Celal Bayar University Journal of Science

Real-Time Word Detection in Turkish Sign Language with Deep Learning

Abdil Karakan[1]*, Yüksel Oğuz[2]

[1]Afyon Kocatepe University, Department of Electrical, 03950 Afyonkarahisar, Türkiye.
[2]Afyon Kocatepe University, Department of Electrical and Electronics, 03200 Afyonkarahisar, Türkiye.
* akarakan@aku.edu.tr
* Orcid No: 0000-0003-1651-7568

Received: February 13, 2025
Accepted: December 21, 2025
DOI: 10.18466/cbayarfbe.1635817

Abstract

Communication occurs when people can mutually understand each other. Hearing-impaired people have great difficulties communicating with the people around them. Hearing-impaired individuals can often understand others through lip reading. However, they often have difficulty expressing themselves to others. The use of sign language has not become widespread around the world. Hearing impaired language; Apart from hearing impaired people, the number of people who know is very low. The study aims to detect the 50 most commonly used words of hearing-impaired individuals in hospitals and especially in emergency services, using deep learning. The study is a word-based detection process, not a letter-based one. In the study, a movement was detected, not a single photograph. For the study, a data set was created using videos taken from different angles of 50 words used in hospitals by 100 volunteers. Grayscale conversion, tilt augmentation, blurring, variability enhancement, noise addition, image brightness change, colour vividness change, perspective change, sizing, and position change were added to the photographs that make up the data set. With these additions, the error that may occur due to any distortion that may occur from the camera is minimized. Thus, the accuracy rate in the detection process with images taken from the camera in real-time has been increased. The created data set was run on the YOLOv8 algorithm. The model achieved an average precision of 95.0% and a mean average precision (mAP) of 95.1%. An accuracy rate of 89.40% was achieved in real-world testing.

Keywords: Sign language, sign language recognition, word detection, deep learning, YOLOv8.

1. Introduction

Sign language is a form of visual communication that people with hearing problems use to express themselves using their hands, facial expressions, and bodies. Sign language has structural rules, just like spoken and written languages. Sign language is not universal. Every country has its own spoken language as well as its own sign language. Similar to spoken languages, sign languages vary across regions. Each country has different structural rules for its own sign language. Although sign languages are influenced by spoken languages, they have a different structure. In sign language, each movement corresponds to a word or letter. Sign language is basically divided into two: word-based and letter-based [1].

Letter-based sign language is also called finger alphabet. Each movement made with the finger alphabet corresponds to a letter in the alphabet. For example, in the Turkish sign language finger alphabet, each movement corresponds to one of the 29 letters in the Turkish alphabet. In communication with the finger alphabet, each letter is expressed using the hands, and words are formed from the letters. The finger alphabet is generally used to express proper nouns and words that are not used very often. It is also used in abbreviations, suffixes, scientific terms, and synonyms when what is meant cannot be expressed exactly [2].

Many studies have been conducted on letter-based sign language detection. These studies are carried out in two ways: hardware or non-hardware. Gloves are mostly used in hardware work. Hand movements are examined with sensors added to the gloves. According to the movement of the hand, the letter equivalent in sign language is determined [3-6]. The use of hardware devices in such studies has not become widespread

because it is both costly and not very useful. Another method for letter-based detection in sign language is image-based detection. In this type of study, hand movement is examined on an image-based basis. Letters are detected according to the movement of the hand. Many different architectures are used in image-based systems. A very high degree of accuracy is achieved [7-12]. However, since only letters are detected in the system, there are many difficulties in communication. To overcome this difficulty, a finger spelling-to-text translation study was carried out. In the study, no direct word detection was made. The word was reached by identifying the syllables one by one. 221 volunteers participated in the study. Images of individual syllables were taken from each volunteer. The data set consists of 13444 samples in total. In the study carried out in real time, the hearing impaired person shows individual syllables. Syllables are detected one by one and words are formed. The study achieved 94% accuracy. Syllable detection was detected through a single frame, not a process [13].

In word-based sign language, body movements, facial expressions, and facial expressions are also important along with hands. In word-based sign language, each gesture corresponds to a word. Many databases have been created for sign language detection at the word level. The detection process is done with these databases. A data set for German sign language was prepared by Ong et al. There are 40 words in this data set. The data set was created with the participation of 15 people. Participants were recorded saying 40 words in front of the camera. Each participant repeated each word 5 times. As a result of the study, an accuracy between 55% and 87% was obtained. In another study, they identified 20 words. There are only 6 participants in this dataset. A total of 840 samples of 20 words were made in the data set. An accuracy rate of 49% was obtained in these studies [14-15].

Athitsos and his colleagues prepared a data set for American Sign Language. There are more than 3300 classes in this dataset. Data set classes were prepared by 6 volunteers. There are 9800 samples in total in the data set. This data set has been used in many studies [16-17]. The prepared data set was also used in computer vision applications. In the study, they used a system that models the transition probabilities between beginning and ending hand signs in word signs. This system detects signs made with one hand and signs made with two hands. A total of 657 signs were recognized, including 333 one-handed and 324 two-handed gestures. As a result of the study, an accuracy between 30.4% and 44.4% was obtained. Kim and his colleagues used the data set prepared by Athitsos in their study. In his studies, he processed images by taking images from the RGB camera. It matched the image taken from the camera based on hand tracking and pose estimation. As a result of the study, they reached an accuracy rate of 53% [18]. In their study, Metaxas et al. identified 5 signs from a data set for American Sign Language. There are a total of 350 samples of 5 signals in the data set. The data set includes both one-handed and two-handed samples. The detection process was made by taking into account the shape, position, and movement of the hand. As a result of the study, an accuracy of 93.3% was obtained [19].

Polish sign language dataset was created by Oszust and Wysocki. There are 30 words in the created data set. The data set was created 10 times by only one participant. There are 300 samples in total in the data set. The created data set was used in classification studies using the full word model and data-driven subsequence model. A Kinect sensor was used for study detection. In classification, the skeleton image of the person is first created. In this image, the positions of the most important joints are determined. Then the hands are detected. Detection is done with the movement of the hands. As a result of the study, an accuracy of 89.3% was obtained. This created data set has been used in different studies such as dynamic time warping and the nearest neighbour technique [20-21].

Ronchetti and his colleagues created an Argentinian sign language dataset called LSA 64. This data set consists of 64 different signals. The data set was created with 10 participants. Each participant repeated 64 different signs in different numbers. The data set consists of 3200 samples in total. With this study, the first comprehensive data set in Argentine sign language was created. This created data set is designed for machine learning. Participants wore specially coloured gloves. In this way, it was possible to focus on the characteristics of the signs. In the study, hand movements, positions, and structures of shapes were examined. The created data set was used to analyse the performance of different methods [22-23].

Chai and his colleagues created the Chinese sign language dataset called DEVISIGN-G. The dataset consists of 432 samples recorded from 8 volunteers for 36 classes. He obtained successful results in his work on hand segmentation [24-25].

In this study, a new word-level data set was created to be used in hospitals and especially emergency services. The data set consists of the 50 most used words in hospitals and emergency departments. The data set was created by repeating 50 words from different angles with 100 volunteer participants. In the continuation of the study, the samples were pre-treated. Here, the samples in video format are first divided into frames, and then these frames are combined sequentially, and turned into sequential single images. The work will be detected in real-time. The detection process will be carried out with the images coming from the camera. For this reason, errors that may occur in the camera will

reduce the accuracy rate. For this reason, grayscale, gradient addition, blurring, variability addition, noise addition, image brightness change, colour vividness change, perspective change, dimensioning and position change were added to all images in the data set. With these additions, the error that may occur due to any distortion that may occur from the camera is minimized. With the operations performed, a total of 48379 samples were reached in the data set. The dataset was run on YOLOV8 architecture. As a result of the study, an accuracy rate of 98.04% was achieved. In real-life applications, the accuracy was 92.45%.

The novelty of this research lies in the creation of a domain-specific Turkish Sign Language dataset tailored for healthcare communication and the optimization of YOLOv8 for real-time inference. While YOLOv8 is used as the backbone, the model parameters and data augmentations were carefully adjusted for hospital and emergency-related gestures, providing a new contextual contribution rather than a purely architectural one.

2. Materials and Methods

This study consists of computer software and a webcam connected to the computer. The study carried out consists of two stages. In the first stage, every movement made by the person standing in front of the camera was accurately recognized using sign language. In the second stage, the result of the recognition process can be reported to the user in written form. For this, YOLOv8, one of the deep learning architectures, was used. Figure 1 shows the general flow chart of the study.

Figure 1. General flow chart of the study

In order to accurately recognize sign language movements using the deep learning method, the system must first be trained. A lot of data is needed to both train and test the system [26,27]. For this reason, the data set containing 50 words used in hospitals and especially in emergency departments was created with images taken from 100 different individuals.

2.1. Dataset

Communication about illnesses is always difficult in hospitals and emergency departments. The pain and weakness caused by the disease make communication very difficult. Communication in hospital emergency departments is even more difficult for hearing-impaired individuals. In such environments, the concept of time is important for the patient to explain his complaint. With the study, 50 words that speech-impaired individuals may frequently need in the hospital environment were determined. A data set was created by taking samples from 100 different individuals in different numbers. The data set created with 100 participants from 50 classes contains a total of 48379 samples. While sampling, volunteers made the sign language gesture from 3 different angles: from the front, from the right, and from the left. Table 1 shows the disease classes and sampling numbers used in the study. The dataset was collected from 100 healthy adult volunteers (52 males and 48 females) aged between 18 and 55 years, all of whom were native users of Turkish Sign Language. Each participant provided written consent prior to data collection. The complete dataset was divided into training (70%), validation (20%), and testing (10%) subsets to ensure balanced class representation.

Table 1. Medical terms used in the dataset and sampling numbers in the data set.

Name of the Disease	Number of Sampling	Name of the Disease	Number of Sampling
Abdomen	928	Heartburn	887
Allergy	943	Hemorrhoids	956
Anorexia	875	Gallstones	963
Asthma	903	Goiter	947
Backache	915	Indigestion	878
Bee Sting	989	Itching	891
Brain	945	Kidney	934
Blood	937	Liver	978
Blood pressure	887	Medicine	971
Bone	909	Mouth	925
Breath	981	Nausea	984
Constipation	992	Pain	933
Cramp	978	Pulse	908
Diabetes	985	Rheumatism	893
Diarrhea	970	Snake Bite	879
Dizziness	968	Sore Throat	966
Earache	935	Spew	978
Epilepsy	900	Stomach Ache	932
Exude	968	Throat	957
Fever	925	Tonsil	981
Flu	897	Toothache	987
Food Poisoning	914	Tremble	875
Fracture	907	Weakness	871
Headache	927	Wound	997
Heart	980	Sputum	927

Celal Bayar University Journal of Science
Volume 22, Issue 1, 2026, p 132-141
Doi: 10.18466/cbayarfbe.1635817

In the proposed method, the data set is first pre-processed. At this stage, video samples were first decomposed into frames, which were then sequentially combined into composite images. Grayscale, tilt addition, blurring, variability addition, noise addition, image brightness change, colour vividness change, perspective change, dimensioning, and position change were added to each individual image. With these additions, the error that may occur due to any distortion that may occur from the camera is minimized. Thus, the accuracy rate in the detection process with images taken from the camera in real-time has been increased. Figure 2 shows the changes applied to the photographs in the data set.

Figure 2. a) Modification of size and position, **b)** Conversion to grayscale, **c)** Slope adjustment**, d)** Application of blur, **e)** Rotation variability, **f)** Introduction of noise, **g)** Adjustment of brightness, **h)** Alteration of colour intensity, **i)** Perspective distortion.

In Figure 2a, the positioning and scaling of database images were adjusted by 25% variation enhancing the model's durability against different camera angles. Figure 2b presents the database images in grayscale. In Figure 2c, a slope adjustment of +15% and -15% has been applied to the images. Figure 2d incorporates random Gaussian blurring to improve resilience against focus variations. In Figure 2e, rotational changes of +15% and -15% have been introduced to strengthen resistance to camera roll. Figure 2f includes noise to enhance robustness against camera-related artifacts. In Figure 2g, brightness levels have been modified by ±15% to improve adaptability to lighting and camera differences. Figure 2h features randomized adjustments to colour intensity. Lastly, Figure 2i applies perspective distortions to increase resistance to variations in camera position, subject alignment, and optical distortions.

2.2. Deep Learning

Deep learning is one of the machine learning methods. In the study conducted by Dechter, the term deep

learning was mentioned for the first time in the field of machine learning. In the study conducted by Aizenberg; the term deep learning as a neural network is mentioned in detail. Deep learning structures are actually multilayer artificial neural networks.

Deep neural networks are capable of performing both feature extraction and classification autonomously. In this respect, deep learning differs from other classification methods. The term deep refers to the number of layers in the network. As the number of layers is increased, the structure of the network becomes deeper. While classical artificial neural networks may consist of only two or three layers; Deep networks can consist of hundreds of layers. Deep neural networks are architectures capable of parallel computation. This parallelism reduces processing time. Otherwise, processing huge data may take weeks. For model training, the YOLOv8m configuration was used with a batch size of 16, learning rate of 0.001, and 100 training epochs. The Adam optimizer and cosine learning rate scheduling were applied. Data augmentation techniques such as random rotation, brightness adjustment, and Gaussian noise were utilized to increase data diversity.

In order to achieve the desired results in deep learning models in a much shorter time, computers with high computational capacity are needed. For example; to develop a driverless vehicle using a deep learning model, millions of images, thousands of hours of videos and a high-performance Graphics Processing Unit (GPU) with a parallel structure are needed. The most important factor why deep learning has become so popular is that it can perform more efficient parallel calculations thanks to its ability to perform calculations with GPU. GPU enables much larger amounts of data to be trained in a much shorter time. When applications run using the Central Processing Unit and GPU are compared, it has been seen that the GPU has approximately 100 times more performance. Deep learning applications; It is used in many fields such as automatic driving, space and defence, medical research, industrial automation and electronics. It has been observed that deep learning models can achieve much more successful results compared to other machine learning techniques.

Deep neural networks; they are parallel structures consisting of input layer, hidden layers, and output layer and are inspired by biological nervous systems. In deep neural networks, layers are connected to each other through neurons, and the output of each layer becomes the input data of the next layer.

3. Results and Discussion

In the study, YOLOv5, YOLOv6, YOLOv7, and YOLOv8 algorithms were used. As a result of the study, the average accuracies of fifty different words were taken. The results of four different algorithms are shown in Table 2.

Table 2. Average accuracy results of four different architectures used in the study.

Name of Architecture	Accuracy (Average %)
YOLOv5	88
YOLOv6	90
YOLOv7	93
YOLOv8	95

In addition to accuracy, the YOLOv8 model achieved a mean precision of 94.6%, recall of 93.8%, and F1-score of 94.2% on the test set. These results indicate a balanced performance across all classes.YOLOv8 enhances the capabilities of previous versions by introducing powerful new features and improvements. This enables real-time object detection on image and video data with improved accuracy and precision. The high accuracy and speed of YOLOv8 makes the computer vision model stand out from previous versions. Unlike previous versions, YOLOv8 can now detect objects in real-time without any delays. YOLOv8 offers a flexible architecture that developers can customize to fit their specific needs. YOLOv8 has new adaptive training capabilities and techniques, such as loss-of-function compensation during training. With YOLOv8's new semantic segmentation and class estimation capabilities, the model can perform basic object detection tasks, detecting colour, texture, and even relationships between objects. With these new and improved features, YOLOv8 performs detection with higher accuracy.

Among the evaluated models, YOLOv8 yielded the highest accuracy and was therefore adopted in this study. In the study, 30 words used in hospitals and emergency service were identified. Figure 4 shows the ROC curve resulting from the study.

The dataset was split into training (70%), validation (20%), and testing (10%) subsets. The YOLOv8m model was trained for 100 epochs using the Adam optimizer with an initial learning rate of 0.001 and a batch size of 16. Early stopping and cosine learning rate scheduling were applied to prevent overfitting. Training was performed on an NVIDIA RTX 4090 GPU (24 GB VRAM) and Intel i9 CPU system. The real-time implementation achieved 27 FPS on average, with an end-to-end latency of approximately 36 ms per frame, confirming real-time performance.

Celal Bayar University Journal of Science
Volume 22, Issue 1, 2026, p 132-141
Doi: 10.18466/cbayarfbe.1635817

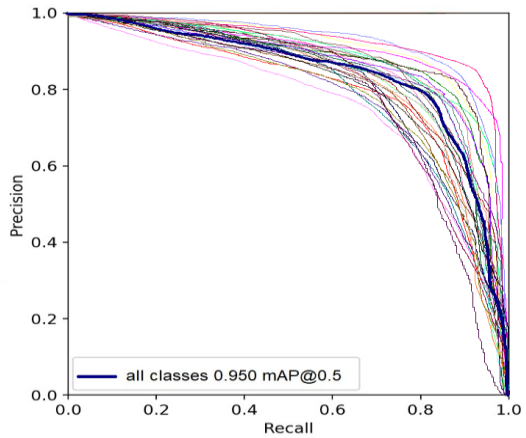

Figure 4. ROC curve showing the classification performance of the YOLOv8 model across 30 hospital-related sign language classes

In the study, 30 words were identified with different accuracy rates. Some words achieved high accuracy. Lower accuracy was achieved for some words. As a result of the study, an average accuracy of 95.00% was revealed. Figure 5 shows the confusion matrix realized in the system.

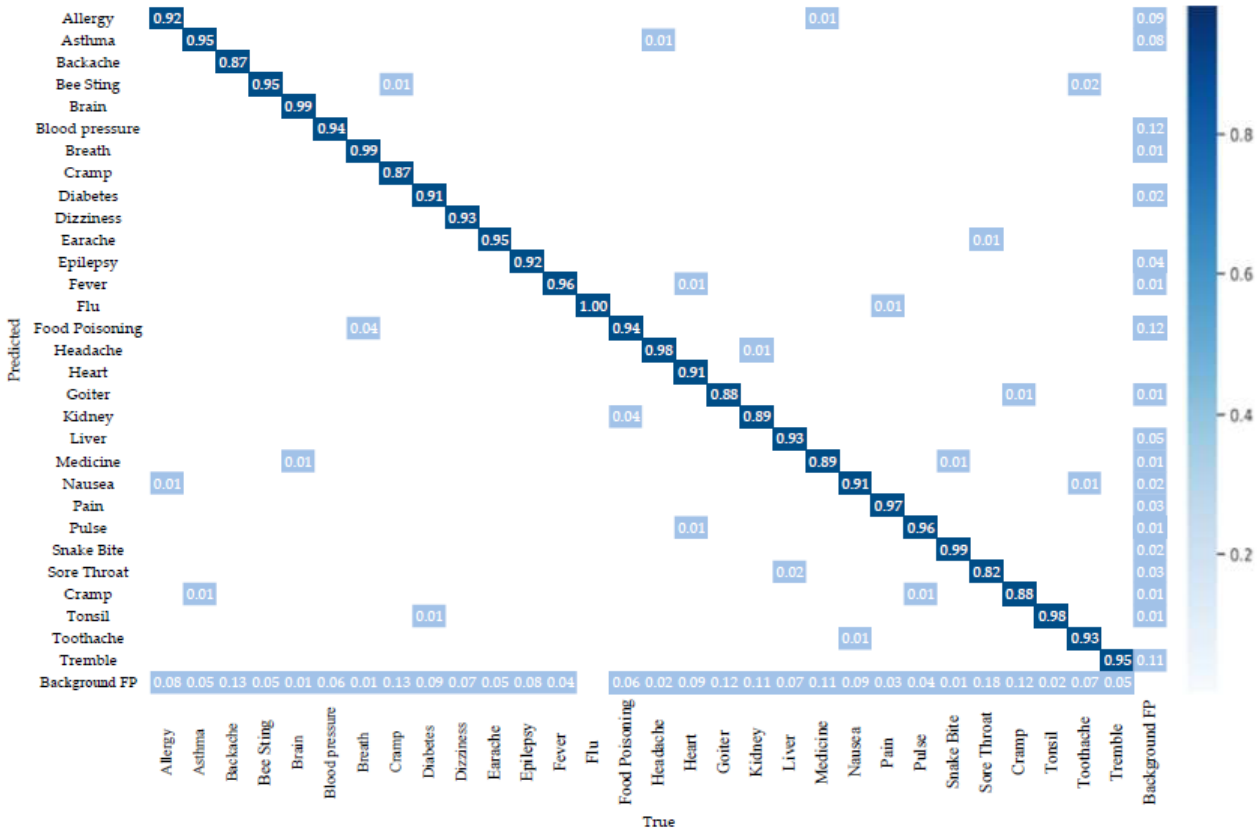

Figure 5. Confusion matrix obtained from YOLOv8 results, illustrating per-class performance.

The study was implemented in a real-life scenario. The application was made with a webcam and computer. The camera image taken on the webcam is divided into 40 different frames. Then, the detection process is carried out in the

YOLOv8 architecture. Figure 6 shows the word headache divided into 40 different squares. Although the dataset initially included 50 sign language words, only 30 of them were evaluated in real-life testing due to time and participant availability constraints. The remaining 20 words were validated through offline model inference to ensure dataset consistency. This explains the difference between the total dataset size and the number of words tested in live application.

Celal Bayar University Journal of Science
Volume 22, Issue 1, 2026, p 132-141
Doi: 10.18466/cbayarfbe.1635817

Figure 6. The word "headache" in sign language divided into 30 different squares

Different numbers of the 30 words in the study were applied to different volunteers. As a result of the application, an average accuracy of 89.4% was obtained. Figure 7 shows the results of the actual application.

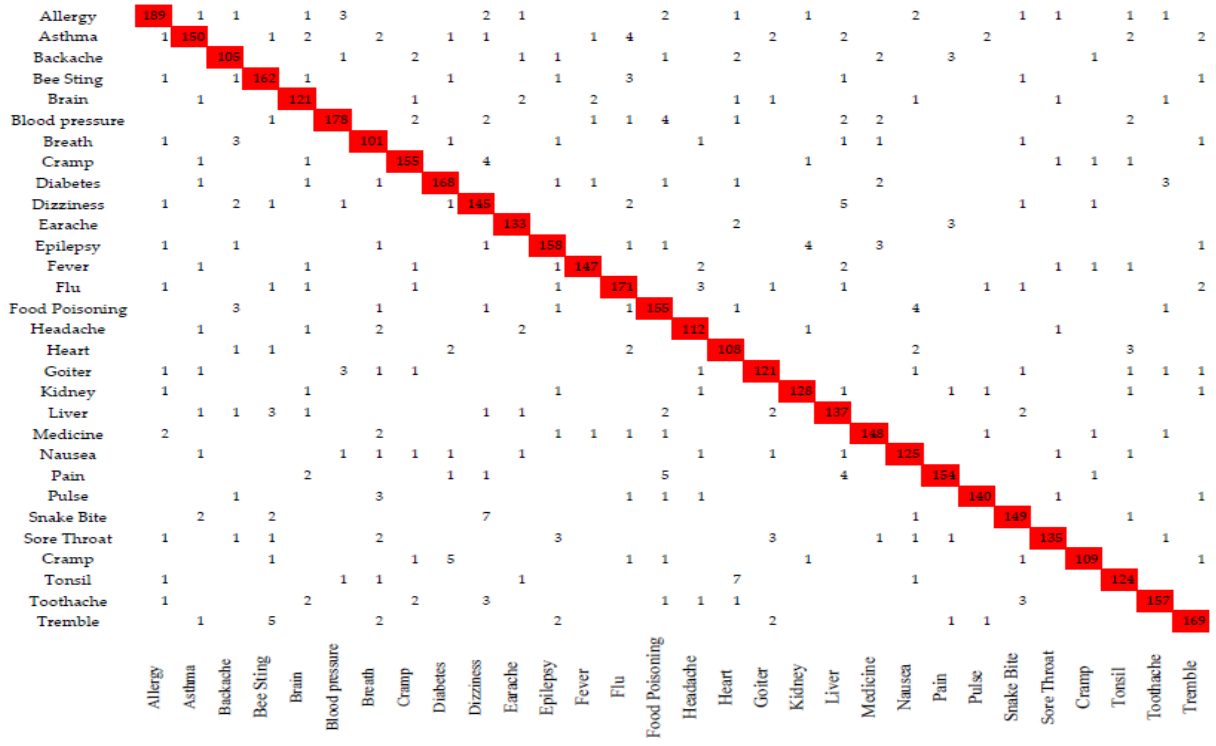

Figure 7. Confusion matrix resulting from YOLOV8 architecture

Table 3. Literature studies

Authors	Architecture	Accuracy
Ameen et al., [28]	Convet	80.00
Beena [29]	Kinect Depth	94.68
Masood et al., [30]	VGG16	96.00
Kim et al., [31]		92.00
Siddique et al., [32]		98.05
Taskiran et al., [33]	AlexNet	42
Shi et al., [34]	CNN	94.00
Shi et al., [35]	RecurentCNN	86.90
Miah et al., [36]	BenSignNet	97.60
Raghuveera et al., [37]	SWM	71.85
Kwolek et al., [38]	ResNet-34	89.00
Shamrat et al., [39]	Custom CNN	99.80
Mannah et al., [40]	Deep Learning	97.67
Gadekallu et al., [41]	CNN	90.00
Aliyev et al.;[42]	MobileNetV2	89.00
Angona et al., [43]	MobileNet	95.71
Özcan et al., [44]	GoogLeNet	88.62
Talukder et al., [45]	YOLOv5x	98.56
Pacal et al., [46]	CNN	98.76
Siddique et al., [47]	YOLOv7	94.91

Compared to other Turkish Sign Language studies, such as ERUSLR [44] and CNN-based approaches [46], the proposed model achieves competitive accuracy while targeting a domain-specific vocabulary (hospital and emergency context). This specialization provides a practical application focus rather than a general language modelling objective.

There are many studies on sign language in the literature. The majority of studies on sign language are on letter identification. Electronic gloves were worn in the first studies to detect letters. Such studies were both costly and not useful. Then, letters were detected using image processing. Table 3 shows studies conducted with sign language.

Many studies have been conducted on letter detection in sign language using image processing. A high level of accuracy was achieved in these studies. However, when only letters are detected in these studies, the operation becomes difficult. It takes a long time to explain a sentence. The detection process is done by looking at only one frame. It takes a lot of time to explain a sentence. As such, letter identification studies cannot be useful.

The study carried out word detection in sign language. Communication is very easy. Especially in hospitals and emergency rooms where time is very valuable. The study is very useful because it detects a movement, not a frame.

The study carried out word detection in sign language. Communication is very easy. Especially in hospitals and emergency rooms where time is very valuable. The study is very useful because it detects a movement, not a frame.

The study determined the most commonly used disease names in hospitals and emergency services. In the study, a dataset was created from videos of 50 words used in hospitals taken from different angles by 100 volunteers. The dataset includes right-handed and left-handed people. A new and original dataset was created for this study. This data was used in four different YOLO architectures. The highest accuracy was achieved in the YOLOv8 architecture with 95%.

4. Conclusion

In the study, a new data set was created containing 50 different words that are frequently used especially in hospitals and emergency services. The dataset was constructed with the assistance of 100 volunteers. Each word was repeated in different numbers and added to the data set. The words were shot from three different angles: from the right, left and front. The created data set was run on YOLOv8 architecture. Different accuracy rates were determined for each word. An average accuracy of 95.0% was achieved in the study. The study was implemented in a real-life scenario. 30 different words were tested in different numbers. The accuracy rate was 89.4%.

The study demonstrates the potential of real-time deep learning systems to support inclusive communication in healthcare. However, the limited vocabulary size restricts model generalization. Future research will focus on expanding the dataset beyond 200 words, incorporating multi-modal cues such as facial expressions and hand trajectories, and exploring hybrid Transformer–YOLO architectures to enhance contextual understanding and robustness in real-world scenarios.

Overall, the findings demonstrate that the proposed YOLOv8-based system can facilitate effective and real-time communication between hearing-impaired individuals and healthcare personnel, offering a promising direction for future assistive technologies.

Author's Contributions

Abdil Karakan: Drafted and wrote the manuscript, performed the experiment, and conducted the result analysis.

Yüksel Oğuz: Assisted in the analytical evaluation of the structure, supervised the experimental process, contributed to the interpretation of results, and supported the preparation of the manuscript.

Ethics

There are no ethical issues after the publication of this manuscript.

References

[1]. Takahashi, T, F, Kishino. 1992. A hand gesture recognition method and its application. *Systems and Computers in Japan*, 23 (3), pp. 38-48. (htpss://doi.org/10.1002/scj.4690230304)

[2]. Grobel, K, Hienz, H. 1996. Fuzzy video-based handshape recognition. *Proceedings of the ACM Symposium on Applied Computing*, pp. 614-618. (https://doi.org/10.1145/331119.331469.)

[3]. Watanabe, K, Iwai, Y, Yagi, Y, Yachida, M. 1999. Recognition of sign language alphabet using coloured gloves. *Systems and Computers in Japan*, 30 (4), pp. 51-61. (https://doi.org/10.1002/(SICI)1520-684X(199904)30:4<51::AID-SCJ6>3.0.CO;2-%23.)

[4]. Wang, H, Leu, MC, Oz, C. 2006. American Sign Language recognition using multi-dimensional Hidden Markov Models. Journal of *Information Science and Engineering*, 22 (5), pp. 1109-1123. (htpss://doi.org/10.6688/JISE.2006.22.5.8.)

[5]. Shanableh, T, Assaleh, K. 2011. User-independent recognition of Arabic sign language for facilitating communication with the deaf community. *Digital Signal Processing: A Review Journal*, 21 (4), pp. 535-542. (htpss://doi.org/10.1016/j.dsp.2011.01.015.)

[6]. Shamrat, FM. 2021. Bangla numerical sign language recognition using convolutional neural network CNNs, *Indonesian Journal of Electrical Engineering and Computer Science,* 23, pp. 405–413. (https://doi.org/10.11591/ijeecs.v23.i1.pp405-413.)

[7]. Rastgoo, R, Kiania, K, Escalerab, S. 2021. Sign Language Recognition: A Deep Survey. *Expert Systems with Applications*, 164, pp. 113794. (htpss://doi.org/10.1016/j.eswa.2020.113794.)

[8]. Yu, S, Jia, S, Xu, C. 2017. Convolutional neural networks for hyperspectral image classification. *Neurocomputing,* 219, pp. 88-98. (https://doi.org/10.1016/j.neucom.2016.09.)

[9]. Nam, Y, Lee, C. 2021. Cascaded convolutional neural network architecture for speech emotion recognition in noisy conditions. *Sensors*, 21(13), pp. 4399. (https://doi.org/10.3390/s21134399.)

[10]. Anantha, RG, Kishore, PVV, Sastry, ASCS, Anil, D, Kiran, K E. 2018. Selfie continuous sign language recognition with neural network classifier, *Lecture Notes in Electrical Engineering,* 434, 31-40. (htpss://doi.org/10.1111/exsy.12197.)

[11]. Rao, GA, Syamala, K, Kishore, PVV, Sastry, ASC. 2018. Deep convolutional neural networks for sign language recognition. *2018 Conference on Signal Processing And Communication Engineering Systems.* pp. 194-197, Vijayawada, India, 04-05 January 2018. (htpss://doir.org/10.1109/SPACES.2018.8316344.)

[12]. Siddique, S, Islam, S, Neon, EE, Sabbir, T, Naheen, IT, Khan, R. 2023. Deep Learning-based Bangla Sign Language Detection with an Edge Device. *Intelligent Systems with Applications,* 18, pp. 200224. (https://doi.org/10.1016/j.iswa.2023.200224 30 March 2023.)

[13]. Jamaladdin, H, Nigar, A, Aykhan, N, Samir, D, Toghrul, T. 2023. Development of a hybrid word recognition system and dataset for the Azerbaijani Sign Language dactyl alphabet, *Speech Communication.* (153) 102960, (https://doi.org/10.1016/j.specom.2023.102960.)

[14]. Ong, EJ, Cooper, H, Pugeault, N, Bowden, R. 2012. Sign language recognition using sequential pattern trees. *Conference on Computer Vision and Pattern Recognition, Washington-USA,* pp. 2200–2207. (https://doi.org/10.1109/CVPR.2012.6247928.)

[15]. Ong, EJ, Koller, O, Pugeault, N, Bowden, R. 2014. Sign spotting using hierarchical sequential patterns with temporal intervals. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Washington-USA,* pp. 1923–193. (http://dx.doi.org/10.1109/CVPR.2014.248.)

[16]. Athitsos, V, Neidle, C, Sclaroff, S, Nash, J, Stefan, A, Yuan, Q, Thangali, A, 2008. The American Sign Language lexicon video dataset, *2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, Alaska-USA,* pp. 1–8. (http://dx.doi.org/10.1109/CVPRW.2008.4563181.)

[17]. Neidle, C, Thangali, A, Sclaroff, S. 2012. Challenges in development of the American Sign Language lexicon video dataset (asllvd) corpus, *Proc.5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon, Language Resources and Evaluation Conference (LREC) 2012, İstanbul-Turkey,* pp. 1-8.

[18]. Kim, JH, Kim, N, Park, H, Park, JC. 2016. Enhanced sign language transcription system via hand tracking and pose estimation, *Journal of Computing Science and Engineering,* 10 (3), 95–101. (http://dx.doi.org/10.5626/JCSE.2016.10.3.95.)

[19]. Metaxas, D, Dilsizian, M, Neidle, C. 2018. Scalable ASL sign recognition using model-based machine learning and linguistically annotated corpora, *8th Workshop on the Representation & Processing of Sign Languages: Involving the Language Community, Language Resources and Evaluation Conference, Miyazaki-Japan*, pp. 1-5, 12 Mayıs, 2018.

[20]. Oszust, M, Wysocki, M. 2013. Polish sign language words recognition with Kinect, *2013 6th International Conference on Human System Interactions (HSI), Gdansk-Poland*, 219–226, 6-8 Haziran, 2013. (http://dx.doi.org/10.1109/HSI.2013.6577826.)

[21]. Oszust, M, Wysocki. M. 2014. Some Approaches to Recognition of Sign Language Dynamic Expressions with Kinect. *Advances in Intelligent Systems and Computing,* vol 300, Hippe Zdzisaw S., Springer Cham, pp. 75-86. (http://dx.doi.org/10.1007/978-3-319-08491-6_7.)

[22]. Kapuscinski, T, Oszust, M, Wysocki, M, Warchol D. 2015. Recognition of hand gestures observed by depth cameras, *International Journal of Advanced Robotic Systems*, 12 (4), 36, pp. 1-15. (http://dx.doi.org/10.5772/60091.)

[23]. Ronchetti, F, Quiroga, F, Estrebou, CA, Lanzarini, LC, Rosete, A. 2016. *LSA64: an Argentinian sign language dataset, CACIC 2016, Roma-Italy,* pp. 1-10, 3-7. (http://dx.doi.org/10.48550/arXiv.2310.17429.)

[24]. Ronchetti, F. 2017. Thesis Overview: Dynamic Gesture Recognition and its Application to Sign Language, *Journal of Computer Science and Technology,* 17, pp. 1–10. (http://dx.doi.org/10.24215/16666038.17.e21.)

[25]. Chai, X, Wang, H, Chen, X. 2014. The devising large vocabulary of Chinese sign language database and baseline evaluations, Technical report VIPL-TR-14- SLR-001. *Key Lab of Intelligent Information Processing of Chinese Academy of Sciences (CAS), Institute of Computing Technology,* 2014. (http://dx.doi.org/10.11999/JEIT221051.)

[26]. Rusul, HH, Rasha, MH, Inaam, SA. 2023. Yolo Versions Architecture: Review, *International Journal of Advances in Scientific Research and Engineering,* Vol. 9, 11, pp.1-20. (https://doi.org/10.31695/IJASRE.2023.9.11.7.)

[27]. Terven, J, Diana-Margarita, C, Julio-Alejandro, R. 2023. A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS, *Machine Learning and Knowledge Extraction,* Vol. 5, 4, 1680-1716. (https://doi.org/10.3390/make5040083.)

[28]. Ameen, C, Vadera, S. 2017. A convolutional neural network to classify American Sign Language fingerspelling from depth and colour images. *Expert Syst.* 34 (3), e12197. (http://dx.doi.org/10.1111/exsy.12197.)

[29]. Beena, M. 2017. Automatic Sign Language Finger Spelling Using Convolution Neural Network: Analysis.

[30]. Masood, S, Thuwal, HC, Srivastava, A. 2018. American signlanguage character recognition using convolution neural network. *Computing and Informatics. Springer Singapore, Singapore,* pp. 403–412.
(http://dx.doi.org/10.1007/978-3-319-16178-5_40.)

[31]. Kim, T, Keane, J, Wang, W, Tang, H, Riggle, J, Shakhnarovich, G, Brentari, D, Livescu, K. 2017. Lexicon-free fingerspelling recognition from video: Data, models, and signer adaptation. *Comput. Speech Lang*. 46, pp. 209–232. (http://dx.doi.org/10.1016/j.csl.2017.05.009.)

[32]. Taskiran, M, Killioglu, M, Kahraman, N. 2018. A real-time system for recognition of American Sign Language by using deep learning. *In: 2018 41st International Conference on Telecommunications and Signal Processing.* TSP, pp. 1–5. (http://dx.doi.org/10.1109/TSP.2018.8441304.)

[33]. Shi, B, Rio, AMD, Keane, J, Brentari, D, Shakhnarovich, G, Livescu, K. 2019. Fingerspelling recognition in the wild with iterative visual attention. *In: 2019 IEEE/CVF International Conference on Computer Vision. ICCV, IEEE,* (http://dx.doi.org/10.1109/iccv.2019.00550.)

[34]. Shi, B, Rio, AMD, Keane, J, Michaux, J, Brentari, D, Shakhnarovich, G, Livescu, K. 2018. American Sign Language fingerspelling recognition in the wild. *In: 2018 IEEE Spoken Language Technology Workshop. SLT, IEEE,* (http://dx.doi.org/10.1109/slt.2018.8639639.)

[35]. Warcho, D, Kapuski, T, Wysocki, M. 2019. Recognition of fingerspelling sequences in polish sign language using point clouds obtained from depth images. *Sensors*, 19 (5), 1078. (http://dx.doi.org/10.3390/s19051078.)

[36]. Raghuveera, T, Deepthi, R, Mangalashri, R, Akshaya, R. 2020. A depth-based Indian Sign Language recognition using Microsoft Kinect, *Sadhana*, 45-34, pp. 1-13.
(http://dx.doi.org/10.1007/s12046-019-1250-6.)

[37]. Kwolek, B, Baczynski, W, Sako, S. 2021. Recognition of JSL fingerspelling using deep convolutional neural networks. *Neurocomputing,* 456, pp. 586–598. (http://dx.doi.org/10.1016/j.neucom.2021.03.133.)

[38]. Mannan, A, Abbasi, A, Javed, AR, Ahsan, A, Gadekallu, TR, Xin, Q. 2022. Hypertuned Deep Convolution Neural Network for Sing Language Recognition. *Computational Intelligence and Neuroscience,* pp. 1–10.
(https://doi.org/10.1155/2022/1450822.)

[39]. Gadekallu, TR, Srivastava, G, Liyanage, M, Iyapparaja, M, Chowdhary, CL, Koppu, S, Maddikunta, PKR. 2022. Hand gesture recognition based on a Harris Hawks optimized convolution neural network. *Comput. Electr. Eng.* 100, 107836. (http://dx.doi.org/10.1016/j.compeleceng.2022.107836.)

[40]. Aliyev, S, Almisreb, AA, Turaev, S. 2022. Azerbaijani sign language recognition using machine learning approach. *J. Phys. Conf. Ser*. 2251 (1), 012007.
(http://dx.doi.org/10.1088/1742-6596/2251/1/012007.)

[41]. Angona, TM. 2020. Automated Bangla sign language translation system for alphabets by means of MobileNet, *TELKOMNIKA (Telecommunication Computing Electronics and Control),* 18, pp. 1292–1301.
(https://doi.org/10.12928/telkomnika. v18i3.15311.)

[42]. Talukder, D, Jahara, F, Barua, S, Haque, MM. 2021. *OkkhorNama: BdSL image dataset for real time object detection algorithms, IEEE Region 10 Symposium,* pp. 1–6. (https://doi.org/10.1109/TENSYMP52854.2021.9550907.)

[43]. Siddique, S, Islam, S, Neon, EE, Sabbir, T, Naheen, IT, Khan, R. 2023. Deep Learning-based Bangla Sign Language Detection with an Edge Device, *Intelligent Systems with Applications*, 18, pp. 200224. (https://doi.org/10.1016/j.iswa.2023.200224 30 March 2023.)

[44]. Özcan, T, Baştürk, A. 2021. ERUSLR a new Turkish sign language dataset and its recognition using hyper parameter optimization aided convolution neural network, *Journal of Faculty Engineering Architecture of Gazi Universty,* 36:1, pp. 527-542. (htpss://doi.org/ 10.17341/gazimmfd.746793.)

[45]. Miah, ASM, Shin, J, Hasan, MAH, Rahim, MA. 2022. BenSignNet: Bengali sign language alphabet recognition using concatenated segmentation and convolutional neural network, *Applied Sciences,* 12(8), pp. 3933.
(https://doi.org/10.3390/ app12083933.)

[46]. Pacal, I, Alaftekin, M. 2023. CNN-Based approaches for automatic recognition of Turkish sign language, *Journal of the Instute of Science Technology,* 13(2), pp. 760-777. (https://doi.org/10.55525/tjst.1073116.)

[47]. Miah, ASM, Shin, J, Hasan, MAH, Rahim, MA. 2022. BenSignNet: Bengali sign language alphabet recognition using concatenated segmentation and convolutional neural network, *Applied Sciences,* 12(8), pp. 3933.