EUROPEAN INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH AND MANAGEMENT STUDIES

VOLUME04 ISSUE05
DOI: https://doi.org/10.55640/eijmrms-04-05-43 Pages: 269-280

UZBEK SIGN LANGUAGE CLASSIFIER BASED ON MACHINE LEARNING

Kayumov Oybek Achilovich
Jizzakh Branch of the National University of Uzbekistan named after Mirzo Ulugbek Jizzakh, Uzbekistan

Kayumova Nazokat Rashitovna
Jizzakh Branch of the National University of Uzbekistan named after Mirzo Ulugbek Jizzakh, Uzbekistan

Xodjabekova Dilnoza Furqat qizi
Jizzakh Branch of the National University of Uzbekistan named after Mirzo Ulugbek Jizzakh, Uzbekistan

ABOUT ARTICLE

Key words: Uzbek Sign Language, Machine Learning, CNN, RNN, Gesture Recognition, Accessibility.

Received: 21.05.2024
Accepted: 26.05.2024
Published: 31.05.2024

Abstract: The "Uzbek Sign Language Classifier Based on Machine Learning" presents a groundbreaking approach to enhancing communication accessibility for the deaf and hard-of-hearing community in Uzbekistan. This research focuses on developing a robust machine learning model to recognize and classify Uzbek Sign Language (UzSL) signs, enabling real-time translation and fostering better integration of sign language users into various aspects of society. Utilizing advanced deep learning techniques, particularly Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), this study addresses the intricate challenges associated with sign language recognition, such as gesture segmentation, feature extraction, and accurate classification. The dataset for this project comprises a comprehensive collection of images and videos of Uzbek sign language gestures. These data points undergo meticulous preprocessing, including normalization and augmentation, to enhance model training. The proposed model architecture leverages CNNs for spatial feature extraction and RNNs for capturing temporal dependencies, ensuring high accuracy in recognizing dynamic sign sequences. Additionally, the integration of transfer learning techniques, employing pre-trained models, significantly improves the model's performance, particularly given the relatively limited size of the dataset. Extensive experimentation and validation are conducted to fine-tune the model's hyper parameters, optimizing its accuracy and robustness. The results demonstrate a promising accuracy rate, highlighting the model's capability to accurately recognize and classify a wide range of Uzbek sign language gestures. Furthermore, this research underscores the importance of developing localized sign language recognition systems, tailored to the specific linguistic and cultural nuances of the target community.

INTRODUCTION

The need for effective communication tools for the deaf and hard-of-hearing community is a global challenge, with significant implications for social inclusion, education, and employment. In Uzbekistan, where Uzbek Sign Language (UzSL) serves as the primary mode of communication for the hearing-impaired, the development of automated sign language recognition systems can play a transformative role in bridging communication gaps. This study focuses on creating an advanced Uzbek Sign Language classifier based on machine learning, leveraging state-of-the-art techniques to provide accurate and real-time translation of sign language gestures.

Sign language is a complex visual language that encompasses hand gestures, facial expressions, and body movements to convey meaning. Each sign language, including UzSL, has unique characteristics influenced by cultural and linguistic contexts. Traditional approaches to sign language recognition have faced limitations due to the intricacies involved in accurately interpreting these gestures. However, the advent of machine learning, particularly deep learning, has opened new possibilities for developing robust sign language classifiers.

Machine learning, a subset of artificial intelligence, involves training models to recognize patterns and make predictions based on data. Within machine learning, deep learning techniques, especially Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), have shown remarkable success in image and sequence recognition tasks. CNNs are particularly effective in extracting spatial features from images, making them well-suited for recognizing hand gestures in sign language. RNNs, on the other hand, excel at capturing temporal dependencies, which is crucial for understanding the sequential nature of sign language.

The proposed UzSL classifier employs a hybrid approach, combining CNNs and RNNs to leverage the strengths of both architectures. The CNN component is responsible for extracting spatial features from individual frames of sign language videos, while the RNN component captures the temporal dynamics of these sequences. This combination allows the model to accurately recognize and classify complex gestures that comprise UzSL.

Data collection is a critical aspect of developing a machine learning-based sign language classifier. For this study, a comprehensive dataset of UzSL gestures was compiled, involving the recording of native signers performing a wide range of gestures. These recordings were then annotated with the corresponding Uzbek words, providing a rich dataset for training the model. Preprocessing steps, including normalization and data augmentation, were applied to enhance the quality and diversity of the training data, which is essential for achieving high model accuracy and generalization.

Transfer learning, a technique that utilizes pre-trained models, was also employed to improve the performance of the UzSL classifier. By fine-tuning models pre-trained on large datasets, the classifier benefits from the extensive feature representations learned from these datasets, thereby enhancing its ability to recognize UzSL gestures with limited training data.

The development of this UzSL classifier has significant implications for the deaf and hard-of-hearing community in Uzbekistan. An accurate and efficient sign language recognition system can facilitate better communication, enabling individuals to interact more effectively in educational, professional, and social settings. This technology can be integrated into various applications, including real-time translation services, educational tools, and communication aids, thereby improving accessibility and quality of life for sign language users.

The Uzbek Sign Language classifier based on machine learning represents a critical advancement in assistive technology. By leveraging the power of deep learning, this study aims to develop a robust system capable of accurately recognizing and translating UzSL gestures. The success of this project not only enhances communication for the hearing-impaired community in Uzbekistan but also sets a precedent for similar efforts in other languages and regions. As machine learning continues to evolve, the potential for developing sophisticated sign language recognition systems will undoubtedly grow, contributing to greater inclusivity and accessibility worldwide.

METHODOLOGY

The development of the Uzbek Sign Language (UzSL) classifier based on machine learning involves a systematic approach encompassing data collection, preprocessing, feature extraction, model design, training, and evaluation. This methodology ensures the creation of a robust and accurate sign language recognition system tailored to the unique characteristics of UzSL.

Data Collection: The foundation of the UzSL classifier is a comprehensive dataset of sign language gestures. The data collection process includes:

- Video Recording: Native UzSL users were recorded performing a diverse range of gestures corresponding to common Uzbek words. Multiple cameras were used to capture the gestures from various angles, providing a rich dataset for model training.

- Annotation: Each video segment was meticulously annotated with the corresponding Uzbek word. This annotation process involved manual labeling by experts to ensure accuracy and reliability.

Data Preprocessing: Preprocessing is crucial for enhancing the quality and consistency of the dataset. The raw video data was converted into image frames and underwent several preprocessing steps:

- Frame Extraction: Videos were split into individual frames at a specified frame rate, resulting in a large collection of static images representing various sign language gestures.

- Resizing: All images were resized to a uniform dimension (e.g., 224x224 pixels) to standardize the input size for the deep learning model.

- Normalization: Pixel values were normalized to a range of [0, 1] to facilitate faster convergence during training.

- Augmentation: Data augmentation techniques such as rotation, flipping, zooming, and shifting were applied to increase the robustness of the model and prevent overfitting by generating a more diverse dataset.

Feature Extraction: Feature extraction involves identifying and extracting significant patterns from the images. Convolutional Neural Networks (CNNs) were employed for automatic feature extraction due to their superior performance in image recognition tasks:

- CNN Architecture: A pre-trained CNN model, such as ResNet-50, was used for feature extraction. ResNet-50 is known for its deep architecture and ability to learn complex features through residual connections, which help mitigate the vanishing gradient problem in deep networks.

 ISSN: 2750-8587

Model Design: The core of the UzSL classifier is a deep neural network designed to learn and classify the extracted features:

- Hybrid Model: The proposed model combines CNNs and Recurrent Neural Networks (RNNs) to leverage the strengths of both architectures. The CNN component extracts spatial features from individual frames, while the RNN component, specifically Long Short-Term Memory (LSTM) networks, captures the temporal dependencies of the gesture sequences.

- Fully Connected Layers: Additional fully connected layers were added on top of the CNN and RNN components to adapt the model for the specific task of sign language recognition. These layers included dropout layers to prevent overfitting and improve generalization.

Model Training: The training process involved several steps to optimize the model's performance:

- Dataset Splitting: The dataset was divided into training, validation, and test sets, typically in a ratio of 70:15:15. This ensured that the model was evaluated on unseen data during the validation and testing phases.

- Loss Function: The categorical cross-entropy loss function was used, as it is suitable for multi-class classification problems.

- Optimizer: The Adam optimizer was employed for its efficiency and ability to handle sparse gradients on noisy problems.

- Hyper parameter Tuning: Hyper parameters such as learning rate, batch size, and the number of epochs were fine-tuned using grid search and cross-validation techniques to achieve the best performance.

Model Evaluation: Evaluating the model's performance involved several metrics to ensure its accuracy and robustness:

- Accuracy: The primary metric used to evaluate the model was accuracy, calculated as the ratio of correctly predicted gestures to the total number of gestures.

- Confusion Matrix: A confusion matrix was used to visualize the performance of the model across different classes, highlighting areas where the model performed well and areas needing improvement.

- Precision, Recall, and F1 Score : These metrics were also calculated to provide a more detailed analysis of the model's performance, particularly in handling imbalanced classes.

Real-Time Implementation: To validate the practical utility of the model, a real-time implementation was developed:

- Live Video Feed: The model was integrated into a system capable of processing live video feeds. The system captured real-time video from a camera, processed the frames, and fed them into the trained model.

Here's a simplified mathematical model for building a classifier for Uzbek Sign Language (USL) based on machine learning:

- X as the input features, representing the characteristics of sign language gestures.

- Y as the output variable, representing the class labels (different signs in Uzbek Sign Language).

- h as the hypothesis function, which maps input features to output labels.

- θ as the parameters of the model.

We can represent the relationship between the input features and the output labels using a supervised learning approach, particularly a classification algorithm.

1. Data Collection and Preprocessing:

   - Collect a dataset consisting of samples of Uzbek Sign Language gestures. Each sample is labeled with the corresponding sign.

   - Preprocess the data, which may include tasks such as normalization, feature extraction (like hand position, movement, etc.), and splitting the dataset into training and testing sets.

2. Model Selection:

   - Choose a suitable machine learning model for classification. Common choices include:

     - Support Vector Machines (SVM)

     - Decision Trees

- Random Forests

- Convolutional Neural Networks (CNNs) for image-based classification.

3. Model Training:

- Use the training dataset to train the selected model. This involves finding the optimal parameters \(\theta \) that minimize the cost function.

- The training process typically involves an optimization algorithm like Gradient Descent or its variants, which iteratively update the parameters to reduce the error between predicted and actual labels.

4. Model Evaluation:

- Evaluate the trained model using the testing dataset to assess its performance.

- Common evaluation metrics for classification tasks include accuracy, precision, recall, F1-score, and confusion matrix.

5. Model Deployment:

- Once the model achieves satisfactory performance, deploy it to classify new, unseen sign language gestures.

- Ensure the deployed model is robust and efficient for real-time or near-real-time classification tasks.

$$\mathrm{h}_{\theta}(\mathrm{X}) = \mathrm{Y}$$

Where:

$\mathrm{h}_{\theta}(\mathrm{X})$ is the hypothesis function, parameterized by θ, which predicts the output labels Y given input features X.

During training, we aim to minimize the cost function, typically represented as the difference between the predicted labels and the actual labels. For instance, in logistic regression, the cost function might be the cross-entropy loss function:

$$J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)}\log\left(h_{\theta}\left(x^{i}\right)\right) + (1-y^{(i)})\log(1-h_{\theta}\left(x^{i}\right))\right]$$

 ISSN: 2750-8587

Where:

- $J(\theta)$ is the cost function.
- m is the number of training examples.
- $y^{(i)}$ is the actual label of the i^{th} training example.
- $h_\theta\left(x^i\right)$ is the predicted probability that $\left(x^i\right)$ belongs to class 1.

The optimization algorithm then adjusts the parameters θ to minimize this cost function, thus improving the model's ability to classify Uzbek Sign Language gestures accurately.

RESULTS

The development of a robust and accurate Uzbek Sign Language (USL) classifier using machine learning techniques holds significant promise for enhancing communication accessibility for the deaf and hard of hearing community in Uzbekistan. In this section, we present the results of our efforts to build such a classifier, outlining the data collection process, preprocessing steps, model selection, training, evaluation, and deployment phases involved in the development process.

Data Collection and Preprocessing:

We began by collecting a comprehensive dataset of sign language gestures commonly used in Uzbek Sign Language. The dataset consists of video recordings of various gestures, each labeled with its corresponding sign. To ensure the quality and diversity of the dataset, we included gestures representing different hand shapes, movements, and facial expressions. Preprocessing the data involved tasks such as normalization, feature extraction, and data augmentation. Normalization techniques were applied to standardize the scale and orientation of the gestures, while feature extraction methods were used to capture relevant characteristics such as hand positions, movements, and facial expressions. Data augmentation techniques, including flipping, rotating, and scaling, were employed to enhance the diversity and robustness of the dataset.

Model Selection:

Several machine learning algorithms were evaluated for their suitability in classifying USL gestures. We considered algorithms such as Support Vector Machines (SVM), Decision Trees, Random Forests, and Convolutional Neural Networks (CNNs). Given the complexity and variability of sign language gestures, we opted for a deep learning approach using CNNs. CNNs are well-suited for learning hierarchical representations from visual data and have shown promising results in various image classification tasks.

Model Training:

The selected CNN model was trained using the preprocessed dataset of USL gestures. The training process involved optimizing the model parameters to minimize a predefined loss function. We employed gradient-based optimization algorithms, such as Stochastic Gradient Descent (SGD) and its variants, to iteratively update the model parameters based on the gradients of the loss function. The goal of training was to enable the model to learn discriminative features from the input gestures and make accurate predictions of the corresponding signs.

Model Evaluation:

Once the model was trained, we evaluated its performance using a separate testing dataset. We assessed the classifier's accuracy, precision, recall, and F1-score to measure its effectiveness in recognizing USL gestures. Additionally, we analyzed the confusion matrix to identify any patterns of misclassification and areas for improvement. The evaluation results indicated that the CNN-based classifier achieved high accuracy and robust performance across a wide range of USL gestures, demonstrating its potential utility in real-world applications.

Model Deployment:

Upon satisfactory evaluation results, the trained USL classifier was deployed for real-world use. We integrated the model into software applications capable of interpreting sign language gestures in real-time. The deployed classifier demonstrated efficient and reliable performance, enabling seamless communication for deaf and hard of hearing individuals in Uzbekistan. The accessibility and usability of the classifier were further enhanced through user-friendly interfaces and integration with assistive technologies.

The development of a USL classifier based on machine learning techniques represents a significant advancement in improving communication accessibility for the deaf and hard of hearing community in Uzbekistan. Through the integration of state-of-the-art machine learning algorithms, comprehensive datasets, and rigorous evaluation methodologies, we have demonstrated the feasibility and effectiveness of automated USL recognition systems. Further research and development in this area hold the potential to transform the lives of individuals with hearing impairments, enabling them to communicate more effectively and participate more fully in society.

CONCLUSION AND DISCUSSION

The development of a robust Uzbek Sign Language (USL) classifier using machine learning techniques represents a significant milestone in the advancement of communication accessibility for the deaf and hard of hearing community in Uzbekistan. In this section, we discuss the implications of our work, highlight key findings, address limitations, and propose future directions for research and development in this field.

The deployment of a USL classifier based on machine learning has profound implications for individuals with hearing impairments in Uzbekistan. By providing automated interpretation of sign language gestures, the classifier facilitates seamless communication between deaf or hard of hearing individuals and the wider community. This enhances social inclusion, educational opportunities, and employment prospects for individuals who rely on sign language as their primary mode of communication. Moreover, the accessibility of the classifier extends beyond traditional communication channels, enabling deaf individuals to interact more effectively with technology, access online resources, and participate in digital platforms.

Our research has demonstrated the feasibility and effectiveness of using deep learning techniques, specifically Convolutional Neural Networks (CNNs), for USL recognition. The CNN-based classifier achieved high accuracy and robust performance across a diverse range of sign language gestures, underscoring the potential of machine learning in addressing the unique challenges of USL interpretation. The comprehensive dataset, rigorous preprocessing steps, and meticulous evaluation methodologies employed in our study contributed to the reliability and generalizability of the classifier, ensuring its suitability for real-world applications.

Despite the promising results, our work is not without limitations and challenges. One limitation is the availability of annotated USL datasets, which are often limited in size and diversity. The scarcity of annotated data poses challenges in training robust classifiers, especially for rare or complex gestures. Additionally, the inherent variability and ambiguity of sign language gestures present challenges in designing classifiers that generalize well to unseen data. Addressing these limitations requires concerted efforts in data collection, annotation, and collaboration with the deaf community to ensure the inclusivity and representativeness of the datasets.

Furthermore, the computational complexity and resource requirements of deep learning models pose challenges in deploying USL classifiers in resource-constrained environments. Real-time interpretation of sign language gestures requires efficient algorithms and optimized hardware architectures to minimize latency and maximize responsiveness. Moreover, ensuring the accessibility and usability of the classifier for individuals with varying levels of hearing impairment necessitates the development of user-friendly interfaces, tactile feedback systems, and assistive technologies tailored to the specific needs of the users.

Moving forward, several avenues for future research and development emerge in the field of USL recognition and interpretation. Firstly, there is a need for the continued expansion and enrichment of annotated USL datasets to encompass a wider range of gestures, contexts, and variations. Collaborative efforts between researchers, educators, and the deaf community can contribute to the creation of more comprehensive and representative datasets.

Secondly, advancements in machine learning algorithms, particularly in the areas of transfer learning, few-shot learning, and multimodal learning, hold promise for improving the robustness and adaptability of USL classifiers. Leveraging pre-trained models, domain adaptation techniques, and multimodal fusion strategies can enhance the performance of classifiers on diverse datasets and in challenging environments.

Additionally, the integration of USL classifiers into wearable devices, smart cameras, and augmented reality systems can extend the reach and accessibility of sign language interpretation in everyday contexts. Real-time feedback mechanisms, interactive interfaces, and personalized learning experiences can further enhance the usability and effectiveness of USL classifiers for deaf and hard of hearing individuals.

In conclusion, the development of a USL classifier based on machine learning represents a transformative step towards fostering inclusivity, empowerment, and communication accessibility for the deaf and hard of hearing community in Uzbekistan and beyond. By leveraging the power of technology, collaboration, and innovation, we can continue to advance the frontiers of USL recognition and interpretation, paving the way for a more inclusive and equitable society for all.

REFERENCES

1. World Health Organization. Safe Listening Devices and Systems: A WHO-ITU Standard; World Health Organization: Geneva, Switzerland, 2019; p. 10.
2. Japan Hearing Instruments Manufacturers Association. JapanTrak 2022; JHIMA: Tokyo, Japan, 2022; p. 99.

 ISSN: 2750-8587

3. Jiang, S.; Sun, B.; Wang, L.; Bai, Y.; Li, K.; Fu, Y. Skeleton Aware Multi-Modal Sign Language Recognition. In Proceedings of the 2021 Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 19–25 June 2021.
4. Hezhen, H.; Wengang, Z.; Houqiang, L. Hand-Model-Aware Sign Language Recognition. In Proceedings of the 35th AAAI Conference on Artificial Intelligence, Virtual, 2–9 February 2021; Volume 35, pp. 1558–1566.
5. Kayumov, O., Kayumova, N., Rayxona, A., & Madina, Y. L. (2023). THE STRATEGIC SIGNIFICANCE OF HUMAN RESOURCE MANAGEMENT IN UZBEKISTAN ENTERPRISES ON THE BASIS OF ARTIFICIAL INTELLIGENCE. International Journal of Contemporary Scientific and Technical Research, 268-272.
6. Turakulov, O. O Kayumov Improving the quality of independent education by creating an interactive intellectual electronic learning resource in higher education institutions. International Journal of Contemporary Scientific and Technical Research.
7. Kayumov, O., Sayfullayeva, I., & Rustamov, I. (2023, May). UZSL CREATING AN INTERACTIVE INTELLECTUAL E-LEARNING “UZSL ONLINE LEARNING PLATFORM” FOR TEACHING UZBEK SIGN LANGUAGE. In International Scientific and Practical Conference on Algorithms and Current Problems of Programming.
8. Kayumov O, Kayumova N, Xodjabekova D. Oʻzbek imo-ishora tilida lotin alifbosini tanib olishning barmoq boʻgʻimi va bilak koordinatalaridan olingan burchak belgilari asosidagi modeli
9. Kayumov, O., & Kayumova, N. (2023). BASED ON MACHINE LEARNING ALGORITHMS TO RECOGNIZE UZBEK SIGN LANGUAGE (UZSL). International Journal of Contemporary Scientific and Technical Research, 1(2), 58-68.
10. Саидова, Х., Каюмов, О., & Каюмова, Н. (2024). Tasvirlarni tanib olish. Новый Узбекистан: наука, образование и инновации, 1(1), 676-679.
11. Xolboʻtayevich, T. O., Achilovich, K. O., & Rashitovna, K. N. (2022). Oʻzbekistonda raqamli iqtisodiyot sohasini rivojlantirish davrida korxonaning intelektual resurslarini boshqarish. International Journal of Contemporary Scientific and Technical Research, 65-68.
12. Ahmad, A., Kayumov, O., & Kayumova, N. (2023). ARTIFICIAL INTELLIGENCE IN THE MANAGEMENT OF INTELLECTUAL RESOURCES OF ENTERPRISES IN THE CONDITIONS OF THE DIGITAL ECONOMY IN UZBEKISTAN. Scientific-theoretical journal of International education research, 1(1), 106-116.