UDC 004.9

AI-BASED SOLUTION FOR SIGN LANGUAGE RECOGNITION

M. E. Dementiev, A. S. Yurieva

The article is presented by the scientific supervisor –
Associate Professor, Candidate of Pedagogical Sciences, Zheltova E. P.

This study presents an AI-based sign language recognition program designed to translate sign language into spoken language, facilitating communication between deaf and hearing communities. The detailed methodology is given using machine learning and deep learning algorithms. Particular attention is given to model building, user interface design and the gesture recognition approach. The evaluation and implications of this program implementation for improving social inclusion are confirmed.

Keywords: artificial intelligence, real-time translation, computer vision, gesture recognition.

Effective communication is a key component of social interaction and integration. For people who are deaf and hard of hearing, sign language serves as a vital means of communication. However, communication barriers persist for people who are unfamiliar with sign language, preventing access to public services, education and employment opportunities. According to the World Health Organization, by 2050, approximately 2.5 billion people are expected to experience some degree of hearing loss, with 700 million requiring rehabilitation [1]. Moreover, 80% of people with hearing loss live in low- and middle-income countries, highlighting the critical need for affordable solutions to facilitate communication between hearing and deaf people. Advances in artificial intelligence (AI) have facilitated significant advances in sign language recognition (SLR), which opens new opportunities for the use of AI-based SLR to address these communication gaps [2-5].

Thus, this research aims to the creation of a tool enabling to facilitate real-time translation of sign language into spoken language, thereby bridging the gap between diverse communicative communities. In the framework of its realization, the following steps have been undertaken:

– to study a topic related to sign language and gestures;

– to collect a dataset of common gestures, to train and test the model on a range of machine learning algorithms, then to select the most appropriate and effective one for the defined task;

– to develop a user interface for convenient interaction with the created model.

For creation of a robust database, the MNIST Sign Language dataset was chosen [2] because it effectively provides an alphabet for common sign language gestures. In addition, a custom dataset including six basic words “Hello”, “Yes”, “No”, “I/me”, “He/she/it”, “Help” frequently used in conversations and emergency scenarios was collected. Data were collected using computer vision tools implemented in the OpenCV library (opencv.org/), with each image preprocessed to 28x28 grayscale format for consistency with the MNIST dataset and stored in CSV format.

This assembled dataset is supposed to be neither exhaustive nor universal in scope. Rather, it serves to illustrate the potential for developing gesture recognition technology through the application of machine learning methods.

The detailed model design under research was as follows:

1. Model development and testing

A series of machine learning algorithms were tested, including logistic regression, k-nearest neighbours (k-NN), support vector machines (SVM), decision trees using the Scikit-learn library (scikit-learn.org), and convolutional neural networks (CNNs) using the TensorFlow library (www.tensorflow.org). CNNs were selected due to their efficiency in handling image data, using localised convolutions to focus on features such as edges, textures and shapes of objects in the image. Moreover, CNNs allow the model to be less prone to overfitting and improve generalisation [3].

The designed CNN model utilised three convolution and three pooling layers and two dense layers, with ReLU activation function throughout and a softmax function on the final dense layer. The trained models were evaluated based on accuracy metrics presented in Table 1.

Table 1. Model Evaluation

Algorithm	*Accuracy metrics*
Logistic Regression	0,925
k-Nearest Neighbors (k-NN):	0,997
Support Vector Machine (SVM)	0,781
Decision Tree	0,692
CNN	0,998

The results of the algorithm evaluation demonstrate that both the convolutional neural network (CNN) and the k-nearest neighbour (KNN) algorithms exhibit comparable and high levels of accuracy. However, the former displays significantly higher speed and performance. Thus, based on the obtained data, the Neural Network based algorithm is the most appropriate choice.

2. User interface development

The user interface was developed using the PyQt5 directory, which facilitates intuitive interaction [4]. OpenCV was employed to capture and preprocess images, and the hands were tracked using the Google MediaPipe hands library [5].

The PyQt-based user interface effectively enables real-time interaction by capturing gestures and interpreting them with high accuracy. Tests validated the seamless integration of hand tracking feature, ensuring responsiveness and usability.

3. Dynamic Gesture Recognition Approach

The video, with the assistance of the OpenCV library, is known to be divided into frames. Therefore, the algorithm is only capable of analysing static images, whereas the domain of gesture language encompasses both static and dynamic elements. So, a novel approach based on sequential analysis of key points of a gesture has been developed to address the problems associated with dynamic gesture recognition, which includes:

– identification of key points capturing the gesture's dynamic characteristics;

– isolated training for each key point;

– individual classification of key points by the algorithm;

– recognition of dynamic gestures if at least 50% of control points appear in sequence in forward or reverse order.

This methodology aims to expand the system capabilities beyond static gesture recognition. Thus, the current program version just implements the possibility of processing static gestures. In the future, the functionality of the program will be extended and the mentioned method of dynamic gesture detection will be implemented. The details on the algorithm implementation and the collected dataset are given on GitHub Repository [6].

To sum up, the implementation of an AI-based sign language recognition system opens up new opportunities for improving communication between deaf and hearing people. The use of machine learning algorithms, particularly convolutional neural networks, enables high accuracy in real-time gesture recognition, helping to overcome barriers to access to public services and enhancing social inclusion. In the future, the development of this technology and the extension of the system's functionality to recognize dynamic gestures will allow for even more effective integration of hearing-impaired individuals into public life, providing them with independence and access to a broader range of services and opportunities.

References

1. Glukhota i poterya slukha. [Deafness and hearing loss] [Electronic Resource] // Vsemirnaya organizatsiya zdravookhraneniya. [World Health Organization] 2 February 2024. – URL: https://www.who.int/ru/news-room/fact-sheets/detail/deafness-and-hearing-loss (last request: 22.10.2024).

2. Yazyk zhestov MNIST. [Sign Language MNIST] [Electronic Resource] // Largest AI ML community. – URL: https://www.kaggle.com/datasets/datamunge/sign-language-mnist (last request: 22.10.2024).

3. *Ivanov G.* 6.1 Svertochnie neironnie seti [Convolutional neural networks] [Electronic Resource] // Yandeks obrazovanie. [Yandex Education]. – URL: https://education.yandex.ru/handbook/ml/article/svyortochnye-nejroseti (last request: 24.10.2024).

4. Rukovodstvo po obnaruzheniyu orientirov ruk. [Hand landmarks detection guide] [Electronic Resource] // Google MediaPipe Studio. – URL: https://ai.google.dev/edge/mediapipe/solutions/vision/handlandmarker#taskdetails (last request: 22.10.2024).

5. Yurieva A.S. BaFoCv [Electronic Resource] // Repoziotory GitHub. [GitHub Repository]. – URL: https://github.com/YurievaAS/BafoCv (last request: 27.10.2024).

Dementiev Mikhail Evgenievich – Bachelor, The Bonch-Bruevich Saint Petersburg State University of Telecommunications, (St. Petersburg, Russia), kantyt3@gmail.com

Yurieva Arina Sergeevna – Bachelor, The Bonch-Bruevich Saint Petersburg State University of Telecommunications, (St. Petersburg, Russia), arishxoxo25@gmail.com

Zheltova Elena Petrovna –scientific supervisor –Associate Professor, Candidate of Pedagogical Sciences, Associate Professor, Foreign Languages Department, The Bonch-Bruevich Saint Petersburg State University of Telecommunications, (St. Petersburg, Russia), elena_gzheltova@mail.ru

РЕШЕНИЕ ДЛЯ РАСПОЗНАВАНИЯ ЖЕСТОВ НА ОСНОВЕ ИСКУССТВЕННОГО ИНТЕЛЛЕКА

М. Е. Дементьев, А. С. Юрьева

Статья представлена научным руководителем –
доц., к. пед. н., Желтовой Е. П.

Данная статья посвящена разработке программы распознавания языка жестов на основе искусственного интеллекта для перевода жестового языка в устную речь с целью облегчения коммуникации между глухими и слышащими сообществами. Подробная методология включает использование алгоритмов машинного обучения и глубокого обучения. Особое внимание уделено построению модели, проектированию пользовательского интерфейса и подходу к распознаванию жестов. Проведена оценка, и последствия внедрения этой программы для улучшения социальной инклюзии подтверждены экспериментально.

Ключевые слова: искусственный интеллект, перевод в режиме реального времени, компьютерное зрение, распознавание жестов.

Список источников и литературы

1. Глухота и потеря слуха [электронный ресурс] // Всемирная организация здравоохранения. 2 февраля 2024. – URL: https://www.who.int/ru/news-room/fact-sheets/detail/deafness-and-hearing-loss (дата обращения: 22.10.2024).

2. Язык жестов MNIST [электронный ресурс] // Largest AI ML community. – URL: https://www.kaggle.com/datasets/datamunge/sign-language-mnist (дата обращения: 22.10.2024).

3. *Иванов Г.* 6.1 Сверточные нейронные сети [электронный ресурс] // Яндекс образование. – URL: https://education.yandex.ru/handbook/ml/article/svyortochnye-nejroseti (дата обращения: 24.10.2024).

4. Руководство по обнаружению ориентиров рук [электронный ресурс] // Google MediaPipe Studio. – URL: https://ai.google.dev/edge/mediapipe/solutions/vision/handlandmarker #taskdetails (дата обращения: 22.10.2024).

5. *Юрьева А. С.* BaFoCv [Электронный ресурс] // Репозиторий GitHub. – URL: https://github.com/YurievaAS/BafoCv (дата обращения: 27.10.2024).

Дементьев Михаил Евгеньевич – бакалавр, Санкт-Петербургский государственный университет телекоммуникаций им. проф. М. А. Бонч-Бруевича, (г. Санкт-Петербург, Россия), kantyt3@gmail.com

Юрьева Арина Сергеевна – бакалавр, Санкт-Петербургский государственный университет телекоммуникаций им. проф. М. А. Бонч-Бруевича, (г. Санкт-Петербург, Россия), arishxoxo25@gmail.com

Желтова Елена Петровна – научный руководитель – доцент, кандидат педагогических наук, доцент кафедры иностранных языков, Санкт-Петербургский государственный университет телекоммуникаций им. проф. М. А. Бонч-Бруевича, (г. Санкт-Петербург, Россия), elena_gzheltova@mail.ru

Статья поступила в редакцию: 31.01.2025; принята к публикации: 04.02.2025.

FOR CITATION:

Dementiev M. E., Yurieva A. S. Reshenie dlya raspoznavaniya zhestov na osnove iskusstvennogo intellekta [Ai-based solution for sign language recognition] // Sociogumanitarnye kommunikacii [Social and humanitarian communications]. 2025. № 1(11). P. 148-151.

ДЛЯ ЦИТИРОВАНИЯ:

Дементьев М. Е., Юрьева А. С. Решение для распознавания жестов на основе искусственного интеллекта // Социогуманитарные коммуникации. – 2025. – № 1(11). – С. 148-151.