UDC 004.932
IRSTI 20.53.19

https://doi.org/10.55452/1998-6688-2025-22-1-94-102

[1]*Tursyn M.S.,
Master, ORCID ID: 0009-0008-2976-0771,
*e-mail: me_tursyn@kbtu.kz

[1]Kazakh British Technical University, Almaty, Kazakhstan

ROLE OF ARTIFICIAL INTELLIGENCE IN SIGN LANGUAGE RECOGNITION

Abstract

Over the decades the increasing computational capability and development of new technologies in the field of artificial intelligence have given us the ability to translate sign language in real time. There exist two main approaches to sign language recognition, the hardware-based approach and the software-based approach. The hardware-based approach relies on using special gloves, Kinect-based devices, and different levels of sensors. On the other hand, one of the approaches to working with sign language is using neural networks, which is the software-based approach. In this work, I observed existing approaches and experimented with machine learning and neural network models for sign language recognition. I got the dataset of Azerbaijani Sign Language, then trained my models based on that dataset, and got the results and metrics. The dataset contained over thirteen thousand samples of signs, which can be used in Kazakh Sign Language. In the end, I discussed the probable opportunity of using the developed models.

Key words: sign language recognition, artificial intelligence, machine learning, neural network, sign language.

Introduction

The main communication between most people is verbal communication, but people with hearing difficulties use non-verbal communication methods like gesture language or sign language. Nowadays over 360 million people with hearing difficulties use sign language to communicate [1]. But the main difficulty is that not everyone has the proper resources to learn sign language. It makes people's lives harder. Advanced artificial intelligence like deep learning, machine learning, and computer vision gave us new opportunities for developing systems that can accurately and efficiently recognize and translate sign language into written or spoken language, which can solve our problems. The potential of these systems to transform communication between individuals who are deaf and those who are hearing is immense. They can bring about a revolution in various settings, including education, healthcare, and social interactions. By providing digital real-time sign language translation, these systems enable instant communication and sharing of information between deaf and hearing individuals. This advancement fosters inclusivity and accessibility, significantly benefiting the deaf community. The paper [2] enlists all existing state-of-the-art approaches in sign language recognition. It starts by classifying into two classes the gesture language recognition: hardware-based approach and software-based approach. The hardware-based approach involves using gloves, Microsoft-provided Kinect, and various sensors. The software-based approach involves using probabilistic, machine learning, and deep learning approaches. In addition to that paper, the papers [3, 4, 5] propose a framework for continuous sign language recognition using deep neural networks. Like these papers I will also try to utilize existing deep neural networks.

Materials and Methods

In recent years, the significance of sign language research and technology applications has grown exponentially, fostering inclusivity and accessibility for the Deaf and Hard of Hearing (DHH) communities. However, a critical gap persists within the Commonwealth of Independent States (CIS), particularly Kazakhstan concerning the absence of a comprehensive sign language database. This makes it difficult to the development of robust sign language recognition systems, educational tools, and communication platforms tailored to the unique linguistic and cultural aspects of the region. The absence of a dedicated sign language database in Kazakhstan contributes to the following issues:

- Limited Research Advancements. The scarcity of a comprehensive sign language database restricts the exploration and development of sign language recognition algorithms, gesture-based interfaces, and other technological solutions aimed at improving communication accessibility.
- Educational Barriers. Educational institutions and programs catering to the DHH in Kazakhstan face challenges in developing effective teaching materials and methodologies without a standardized sign language database. This hampers the educational experience and opportunities for this community.
- Communication Inequality. Without a centralized database, creating inclusive communication tools and platforms tailored to the linguistic nuances of sign languages becomes challenging. This worsens the communication gap between DHH individuals and the wider society.

For that work, I will use the Azerbaijani Sign Language dataset, AzSL [6], cause Kazakhstan, Azerbaijan, and Russia were part of the USSR, and our sign language is similar [7]. The AzSL dataset consists of 13 444 samples gathered through the efforts of 221 volunteers and is now accessible to the sign language recognition community [6]. For my NN model, a full-fledged dataset with all the needed features is only available in this AzSL dataset in Kaggle. As shown in Figure 1, the dataset consists of images for each sign. Based on those images the NN model will be trained. On the other hand, there exists a Kazakh-Russian Sign Language dataset, K-RSL [7]. The main difficulty of using this dataset is that it contains only the video of signs, which is more focused on in-context signs with facial mimics. So, I will try to develop my own neural network and machine learning models to solve this problem.



Figure 1 – Images in the AzSL dataset

Previously the dataset files were shown in Figure 1. Processing all of the images and getting the necessary 21 points were conducted in Google Collab using the MediaPipe framework. That framework helps to identify points that are necessary in training my models. Figure 2 shows the result of the identified points of hand.

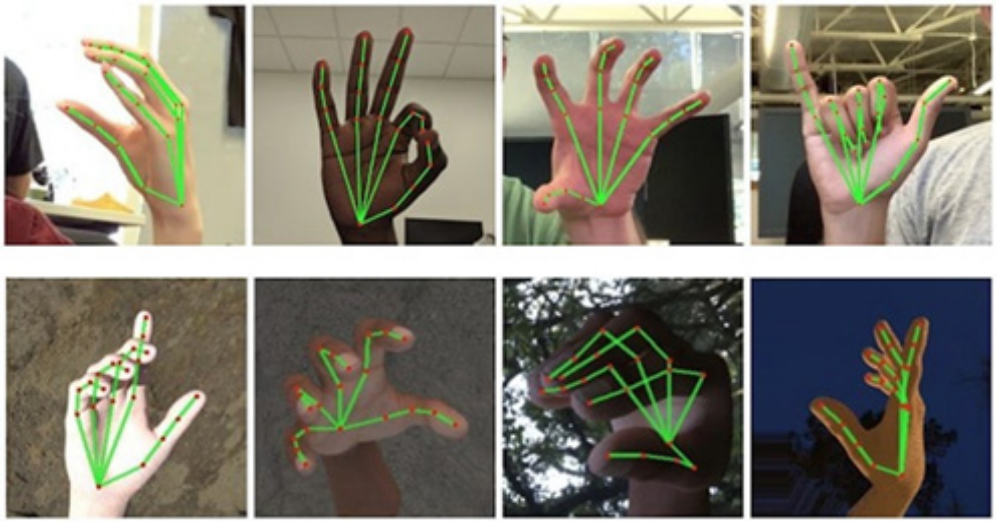


Figure 2 – Points in hands

In the first iteration of the work, I trained my model only on 5 letters (A, B, C, E, H). All identified points were stored in a .csv file with the help of Python programming language and MediaPipe framework. All code is available in my GitHub [8]. Each row represented the data on one hand, the 21 points of X and Y coordinate in the image. Figure 3 shows the stored .csv file.

```
In [2]: dataset = 'landmarks.csv'
        df = pd.read_csv(dataset)
        df
```

Out[2]:

	0	0.646450281	0.565575242	0.512143493	0.536735415	0.386336446	0.483176112	0.333536506	0.419391036	0.395050436	...
0	0	0.410841	0.592929	0.560942	0.547451	0.691772	0.488442	0.733234	0.420502	0.667936	...
1	0	0.646450	0.565575	0.512143	0.536735	0.386336	0.483176	0.333537	0.419391	0.395050	...
2	0	0.646450	0.565575	0.512143	0.536735	0.386336	0.483176	0.333537	0.419391	0.395050	...
3	0	0.487851	0.621440	0.305564	0.538390	0.189280	0.424867	0.177141	0.319205	0.288683	...
4	0	0.581802	0.755612	0.426105	0.725946	0.300148	0.664614	0.272424	0.596152	0.351902	...
...	...	...	...	...	...	...	...	...	...	...	...
1647	4	0.372567	0.390912	0.258315	0.469927	0.194400	0.538039	0.142778	0.602461	0.087905	...
1648	4	0.388814	0.216018	0.263187	0.268460	0.194008	0.324481	0.138569	0.378185	0.064274	...
1649	4	0.480436	0.277056	0.333318	0.331358	0.249772	0.398042	0.173027	0.462483	0.100165	...
1650	4	0.538669	0.310676	0.388489	0.365084	0.307966	0.440927	0.229645	0.506342	0.150856	...
1651	4	0.334309	0.334511	0.461677	0.339111	0.600933	0.377585	0.710551	0.392854	0.789591	...

1652 rows × 43 columns

Figure 3 – Processed images in .csv file

Support Vector Machine (SVM) is a supervised learning algorithm in machine learning, that is used in classification and regression tasks [9]. SVMs are applicable for binary classification problems but also can be used in multiclassification problems, which is our case. The objective of an SVM algorithm is to identify the optimal line, known as a decision boundary or hyperplane, that effectively separates data points belonging to different classes. This hyperplane concept extends to high-dimensional feature spaces. The primary goal is to maximize the margin, defined as the distance between the hyperplane and the nearest data points from each class. To practically use the SVM model I used sklearn library for Python. Trained on our gestures dataset, chose the RBF as kernel function, then got metrics, Figure 4. All code is available in my GitHub repository [8].

```
Classes in dataset:  [0 1 2 3 4]
A +
B +
C +
E +
H +
Accuracy: 0.94
Classification Report:
              precision    recall  f1-score   support

           0       0.98      0.97      0.98       117
           1       1.00      0.72      0.84        29
           2       0.85      0.97      0.91        95
           3       0.96      0.90      0.93        49
           4       1.00      0.98      0.99        41

    accuracy                           0.94       331
   macro avg       0.96      0.91      0.93       331
weighted avg       0.94      0.94      0.94       331
```

Figure 4 – Metrics of trained SVM model

A feedforward neural network represents one of the most basic forms of artificial neural networks. In this architecture, data flows unidirectionally – progressing from the input nodes, traversing any hidden nodes, and concluding at the output nodes. Unlike more complex networks such as recurrent neural networks and convolutional neural networks, there are no cycles or loops within the structure of a feedforward neural network. These networks, being the earliest form of artificial neural networks developed, are characterized by their simplicity and the absence of recurrent connections [10]. The structure of a feedforward neural network comprises three layers: the input layer, hidden layers, and the output layer. Each layer is composed of units, referred to as neurons, and these layers are linked through weights. At Figure 5 you can see the metrics of NN model's metrics.

```
Classification Report:
              precision    recall  f1-score   support

         A +       0.83      0.98      0.90       132
         B +       0.00      0.00      0.00        36
         C +       0.61      0.96      0.75       127
         E +       0.00      0.00      0.00        65
         H +       0.95      1.00      0.97        54

    accuracy                           0.74       414
   macro avg       0.48      0.59      0.52       414
weighted avg       0.57      0.74      0.64       414
```

Figure 5 – Classification report of first NN model

As you can see at B letter all metrics are 0. After that I decided to adjust my model's architecture. I changed output feature size and added normalization layer to stabilize and accelerate the training process, Figure 6.

```
CustomModel(
  (fc1): Linear(in_features=42, out_features=50, bias=True)
  (batch_norm1): BatchNorm1d(50, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
  (dropout1): Dropout(p=0.5, inplace=False)
  (fc2): Linear(in_features=50, out_features=20, bias=True)
  (batch_norm2): BatchNorm1d(20, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
  (dropout2): Dropout(p=0.3, inplace=False)
  (fc3): Linear(in_features=20, out_features=5, bias=True)
)
```

Figure 6 – Second NN architecture

Now it shows better results. Almost each letter predicted correctly, see Figure 7.

Classification Report:

	precision	recall	f1-score	support
A +	0.98	0.98	0.98	132
B +	0.97	0.94	0.96	36
C +	0.97	0.97	0.97	127
E +	0.98	0.97	0.98	65
H +	0.98	1.00	0.99	54
accuracy			0.98	414
macro avg	0.98	0.97	0.97	414
weighted avg	0.98	0.98	0.98	414

Figure 7 – Classification report of second NN

So, I developed three models for hand gesture prediction. The first model was the SVM model, and the other two models were custom neural network models. In the experiment with 5 letters A, B, C, E, and H the best results showed the second custom NN model with 0.98 points of accuracy. With this ability of training and testing these models I can develop the whole system for sign language recognition.

Results and Discussion

Despite using conventional neural network architectures, the work presented in the paper [11] proposes a distinct system for dynamic hand gesture recognition. This system incorporates multiple deep learning architectures specifically designed for recognizing hand feature fragments. The evaluation of this system is conducted on a challenging dataset comprising 40 dynamic hand gestures performed by 40 subjects in uncontrolled environments. The results demonstrate superior performance compared to state-of-the-art approaches. The proposed system effectively combines local hand shape features with global body configuration features, making it well-suited for intricate structured hand gestures found in sign language. The study employs a framework for hand region detection and estimation, a robust face detection algorithm, and the theory of body parts ratios for gesture space estimation and normalization. Fine-grained hand shape features are learned using two 3DCNN instances, while coarse-grained global body configuration features are learned using MLP and autoencoders, which aggregate and globalize the extracted features. Classification is performed using the SoftMax function, and domain adaptation is employed to reduce training costs. Potential

future work includes exploring alternative strategies for modeling the temporal aspect, optimizing the length of input clips, and testing the system's real-time hand gesture recognition capabilities.

In contrast to the beforementioned papers, the paper [12] focuses on the necessity of an interdisciplinary approach to sign language processing. The field requires expertise in various domains, including computer vision, natural language processing, human-computer interaction, linguistics, and Deaf culture. The paper presents the outcomes of a two-day workshop involving 39 domain experts from diverse backgrounds. It addresses three key questions: the insights gained from an interdisciplinary perspective, the major challenges faced by the field, and calls to action for the research community. The paper underscores the significance of understanding Deaf culture and sign language linguistics, reviews the current state-of-the-art, identifies urgent challenges, and emphasizes the need for more data to enhance sign language processing systems. Overall, the paper aims to provide orientation to both computer science and non-computer science readers in the field, foster interdisciplinary collaborations, and guide research priorities in sign language processing.

Other than the approach of using CNN there exists a wide variety of other solutions in sign language recognition. In the paper [13], the authors suggested a methodology using Transformers. The paper emphasized the cons of Vision Transformer (ViT) versus CNN. The rise of the ViT presents a formidable challenge to CNNs, the prevailing technology in computer vision for various image recognition tasks. ViT models outperform CNNs in computational capabilities, efficiency, and accuracy, as indicated in [14]. While transformer architectures have established themselves as the gold standard in natural language processing, their adoption in computer vision has been limited. Attention mechanisms are typically used in conjunction with CNNs or as substitutes for specific convolution features, maintaining the original structure. However, the transformer encoder breaks away from these dependencies inherent in CNNs. This allows the standard transformer architecture to be directly applied to sequences of image patches, proving surprisingly effective and accurate in image classification tasks. As a result, ViT demonstrates notable advantages over traditional CNNs in the realm of computer vision.

The paper also provides a comparison of different approaches using CNN, ANN, DeepCNN, SVM, Multimodal Transformer, and pre-trained models like ResNet50, and EfficientNet B4.

Another paper that describes different approaches in sign language recognition is [15]. The paper delves into the application of Recurrent Neural Networks (RNNs), specifically focusing on Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRUs), for the recognition of sign language gestures. It provides an in-depth exploration of the architectural aspects of LSTM and GRU, highlighting their ability to effectively capture long-term dependencies in sequential data, particularly in the context of sign language gestures. The authors elaborate on the various datasets utilized for model training and describe the preprocessing techniques implemented to enhance model accuracy. Additionally, the paper examines the diverse evaluation metrics employed to gauge the model's performance. In conclusion, the paper suggests that employing LSTM and GRU in sign language recognition shows promise and has the potential to significantly enhance the accuracy of recognition systems.

Using only models and combining them can solve the task of recognition, but optimization drawbacks also should be solved like in the paper [16]. To assess the quality of translations, authors employ BLEU and rBLEU metrics. Their examination through an ablation study reveals substantial impacts on the model's performance arising from optimizers, activation functions, and label smoothing. That paper's endeavors are directed toward improving the capture of visual features, optimizing decoder utilization, and incorporating pre-trained decoders to enhance translation outcomes. Other than that, the paper [17] suggests a methodology not only for the translation of sign language but also for generating sign language from text. This paper introduces USLNet, an unsupervised sign language translation and generation network inspired by the success of unsupervised neural machine translation (UNMT). USLNet leverages abundant single-modality data (text and video) without parallel sign language data. The model consists of single-modality reconstruction modules (text and video) and cross-modality back-translation modules. Unlike text-based UNMT, USLNet addresses cross-modality challenges, such as length and feature dimension mismatches, using a sliding window

method. USLNet is the first unsupervised model capable of generating both natural language text and sign language video in a unified manner. Experimental results on BBC-Oxford Sign Language dataset (BOBSL) and Open-Domain American Sign Language dataset (OpenASL) datasets demonstrate competitive performance compared to supervised baseline models, highlighting its effectiveness in sign language translation and generation.

Sign language recognition also needs some aspects of security as emphasized in the paper [18]. This study addresses the challenges of data scarcity and privacy concerns in sign language translation (SLT). Due to a lack of aligned captions, existing sign language data on the web is often unsuitable for training supervised models. Moreover, privacy risks associated with large-scale web-scraped datasets containing biometric information need to be addressed in the development of SLT technologies. The proposed two-stage framework, SSVP-SLT, combines self-supervised video pretraining on anonymized, unannotated videos with supervised SLT finetuning on a curated parallel dataset. SSVP-SLT achieves state-of-the-art performance on the How2Sign dataset. The study discusses the advantages and limitations of self-supervised pretraining and facial obfuscation for SLT, providing insights from controlled experiments.

So, for that moment I developed a feedforward neural network model with input size of 42 and with the feature output size of 5, used batch normalization function and liner activation function for better performance in recognition which is described in paper about feedforward neural networks [20]. Then had pretty good results on accuracy 0.98 of recognition. Additionally, I developed the SVM model with linear function as a kernel function for better performance which described in paper about SVM performance in image recognition [21]. In that case I got also good results, but the accuracy 0.94 was lower than the approach with feedforward neural network. So, these models are applicable for creating systems for sign language recognition, which can have features like translating sign language bidirectionally and generating sign language from natural language. This system can be useful in the fields like education and medicine.

Conclusion

This paper reviewed existing approaches in the field of sign language recognition and explored the use of Feedforward Neural Network (FNN) and Support Vector Machine (SVM) models as an appropriate approach. By training a FNN and SVM models for sign language recognition and analyzing the accuracy evolution, significant progress was achieved in interpreting sign language gestures. The results obtained from the trained model demonstrate its effectiveness in recognizing and interpreting sign language gestures. The implications of this research go beyond the successful training of the models. The opportunities presented by this model are promising and diverse. The real-time translation of sign language gestures can greatly enhance communication and inclusivity for individuals who are deaf or hard of hearing. The models can serve as an educational tool, assisting learners in mastering sign language or improving their signing skills. Additionally, the integration of sign language recognition into public spaces and assistive technologies can significantly improve accessibility and facilitate human computer interaction. The applications of the models for sign language recognition are vast and extend into various domains. Future research can delve deeper into optimizing the model's performance, exploring novel architectures, and expanding its capabilities to capture the nuances of different sign languages.

REFERENCES

1 Dubey S., Suryawanshi S., Rachamalla A., and K. Madhu Babu. Sign language recognition. International Journal for Research in Applied Science & Engineering Technology (IJRASET) https://doi.org/10.22214/ijraset.2023.48586

2 Farooq U., Rahim M.S.M., Sabir,N. et al. Advances in machine translation for sign language: approaches, limitations, and challenges. Neural Comput & Applic, 2021, vol. 33, pp. 14357–14399. https://doi.org/10.1007/s00521-021-06079-3

3 Amrutha K. and P. Prabu. ML Based Sign Language Recognition System, 2021 International Conference on Innovative Trends in Information Technology (ICITIIT), Kottayam, India, 2021, pp. 1–6, https://doi.org/10.1109/ICITIIT51526.2021.9399594.

4 Dr. Sabeenian R.S., Bharathwaj S., Aadhil M. Sign Language Recognition Using Deep Learning and Computer Vision. Journal of Advanced Research in Dynamical and Control Systems, 202, vol. 12, pp. 964–968. https://doi.org/10.5373/JARDCS/V12SP5/20201842.

5 Adithya V., Rajesh R. A Deep Convolutional Neural Network Approach for Static Hand Gesture Recognition, Procedia Computer Science, 2020, vol. 171, pp. 2353–2361. https://doi.org/10.1016/j.procs.2020.04.255.

6 Hasanov Ja., Alishzade N., Nazimzade A., Dadashzade S., Tahirov T. Development of a hybrid word recognition system and dataset for the Azerbaijani Sign Language dactyl alphabet. Speech Communication, 2023, vol. 153, p. 102960. https://doi.org/10.1016/j.specom.2023.102960.

7 Imashev A., Mukushev M., Kimmelman V., Sandygulova A. A Dataset for Linguistic Understanding, Visual Evaluation, and Recognition of Sign Languages: The K-RSL, 2023, pp. 631–640. https://doi.org/10.18653/v1/2020.conll-1.51.

8 Source code of experiments in GitHub. https://github.com/MeTuA/Sign-Language-Recognition-with-Mediapipe/tree/main.

9 Tabsharani F., techtarget.com, August 2023, Accessed 23 February 2024, https://www.techtarget.com/whatis/definition/support-vector-machine-SVM.

10 Whitfield B., Feedforward Neural Networks: A Quick Primer for Deep Learning, August 2022, builtin.com, https://builtin.com/data-science/feedforward-neural-network-intro.

11 Al-Hammadi M. et al. Deep Learning-Based Approach for Sign Language Gesture Recognition With Efficient Hand Gesture Representation, in IEEE Access, 2020, vol. 8, pp. 192527–192542. https://doi.org/10.1109/ACCESS.2020.3032140.

12 Bragg D., Verhoef T., Vogler Ch., Morris M., Koller O., Bellard M., Berke L., Boudreault P., Braffort A., Caselli N., Huenerfauth M., Kacorri H. Sign Language Recognition, Generation, and Translation: An Interdisciplinary Perspective, 2019, pp. 16–31. https://doi.org/10.1145/3308561.3353774.

13 Kothadiya D., Bhatt Ch., Saba T., Rehman A. SIGNFORMER: DeepVision Transformer for Sign Language Recognition. IEEE Access., 2023, pp. 1–1. https://doi.org/10.1109/ACCESS.2022.3231130.

14 Dosovitskiy A., Beyer L., Kolesnikov A., Weissenborn D., Zhai X., Unterthiner T., Dehghani M., Minderer M., Heigold G., Gelly S., Uszkoreit Ja., Houlsby N. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. https://doi.org/10.48550/arXiv.2010.11929.

15 Tabsharani F., techtarget.com, August 2023, Accessed 23 February 2024, https://www.techtarget.com/whatis/definition/support-vector-machine-SVM.

16 Roy P., Han J., Chouhan S., Thumu B. American Sign Language Video to Text Translation, 2024. https://doi.org/10.48550/arXiv.2402.07255.

17 Zhengsheng G., Zhiwei H., Wenxiang J., Xing W., Rui W., Kehai C., Zhaopeng T., Yong X., Min Z. https://doi.org/10.48550/arXiv.2402.07726.

18 Rust Ph., Shi B., Wang S., Camgöz N.C., Maillard Je. Towards Privacy-Aware Sign Language Translation at Scale https://doi.org/10.48550/arXiv.2402.09611.

19 Lee C.C., Rahiman M.H.F, Rahim R.A. and Saad F.S.A. A Deep Feedforward Neural Network Model for Image Prediction. Journal of Physics: Conference Series 1878, 2021, p. 012062. https://doi.org/10.1088/1742-6596/1878/1/012062.

20 Bartłomiej M., Kamil M. and Paweł C. Symposium for Young Scientists in Technology, Engineering and Mathematics, https://api.semanticscholar.org/CorpusID:235693217.

21 Teja Gontumukkala S.S., Varun Godavarthi Y. S., Ravi Teja Gonugunta B.R., Subramani R. and K. Murali. 12th International Conference on Computing Communication and Networking Technologies (ICCCNT), https://doi: 10.1109/ICCCNT51525.2021.9579803.

[1]*Тұрсын М.Ш.,
магистр, ORCID ID: 0009-0008-2976-0771,
*e-mail: me_tursyn@kbtu.kz

[1]Қазақстан-Британ техникалық университеті, Алматы қ., Қазақстан

ЖАСАНДЫ ИНТЕЛЛЕКТТІҢ ЫМ ТІЛІН ТАНУДАҒЫ РӨЛІ

Аңдатпа

Соңғы онжылдықтарда есептеу қабілетінің артуы мен жасанды интеллект технологияларының қарқынды дамуы ым тілін нақты уақыт режимінде тану және аудару мүмкіндіктерін айтарлықтай кеңейтті. Ым тілін танудың екі негізгі тәсілі бар: аппараттық және бағдарламалық жасақтамаға негізделген әдістер. Аппараттық тәсіл арнайы қолғаптарды, Kinect негізіндегі құрылғыларды және әртүрлі сенсорлық технологияларды пайдалануға сүйенеді. Бағдарламалық тәсіл, өз кезегінде, нейрондық желілер мен машиналық оқыту алгоритмдерін қолдануға негізделген. Бұл зерттеуде қолданыстағы әдістерді талдап, ым тілін тану үшін машиналық оқыту және нейрондық желі үлгілерімен эксперименттер жүргізілді. Зерттеу барысында әзірбайжан ым тілінің деректер жинағы пайдаланылып, осы деректер негізінде үлгілерді үйрету жүргізілді. Деректер қоры қазақ ым тілінде қолдануға болатын он үш мыңнан астам белгіні қамтыды. Алынған нәтижелер мен көрсеткіштерді талдау негізінде әзірленген модельдердің қолдану мүмкіндіктері қарастырылды. Зерттеу нәтижелері ым тілін танудың тиімділігін арттыруға және оны түрлі қолданбалы салаларда пайдалану перспективаларын кеңейтуге ықпал етуі мүмкін.

Тірек сөздер: ым тілін тану, жасанды интеллект, машиналық оқыту, нейрондық желі, ым тілі.

[1]*Турсын М.Ш.,
магистр, ORCID ID: 0009-0008-2976-0771,
*e-mail: me_tursyn@kbtu.kz

[1]Казахстанско-Британский технический университет, г. Алматы, Казахстан

РОЛЬ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В РАСПОЗНАВАНИИ ЯЗЫКА ЖЕСТОВ

Аннотация

На протяжении десятилетий растущие вычислительные возможности и развитие новых технологий в области искусственного интеллекта дали нам возможность переводить язык жестов в режиме реального времени. Существует два основных подхода к распознаванию языка жестов: аппаратный и программный. Аппаратный подход основан на использовании специальных перчаток, устройств на базе Kinect и датчиков различного уровня. С другой стороны, одним из подходов к работе с языком жестов является использование нейронных сетей, то есть подход, основанный на программном обеспечении. В этой работе я изучил существующие подходы и поэкспериментировал с машинным обучением и моделями нейронных сетей для распознавания языка жестов. Я получил набор данных по азербайджанскому языку жестов, затем обучил свои модели на основе этого набора данных и получил результаты и показатели. Набор данных содержал более тринадцати тысяч образцов жестов, которые могут быть использованы в казахском жестовом языке. В конце я обсудил вероятную возможность использования разработанных моделей.

Ключевые слова: распознавание языка жестов, искусственный интеллект, машинное обучение, нейронная сеть, язык жестов.

Article submission date:16.04.2024