Journal of Theoretical and Applied Information Technology
15th October 2022. Vol.100. No 19

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

COMPARATIVE ANALYSIS OF CLASSIFICATION METHODS OF THE DACTYL ALPHABET OF THE KAZAKH LANGUAGE

NURZADA AMANGELDY [1], SAULE KUDUBAYEVA [1], BIBIGUL RAZAKHOVA [1], MUKANOVA ASSEL[2] , TURSYNOVA NAZIRA[1]

[1] Faculty of Information Technologies, L.N. Gumilyov Eurasian National University, Nur-Sultan 010008, Kazakhstan;

[2] Higher School of Information Technology and Engineering, Astana International University, Nur-Sultan 010000,Kazakhstan

E-mail: nurzadaamangeldy@gmail.com , saule.kudubayeva@gmail.com, razakhova_bsh@enu.kz, ntursynova000@gmail.com, [2] asiserikovna@gmail.com

ABSTRACT

Kazakh Sign Language (KSL) is relatively new, and there is still no full-fledged system that can automatically recognize the Kazakh language's dactylic alphabet since Russian letters in the Kazakh alphabet have been identified until today. In order to prove that the Kazakh sign language can exist as a separate sign language, a comparison of the form of the display was carried out in terms of configuration (arm /forearm), place of execution (localization), the direction of movement, nature of movement and component that cannot be performed manually (facial expression and articulation) of Kazakh, Russian, English, Turkish sign languages. As a result of the study, it was proved that the Kazakh sign language can exist as a separate sign language by observing the forms of demonstration of 1050 words in 4 languages. Further, a comparative analysis of classification methods such as support vector machines, ensemble classifiers and nearest neighbor classifiers of dactyl alphabet letters was carried out to identify the optimal method.

Keywords: *Sign Language; Hand Shape; Palm Definition Model; Mediapipe Hands; Support Vector Machines; Ensemble Classifiers; Nearest Neighbor Classifiers Pattern; Recognition; Multiple Classification.*

1. INTRODUCTION

In order to prove that the Kazakh sign language can exist as a separate sign language, it is necessary to conduct a comparative analysis with Russian, English, and Turkish. In Kazakhstan, the number of dead or hard of hearing is about 1.6% of the total population, which is about 300 thousand people. Sign language differs from spoken speech, which emphasizes sound and verbal aspects.

Sign language (SL) is a non-verbal communication system between hearing people and people with hearing problems. For the latter, it is used as the main method of communication in which almost every word can be found as a sign correspondence [1,2]. The basic unit of sign language is a gesture characterized by iconicity, i.e., the ability to designate an object by demonstrating, with the help of hand movements, facial expressions and articulation of the face, head turns, etc., visual parameters of the object.

With the help of sign language, in most cases, it is impossible to convey first and last names and foreign, technical, and medical definitions. As a result, the deaf extensively use the dactyl alphabet along with sign language. The grammar of the dactyl language is similar to the native language of the deaf. Dactylology is correctly called writing with fingers in the air: it perceives visually and as written speech obeys all the spelling rules. But not punctuation: exclamation and question are conveyed by appropriate facial expressions; dot and ellipsis by a pause; dash, colon, and other punctuation symbols,

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

but also have their designation, are not indicated in dactylic writing.

There are two sign languages in Kazakhstan, Kazakh Calcifying Sign Language (KCSL) and Kazakh Sign Language (KSL). KCSL converts spoken Kazakh into sign language and follows the grammatical structure of Kazakh spoken language with prefixes and suffixes. Unlike KCSL, KSL translates one word from the Kazakh spoken language according to its context, followed by an expression showing the events taking place.

In [3], the Kazakh-Russian sign language (KRSL) was studied, the sign language used in the Republic of Kazakhstan. They write that KRSL is closely related to the Russian Sign Language (RSL) and some other sign languages of the former Soviet Union. Kazakhstan was once under the influence of the Russian Empire. It then was part of the Soviet Union (until 1991), whose centralized language policy also led to the spread of RSL in the Soviet republics. They considered the Kazakh SL as part of the RSL.

Moreover, linguists and sign language interpreters should also consider Kazakh Sign Language, as it is necessary to consolidate its status as a separate sign language and prevent its extinction [4].

The analysis of scientific articles revealed that the comparison of the form of demonstration of Kazakh sign words with other sign languages was not reflected in these works. Also, since the Kazakh sign language was not considered as a separate SL, many scientists recognized only Russian letters in the Kazakh alphabet [5]. What gives rise to the following questions from the scientific community:

1) Can Kazakh sign language exist as a separate SL?

2) Which optimal classifier should be used to classify dactylic letters of the Kazakh language?

2. COMPARISONS OF KSL WORDS BY FORM OF DEMONSTRATION FORM WITH OTHER SL

During the comparative analysis [6], the resources were studied surdo.kz [7] and spreadthesign.com [8] to compare the forms of demonstration of words in Kazakh, English, Russian and Turkish. Resource spreadthesign.com was used to conduct the study, as it is convenient to track the form of demonstration of the same word in many languages (Table 1). The Kazakhstani scientists, together with the scientists of ISC RAS, created the first electronic dictionary of the Kazakh sign language on the website www.surdo.kz, as well as an electronic textbook on the Kazakh fingerprint (letter) sign language [9].

Table 1. Initial table

№	Words	KAZ	RUS	ENG	TUR
1.	Articulation	+	+	-	-
2.	Disease	+	+	-	-
3.	Head	+	+	-	-
4.	Face	+	+	+	+
5.	Body	+	+	-	+
6.	Health	+	+	+	-
10.	Heart	+	+	+	+
11.	Infectious disease	*	*	-	
12.	Shoes	-	-	-	-
13.	Women's hat	-	-	-	-
14.	Skirt	+	-	-	+
15.	Mask	-	-	-	-
16.	Boots	-	-	-	-
17.	Clothing	-	-	-	-
18.	Costume	-	-	-	-
29.	Hat	+	-	-	+
20.	Gloves	-	-	-	-
21.	Fashion	-	-	-	-
22.	Towel	-	-	-	-
23.	Heel	-	-	-	-
24.	Sport shirt	-	-	-	-

+ provided that everyone's presentation form is the same;

- in the absence of absolute similarity;

* - very similar, but there is a slight change at the end or beginning

For example, in the table above, the words "Shoes," "Gloves," "Fashion," "Towel," and "Heel" in all four languages do not have the same form of demonstration. The words "Articulation," "Disease, "Head, "Face," "Body," Health," and" Heart" are simultaneously demonstrated in Russian and Kazakh. The word "Infectious disease" is very similar, but there is a slight change at the end or beginning. In turn, the words "Skirt" and "Hat" are identical in Kazakh and Turkish.

Of the 1,050 words studied, about 300, 38 percent are words with no analogs in the form of demonstration in any language—an excerpt from a table of words that are not similar to any language (Table 2).

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 2. Table of dissimilar words

№	Words	KAZ	RUS	ENG	TUR
3	Autumn	-	-	-	-
4	Dark	-	-	-	-
5	Size	-	-	-	-
6	Hours/Watch	-	-	-	-
7	Second	-	-	-	-
8	Wednesday	-	-	-	-
9	July	-	-	-	-
10	Cargo	-	-	-	-
11	Solid	-	-	-	-
12	Narrow	-	-	-	-
13	Gold	-	-	-	-
14	Album	-	-	-	-
15	Parsley	-	-	-	-
16	Bomb	-	-	-	-
17	Good	-	-	-	-
18	Close	-	-	-	-
19	Harm	-	-	-	-
20	Next	-	-	-	-

Words that are not similar to any language, only in the Kazakh language, can prove that the Kazakh sign language is separate.

At the same time, there are words whose forms of demonstration are identical in all languages. They accounted for 15 percent of the total number of words (Table 3).

Table 3. Words that are demonstrated in all languages in the same way

№	Words	KAZ	RUS	ENG	TUR
3	Soup	+	+	+	+
4	Boat	+	+	+	+
5	Pride	+	+	+	+
6	Believe	+	+	+	+
7	Court	+	+	+	+
8	Basketball	+	+	+	+
9	Dance	+	+	+	+
10	Box	+	+	+	+
11	Volleyball	+	+	+	+
12	Stadium	+	+	+	+
13	Tennis	+	+	+	+
14	Paint	+	+	+	+
15	Intervention	+	+	+	+
16	Cancel	+	+	+	+
17	Hearing	+	+	+	+
18	High	+	+	+	+
19	Meeting	+	+	+	+
20	Queue	+	+	+	+

In conclusion, we can give a diagram of the similarity of words in the Kazakh sign language with three studied languages. As seen from the chart, we see that 50% of the words considered were identical to Russian gestures. Although, in turn, the similarity of Russian words with English was 30 percent. We see that 38% of all words are not similar to other sign languages and correspond only to the Kazakh language. So 9% of the words were very similar in Russian, but there was a slight change at the end or beginning of the demonstration. Kazakh words were also identified, which did not have translations in any sign language, for example: Күйкелек, Кебіру, Жібеуші , Жирен, Ақсақал, Тоқал, Бәйбіше, Отағасы.

2.1 The Kazakh Dactylic Alphabet

The Kazakh Dactyl alphabet is based on the Cyrillic alphabet and includes 42 letters, 10 of which reflect the specific sounds of the Kazakh language: ә, і, ө, ұ, ү, ғ, қ, һ, ң. Letters в, ф, ц, ч, щ, ь, ъ, е, э, ю, я are used only when writing words of foreign (mostly Russian) origin.

The sound “ә” resembles the Russian sound “а” in an unstressed position after a soft consonant (Figure 1). Example as English "language."

Figure 1. Demonstration of Ә and А

The sound “ө” is a soft variation of the sound “о” (Figure 2). This sound is pronounced roughly like English "those."

Figure 2. Demonstration of О and Ө

The sound “ұ” is pronounced firmly and super-briefly (Figure 3). This sound has no analogs in other languages but resembles a sound intermediate between “о” and “у”. The sound “у” is a soft variety of “ұ”, pronounced softly and briefly.

Figure 3. Demonstration of У and Ұ

The sound “ы”, although it coincides in spelling with the sound “ы” in Russian, its pronunciation is significantly different (Figure 4). The sound “і” is a soft variety “ы”. Just like “ы”, it is pronounced super-briefly and indistinctly, but unlike it, when pronouncing the sound “і”, the tongue is advanced.

Figure 4. Demonstration of І and Ы

The sound “қ” on the letter is indicated by the letter “қ”. Its pronunciation is similar to “х” in Russian, but unlike it, the consonant “қ” is explosive, and “х” is slit (Figure 5).

Figure 5. Demonstration of Қ

The sound “ғ” on the letter is indicated by "ғ." The articulation “ғ” is analogical “қ”, but unlike it, “ғ” is pronounced with the participation of the voice. This is how ‘қ” is pronounced in some areas(Figure 6).

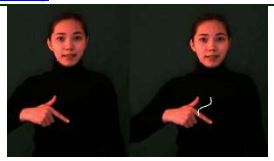


Figure 6. Demonstration of Г and Ғ

There are very few words with the sound “h” in the Kazakh language(Figure 7). This sound is found in borrowings from Oriental languages or in interjections: kaharman "hero," gauhar "jewel," ah "ah."

Figure 7. Demonstration of Г and Ғ

The sound is formed by the close contact of the back of the tongue with the posterior palate. A similar sound is available in English and German. Cf. German. lang "long,"sing," etc.

As a result of the comparative analysis, it can be said that, despite the 50 percent similarity with the Russian sign language, the Kazakh sign language is separate. Since the vocabulary of sign languages is many times smaller than the vocabulary of natural languages and deaf and hard of hearing people communicating with each other continue to create new words, adapting them to the conversation, it can be concluded that the Kazakh sign language is a separate sign language with its specifics.

3 CLASSIFICATION OF THE DACTYL ALPHABET OF THE KAZAKH LANGUAGE

3.1 Features selection

MediaPipe Hand is a model that receives a full frame as input and outputs high-precision three-dimensional key points of the arm as a response detected by the palm detector.

The palm detection model takes an input frame as a NumPy array and suggests different models depending on the received complexity received.

The defined palm is a list of 21 base coordinates, each point consisting of x, y, and z values, x and y being adjusted to the width and height of the image

Journal of Theoretical and Applied Information Technology
15th October 2022. Vol.100. No 19

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

respectively [0, 0, 1.0], and z - to the depth of the point. The depth of the forearm is the reference point, and the smaller the value, the closer the orientation to the camera. The z value uses roughly the same scale as x.

After receiving the coordinates of 21 joints of the human palm, these values are transmitted to real coordinates of the virtual three-dimensional virtual world. The orientation of the hand is determined based on magnetic positioning (Figure 8). There is terminology for the coordinate system of digital three-dimensional space. It is non--standard, although these concepts help programmers create 3D applications and games. Here, X is the width, Y is the height, and Z is the depth compared to the display.

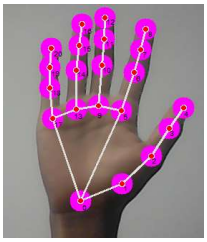


Figure 8. The 21 points of the human palm

3.2 Classification

After receiving the data set, a comparative analysis of classification methods was carried out, such as support vector machines, ensemble classifiers and nearest neighbor classifiers of the letters of the dactyl alphabet to identify the optimal method.

Support vector machines. The main idea of the classifier on support vectors is to build a separating surface using only a small subset of points lying in the zone critical for separation. In contrast, the rest of the correctly classified observations of the training sample outside this zone are ignored.

The objective of the [10] research was to recognize the hand gestures of Arabic Sign Language(ArSL) words using two depth sensors. The researchers developed a model to examine 143 signsgestured by 10 users for 5 ArSL words (the dataset). The sensors captured depth images of theupper human body, from which 235 angles (features) were extracted for each joint and betweeneach pair of bones. The dataset was divided into a training set (109 observations) and a testing set(34 observations). The support vector machine (SVM) classifier was set using different parameterson the gestured words' dataset to produce four SVM models, with linear kernel (SVMLD andSVMLT) and radial kernel (SVMRD and SVMRT) functions. The testing results for the SVMLD, SVMLT, SVMRD, and SVMRT models achieved97.059%, 97.059%, 94.118%, and 97.059% accuracy.

The paper [11] proposed a system to recognize hand gestures based on Indonesian Sign Language Standard. Their system uses Myo Armband as hand gesture sensors. Myo Armband has 21 sensors to express the hand gesture data as Mediapipe. Recognition process uses a Support Vector Machine (SVM) to classify the hand gesture based on the dataset of Indonesian Sign Language Standard. SVM yields the accuracy of 86.59% to recognize hand gestures as sign language.

1 step. Setting all the parameters of the classifier and training them.

```
template = templateSVM(...
    'KernelFunction', 'polynomial', ...
    'PolynomialOrder', 2, ...
    'KernelScale', 'auto', ...
    'BoxConstraint', 1, ...
    'Standardize', true);
classificationSVM = fitcecoc(...
    predictors, ...
    response, ...
    'Learners', template, ...
    'Coding', 'onevsone', ...
    'ClassNames', categorical({'A2'; 'G2'; 'H'; 'I'; 'K2'; 'N2'; 'O2'; 'U2'; 'U3'}));
```

Step 2. Creating the results structure using the predict function.

```
% Create the result struct with predict function
predictorExtractionFcn = @(t) t(:, predictorNames);
svmPredictFcn = @(x) predict(classificationSVM, x);
trainedClassifier.predictFcn = @(x) svmPredictFcn(predictorExtractionFcn(x));
```

Step 3. Extracting predictors and response (Figure 9).

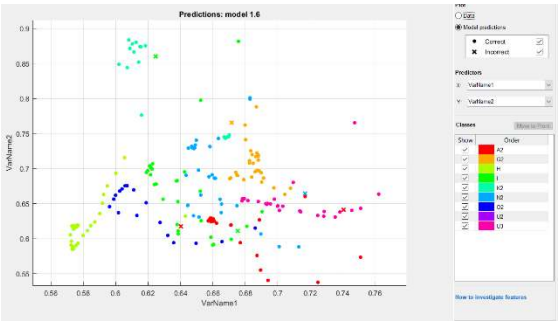


Figure 9. Predicton Model 1.

In this study, SVM with OVO showed 98.7 accuracy for specific sounds of the Kazakh language.

Ensemble classifiers. These are groups of algorithms that use several machine learning methods at once and correct each other's mistakes. These include such classifiers as Random Forest and XGBoost. Boosting is when algorithms are trained

Journal of Theoretical and Applied Information Technology
15th October 2022. Vol.100. No 19

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

sequentially, while everyone pays special attention to the errors of the previous one.

Su et al. [12] proposed a random forest-based Chinese Sign Language (CSL) sub-word recognition method using an improved decision tree to increase the probability of ob-taining the correct result from each decision tree in random forests. Based on the recognition results of 121 frequently used CSL sub-words, the superior performance of the random forest method in terms of accuracy and reliability was tested. Results with a recognition accuracy of 98.25% were obtained.

Brock et al. [13] proposed a system in which they introduced a two-stage pipeline based on two-dimensional body connec-tion positions extracted from RGB camera data. First, the system divides the signed expression data stream into meaningful word segments based on a frame-by-frame binary random forest. Each segment is then converted into an image-like form and classified using a convolutional neural network. The proposed system is then evaluat-ed on a data set of continuous Japanese gesture language sentence expressions with variations of non-manual expressions. By exploring a variety of data representations and network parameters, we can distinguish verbal segments of specific non-manual intonations from the underlying body joint motion data with 86% accuracy.

1 step. Setting all the parameters of the classifier and training them.

```
template = templateTree(...
    'MaxNumSplits', 20);
classificationEnsemble = fitcensemble(...
    predictors, ...
    response, ...
    'Method', 'AdaBoostM2', ...
    'NumLearningCycles', 30, ...
    'Learners', template, ...
    'LearnRate', 0.1, ...
    'ClassNames', categorical({'A2'; 'G2'; 'H'; 'I'; 'K2'; 'N2'; 'O2'; 'U2'; 'U3'}));
```

Step 2. Creating the results structure using the predict function.

```
predictorExtractionFcn = @(t) t(:, predictorNames);
ensemblePredictFcn = @(x) predict(classificationEnsemble, x);
trainedClassifier.predictFcn = @(x) ensemblePredictFcn(predictorExtractionFcn(x));
```

Step 3. Extracting predictors and response

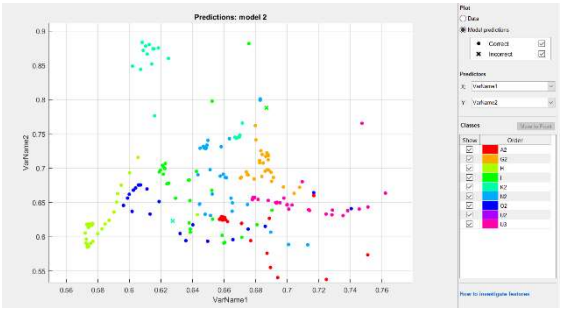


Figure 10. Predicton Model 2.

Results with a recognition accuracy of 98 % were obtained for specific sounds of the Kazakh language.

Nearest neighbor classifiers. Learning with a teacher is when a machine has a teacher who knows which answer is correct. This means that the source data has already been placed (sorted) correctly, and the device only needs to determine the object with the desired feature or calculate the result. Some works 31.35.45.84 use CNN and Naïve Bayes for gesture recognition.

1 step. Setting all the parameters of the classifier and training them.

```
classificationKNN = fitcknn(...
    predictors, ...
    response, ...
    'Distance', 'Euclidean', ...
    'Exponent', [], ...
    'NumNeighbors', 1, ...
    'DistanceWeight', 'Equal', ...
    'Standardize', true, ...
    'ClassNames', categorical({'A2'; 'G2'; 'H'; 'I'; 'K2'; 'N2'; 'O2'; 'U2'; 'U3'}));
```

Step 2. Creating the results structure using the predict function.

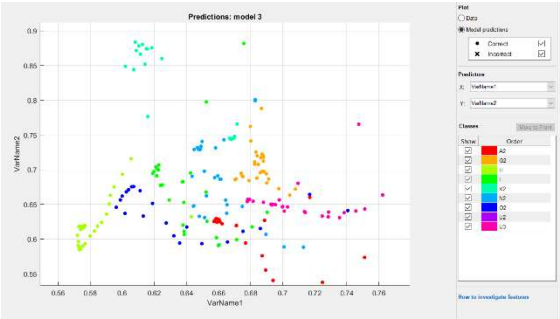


Figure 11. Predicton Model 3.

The testing results for the KNN models achieved 87 % accuracy.

In this study, SVM with OVO showed better results than KNN and Boosting trees for the test dataset. The reason may be that the data contains many features (63), which makes the model complex.

All 41 classes of the dactyl alphabet were also checked of SVM with OVO .

Journal of Theoretical and Applied Information Technology
15th October 2022. Vol.100. No 19

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

4. CONCLUSION

In this paper, the two questions presented in the introduction are given complete and factual answers. Questions:

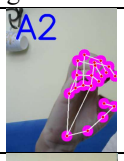
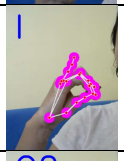
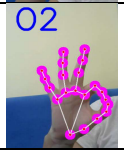
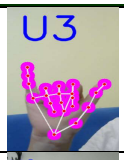
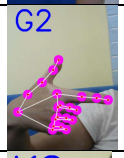
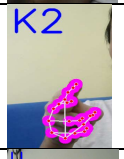
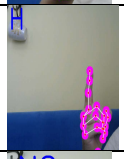
Can Kazakh sign language exist as a separate SL?

This article presents a comparative analysis of Kazakh, Russian, English, and Turkish sign languages to prove that the Kazakh sign language can exist as a separate sign language. A study of the form of the display was conducted in terms of configuration (arm/forearm), place of execution (localization), the direction of movement, nature of the movement, and components that cannot be performed manually (facial expression and articulation). As a result of the study, it was proved that the Kazakh sign language can exist as a separate sign language by observing the forms of demonstration of 1050 words in 4 languages.

Which optimal classifier should be used to classify dactylic letters of the Kazakh language?

In this work, a classification of letters was compiled, characteristic only of the Kazakh language, and their comparative analysis was carried out using three different models. As a result of a comparative study of SVM with the OVO model, it showed a high result. Despite a large amount of input data, the classification result shows an indicator above 98 percent, and the system works correctly (Table 4).

Table 4. Results of recognition specific sounds of the Kazakh language.

Specific sounds of the Kazakh language	Recognizing hand gestures
Ә	
І	
Ө	
Ү	
Ұ	
Ғ	
Қ	
Һ	
Ң	

REFERENCES:

[1] Dimskis, L.S. We Study Sign Language; Academia Publishing Center: Moscow, Russia, 2002; P. 128.

[2] Zaitseva, G.L. Signed Speech. In Dactylology; Vlados: Moscow, Russia, 2000; P. 192.

[3] Kimmelman, V., Imashev, A., Mukushev, M., Sandygulova, A. (2020). Eyebrow Position In Grammatical And Emotional Expressions In Kazakh-Russian Sign Language: A Quantitative Study. Plos One 15 (6).

[4] Kudubayeva, S.; Amangeldy, N.; Sundetbayeva, A.; Sarinova, A. The Use Of Correlation Analysis In The Algorithm Of Dynamic Gestures Recognition In Video Sequence. In Proceedings Of The 5th International Conference On Engineering And Mis, Pahang, Malaysia, 6–8 June 2019. Https://Doi.Org/10.1145/3330431.3330439.

[5] Kenshimov, C.; Buribayev, Z.; Amirgaliyev, Y.; Ataniyazova, A.; Aitimov, A. Sign Language Dactyl Recognition Based On Machine Learning Algorithms. East. -Eur. J. Enterp. Technol. 2021, 4, 58–72.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[6] Amangeldy, N.; Kudubayeva, S.A.; Tursynova, N.A.; Baymakhanova, A.; Yerbolatova, A.; Abdieva, S. Comparative Analy-Sis On The Form Of Demonstration Words Of Kazakh Sign Language With Other Sign Languages. Turklang 2022, 113 .

[7] Www.Surdo.Kz

[8] Https://Spreadthesign.Com/En.Gb/Search/

[9] Kudubayeva, S., Zhussupova, B., & Aliyeva, G. (2019). Features Of The Representation Of The Kazakh Sign Language With The Use Of Gestural Notation: Compiling The Dictionary Of Gestures Of The Ksl Based On The Notation Of L.S. Dimskis. Acm International Conference Proceeding Series.

[10] Almasre, M. A., & Al-Nuaim, H. (2020). A Comparison Of Arabic Sign Language Dynamic Gesture Recognition Models. Heliyon, 6(3). Https://Doi.Org/10.1016/J.Heliyon.2020.E03554

[11] Anwar, A., Basuki, A., & Sigit, R. (2020). Hand Gesture Recognition For Indonesian Sign Language Interpreter System With Myo Armband Using Support Vector Machine. Klik - Kumpulan Jurnal Ilmu Komputer, 7(2). Https://Doi.Org/10.20527/Klik.V7i2.320

[12] Su, R.; Chen, X.; Cao, S.; Zhang, X. Random Forest-Based Recognition Of Isolated Sign Language Subwords Using Data From Accelerometers And Surface Electromyographic Sensors. Sensors 2016, 16, 100, Https://Doi.Org/10.3390/S16010100.

[13] Brock, H.; Farag, I.; Nakadai, K. Random Forest-Based Recognition Of Isolated Sign Language Subwords Using Data From Accelerometers. 2016, 9–11. Https://Doi.Org/10.3390/S16010100