

Deep Neural Network Based Algorithm for Recognition of Static Signs of Polish Sign Language

Wiktor Barańczyk^a and Piotr Duch^b

Institute of Applied Computer Science, Lodz University of Technology, Bohdana Stefanowskiego 18, Lodz, Poland

Keywords: Sign Language Recognition, Deep Learning, Computer Vision.

Abstract: Developing sign language recognition algorithms is important for promoting accessibility and inclusion for deaf and hard-of-hearing individuals, improving education, and advancing technological applications in various fields. This paper presents a novel approach for recognizing static signs of Polish Sign Language using characteristic points and deep neural networks. As an input to the deep neural network the distances between every landmark of hands, elbows, and shoulders were used. The study focused on exploring the effectiveness of using deep learning techniques for sign recognition. The proposed algorithm was evaluated on two publicly available databases (NUS and LSA16) and achieved higher or comparable accuracy to other algorithms. Additionally, it was tested on a collected database of photographs of 24 people. The proposed algorithm achieved 96.45% accuracy, 96.15% recall, and 96.66% precision.

1 INTRODUCTION

It is estimated that in the world live about 70 million deaf people, more than 80% of whom live in developing countries. More than 300 different sign languages are in use¹. For Deaf people, phonetic languages used in their country are foreign to them. That situation does not only cause problems in social life but also in official matters, such as problems with understanding documents or contracts.

In Poland, Deaf people use Polish Sign Language (PSL). Currently, the estimated number of PSL users is approximately 50,000. Despite its significant divergence from spoken Polish, PSL was not officially recognized as a natural language for many years. In 2011, the Polish Government passed a law on sign language and other communication methods, which granted the Deaf community the legal right to request interpreter services when interacting with public administration (Linde-Usiekniewicz et al., 2016).

In this work, we propose a method for recognizing static signs of Polish Sign Language using image processing. The presented algorithm is an integral component of our broader research framework,


enabling the identification and classification of PSL signs. This is an important contribution to the field, as there are few articles related to recognizing PSL signs. The approach involves the use of a database that was collected with the help of both deaf and non-deaf people, which ensures that the database is representative of the diverse range of signing styles and handshapes that exist in PSL. In the conducted experiments, we used a single color camera situated in front of the signer. There is no need to wear gloves or to use special equipment.

2 LITERATURE ANALYSIS

Sign languages are natural languages that rely on hands and body movements, facial expressions, and gestures rather than voice to communicate. The main users are Deaf individuals, hearing-impaired people, or those who are unable to speak due to physical conditions.

Many hearing people think sign languages are visual representations of the spoken languages used in a respective country, with hand movements replacing vocalization. Another misunderstanding is the belief that only one universal sign language exists. Both of them are incorrect. Each sign language has its own grammar, words, and syntax, which differs from other languages. Of course, similarities may exist between

^a  <https://orcid.org/0009-0007-9377-0950>

^b  <https://orcid.org/0000-0003-0656-1215>

¹ <https://www.un.org/en/observances/sign-languages-day>

different sign languages². Moreover, due to the isolated nature, sign languages have many developed regional variations counted to the same language. It is similar to the dialects or regional accents in spoken languages³.

Signs can be classified into two categories based on observable features: static and dynamic signs. Static signs are those in which the hand position does not change during the duration of the sign. In Polish Sign Language, static signs are used mainly to describe alphabet letters and digits. On the other hand, dynamic signs involve hand position and finger arrangement changes over time. Despite the shape of fingers, hand trajectory, orientation, and sometimes the coordinated movement of both hands matter in those signs.

The paper focuses on recognizing static signs in Polish Sign Language. Although PSL is one of the biggest minority languages in Poland, it still remains under-researched. In 2010, the first academic unit dedicated to research on PSL grammatical and lexical properties was established within the University of Warsaw (Linde-Usiekniewicz et al., 2014).

For further experiments, 27 static signs from Polish Sign Language were selected, including only the static characters from the manual alphabet, along with numbers and some simple words.

2.1 Sign Language Recognition

Several methodologies are used in sign language recognition, with the three primary approaches being sensor-based, vision-based, and hybrid methods.

The sensor-based approach is mainly based on special equipment, such as accelerometers (Fatmi et al., 2019), surface electromyography (sEMG) sensors (Jiang et al., 2017), gloves (Wen et al., 2021), or even Kinect (Raghuveera et al., 2020). The recognition of the signs is a two-step process. In the first step, the signal from the sensor is captured, and in the second step, the data are processed using different approaches.

Vision-based methods use cameras to capture information about specific signs. Compared to the sensor-based techniques, this approach is less demanding. A user requires only a camera, which can be found in many daily-used devices like computers or smartphones. However, the main limitations of using only the cameras are sensitivity to the different lighting or lack of depth perception. Sometimes the multiple cameras approach is used (Erol et al., 2007).

²https://en.wikipedia.org/wiki/Sign_language

³<https://www.british-sign.co.uk/what-is-british-sign-language/>

The hybrid approach combines both previous systems. The input data are collected from multiple sensors, like leap motion and camera (ElBadawy et al., 2015).

In the vision-based approach, one of the most popular methods is based on convolutional neural networks (CNN) in various architectures. The CNN trained to recognize 200 different signs of Indian Sign Language achieved an 88.98% recognition rate (Rao et al., 2018). Another approach used DenseNet121 architecture, achieving 80% accuracy on a dataset comprising 24 static signs of the American Sign Language alphabet (Kołodziej et al., 2022).

Several studies have utilized classic CNN architectures with features extraction and classification parts, processing either RGB or grayscale images (Jeny et al., 2021; Eid and Schwenker, 2023). Some of these approaches have also used depth maps (Kang et al., 2015) or RGB-D data (Elboushaki et al., 2020). In other research, architectures combining detection and recognition like You Only Look Once architecture (Daniels et al., 2021; Sarma et al., 2021) and Single Shot Detector architectures (Rastgoo et al., 2020b; Rastgoo et al., 2020a) have been explored.

More recent advances include attention mechanism (Bhaumik et al., 2023) or Vision Transformers (Tan et al., 2023). Another vision-based approach involves a multi-step process. Key characteristic points and skeletal structures are first extracted from images, followed by sign recognition using classifiers (Jiang et al., 2021; Bajaj and Malhotra, 2022).

2.2 Polish Sign Language Recognition

PSL is relatively understudied compared to other sign languages like American Sign Language or British Sign Language. Specifically, only a few articles describe algorithms for recognizing signs of PSL, highlighting the need for further research in this domain.

One of the first approaches to recognizing isolated PSL words was based on analyzing images captured from the canonical stereo system. This method utilized information on the hand position and its 3D position relative to the face. The study employed a database of 101 words signed by two individuals (Kapuscinski and Wysocki, 2005).

Another approach utilized image analysis for PSL word recognition through a graph-based representation of hand posture. In this method, the hand was detected using a Gaussian distribution model of skin colour and morphological operations. The contour of the hand was then extracted to construct a graph for recognition. The method achieves an accuracy of 94.3% (Flasiński and Myśliński, 2010).

Another work explores also the recognition of sequences of PSL letters. One of the proposed approaches used input from Kinect, which was analyzed with hidden Markov models to classify postures corresponding to specific letters by analyzing their transitions, yielding a recognition accuracy of 75.2%. The experiments were conducted with a database comprising three individuals and 20 sequences (Warchoł et al., 2019).

Additionally, for the PSL letter recognition, a prototype of a glove employing textronic elements was proposed, achieving an accuracy of 86.5% (Korzeniowska et al., 2022).

3 PROPOSED ALGORITHM

Three crucial problems must be addressed when evaluating sign language recognition methods.

The first issue concerns the division of data into training and testing sets. Neural networks can easily adapt to the shape and proportions of the hands presented in the training. The dataset should be partitioned based on individuals rather than the number of images to ensure reliable testing. This means that images of individuals present in the training set should not appear in the testing set.

The second problem is related to capturing the spatial positioning of the hands in relation to the signer and their surroundings, which is crucial for many words in sign languages. CNNs also need extensive labelled data to achieve the best performance. Obtaining a database that is large enough to contain diverse PSL signs is difficult due to the limited availability of data, expert labelling requirements, and individual differences among users in signing styles. PSL signs exhibit significant intra-class variations, where different individuals may show the same sign in distinct ways due to personal preferences. CNNs perform better on tasks that are characterized by relatively stable and consistent visual patterns — intra-class variations might greatly impact their performance. A promising approach may involve using characteristic points and skeleton structures or CNN-like architectures combined with these features.

The third challenge is that the limited of available Deaf individuals for dataset collection is limited. A diversity of hand shapes is essential for good generalization of the problem, but only a few PSL users are willing to help collect the dataset. Additionally, variations and differences in signs between PSL users further complicate the task.

3.1 Database

No PSL sign database fulfilled the conditions required for this study. Many sign language databases only contain images of hands, which do not allow for evaluating the hand's position relative to the signer and do not reflect typical Deaf communication. As a result, a new database was self-collected. However, this dataset will not be publicly available due to the lack of participant consent.

It is important to note that the dataset mostly contains photos of hearing individuals. Since hearing individuals were generally unfamiliar with PSL, they were conducted by someone who knew the signs. This approach was adopted to try to solve the third problem - data scarcity and diversity. Using hearing individuals presents an advantage, as their availability is almost unlimited, making it easier to collect a proper amount of data.

Each individual was photographed on three different backgrounds. For each background, 27 photos were taken - one for each sign. In the summary, 81 images. The subjects were visible from at least the knees to the top of the head, and both hands were fully visible. Images where all critical hand points were not detected were excluded from the dataset.

The database contains photos of 24 people – 3 Deaf and 21 hearing. Images of Deaf individuals were extracted from videos filmed with the help of the Polish Deaf Association in Lodz. The images were categorized into 27 groups, each corresponding to one specific word. Augmentation techniques applied to the images were:

- Mirror reflection (signs used in the research were "hand invariant"),
- Resizing hearing individuals's photos to 80%, 60%, 40%, and 20% of the original size,
- Framing Deaf individuals' images to the same 4:3 proportions as those of hearing participants
- Rotating each image by 1 or 2 degrees, both clockwise and counterclockwise.

Images of Deaf individuals were extracted from video, which has a lower resolution compared to the photos of hearing participants. Therefore, further resizing them was avoided to maintain data quality. The final database contained 90549 images. However, after excluding images where not every key point was detected, 77933 photos were used for training and testing.

3.2 Algorithm Description

The proposed algorithm comprises several stages. Initially, characteristic points of the human posture are detected, after which those relevant to the approach are selected. Distances between the selected points are calculated and normalized. These distances are the input to the neural network, which outputs information about recognized sign (Figure 1).

The first stage of the algorithm involves detecting key body landmarks. For this purpose, the Holistic MediaPipe Solution⁴ is used, which detects 75 landmarks—33 pose landmarks and 21 hand landmarks—for each hand). Therefore, from the 75 detected landmarks, 46 are selected for further processing: 42 representing hand poses and 4 representing the positions of the shoulders and elbows.

Following this, 1035 Euclidean distances between each point, based on their x and y coordinates, are computed and normalized using the following equation:

$$value_{scaled} = \frac{value - value_{min}}{value_{max} - value_{min}} \quad (1)$$

where:

- $value$ - currently scaled value,
- $value_{max}$ - maximum value in a column to which belongs scaled value,
- $value_{min}$ - minimum value in a column to which belongs scaled value.

It is worth noting that using distances between key hand and body points in sign language recognition has not been previously presented in the literature reviewed by the authors.

3.3 Neural Network Architecture

A deep neural network (DNN) was used to recognize the selected signs. The input to the network consists of the distances between the characteristic points. The optimal architecture, learning rate, and optimizer were selected using grid search algorithm. In the result proposed model comprises six fully connected layers with 1035, 512, 256, 128, 64, and 27 neurons, respectively. Each layer, except the final one, utilizes the LeakyReLU activation function, while the output layer uses the softmax function. Categorical cross-entropy is used as a loss function, and Adamax is used as an optimizer with a learning rate of $1e-4$. The neural network was trained for 1000 epochs. This number of epochs was chosen because the algorithm initially learned quickly, reaching approximately 90% accuracy. However, the model appeared to struggle with

⁴<https://google.github.io/mediapipe/solutions/holistic>

Table 1: Performance of the proposed pipeline and selected approaches from the literature on the NUS dataset.

Model	Accuracy
(Eid and Schwenker, 2023)	93.5%
(Tan et al., 2021)	96.75%
(Bhaumik et al., 2023)	97.78%
Our	97.87%
(Tan et al., 2023)	99.85%

the details of the signs. Prolonged training allows the model to focus on details, leading to improved performance.

4 RESULTS

The research was divided into three parts. The first part focused on finding the best architecture and hyperparameters, which provided the foundation for the next experiments. The second part involved comparing the proposed algorithm with other approaches. Lastly, a detailed analysis of the results on the collected dataset was conducted.

4.1 Comparing with Other Approaches

To better demonstrate the effectiveness of the proposed approach, the algorithm was evaluated on publicly available datasets and compared with models described in the literature.

4.1.1 Results on NUS Hand Posture Dataset

The NUS hand posture dataset was used to better compare with other approaches. This dataset includes images of 10 different hand shapes captured from 40 individuals. Each hand shape was captured five times for every subject, resulting in 200 images per class and a total of 2000 images in the dataset. The dataset does not provide user independence.

Due to the relatively small sample size, the batch size was reduced to 512 for training. The model was evaluated using 5-fold cross-validation. The model achieved 97.47%, 97.97%, 97.97%, 98.73% and 97.22% accuracy across the folds, with an average of 97.87%. A comparison with other approaches on this dataset is presented in Table 1.

The proposed model demonstrates excellent performance compared to other models. While the model presented in (Tan et al., 2023) achieves significantly better accuracy, it utilizes Vision Transformer architecture. The Transformer model has 86 million parameters, which is more than nine times the number of parameters in the proposed pipeline. This large num-

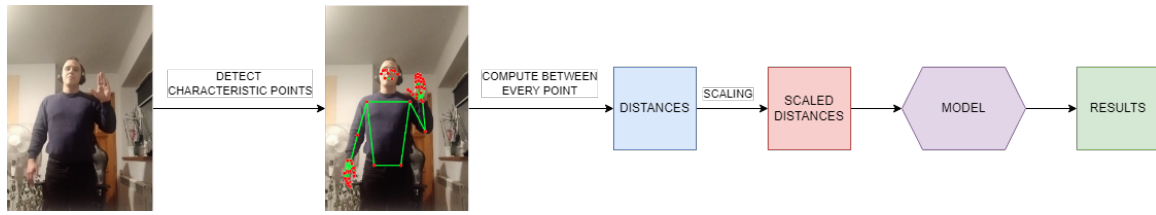


Figure 1: Diagram of the algorithm for the recognition of static signs in Polish Sign Language.

ber of parameters may make the Transformer model unsuitable for mobile devices.

4.1.2 Results on LSA16 Dataset

The LSA16 dataset contains images of 16 hand shapes taken from 10 subjects, with each hand shape performed five times, resulting in a total of 800 images. This dataset ensure user-independence. Several models were evaluated in (Quiroga et al., 2017), from which citation is provided. The authors focused only on the right hand for recognition; thus, for a fair comparison, only distances between characteristic points detected on the right hand were utilized. The batch size was reduced to 128 due to the limited number of samples. The model was tested using 5-fold cross-validation, with each testing pair comprising one male and one female subject. In (Quiroga et al., 2017) 100 runs of stratified randomized sub-sampling cross-validation were conducted. However, it is important to note that some runs may not guarantee user independence, making the task easier in that paper compared to the current approach. The proposed model achieved 97.64%, 98.39%, 96.06%, 94.78% and 94.44% accuracy, which is 96.26% on average. Comparison with other approaches on this dataset is shown in Table 2.

Table 2: Performance of the proposed pipeline and selected models on LSA16 dataset.

Model	Accuracy
Inception (Szegedy et al., 2015)	91.98%
ResNet (He et al., 2016)	93.49%
AllConvolutional (Springenberg et al., 2014)	94.56%
LeNet (LeCun et al., 1998)	95.78%
VGG16 (Simonyan and Zisserman, 2014)	95.92%
Our	96.26%

4.2 The Detailed Analyze of Result on Collected Dataset

The proposed algorithm was evaluated in two different scenarios: performance for "known" and "un-

known" hands. Three metrics were used for evaluation: accuracy, recall, and precision. These metrics help evaluate the performance of a neural network in classification tasks, as they provide insights into different aspects of the network’s performance and help identify areas where the network can be improved. Additionally, a confusion matrix was presented to provide more detailed information about the algorithm’s behaviour and classification outcomes.

4.2.1 Model Performance for "known" Hand

The first experiment focus on comparing the performance of the proposed algorithm tested on the "known" and "unknown" hands. Two models with the same architecture and learning rate (1e-4) were trained for 100 epochs. The first model, tested on "known" hands, was trained using the entire dataset and randomly split into training and testing groups. The training group contained 61383 images, while the remaining images were used for testing. It is the same distribution as in the "unknown" hand scenario. In the second model, the data were split according to the individuals performing the signs, ensuring that the individuals in the training set were distinct from those in the testing set.

The model tested on "known" hands achieved 97.35% accuracy, 96.82% recall, and 97.98% precision. In contrast, the model tested on "unknown" hands produced results of 94.64% accuracy, 93.60% recall, and 95.51% precision. The results highlight the impact of using different individuals’ images in the training and testing set. Using the same "hands" for training and testing can lead to an increase in performance of nearly three percentage points.

4.2.2 Test with only Deaf’s Images and Hearing Images

Due to different methods of photo collection, it is important to evaluate the model’s performance on images of Deaf individuals separately from those of hearing individuals. For the Deaf participants’ images, the model achieved 78.96% accuracy, 78.10% recall and 79.71% precision. In contrast, for the hearing participants’ images, the model achieved 96.95%

Table 3: Confusion matrix fragment which shows results for signs which are the most likely to be mistaken. "0\$0" is a label of sign which means "0" or "o".

		Predicted										
		0\$0	3	4	5	a	ja	n	s	t	ty	
Actual	0\$0	664	0	0	0	0	0	0	78	13	0	
	3	0	558	1	0	0	0	55	0	0	0	
	4	0	18	465	31	0	0	0	5	65	2	
	5	0	2	58	519	0	0	0	2	10	0	
	a	0	0	0	0	587	0	0	0	0	0	
	ja	0	0	0	0	0	501	0	0	0	0	
	n	0	59	3	0	0	0	537	0	0	20	
	s	57	0	0	0	1	0	0	550	23	0	
	t	55	0	0	0	0	0	0	21	558	0	
	ty	43	0	0	0	83	62	0	0	0	491	

accuracy, 96.66% recall and 97.14% precision. The observed differences in results are caused by variability in the Deaf individuals' photos, which were extracted as frames from videos. This led to situations where the subjects were not centred in the frame, unlike the individuals in the training set. Additionally, sometimes the Deaf people were seated, which was not presented in training data. This highlights a limitation of the current model, which can probably be overcome by adding more varied data in the training phase.

4.2.3 Confusion Matrix

A confusion matrix was generated to analyze the results better. Based on the analysis of this matrix, it can be noticed that the model struggles with certain signs. A fragment of the confusion matrix showing the most interesting cases is presented in Table 3.

It can be seen that the algorithm has the most difficulty with the following groups of signs: "0\$0", "s", "t" and "ty" (you in Polish) "3" and "n", "4" and "5", "a" and "ty", and "ja" (I in Polish) and "ty". It is caused by the similarity of these signs and the lack of depth information in the photos.

The model has problems with differentiating between signs "0\$0", "s", "t" and "ty". These signs are shown in Figure 2. Although the algorithm correctly recognizes the actual signs in most cases, it makes errors in about 10% of them. The confusion between the first three signs occurs because they only differ in the position of the index finger's tip relative to the thumb. Mistakes between "0\$0" and "ty" are probably caused by a lack of depth.

Another common error is between signs "3" and "n". The algorithm makes mistakes in about 10%. The signs have an almost identical hand shape; in the sign "3", the fingers are apart from each other, while in the sign "n" they are joined together.

An unusual situation arises between signs "4" and "5". These signs differ by only one finger; in "4", the person shows 4 fingers, and in "5", shows five fingers. Occasionally, during testing, individuals present an alternative hand shape for "4" that was not present in the training set. During training, the number "4" is represented without the little finger, while some examples in the test set represent it without the thumb. Despite this, the model performs well on "4" and "5".

The confusion between the signs "a" and "ty" and the signs "ja" and "ty" is mainly due to a lack of depth information, even though these signs are visually quite different. These signs involve a clenched hand and differ primarily in orientation.

Based on this, it can be concluded that the model has the most difficulty with signs that differ only in distances between fingers and when the signs are visually similar in situations where depth information is lacking.

5 CONCLUSION

In the paper, we presented a method for recognizing static signs in Polish Sign Language. The proposed method is based on image processing and does not require any specialized equipment but can be used with a simple camera. The deep neural network recognizes presented signs based on the detected hands' landmarks. We used our dataset, containing signs presented by Deaf people, for whom it is the primary means of communication with the world, and by people not experienced with PSL. For the experiments, 27 signs were selected and presented by 24 individuals. The proposed method achieved high recognition rates, comparable to the best results in the literature.

Our algorithm was further tested against other approaches using two publicly available datasets: NUS

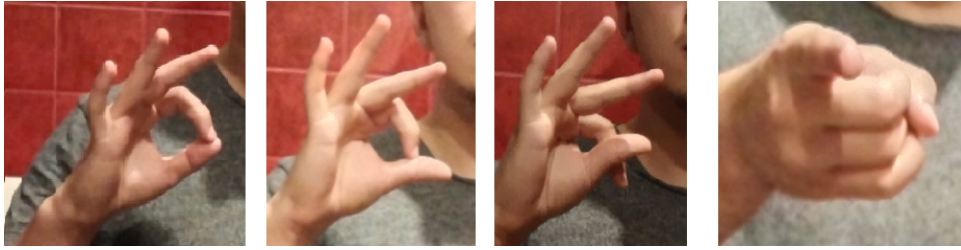


Figure 2: Respectively sign "0\$0", sign "s", sign "t" and sign "ty".

and LSA16. The comparison results show that our solution performs similarly or better than those presented in the literature. Notably, our model has significantly fewer parameters than proposed in articles, making it more suitable for deployment on mobile devices.

Sign language recognition is a challenging topic. Even recognizing simple PSL signs from pictures is difficult. However, this research presents the first step in developing a proper translator for sign languages. Sign languages typically construct sentences through simultaneous combinations of signs, not just sequential gestures. Furthermore, the complexity of sign language interpretation also includes the relative positioning of the hands and spatial configurations. Future work will explore using recursive networks or transformers to recognize sequences of movements that form signs.

The results obtained from the proposed algorithm serve as a foundation for the subsequent phase of our research, which focused on the recognition of dynamic gestures. Ultimately, we aim to use dynamic gesture recognition for an algorithm designed to translate between PSL and spoken Polish.

ETHICS

Data collection was conducted following informed consent procedures, with access to the data restricted solely to the authors. Third-party access is limited to files containing coordinates and distances, while anonymity of individuals involved in the data collection is maintained after contact with the authors. The images generated or analyzed during this study are not publicly available to ensure the protection of personal data.

ACKNOWLEDGEMENTS

This work was financed by the Lodz University of Technology, Faculty of Electrical, Electronic, Com-

puter and Control Engineering as a part of statutory activity (project no. 501/2-24-1-2).

REFERENCES

- Bajaj, Y. and Malhotra, P. (2022). American sign language identification using hand trackpoint analysis. In *International Conference on Innovative Computing and Communications: Proceedings of ICICC 2021, Volume 1*, pages 159–171. Springer.
- Bhaumik, G., Verma, M., Govil, M. C., and Vipparthi, S. K. (2023). Hyfinet: hybrid feature attention network for hand gesture recognition. *Multimedia Tools and Applications*, 82(4):4863–4882.
- Daniels, S., Suciati, N., and Fathichah, C. (2021). Indonesian sign language recognition using yolo method. In *IOP Conference Series: Materials Science and Engineering*, volume 1077, page 012029. IOP Publishing.
- Eid, A. and Schwenker, F. (2023). Visual static hand gesture recognition using convolutional neural network. *Algorithms*, 16(8):361.
- ElBadawy, M., Elons, A. S., Sheded, H., and Tolba, M. F. (2015). A proposed hybrid sensor architecture for arabic sign language recognition. In *Intelligent Systems' 2014: Proceedings of the 7th IEEE International Conference Intelligent Systems IS'2014, September 24-26, 2014, Warsaw, Poland, Volume 2: Tools, Architectures, Systems, Applications*, pages 721–730. Springer.
- Elboushaki, A., Hannane, R., Afdel, K., and Koutti, L. (2020). Multid-cnn: A multi-dimensional feature learning approach based on deep convolutional networks for gesture recognition in rgb-d image sequences. *Expert Systems with Applications*, 139:112829.
- Erol, A., Bebis, G., Nicolescu, M., Boyle, R. D., and Twombly, X. (2007). Vision-based hand pose estimation: A review. *Computer Vision and Image Understanding*, 108(1-2):52–73.
- Fatmi, R., Rashad, S., and Integlia, R. (2019). Comparing ann, svm, and hmm based machine learning methods for american sign language recognition using wearable motion sensors. In *2019 IEEE 9th annual computing and communication workshop and conference (CCWC)*, pages 0290–0297. IEEE.
- Flasiński, M. and Myśliński, S. (2010). On the use of graph parsing for recognition of isolated hand pos-

- tures of polish sign language. *Pattern Recognition*, 43(6):2249–2264.
- He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.
- Jeny, J. R. V., Anjana, A., Monica, K., Sumanth, T., and Matha, A. (2021). Hand gesture recognition for sign language using convolutional neural network. In *2021 5th International Conference on Trends in Electronics and Informatics (ICOEI)*, pages 1713–1721. IEEE.
- Jiang, S., Lv, B., Guo, W., Zhang, C., Wang, H., Sheng, X., and Shull, P. B. (2017). Feasibility of wrist-worn, real-time hand, and surface gesture recognition via semg and imu sensing. *IEEE Transactions on Industrial Informatics*, 14(8):3376–3385.
- Jiang, S., Sun, B., Wang, L., Bai, Y., Li, K., and Fu, Y. (2021). Skeleton aware multi-modal sign language recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3413–3423.
- Kang, B., Tripathi, S., and Nguyen, T. Q. (2015). Real-time sign language fingerspelling recognition using convolutional neural networks from depth map. In *2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR)*, pages 136–140. IEEE.
- Kapuscinski, T. and Wysocki, M. (2005). Recognition of isolated words of the polish sign language. In *Computer Recognition Systems: Proceedings of the 4th International Conference on Computer Recognition Systems CORES'05*, pages 697–704. Springer.
- Kołodziej, M., Szypuła, E., Majkowski, A., and Rak, R. (2022). Using deep learning to recognize the sign alphabet. *Przegląd Elektrotechniczny*, 98.
- Korzeniewska, E., Kania, M., and Zawisłak, R. (2022). Textronic glove translating polish sign language. *Sensors*, 22(18):6788.
- LeCun, Y., Bottou, L. o., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324.
- Linde-Usiekiewicz, J., Czajkowska-Kisil, M., Łacheta, J., and Rutkowski, P. (2014). A corpus-based dictionary of polish sign language (pjm). pages 365–376.
- Linde-Usiekiewicz, J., Czajkowska-Kisil, M., Łacheta, J., and Rutkowski, P. (2016). Korpusowy słownik polskiego języka migowego/corpus-based dictionary of polish sign language.
- Quiroga, F., Antonio, R., Ronchetti, F., Lanzarini, L. C., and Rosete, A. (2017). A study of convolutional architectures for handshape recognition applied to sign language. In *XXIII Congreso Argentino de Ciencias de la Computación (La Plata, 2017)*.
- Raghuvvera, T., Deepthi, R., Mangalashri, R., and Akshaya, R. (2020). A depth-based indian sign language recognition using microsoft kinect. *S = adhan = a*, 45:1–13.
- Rao, G. A., Syamala, K., Kishore, P., and Sastry, A. (2018). Deep convolutional neural networks for sign language recognition. In *2018 conference on signal processing and communication engineering systems (SPACES)*, pages 194–197. IEEE.
- Rastgoo, R., Kiani, K., and Escalera, S. (2020a). Hand sign language recognition using multi-view hand skeleton. *Expert Systems with Applications*, 150:113336.
- Rastgoo, R., Kiani, K., and Escalera, S. (2020b). Video-based isolated hand sign language recognition using a deep cascaded model. *Multimedia Tools and Applications*, 79:22965–22987.
- Sarma, N., Talukdar, A. K., and Sarma, K. K. (2021). Real-time indian sign language recognition system using yolov3 model. In *2021 Sixth International Conference on Image Information Processing (ICIIP)*, volume 6, pages 445–449. IEEE.
- Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- Springenberg, J. T., Dosovitskiy, A., Brox, T., and Riedmiller, M. (2014). Striving for simplicity: The all convolutional net. *arXiv preprint arXiv:1412.6806*.
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9.
- Tan, C. K., Lim, K. M., Chang, R. K. Y., Lee, C. P., and Alqahtani, A. (2023). Hgr-vit: Hand gesture recognition with vision transformer. *Sensors*, 23(12):5555.
- Tan, Y. S., Lim, K. M., and Lee, C. P. (2021). Hand gesture recognition via enhanced densely connected convolutional neural network. *Expert Systems with Applications*, 175:114797.
- Warchoń, D., Kapuściński, T., and Wysocki, M. (2019). Recognition of fingerspelling sequences in polish sign language using point clouds obtained from depth images. *Sensors*, 19(5):1078.
- Wen, F., Zhang, Z., He, T., and Lee, C. (2021). Ai enabled sign language recognition and vr space bidirectional communication using triboelectric smart glove. *Nature communications*, 12(1):5378.