this study used a combination of Weighted Simple Matching Coefficient (WSMC) and K-Nearest Neighbors (KNN). Later combination between Weighted Simple Matching Coefficient (WSMC) and K-Nearest Neighbors (KNN) is called by Weighted K-NNN. For the experiment, eighteen sign language gestures were tested. Tree kinds of K-NN was applied in testing; K-NN; locally weighted K-NN; and globally weighted K-NN. The recognition accuracy of this system, although evaluated with a limited vocabulary, presents very promising result with value 88% by using locally weighted K-NN.**

Keywords—Dynamic Sign Language; Weighted K-NN; Dynamic Feature.

I. INTRODUCTION

Disability is a condition where people unable to carry out a certain activities or act as a normal person which caused by a condition of impairment (loss/inability). People with disabilities are the largest minority group in the world, of which 80% of them live in developing countries including Indonesia. In Indonesia, susenas 2012 conducted by data and information centre in Ministry of Health stated that the number of person with a disability is 2.45% of the total population in Indonesia. Which are people with hearing disability are the second largest [1].

Usually, people with the hearing problem also have problem with communication. Sign language is one of the solutions to solve the problem.Therefore, in 1994, Ministry of Culture and Education released SIBI (*Sistem Isyarat Bahasa Indonesia*). SIBI is formal Indonesian Sign Language. It consists of various finger position and hand gesture movement to represent Indonesian vocabulary [2].

However, SIBI is less popular. Only limited people can understand it. So the communication area of them who suffer from hearing disability is limited. A system which able to recognise SIBI is required. Research about recognising sign language are conducting in many countries [3] [4] [5].

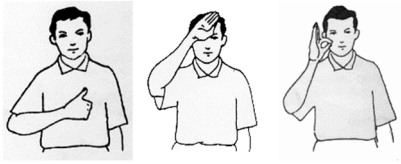
In Indonesia, research about recognising Indonesian Sign Language started in 2013 by Rakun et al. They combine depth image and skeleton data in their research [2]. Khotimah et al. conducted another research on SIBI recognition. However, they only focus on static sign language, therefore, the sign languages which can be translated are limited [6]. However, some sign languages have dynamic gestures. Therefore, this research proposed dynamic feature extraction for recognising SIBI. In another side, the features extracted in this research are categorical features. And classifying categorical features are more complex than classifying numerical features. KNN is suitable for classifying categorical features. However, normal K-NN assume every attribute has equal contribution. In the reality, some attributes may have more contribution than others. Therefore, in this study, we used weighted K-NN, combination between Weighted Simple Matching Coefficient (WSMC) and K-Nearest Neighbors (KNN), for the classification process.

The organisation of this paper is as follows. In section II and III, we present the literature review and methods that needed for developing the application. Section IV discusses the experiment and analysis. The last, in section V conveys the conclusion of this proposed research.

II. LITERATURE REVIEW

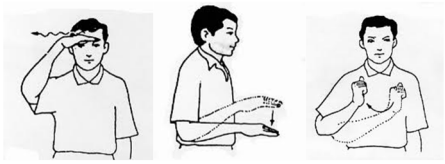
This part discusses the research related to the areas of Indonesian Sign Language, Dynamic Feature Extraction, WMSC and KNN.

A. Indonesian Sign Language



(1) (2) (3)

Fig. 1 SIBI with static gesture



(1) (2) (3)

Fig. 2 SIBI with dynamic gesture

Indonesian Sign Language is popular with SIBI (*Sistem Isyarat Bahasa Indonesia*). SIBI has two kinds of gestures, static gesture and dynamic gestures. A static gesture is a gesture that does not involve any movement in its delivery. The example was shown in Fig.1.While dynamic gesture has movements in its delivery which is shown in Fig. 2.

B. Dynamic Feature Extraction

Dynamic features are characteristics that can describe the trajectory of hand gestures. Chen et al. [7] extracted dynamic features from orientation angle of hand gestures to XOZ plane system. The hand gestures were captured by Kinect. In each frame, the position of hand skeleton $H(x_t,y_t,z_t)$ was received and saved. The trajectory in x-axis and y-axis in two sequence frames was calculated by using (1) and (2) consecutively. Then the orientation of hand actions was computed using (3).

$$\Delta x = x_t - x_{t-1} \tag{1}$$

$$\Delta y = y_t - y_{t-1} \tag{2}$$

$$\alpha_t = \begin{cases} \arctan\left(\frac{\Delta y}{\Delta x}\right) * \left(\frac{180}{\pi}\right) + 180, & \Delta x < 0 \\ \arctan\left(\frac{\Delta y}{\Delta x}\right) * \left(\frac{180}{\pi}\right) + 360, & \Delta y < 0 \\ \arctan\left(\frac{\Delta y}{\Delta x}\right) * \left(\frac{180}{\pi}\right), & \Delta x > 0, \Delta y \geq 0 \end{cases} \tag{3}$$

Where x_t is the position of hand skeleton in the x-axis in time *t*, x_{t-1} is the position of hand skeleton in the x-axis in time *t-1*, y_t is the position of hand skeleton in the y-axis in time *t* and y_{t-1} is the position of hand skeleton in the y-axis in time *t-1*.

The example of orientation computed in ten frames movement of one hand skeleton was shown in Table 1. Given the trajectory direction, the features were calculated by transforming the orientation into a serial category by using transformation rule that was shown in Fig.3 and was explained in Table 2.

TABLE 1 THE EXAMPLE OF HAND SKELETON POSITION AND ITS TRAJECTORY ORIENTATION

Frame(t)	x	y	Δx	Δy	αt
1	1	5	0	0	0
2	2	6	1	1	23
3	4	9	2	3	28
4	6	4	2	-5	326
5	7	5	1	1	23
6	9	8	2	3	28
7	7	13	-2	5	350
8	10	17	3	4	27
9	12	18	2	1	13
10	13	19	1	1	23

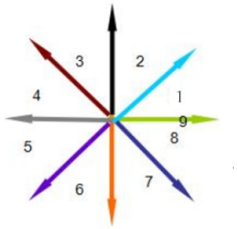


Fig. 3 Transformation Rule

TABLE 2 TRANSFORMATION RULE

Rule No.	αt	Feature Value
1	αt = 0	9
2	αt > 314	8
3	αt > 269	7
4	αt > 224	6
5	αt > 179	5
6	αt > 134	4
7	αt > 89	3
8	αt > 44	2
9	αt > 0	1

C. WSMC (Weighted Simple Matching Coefficient)

WSMC is a statistical method which used for comparing the similarity and differences of every attribute in data sets. In WSMC, every attribute is not compared equally, instead, each has its own weight which will influence classification process. Attribute with higher weight will give greater contribution to the classification than those with lower weight. This method is sufficient to classify categorical data [8].

There are two kinds of WSMC; global WSMC ($WSMC_{global}$) and local WSMC ($WSMC_{local}$). The difference between the two is the process of computing the weight. In $WSMC_{global}$, the weight is computed globally. Here, each attribute correspondence to weight vector $\varpi = (\omega_1, \dots, \omega_d, \dots, \omega_D)$ where *D* is the number of the attribute. A distance between two data is computed in (4).

$$WSMC_{global}(x_i, x_j, \varpi) = \sum_{d=1}^{D} \omega_d \times I(x_{id} \neq x_{jd}) \tag{4}$$

Where *I*(.) is an indicator function. If $x_{id}=x_{jd}$ then $I(x_{id}=x_{jd}) = 1$ and if not $I(x_{id}\neq x_{jd}) = 0$, and ω_d is unique weight of attribute *d*. ω_d is calculated by using *Global Gini diversity index* (*GG*) that shown in (5).

$$\omega_d^{(GG)} = e^{\frac{M}{M-1}\sum_{s_d \in S_d} p(s_d) \times GG(s_d)} \tag{5}$$

Where s_d is the category value of attribute *d*. *M* is the number of the class in the data set. $p(s_d)$, which is computed using (6), is the degree of influence of attribute s_d to all data. Its value is between 0 and 1, $0 \leq p(s_d) \leq 1$.

$$p(s_d) = \frac{1}{N}\sum_{(x,y)\in tr} I(x_d = s_d) \tag{6}$$

Where *N* and *tr* are the number of data. And $GG(s_d)$, Global Gini diversity index of category s_d in attribute *d* can be computed by using (7) and $p(m|s_d)$ is defined in (8).

$$GG(s_d) = 1 - \sum_{m=0}^{M} [p(m|s_d)]^2 \tag{7}$$

$$p(m|s_d) = \frac{\sum_{(x,y)\in c_m} I(x_d = s_d)}{\sum_{(x,y)\in tr} I(x_d = s_d)} \quad (8)$$

In the local method, $WSMC_{local}$, the goal is giving weight based on the class of each attribute. Thi weight indicates a contribution of each attribute to the different class. In this method, each class has its own weight vector. The weight vector of class m is $\boldsymbol{w}_m = (\boldsymbol{w}_{m1}, \dots, \boldsymbol{w}_{md}, \dots, \boldsymbol{w}_{mD})$ with $0 \leq \boldsymbol{w}_{md} \leq 1$ and $d = 1,2,\dots D$. $WSMC_{local}$ distance between two instances is shown in (9). $w_{md}^{(LG)}$, local weight computed by Local Gini diversity index $LG(m,d)$, is shown in (10). C_m is the m^{th} class and $|C_m|$ is the number of sample in class m.

$$WSMC_{local}(x_i, x_j, w_m) = \sum_{d=1}^{D} w_{md}^{(LG)} \times I(x_{id} \neq x_{jd}) \quad (9)$$

$$w_{md}^{(LG)} = e^{\frac{|S_d|}{|S_d|-1} \times LG(m,d)} \quad (10)$$

$$LG(m,d) = 1 - \sum_{s_d \in S_d} [p(s_d|m)]^2 \quad (11)$$

$$\boldsymbol{p}(s_d|m) = \frac{1}{|C_m|} \sum_{(x,y)\in C_m} I(x_d = s_d) \quad (12)$$

D. KNN (K- Nearest Neighbors)

K-nearest neighbors (KNN) is an algorithm that can be used for classification. Besides for classification, this algorithm also can be used for regression. In classification purpose, KNN works by saving all information of training data and classify testing data by computing the similarity between the testing data and training data. Testing data will be labelled based on the label of its k nearest neighbour [9]. Where k is a small positive number.

III. METHODS

This research used Kinect 2.0 to collect data set. The process of collecting data set was shown in Fig.4. Each sign language gesture is recorded for 40 frames. Four skeletons point position in each frame is captured and saved. These skeletons are Hand Right (HR), Hand Left (HL), Neck (N), and Spine Mid (SM). Their position is shown in Fig.5.

A. Feature Extraction

Two kinds of features are used in this research, dynamic features that characterized the movements of hand gestures and hand position which illustrated the position of the hand, right and left hand, during giving sign language.

In dynamic feature extraction process, the position of left palm skeleton $LH(x_t, y_t, z_t)$ and right palm skeleton $RH(x_t, y_t, z_t)$ in each frame is received and saved. This process is repeated for 40 frames. Then the difference position of RH and LH in x-axis and y-axis between frame t and $t+2$ is computed using (1) and (2) consecutively. Their directions are computed using (3) and are quantized by using rule shown in Table 2.The result is 20 features for each skeleton. However, the first 10% features are omitted because unstable.Therefore only 18 features remain. Because the signer used two hands, as a result 36 features are received from dynamic feature extraction process.

For hand position features, this research divides the position of the hand into three groups: top area (Area 1), middle area (area 2), and bottom area (Area 3). Those groups can be seen in Fig. 6. In this research, the hand position features only characterise the position of RH or LH position on the y-axis at frame 20. Rules in Table 3 is used to get this features. Where N refers to neck skeleton and SM refers to spine mid skeleton. For summary, 38 features are used in this research. The first eighteen features are dynamic features of the right hand; the second 18 features are left-hand dynamic features; one feature of right-hand position; and the last one feature is left-hand position. Feature number 1 to 36 are categorical feature with value from 1 to 9. Feature number 37 and 38 are also categorical feature with value either area 1; area 2; or area 3.

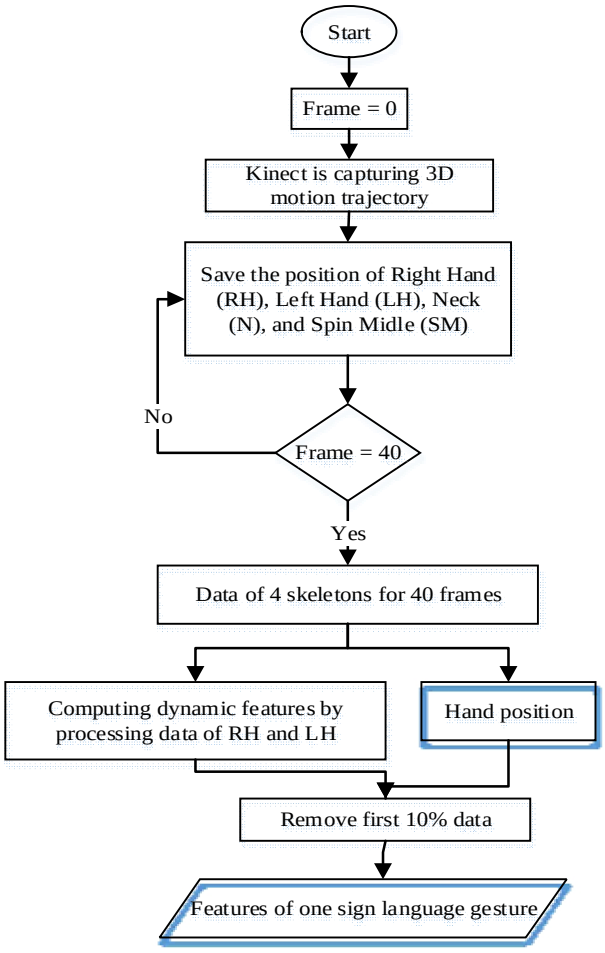


Fig. 4 Collecting Data Set Process

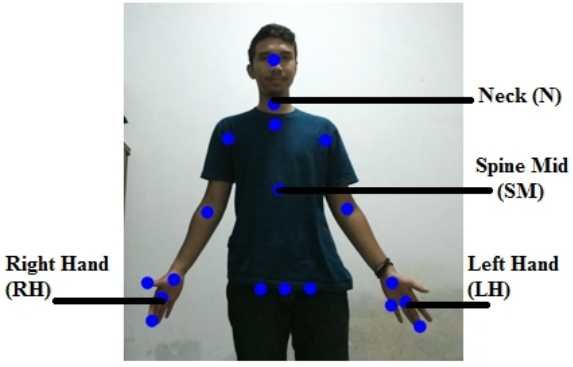


Fig. 5 Skeletons Position

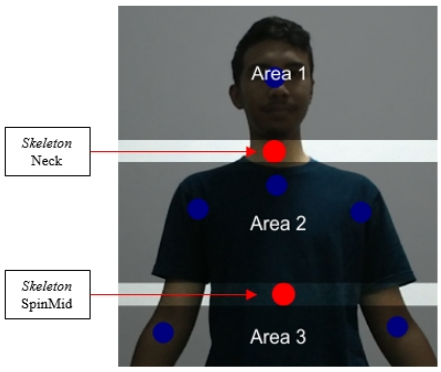


Fig. 6 Hand position group area

TABLE 3 RULES FOR HAND POSITION AREA

No	Rules	Result
1	*RH* position > *N* position	Right hand is in area 1
2	*LH* position > *N* position	Left hand is in area 1
3	*RH* position < *N* position AND *RH* position > *SM* position	Righ hand is in area 2
4	*LH* position < *N* position AND *LH* position > *SM* position	left hand is in area 2
5	*RH* position < *SM* position	Right hand is in area 3
6	*LH* position < *SM* position	Left hand is in area 3

B. Classification using K-NN with WSMC

K-NN (K-Nearest Neighbors) with WSMC is an improved method of k-NN with SMC. Instead of using the same weight in each attribute, k-NN with WSMC are using a different weight which contributes to the classification process. There are two kinds of weighting in k-NN with WSMC utilized in this research: Global Weighting (GG) and Local Weighting (LG) that are computed using (5) and (10).

In training process, given the training data, the global weight and the local weight of each attribute are computed by using (5) and (10). Their value is saved in two different matrices. The size of global weight matrix is *1xD*. Whereas, the size of local weight matrix is *MxD*. Where *M* and *D* is the number of class and the number of attributes consequently.

In the testing process, the distance between the test data and each training data is computed using (4) for global weight method and is computed using (9) for local weight method. The result is sorted ascendingly then *k* nearest neighbours were selected. The label of the test data is from voting of the selected neighbours. The algorithm of testing k-NN with WSMC was shown in Fig.7.

Input : *tr* (training data), *k*, test data z = (x,y)
Output : *y* (label prediction)
Begin
If local weighting
Compute $WSMC_{local}$ *(z, x_i, w)*, between test data and all training data (x_i) using (9)

If global weighting
Compute $WSMC_{global}$ *(z, x_i, w)*, between test data and all training data (x_i) using (4)
end
Select *k* nearest neighbours from training data (NNz)
//do voting for label prediction
Output y = $\text{argmax}_m \sum_{(x_j,y_j)\in NN_z} I(m = y_j)$
End

Fig. 7 Testing K-NN with WSMC

IV. EXPERIMENT AND ANALYSIS

A. Data Set

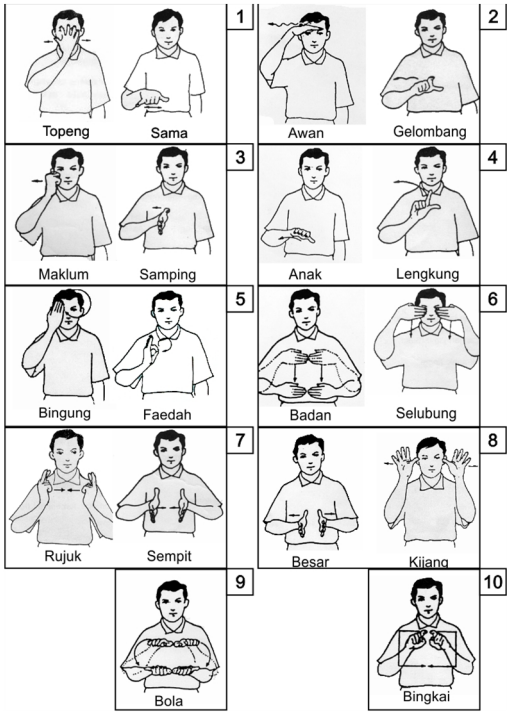


Fig. 8 The testing data class

The data set used in this research was collected by using Kinect. Eighteen sign languages were collected from a people. All of them have dynamic features. These are Mask (*Topeng*), Same (*Sama*), Cloud (*Awan*), Wave (*Gelombang*), Informed (*Maklum*), Side (*Samping*), Child (*Anak*), Curve (*Lengkung*), Confused (*Bingung*), Benefit (*Faedah*), Body (*Badan*), Sheath (*Selubung*), Reconciliation (*Rujuk*), Narrow (*Sempit*), Big (*Besar*), Deer (*Kijang*), Ball (*Bola*), and Frame (*Bingkai*). The reason for choosing them is because they have variety in gestures. Such as:

one handed movement forming

- a straight line against x-axis (e.g., mask and same),
- a straight line of two directions against x-axis (e.g., informed and side),
- arch (e.g., child and curve),

- waves (e.g., cloud and wave),
- a circle (e.g., confused and benefit),

two handed movements forming

- a straight line against y-axis (e.g., body and sheath),
- a straight line and two hands approaching each other in x-axis (e.g., reconciliation and narrow),
- a straight line and the hands stay away each other against the x-axis (e.g., big and deer),
- a circle (e.g., ball),
- a square (e.g., frame).

Their gestures were shown in Fig.8. Each sign was repeated for 20 times. Totally, 360 data were collected.

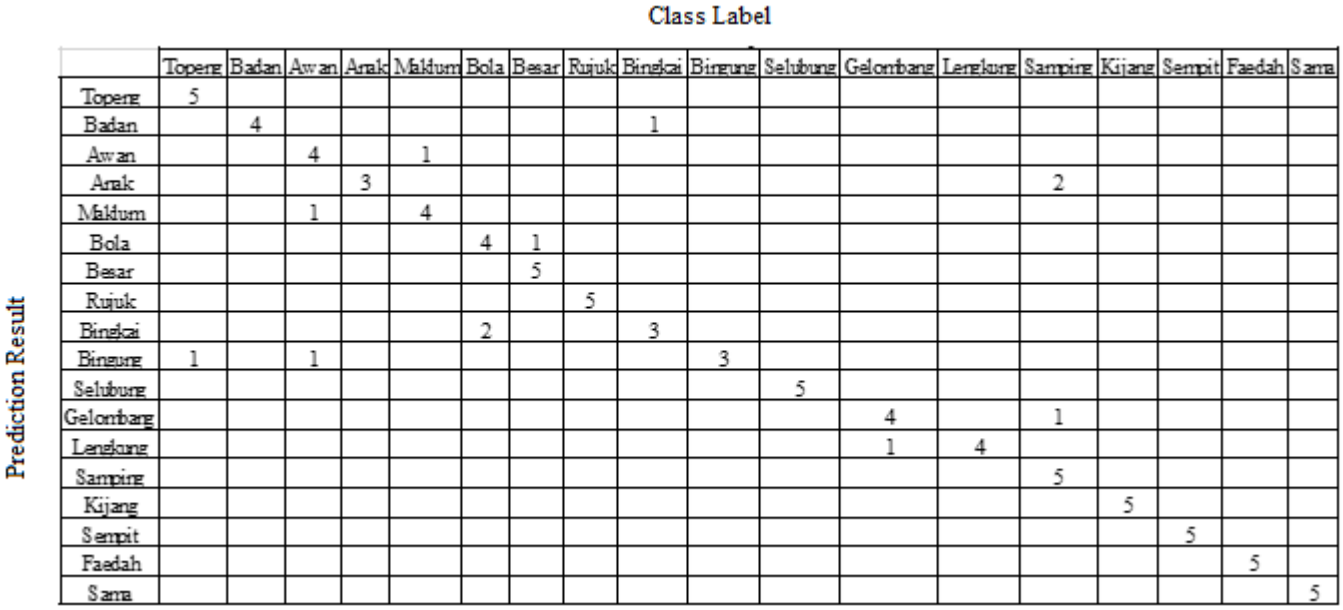


Fig. 9 Confusion Matrix of Testing using local weighting

B. *Experiments Result*

For experiment 80% of the data were used for training and others are for testing. Tree kinds of K-NN was applied in testing; K-NN; locally weighted K-NN; and globally weighted K-NN. Locally weighted K-NN give the highest accuracy around 88%. The datail result is shown in confusion matrix Fig.9. While the accuracy of this study using globally weighted K-NN and KNN aare 74% and 71% respectively. Local weighting is better than global weighting in this study.

C. *Analysis*

Some reasons that may affect the classification are:

1. Some sign languages have similar gesture which makes classification worse. For example gesture of frame and ball, side and wave. Their positions are nearly the same. Their different only one gesture is straight while another is curving, whereas, Kinect is a very sensitive tool. It means if a user intends to make a straight movement but their hand vibrates, their movement can be interpreted as either curve movement or wave movement. And this problem makes misclassification occurred.
2. The function *I*(.) in WSMC distance compute a distance between two data with 0 or 1. 0 if the data is different and one otherwise. In another side, the value of each attribute has closeness meaning. For example feature value ‘area 1’ is closer to feature value ‘area 2’ than to feature value ‘area 3’.
3. Consequence features may have a relationship because they are computed base on time point. The speed of giving sign among people is different. However, KNN with WSMC ignored the relationship between consecutive features. An algorithm that cares with the relation between serial features may improve the accuracy.

V. CONCLUSION

This research aims to contribute to the improvement of a communication system of people with speech and/or hearing disabilities. The recognition module, although evaluated with a limited vocabulary, presents promising result with accuracy 88%. However, some improvement can be done to raise the accuracy. Those improvements include adding other features, using distance that cares to closeness of the categorical value and using algorithm that cares about relationship among sequence features

ACKNOWLEDGMENT

This research was supported by a research grant for "Penelitian Pemula" in 2017 from Institut Teknologi Sepuluh Nopember Surabaya.

REFERENCES

[1] A. Diana, Mujaddid, F. A. Prasetyo, and D. Budijanto, "Jendela Data dan Informasi Kesehatan," vol. II, 2014.

[2] E. Rakun, M. Andriani, I. W. Wiprayoga, K. Danniswara, and A. Tjandra, "Combining depth image and skeleton data from Kinect for recognizing words in the sign system for Indonesian language (SIBI [Sistem Isyarat Bahasa Indonesia])," in *2013 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, 2013, pp. 387–392.

[3] M. Maraqa, F. Al-Zboun, M. Dhyabat, and R. A. Zitar, "Recognition of Arabic Sign Language (ArSL) Using Recurrent Neural Networks," *J. Intell. Learn. Syst. Appl.*, vol. 04, no. 01, pp. 41–52, 2012.

[4] S. G. Moreira Almeida, F. G. Guimarães, and J. Arturo Ramírez, "Feature extraction in Brazilian Sign Language Recognition based on phonological structure and using RGB-D sensors," *Expert Syst. Appl.*, vol. 41, no. 16, pp. 7259–7271, Nov. 2014.

[5] C. Sun, T. Zhang, B. K. Bao, C. Xu, and T. Mei, "Discriminative Exemplar Coding for Sign Language Recognition With Kinect," *IEEE Trans. Cybern.*, vol. 43, no. 5, pp. 1418–1428, Oct. 2013.

[6] W. N. Khotimah, Y. A. Susanto, and N. Suciati, "Combining Decision Tree and Back Propagation Genetic Algorithm Neural Network for Recognizing Word Gestures in Indonesian Sign Language using Kinect," *J. Theor. Appl. Inf. Technol.*, vol. 95, no. 2, pp. 292–298, Jan. 2017.

[7] Y. Chen, B. Luo, Y. L. Chen, G. Liang, and X. Wu, "A real-time dynamic hand gesture recognition system using kinect sensor," in *2015 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, 2015, pp. 2026–2030.

[8] L. Chen and G. Guo, "Nearest neighbor classification of categorical data by attributes weighting," *Expert Syst. Appl.*, vol. 42, no. 6, pp. 3142–3149, Apr. 2015.

[9] A. Suárez Sánchez, F. J. Iglesias-Rodríguez, P. Riesgo Fernández, and F. J. de Cos Juez, "Applying the K-nearest neighbor technique to the classification of workers according to their risk of suffering musculoskeletal disorders," *Int. J. Ind. Ergon.*, vol. 52, pp. 92–99, Mar. 2016.