

# A Survey of Methods for Hand Gesture Recognition for Sign Language Methods: Research Gap, Trends, Challenges and Future Directions

Okorie Emmanuel O<sup>1\*</sup>, Nachamada V Blamah<sup>1</sup> and Gideon Dadik Bibu<sup>2</sup>

<sup>1</sup>Department of Computer Science, University of Jos, Nigeria

<sup>2</sup>Department of Computer Science, Higher Colleges of Technology, UAE

## \*Corresponding Author

Okorie Emmanuel O, Department of Computer Science, University of Jos, Nigeria.

Submitted: 2025, Aug 05; Accepted: 2025, Sep 30; Published: 2025, Oct 09

**Citation:** Okorie E. O., Blamah, N. V., Bibu, G. D. (2025). A Survey of Methods for Hand Gesture Recognition for Sign Language Methods: Research Gap, Trends, Challenges and Future Directions, *Curr Trends Mass Comm*, 4(3), 01-26.

## Abstract

Recent advancements in gesture and sign language recognition have been categorized into non-vision and vision-based techniques. The use of sensors, wearable gloves, microcontrollers, deep learning, computer vision and recently virtual and augmented reality have made this research area an interesting one. This paper presents a review of trends and techniques used in recent works to address the problem of gesture for sign language recognition. The objectives of this study are to critically review the state of the art non-vision and vision-based approaches of gesture and sign language recognition, observe the trends of recent works, identify challenges in model, design, algorithms and suggest possible and potential future research directions. 110 relevant papers spanning from the year 1998 to 2022 are surveyed. The findings could aid future research plans, while the suggested ideas could help researchers better design and build gesture and sign language recognition systems to support the communication of people with speech and hearing impairment.

**Keywords:** Computer Vision, Deep learning, Convolutional Neural Networks (CNNs), Human Computer Interactions, Gesture Recognitions, Sign Languages, Algorithms, Trends

## 1. Introduction

It is estimated that 430 million people or over 5% of the world's population suffer from hearing and speech impairments [1]. 2.5 billion people are projected to have some degree of hearing loss and at least 700 million will require hearing rehabilitation by the year 2050. Hearing impairment can be categorized into congenitally (people that become deaf from birth) and adventurously (people who though not born deaf but became deaf later in life due to accident or illness). V. Nwadinobi went further to explain causes of hearing impairments [2]. The hearing and speech impaired people irrespective of their disability must have a way to communicate with people and sign language and gestures recognition comes to the rescue [2]. However, gestures and signs are the common means of communication among hearing and speech impaired and normal persons. Although gestures and signs are categorized under non-verbal communication, they deliver the messages effectively to the other end of the communication especially within the hearing-

impaired community [3]. Speech and hearing-impaired people are now considered a large segment in many countries; hence, they need an effective and easy means of communication between them and normal people. This leads to the emergence of sign language recognition systems. Sign language is a visually oriented, natural, non-verbal communication medium, which is used by millions of hearing-impaired people around the globe as their first language. According to the British Deaf Association, there are over 151,000 people in Britain who use sign language. The two main components of sign language are finger-spelling (postures) and dynamic hand movements (gestures) [4].

Several published papers on sign language recognition and methods, dataset, and country can be found in the literature. M. Al-Hammadi et al categorized them into vision-based approach and a non-vision-based approach [5]. In 2021, surveyed the latest developments in gesture recognition technology for videos based

---

on deep learning, the reviewed methods were broadly categorized into three groups based on the type of neural networks used for recognition: stream convolutional neural networks, 3D convolutional neural networks, and Long-short Term Memory (LSTM) networks [6]. This paper critically reviews state of the art research on gesture/sign language, their performance accuracy, techniques, country sign language, datasets, framework base on vision and non vision based approach within the timeframe of early 1998 to 2022. In addition, it also presents trend of recent methods available in public domain for sign recognition. A trend analysis gives a run-through of identified gaps from the current work and suggests directions for bridging the identified gaps in future research which is one of the aims of this paper. Also, a taxonomy conveys the relationship between current works and classifies them based on identified strengths and weakness, and gives suggestions that can help researchers and others easily comprehend the topic [7]. This review paper aims to give new researchers in this domain a sense of the current research attainment. It also aims to provide, suggest more robust and reliable communication systems for more effective communication between the hearing/speech impaired with the normal people in the community.

This paper aims to critically review the trends of the gesture/sign language recognition methods, while the objectives are as follows:

- Showcase the research efforts addressing the communication barrier between the hearing/speech impaired people and the normal people using both vision and non-vision approaches.
- Identify the limitations of the existing methods and.
- Suggest possible future research directions that are oriented towards developing effective communication systems for hearing/speech impaired with the normal people.

### 1.1. The Significance and Contribution of this Study

The contributions of this study to the body of research knowledge include:

- A comprehensive and critical analytic review of state-of-art gesture/sign language recognition methods, their trends, challenges, and suggestions of possible future directions that can help researchers bridge the identified gaps towards the achievement of nearly perfect communication system for hearing/speech impaired people.
- Alleviate the social and economy burden on relations and friends of hearing/speech impaired people in the society. By exposing them to researches been made to help their loved ones and the possibilities of them becoming useful to themselves, family and society at large.
- Help the hearing/speech impaired contribute significantly to the society using their skills, intelligence and ability assuming there is no longer communication barrier.

This paper is structured as follows: Section II reviews the non-vision-based approaches for gesture/sign language recognition. Section III reviews the Vision based approaches for gesture/sign language recognition. Section IV details the evaluation metrics and datasets mostly used in the reviewed literature. Section VI focuses on the key knowledge gaps in the literature. Section VII

presents an analysis of the trend, challenges, and future direction of gesture/sign language recognition for effective communication. Section VIII concludes the review.

A total of 110 journal and conference papers were sourced and reviewed in this study

## 2. Non-Vision-Based Approaches For Gesture /Sign Language Recognition

In non-vision-based approach of sign language recognition, hand gesture data are collected via interfacing devices such as data gloves, motion sensors, and position trackers [5]. In this approach no camera is used, microcontrollers, kinect sensor and other sensors are used to get hand movements and translated appropriately and these were the earlier sign language recognition approaches before the adoption of deep learning and convolutional neural network (CNN) [5].

### 2.1. Hidden Markov Model (HMM)

Presented an experimental system for recognizing manual gestures of ASL [8]. The system consists of modules for hand detection, tracking, features extraction and HMM classifier [9]. Proposed Chinese sign language recognition based on trajectory modeling with hidden Markov models (HMMs). The authors normalized and re-sampled the raw trajectory data and partition the trajectory into multiple segments. A new curve feature descriptor based on shape context was proposed to represent each trajectory segment and the hidden Markov model was used to model each isolated sign word for recognition [10]. Proposed an entropy-based K-means algorithm to evaluate the number of states in the HMM model with an entropy diagram for the recognition of home-service-related Taiwan sign language words. Four real datasets are utilized to verify the developed entropy-based K-means algorithm, a data-driven method is given to combine the artificial bee colony algorithm with the Baum–Welch algorithm to determine the structure of HMM. The recognition system is established by 11 HMM models, and the cross-validation demonstrates an average recognition rate of 91.3%. In two real-time hidden Markov model-based systems for recognizing sentence-level continuous American Sign Language (ASL) using a single camera to track the user’s unadorned hands was proposed [11]. The first system observes the user from a desk mounted camera and achieves 92 percent word accuracy. The second system mounts the camera in a cap worn by the user and achieves 98 percent accuracy (97 percent with an unrestricted grammar). Both experiments use a 40-word lexicon [12]. introduced a discriminative hidden-state approach for the recognition of human gestures and demonstrated its utility both in detection and in a multi-way classification formulation. The evaluated method results showed that HCRFs (hidden state conditional random field) outperform both CRFs and HMMs (Hidden Markov model) for certain gesture recognition tasks. For arm gestures, the multi-class HCRF model outperforms HMMs and CRFs (Conditional random fields) even when long range dependencies are not used, demonstrating the advantages of joint discriminative learning [13]. Proposed a framework for recognizing valid sign segments

---

and identifying movement epenthesis, it utilizes a single HMM threshold model, per hand, to detect movement epenthesis. A technique to utilize the threshold model and dedicated gesture Hidden Markov Models (HMMs) to recognize gestures within continuous sign language sentences was also proposed. The system achieved a gesture detection ratio of 0.956 and a reliability measure of 0.932 when spotting 8 different signs from 240 video clips.

## 2.2. Support Vector Machine (SVM)

proposed a support vector machine-based recognition framework which uses a combination of eigenspace size function and Hu moments features to classify different hand postures [14]. The results of a user independent evaluation of the recognition framework showed the system had a ROC (receiver operating characteristic curve) AUC (Area under the Curve) of 0.973 and 0.935 when tested on the ISL (Indian Sign Language) data set and the Treisch data set respectively [15]. proposed a system based on the combination of Spatio-Temporal local binary pattern (STLBP) feature extraction technique and support vector machine classifier. The system takes a sequence of sign images or a video stream as input, and localize head and hands using IHLS color space and random forest classifier. A feature vector is extracted from the segmented images using local binary pattern on three orthogonal planes (LBP-TOP) algorithm which jointly extracts the appearance and motion features of gestures. The obtained feature vector is classified using support vector machine classifier. The method achieved a 99.5% accuracy.

## 2.3. Mobile Apps

A chat application (Chat Assist) was proposed by [16]. The developed system was based on Sinhala Sign language and consists of four main components: text messages are converted to sign messages, voice messages are converted to sign messages, sign messages are converted to text messages and sign messages are converted to voice messages. Google voice recognition API was used to develop speech character recognition for voice messages [17]. presented an android-based mobile application for learning sign language called LEARN SIGN. With the incorporation of AR (Augmented Reality) and speech detection technologies, this system can also help people with disabilities, especially the deaf and mute community, to communicate with non-disabled people and vice-versa

## 2.4. Other Algorithms Techniques

Conducted an American Sign Language (ASL) recognition experiment on Kinect sign data using Dynamic Time Warping (DTW) for sign trajectory similarity and Histogram of Oriented Gradient (HOG) for hand shape representation [18]. The experiment achieving an 82% accuracy in ranking signs in the 10 matches. In addition to improved sign recognition accuracy, the authors proposed a simple RGB-D alignment tool that can help roughly approximate alignment parameters between the color (RGB) and depth frames [19]. proposed a framework for automatically learning a large number of signs from sign language-interpreted TV broadcasts, the method exploits co-occurrences of mouth and hand motion to substantially improve the 'signal-to-noise' ratio in the

correspondence search. The authors also developed a multiple instance learning method using an efficient discriminative search, which determines a candidate list for the sign with both high recall and precision. In three automatic recognitions of BSL (British Sign Language) signs from continuous signing video sequences were presented [20]. first: automatic detection and tracking of the hands using a generative model of the image, second: automatic learning of signs from TV broadcasts of single signers, using only the supervisory information available from subtitles and lastly: discriminative signer-independent sign recognition using automatically extracted training data from a single signer [21]. presented an algorithm for extracting and classifying two-dimensional motion in an image sequence based on motion trajectories on 40 recognized American sign languages. A multiscale segmentation is performed to generate homogeneous regions in each frame. Regions between consecutive frames are then matched to obtain two-view correspondences. Affine transformations are computed from each pair of corresponding regions to define pixel matches. Pixels matches over consecutive image pairs are concatenated to obtain pixel-level motion trajectories across the image sequence. Motion patterns are learned from the extracted trajectories using a time-delay neural network, using an ensemble of trajectories helps achieve high recognition rates [22]. proposed an easy-to-use and inexpensive approach to recognize single handed as well as double handed gestures accurately. A real time hand gesture recognition of Indian sign language system is proposed by means of Camshift method, HSV color model and Genetic algorithm and achieved high recognition accuracy. A statistical recognition approach performing large vocabulary continuous sign language recognition across different signers is presented in [23]. It focused on tracking, features, signer dependency, visual modelling and language modelling and enumerated the impact of multimodal sign language features describing hand shape, hand position and movement, inter-hand-relation and detailed facial parameters, as well as temporal derivatives. In terms of visual modelling the paper evaluated non-gesture-models, length modelling and universal transition models. Signer-dependency is tackled with CMLLR adaptation and was further improved by employing class language models, the achieved result on two datasets can be found in [23]. Proposed a serial particle filter with feature co-variance matrix representation for isolated sign language recognition [24]. The background model is constructed via the fusion of median and mode filters on the entire video sequence to better detect the hands, based on the background model, the foreground is extracted and passed on to the proposed serial particle filter to enable the tracking of the hand trajectories and the hand regions based on the trajectories are extracted and the feature covariance matrix is computed. The sign gesture recognition based on the proposed methods yields an 87.33% recognition rate for the American Sign Language [25]. Presented an approach to detecting and recognizing gestures in a stream of multi-modal data. This approach combines a sliding window gesture detector with features drawn from skeleton data, color imagery, and depth data produced by a first-generation Kinect sensor and achieved a Jaccard Index score of 0.834 on the ChaLearn-2014 Gesture Recognition Test dataset. SIFT (scale invariance Fourier transform) algorithm was proposed in to rec-

---

ognize Indian Sign Language, the images are of the palm side of right and left hand and are loaded at runtime [26]. The method was developed with respect to a single user. The real time images are captured first and then stored in a directory and recently captured image and feature extraction will take place to identify which sign has been articulated. 95% accuracy was achieved by the proposed algorithm for 9 alphabets images captured at every possible angle and distance. Presented a gesture recognition setup capable of recognizing and emphasizing the most ambiguous static single-handed gestures [27]. Performance of the proposed scheme was tested on the alphabets of American Sign Language (ASL). Segmentation of hand contours from image background is carried out using two different strategies; skin color as detection cue with RGB and YCbCr color spaces, and thresholding of gray level intensities. A novel, rotation and size invariant, contour tracing descriptor is used to describe gesture contours generated by each segmentation technique. Performances of K-Nearest Neighbor (k-NN) and multiclass Support Vector Machine (SVM) classification techniques are evaluated to classify a particular gesture. Gray level segmented contour traces classified by multiclass SVM achieved accuracy up to 80.8% on the most ambiguous gestures of ASL alphabets with overall accuracy of 90.1%. Linear Discriminant Analysis (LDA) algorithm was used for gesture recognition and recognized gesture converted into text and voice format by [28]. The proposed system contains four modules such as: pre-processing and hand segmentation, feature extraction, sign recognition and sign to text, 26 hand gestures of Indian sign language was used for the experiment in MATLAB [29]. surveyed various techniques, methods and algorithms related to the gesture recognition such as Anticipated Static Gesture Set, Hand Segmentation Using HSV Color Space and Sampled Storage Approach, Hand Tracking and Segmentation (HTS) Algorithm that provides segmentation of given input sent for recognition without any noise (Segmentation algorithms). The Hand gesture recognition models such as Hidden Markov Model, YUV color space model, 3D model and Appearance model that will detect input and process them for recognition.

## 2.5. Embeded Systems and Gloves

Conversion of sign language into human hearing voice was proposed in [30]. The authors used non-vision-based technique for the conversion of voice into sign language. A technique called artificial speaking mouth for dumb people was proposed by, the system is based on the MEMS SENSOR. All sign messages are kept in a database and every template used are derived from the database [31]. For every action(gestures) the MEMS SENSOR get accelerated and give the signal to the MC (Microcontroller). The MC matches the motion with the database and produces a speech signal that is played via the speaker. The system also includes a text to speech conversion (TTS) block that interprets the matched gestures. Presented a glove-based sign-to-text/voice translating system for deaf and dumb people [32]. The glove represents the Arabic sign language letters as a text on LCD and outputs an audio through a speaker. The system was designed using Arduino board, programmed, implemented and tested with a very good result. While used PIC Microcontroller to design a wireless glove to aid communication between hearing and speech impaired by

converting signs/gestures into audio, Click or tap here to enter text. used the same Microcontroller to design what they called artificial mouth which is based on the motion sensor [33-36]. The microcontroller matches the motion with previous gestures stored in the database and produces the speech signal played via a speaker [36].

Used flex sensor technology to improve communication between hearing and speech impaired. It was developed to translate different signs including Indian sign language to text as well as voice format. Flex sensors are placed on hand gloves which pick up gestures and transmit that to text data with the help of analog to digital convertor and microcontrollers. This converted text data is then sent wirelessly via Bluetooth to a cell phone which runs text-to-speech software and the incoming message is converted to voice [37].

Also used a Flex Sensor and Raspberry Pi to design a hand glove to accurately translate ASL to text and speech while adopting a non-vision based communication system [38]. Proposed an embedded system consisting of wearable sensing gloves along with flex sensors which were used to sense the motion of the fingers (ISL). Flex sensors and accelerometer were used as sensor, these sensors were mounted on the gloves, the movement include the angle tilt, rotation and direction changes, the signals are processed by a microcontroller (AVR) and playback voices are generated indicating signs through speaker. A smart glove and flex sensor for sign language was proposed in [39]. The proposed approach is based on detection of the finger movements and hand gestures to identify gesture using signal processing kit in LabVIEW software and a data acquisition device (NI USB 6008 DAQ card). The processed signal is used to identify signs shown and concatenate the letters into suitable words and also present the word in audio format. The code implementation is done in LabVIEW platform for twenty-six letters and concatenation of letters up to 6 letters according to American Sign Language. Three solutions were modulated to be in a single unique system by using Raspberry Pi [40]. For visually impaired people the system processes image to text and text-to-speech is given by the Tesseract OCR (online character recognition), for hearing impaired people an app is used to make them understand what a person says by displaying the message on the screen and lastly, vocally impaired people can convey their message by text so the other persons can hear the message in a speaker [41]. Presented work on understanding Vietnamese Sign Language (VSL) through the use of MEMS accelerometers. The system consists of six ADXL202 accelerometers for sensing the hand posture, a BASIC Stamp microcontroller, and a PC for data acquisition and recognition of sign language. The classification process is done by a fuzzy rule-based system on the preprocessed data [42].

Discussed the key responsibility of the sign language translator for the betterment of interaction between normal people and deaf and dumb community through Human to Computer Interaction (HCI). Hand Segmentation Using Lab Color Space (HSL) by extracting the 'a' component of the LAB image was used to evaluate the

---

interactivity of users and the feature extraction was done with the help of Generalized Hough Transform technique. System was able to recognize 31 Tamil Language Alphabets [43]. proposed the use of RF sensors for HCI applications serving the Deaf community. Multi-frequency RF sensor network is used to acquire non-invasive, non-contact measurements of ASL signing irrespective of lighting conditions. ASL data are investigated using machine learning (ML) with the Short-Time Fourier Transform. Using the minimum redundancy maximum relevance(mRMR) algorithm, an optimal subset of 150 features is selected and input to a random forest classifier to achieve 95% recognition accuracy for 5 signs and 72% accuracy for 20 signs.

### 3. Vision Based Approach For Gesture/Sign Language Recognition

Vision-based approach overcomes the downsides of non-vision-based approach of gesture/sign recognition by collecting the data via cameras and imaging sensors. However, research works using this approach have encountered many challenges that degrade the performance of existing systems such as lighting inconsistency, motion blur, background clutter, and hands occlusion [5].

#### 3.1. Conventional Techniques

To overcome the problem of real time gesture communication among the hearing and speech impaired people with normal people [44] developed a friendly, cost-effective system. The proposed system captures a hand gesture using the high-definition Pi camera. Image processing of captured gesture is done on Raspberry Pi 2. Amplified audio corresponding to each processed gesture is the final output. In 2019, went further design a single device solution that is simple, fast, accurate and cost effective [45]. Furthermore, image to text conversion and speech synthesis is done, converting it into an audio format that reads the extracted text translating documents, books and other available materials in daily life. For the audibly challenged, the input is in form of speech taken in by the microphone and recorded audio is then converted into text which is displayed in the form of pop-up windows for the user on the screen of the device. The vocally impaired are aided by taking the input by the user as text through the built in customized on-screen keyboard where text is identified, text into speech conversion is done and the speaker gives the speech output. A Real-time vision-based system was proposed in to assist hearing and speech impaired people [46]. It is built based on the Raspberry Pi with camera module and programmed with Python programming Language supported by Open-Source Computer Vision (OpenCV) library. It also contains a 5 inch Resistive HDMI Touch screen for input/output data. The Raspberry Pi embeds with an image-processing algorithm called hand gesture, which monitors an object (hand fingers) with its extracted features.

In hand gesture/sign language was converted into voice (audio) [47]. Image processing was used for hand gesture recognition in the system. Using camera to get images of hand, and then pre-process those images by color splitting, morphological processing and feature extraction. Lastly, template matching was used to realize the hand gesture recognition. The recognized image is processed

by the hardware (system) and converted to voice. An innovative communication system framework for deaf, dumb and blind people in a single compact device was proposed in [48]. The technique helps a blind person to read a text and was achieved by capturing an image through a camera which converts a text to speech it also provides a way for the deaf people to read a text by speech to text (STT) conversion technology. It provides a technique for hearing impaired people using text to voice conversion. The system is provided with four switches and each switch has a different function. A blind person is able to read words using Tesseract OCR (Online Character Recognition), the hearing impaired can communicate their message through text which is read out by espeak, the hearing impaired is able to hear others speech from text. All these functions were implemented with Raspberry Pi

#### 3.2. Deep Learning-Based Techniques (CNNs)

Created and implemented a real time sign language detector with improved communication between the deaf and the general populace using American Sign Language (ASL) [49]. Based on a Convolutional Neural Network (CNN) and utilized a pre-trained SSD Mobile net V2 architecture, which was trained on the researcher generated dataset. The data is made up of collection of over 2000 images, around 400 for each of its classes, it contains a total of 5 symbols (Hello, Yes, No, I Love You and Thank You). The system is able to recognize selected Sign language with accuracy of 70-80% without a controlled background with small light. A web camera was used for capturing images of hand gestures with the help of OpenCV, other tools used include TensorFlow, Tensor Object detection API and LabelImg. Its sign recognition accuracy for Yes (88.7%), No (88.6%), Thank You (84,1%), Hello (91.0%) and I Love You (82.4%). The model however, has some limitations such as environmental factors like low light intensity and uncontrolled background which caused decrease in the accuracy of the detection. Another Realtime communication for speech and hearing-impaired people was proposed in [50]. It is a real-time communication system built using advancement in image processing, deep learning (CNN) and computer vision that provides real-time sign language to text and text to sign language conversion. It is also a two-way communication system allowing communication with hearing impaired and normal people, the system is able to interpret alphabets, numbers and words in the Indian sign language and predicted 17600 test images in 4 seconds with an average predictions time of 0.000805 with an accuracy of 99%. In order to facilitate communication among speech and hearing-impaired person proposed the extraction of hand and body from video sequences [51]. LSA64 dataset, a large Argentinian sign language dataset consisting of 10 subjects with a total of 64 different commonly used signs were used and was split randomly in a training set consisting of 80% of the sample and a test set consisting of the remaining 20% of the sample data. The research was implemented in Keras-Tensorflow framework and trained using the Adam optimizer with batch of 32 and learning rate equal to 0.0001. The model was pretrained on ImageNet VGG-19 network of up to conv4\_4 as feature extractor for hand skeleton detection, and the first 10 layers of the same network was employed for body skeleton detections. It employed linear dynamic system (LDS) histograms and

four stream deep neural networks that consist of stacked LSTM layers [52]. The experimentation on LSA64 dataset showed that SLR (sign language recognition system) out-performs other vision based SLR approached reviewed by the authors despite difficulties in extracting accurate skeletal data due to occlusions [53]. Proposed a method for recognition of hand gestures in sign language vocabulary. This was based on an efficient deep convolutional neural network architecture. The method was tested on two publicly available datasets: NUS (National University of Singapore) hand posture dataset and American fingerspelling a datasets. The model avoids the tedious and computationally complex feature extraction

phase of the traditional recognition approach by using CNN to recognize the static hand gestures. The convolutional layers contain unit called feature maps and each of them was connected to the local patches in the previous layer through filter banks. NUS dataset was divided into five subsets including 40 sample images of each gesture class. The classifier was trained with any four subsets and the remaining subset used for testing. The experiment was repeated five times in a similar manner until each of the subset was used for development and testing. The classification result evaluated using the average accuracy, precision, recall and F1-score values is given in Table 1

Accuracy	Precision	Recall	F1-Score
94.7±0.80 %	94.96±1.20 %	94.85±1.30 %	94.26±1.70 %

**Table 1: Interpretation of the Classification Performance of the Proposed CNN Model on NUS Hand Posture Dataset Using Statistical Measures**

Second dataset is the American fingerspelling A dataset which contains 24 letters of the ASL alphabet excluding the letters 'j' and 'z' (since they involve motion). The images were captured in five different sessions, with different users in similar lighting conditions

in the presence of complex background objects. The classification result evaluated using the average accuracy, precision, recall and F1-score values is given in Table 2

Accuracy	Precision	Recall	F1-Score
99.96±0.04 %	99.96±0.04 %	99.96±0.04 %	99.96±0.04 %

**Table 2: Interpretation of the Classification Performance of the Proposed CNN Model on ASL Fingerspelling Dataset Using Statistical Measures**

Proposed a novel approach for video-based continuous sign language recognition (CLSR), the method leverages text information to model intra-gloss dependencies and create more descriptive video-based latent representations that improves recognition accuracy [54]. It consists of a CNN for spatial feature extractions, stacked 1D temporal convolutional layers (TCL) for short-term temporal modelling and a bidirectional long short-term memory (BLSTM) unit for global context learning. A new approach for the alignment of video and text embeddings using a join function was also proposed by the authors. This approach was evaluated on three challenging sign language recognition datasets namely RWTH-Phoenix-Weather-2014, RWTHPhoenix-Weather-2014T, and CS and when compared with several state-of-the-art approaches the experimental results on the three most widely used CSLR datasets demonstrated the ability of the proposed method to provide highly accurate CSLR results.

this improved the accuracy of hand location [55]. The paper studied hand locating and sign language recognition of common sign language based on neural network and the main research contents include: 1. A hand locating network based on the Faster R-CNN to recognize the sign language video or the part of the hand in a picture and result handed to subsequent processing; 2. A 3D CNN feature extractions network and a sign language recognition framework based on the long and short time memory (LSTM) coding and decoding network was constructed for sign language images of sequence sequences; 3. The paper combined hand locating network, 3D CNN feature extraction network and LSTM encoding and decoding to build the recognition algorithm to solve the problem of RGB sign language image or video recognition in practical problems. Training was done using a data set of 40 common words and 10,000 sign language images with Stochastic Batch Gradient Descent (SGD) optimizer. It also compared the accuracy of hand locating of Faster R-CNN, Fast R-CNN and YOLO and Faster R-CNN performed better as shown in the Table 3.

The Faster R-CNN model of Convolutional Neural Network was proposed in the model used CNN feature for target recognition as

Methods	mAP(%)	Right hand	Left Hand	Both Hand
YOLO	83.2	80.5	81.7	87.3
Fast R-CNN	89.0	86.1	88.4	92.5
Faster R-CNN	91.7	89.2	89.8	96.2

**Table 3: Detection Results of Each Algorithm**

An attention-based 3D-Convolutional Neural Networks (3D-CNNs) for SLR was also proposed in [56]. The framework has two advantages: 3D convolutional networks learn spatio-temporal features from raw video without prior knowledge, and attention mechanism helps to select the clue. During training of 3D-CNN for capturing spatio-temporal features, spatial attention was incorporated in network to focus on the areas of interest after feature extraction, temporal attention was utilized to select the significant motions for classification. The authors evaluated the proposed methods on two large scale sign language datasets. The first one being a Chinese Sign Language (CSL) dataset that consists of 500 categories and the other the ChaLearn14 benchmark [57]. In a sign language recognition system using deep learning was presented. A 3D-CNN was used for recognizing process and the images received from Kinect sensor with a recognition accuracy of 91.23%.

To create a vision-based application which offers sign language translation to text proposed a model that takes video sequences and extracts temporal and spatial features from them [58]. This they

achieved with Inception, a CNN for recognizing spatial features and RNN (Recurrent Neural Network) to train temporal features. The dataset used however, is custom own generated American sign language dataset. After completion of the training steps, the model reported 99% accuracy on the training set even though the paper proposed that Capsule Networks will yield better results as against Inception it used [59]. Proposed a Capsule network for training and testing processes, it was compared with LeNet which is one of the first successful and currently used model of deep learning. The recognition of sign language characters system using images of letters from America sign language was the aim of the research. The result of the study showed that capsule networks (88% accuracy on the test set) are useful for sign language character recognition and produced more successful and better result than LeNet (82% accuracy on the test set) on MNIST sign language dataset obtained from Kaggle. The success of the capsule was increased to 95% by augmenting the data over the training data. The dataset consists of 24 classes of English alphabet beside Letters J and Z because it requires movement representation. Table 4 shows Metrics of the models.

Metrics of models	Metrics			
	Accuracy	Precisions	Recall	F-score
LeNet	82.19%	81.24%	81.82%	80.95%
CapsNet	88.93%	84.48%	89.04%	86.41%
CapsNet augmented	95.08%	91.11%	95.63%	93.22%

**Table 4 : Metrics of the Models**

Another capsule-based deep neural network sign posture translator for an American Sign Language (ASL) fingerspelling (posture) was proposed in [60]. The performance validation showed that the approach can successfully identify sign language, with accuracy of 99%, the developed capsule network architecture does not require a pre-trained model. The framework uses a capsule network with adaptive pooling which is the key to its high accuracy. The framework is not limited to sign language understanding, but it has scope for non-verbal communication in Human-Robot Interaction (HRI) also [61]. went further to demonstrate a user-friendly approach towards Bangla sign language by converting its sign language into text through customized region of interest (ROI) segmentation and CNN. 5 Sign gestures were trained using custom image dataset and was implemented in a Raspberry board for portability. The researched showed that using ROI selection approach showed better outcome than conventional approached in terms of accuracy and real time detection from video streaming through webcam.

A signer independent deep learning-based method for building an Indian sign language (ISL) static alphabet recognition system was proposed by [62]. The research implemented a CNN architecture for ISL static alphabet recognition from the binary silhouette of signer hand region, a custom dataset consisting of 24 ISL static alphabet was used for training with 98.64% accuracy. Training accuracy of 99.93% and validation accuracy of 98.64% [63]. Presented an application to translate alphabets of Indian Sign language in

real time. Custom dataset was generated for model training. The application works in real time and in varying backgrounds and user can type alphabets by doing corresponding gestures in front of a webcam. The research used resnet18 model for training by unfreezing last few layers and using differential learning rates after making it clear that different versions of Resnet can be used but resnet18 works best as it is the lightest network with better efficiency when working in real time application. Another real-time system which can convert Indian Sign Language (ISL) to the text was proposed in and was based on handcrafted feature [64]. The author introduced Deep learning approach to classify signs using the convolutional neural network, this was done in two phases. Making a classifier model using the numeral signs using the Keras implementation of convolutional neural network using python was the first phase. In phase two another real-time system which used skin segmentation to find the Region of Interest in the frame which shows the bounding box. The segmented region is feed to the classifier model to predict the sign. The system attained an accuracy of 99.56% for the same subject and 97.26% in the low light condition [65]. proposed the recognition of Indian sign language gestures using convolutional neural networks (CNN). Selfie mode continuous sign language video was the capture method used in the work, where a hearing-impaired person can operate the SLR mobile application independently. Due to non-availability of datasets on mobile selfie sign language, the authors created custom own dataset with five different subjects performing 200 signs in 5 different viewing angles under various background

environments. Each sign occupied for 60 frames or images in a video. CNN training is performed with 3 different sample sizes, each consisting of multiple sets of subjects and viewing angles. The remaining 2 samples were used for testing the trained CNN. Different CNN architectures were designed and tested with the selfie sign language data to obtain better accuracy in recognition. 92.88% recognition rate was achieved when compared to another classifier models reported on the same dataset In 2020 tried to do a real time translation of hand gestures into equivalent English text [66]. This system takes hand gestures as input through video and translates it to text which could be understood by a non-signer by implementing the use of CNN for hand gesture classification. For hand detection the authors used YOLO and VGGNet for gesture classification of Indian Sign language.

combined hand-crafted features and deep learning methods to classify signs, it applied the skin color based YCbCr segmentation method and local binary pattern for accurate shape segmentation and for texture features or local shape information [67]. The transfer learning framework (VGG-19) was fine tuned to obtain the features that are then fused with hand crafted features by serial based fusion technique, these features were finally given to support vector machine (SVM) classifier to classify the signs. ASL Finger Spelling benchmark dataset which consist of both color and depth images which were obtained from 5 different users consisting of 24 static signs excluding letters j and z. 98.44% accuracy was obtained by the proposed system by generating 0.0568 loss [68]. Proposed a Deep Learning based sign language recognition frame-work. The method is built to recognize static hand signs of 37 Bengali signs with total of 1147 images. A 96.33% recognition rate on training dataset and 84.68% on the validation dataset was achieved using Deep convolutional neural networks (DCNN) while utilizing the features from a pretrained (VGG16) network [69]. Proposed another Bengali sign gesture using convolutional neural networks. A large publicly available Bengali sign language dataset was used which consists of 24168 samples (basic characters: 18745 and numerals: 5423), CNN was used to recognize and classify hand image on the screen and then categorized the hand skeletal features extracted from the image into a standard communicative meaning. 98.75% accuracy was reached on the proposed model. On another research on Bengali Sign language aim to construct a model to recognize Bengali Character Language using deep learning, CNN was used to train individual signs [70]. For individual signs, a dataset was constructed named Bengali Ishara-Lipi, the model was trained using 5760 preprocessed images and tested by 1440 images. according to the authors the model gained 92.7% accuracy

to recognize Bengali alphabetical sign language [71]. proposed a novel Convolutional Neural Network (CNN) model for the recognition of the Bengali sign alphabets from the Ishara-Lipi dataset and the model achieved an overall accuracy of 99.22%. To address the problem of Sensor-based methods for sign recognition

Proposed a custom DNN model for recognition of English language alphabets using convolutional Neural Network [72]. The proposed DNN extracts features automatically from input gestures and classify them. The dataset used consists of images of hand gestures for English Sign Language (ESL) obtained from Kaggle website, the data is made of color images of sign gestures representing English alphabets and additional symbols such as space, delete and nothing. The proposed system used a three layers deep CNN for hand gesture recognition system and achieved a peak accuracy of 100% for training process and 82% for validation process while test accuracy was 70%. These authors developed a sign language fingerspelling alphabet identification system by using image processing technique, supervised machine learning and deep learning [73]. 24 alphabetical symbols were presented by several combinations of static gestures excluding letter J and Z gestures. Histogram of Oriented Gradients (HOG) and Local Binary Pattern (LBP) features of each gesture were extracted from training images and Multi-class Support Vector Machines (SVMs) applied to train the extracted data and finally an end-to-end CNN architecture applied for training the dataset. The authors concluded that result from CNN and CNN-SVM models (97:08% and 98:30%) prove that by implementing CNN as a standalone feature extractor, better result could be obtained than using an end-to-end CNN architecture. In addition, due to similarity between Sign Language Recognition and Action Recognition (Sign Language Recognition Using Modified Convolutional Neural Network Model) implemented an i3d inception model to sign language recognition with transfer learning method. 100% accuracy was achieved on a training set of 10 words and 10 signers with 100 classes. However, the validation accuracy was low and the model was too overfit. A modified LSTM model for continuous sign language recognition proposed by models continuous sequences of gestures using a dataset of 35 isolated signs words [74]. This was based on splitting continuous signs into sub-units and modeling them with neural networks, the proposed system was tested with 942 sign sentences of ISL and the average accuracy of 72.3% and 89.5% was recorded on signed sentences and isolated signs respectively. The performance of the system was also compared with traditional LSTM and the result is shown in the table 5 below.

Model	Sign word Recognition	Sign sentence Recognition
Traditional LSTM	68.60%	53.20%
Proposed	89.50%	72.30%

**Table 5: Comparative Performance Analysis Between the Proposed and Traditional LSTM Model**

Classified RGB images of static letter hand posed in Sign language using CNN with Densely connected Convolutional Neural Net-

works (DenseNet) [75]. The proposed network achieved 90.3% accuracy after training on own custom dataset of ASL with a pre-

---

diction rate of 50 to 100Hz. For fingerspelling translator based on skin segmentation and machine learning algorithms for ASL [76]. Proposed YCbCr space for video coding and chrominance information for modeling the human skin color. Skin-color distributions was model as a bivariate normal distribution in CBCr Plan, CNN was used to extract features from the Images while AlexNet (pretrained neural network) Deep learning methods used to train a classifier to recognize Sign language. The tested methods achieved a test accuracy of 94% on custom datasets of ASL. A system based on skin-color modelling technique was also proposed by the skin-color range extracts pixels (hand) from non-pixel (background) [77]. The images were fed into the CNN for classification while Keras was used for image training. The author achieved 99% training accuracy with testing accuracy of 90.4% in letter recognition, 93.44% in number recognition and 97.52% in static word recognition, obtaining an average of 93.667% based on the gesture recognition with limited time. Each system was trained using 2,400,  $50 \times 50$  images of each letter/number/word gesture of ASL. Proposed a system to recognize Russian letters presented as static signs in Russian Sign Language [78]. Own custom dataset was used (RSL dactyl), it also adopted LetNet-like and QuadroCovPoolNet models. utilized transfer learning and fine turning deep CNN to improve the accuracy of 32 hand gestures from Arabic sign language [79]. The proposed method worked by creating models matching the VGG16 and ResNet152, the pretrained model weights were loaded into each network layers. 2D images of Arabic sign language were feed the network and 99% accuracy achieved by the authors. A framework for converting sign language to emotional speech by deep learning was proposed by deep neural network (DNN) model was adopted to extract the features of sign language and facial expression [80]. Two support vector machines (SVM) were trained to classify the sign language and facial expression for recognizing the text of sign language and emotional tags of facial expression. The author also trained a set of DNN-based emotional speech acoustic models by speaker adaptive training with multi-speaker emotional speech corpus and DNN-based emotional speech acoustic models, tags were finally selected to synthesize emotional speech from the text recognized from the sign language. The objective test of the framework showed that the recognition rate for static sign language was 90.7% while the recognition rate of facial expression achieved 94.6% on extended Cohn-Kanade database (CK+) and 80.3% on the Japanese Female Facial Expression (JAFFE) databases respectively. An intelligent recognition of static, manual and non-manual Hausa sign language (HSL) using a Particle Swarm Optimization (PSO) to enhanced Fourier descriptor was proposed in [81]. A vision-based approach was used, a Red Green Blue (RGB) digital camera was used for image acquisition and Fourier descriptors used for features extractions. The features extracted were enhanced by PSO and fed into artificial neural network (ANN) for classification. The authors achieved a high average recognition accuracy of 93.9%.

A simple deep neural network architecture called model E was proposed into recognize ASL hand gesture. The dataset used were collected from Kaggle.com ASL datasets [82]. The authors attempted comparing the accuracy between model E and AlexNet

by adjusting the kernel size and the number of epochs for each model and model E out performed AlexNet with 96.82% accuracy [83]. Proposed a method to create an ISL dataset using a webcam and SSD MobileNet v2 320x320 pre-trained model with TensorFlow. The developed system showed an average confidence rate of 85.45%. Though the system achieved a high average confidence rate, the dataset used for training is small in size and limited.

Studied Bangladeshi Sign language (BSL) recognition based on fingertip position [84]. It considered relative tip positions of five fingers in two-dimension space and position vectors were used to train artificial neural network (ANN) for sign recognition. The proposed method was tested on a prepared data set of 518 images of 37 signs and achieved 99% recognition rate. A dynamic hand gesture recognition using multiple deep learning architectures for hand segmentation, local and global feature representations, and sequence feature globalization and recognition proposed by was evaluated on a very challenging dataset (isolated words and phrases from common expressions in Saudi Sign Language (SSL)) that consists of 40 dynamic hand gestures performed by 40 subjects in an uncontrolled environment [5]. Two 3DCNN instances were used separately for learning the fine-grained features of the hand shape and the coarse-grained features of the global body configuration [85]. Proposed another efficient deep convolutional neural network (3DCNN) approach for hand gesture recognition. It employed transfer learning to beat the scarcity of a large labeled hand gesture dataset. The authors evaluated three gesture datasets from color videos: 40,23 and 10 classes from the dataset. The proposed approach obtained recognition rates of 98.12%, 100% and 76.67% on the three datasets respectively for the signer dependent mode while 84.38%, 34.9% and 70% recognition rate were Obtained on the three datasets respectively for the signer-independent mode using 3DCNN for hand gesture recognition.

Proposed a system for alphabetic Arabic sign language recognition using depth and intensity images acquired from SOFTKINECT sensor(camera), the method does not require any extra gloves or any visual marks. Local features from depth and intensity images are learned using unsupervised deep learning method called PCANet [86]. The extracted features are then recognized using linear support vector machine classifier. The proposed method performance was evaluated on dataset of real images captured from multi-users, the authors also performed separate experiment using combination of depth and intensity images and using depth and intensity images separately. The result obtained showed that the performance of the system improved by combining both depth and intensity information which produced an average accuracy of 99.5%. With a large synchronous dataset of 18 BSL (British Sign Language) gestures collected from multiple subjects benchmarked and compared two deep neural networks [87]. The vision model was implemented with CNN and optimized with Artificial Neural Network topology, the two best networks fused for synchronized processing and achieved overall results of 94.44%. The hypothesis was further supported by application of the three models to a set of completely unseen data where a multimodality approach achieved the best results relative to the single sensor method.

---

When transfer learning with the weights trained via British Sign Language, all three models outperform standard random weight distribution when classifying American Sign Language (ASL), and the best model overall for ASL classification was the transfer learning multimodality approach, which scored 82.55% accuracy. Presented a weakly supervised framework with deep neural networks for vision based continuous sign language recognition [88]. The approach addressed the mapping of video segments to glosses by introducing recurrent CNN for spatio-temporal features extraction and sequence learning [89]. Proposed a continuous sign language (SL) recognition framework with deep neural networks, which directly transcribes videos of SL sentences to sequences of ordered gloss labels. The architecture adopts deep convolutional neural networks with stacked temporal fusion layers as the feature extraction module, and bi-directional recurrent neural networks as the sequence learning module using limited dataset. The research also contributed to multimodal fusion of RGB images and optical flow in sign language and the evaluation performed on two challenging SL recognition benchmarks outperforms the state of the art with 15%.

Convolutional neural networks have been employed in to recognize sign language gestures [90]. The image dataset used consists of static sign language gestures captured on an RGB camera, Pre-processing was performed on the images, which then served as the cleaned input. Results obtained by retraining and testing sign language gestures dataset on a convolutional neural network model using Inception v3 was above 90% for validation accuracy.

Seq2Seq (sequence to sequence neural network model) learning model for SL communication interfaces was introduced and evaluated by for recognition and generation of signed sentences [91]. An encoding of the SL annotations and conducted experiments on the network structure to define a most accurate translation model was implemented and the study proved the network trainable and possibly applicable in real-life with an extended dataset, which can be tested for deployment in virtual translation assistants.

Proposed a robust deep learning-based method for sign language recognition, the approach represents multimodal information (RGB-D) through texture maps to describe the hand location and movement with an intuitive method to extract a representative frame that describes the hand shape [92]. This information served as inputs to two three-stream and two-stream CNN models to learn robust features capable of recognizing a dynamic sign. The authors conducted the experiment on two sign language datasets (Brazilian Sign Language) and went further to compare the results with state-of-the-art SLR methods and their results proved superior due to the combination of texture maps and hand shape for SLR tasks.

In 2019 proposed and implemented a novel yet deep convolutional neural network to classify and recognize Ghanaian Sign Language and attained an accuracy of 96.0% [93]. The authors leveraged (VGG-16 and VGG-19) transfer learning techniques by fine-tuning state-of-the-art network architectures pre-trained on the ImageNet database and improved the accuracy with a reported increase of 3.1%. The dataset used for evaluation of the

proposed CNN was created by the authors due to non availability of public Ghanaian Sign language dataset [94]. In 2019 proposed a real time sign language interpretation of hand gestures based on deep convolutional neural networks with focus on development of a cost-effective and efficient hardware prototype for communication ease with deaf and dumb people. The proposed sign language interpreter system is based on a deep CNN and uses open-source framework like Keras and TensorFlow. The dataset was prepared by collecting 4300 images for each of the 29 classes. 8 different backgrounds were incorporated in each of those images.

proposed a novel method for SLR using Real-Sense in which camera device was used to detect and track the location of hands in a natural way [95]. The authors built a deep neural network (DNN) based on Real-Sense to recognize different signs. The DNN takes the 3D coordinates of finger joints as input directly without using any handcrafted features. To demonstrate the effectiveness of RealSense, they collected two datasets Real-Sense and Kinect respectively, then built DNNs based on each dataset for recognition. presented a real time system for hand gesture recognition on the basis of detection of some meaningful shape-based features like orientation, center of mass (centroid), status of fingers, thumb in terms of raised or folded fingers of hand and their respective location in image [96]. This approach depended on shape parameters of the hand gesture and does not consider any other means of hand gesture recognition like skin color, texture because image-based features are extremely variant to different light conditions and other influences. This was achieved using CNN.

Presented a new approach to learning a frame-based classifier using weakly labelled sequence data by embedding a CNN within an iterative EM algorithm [97]. This allows the labeling of vast amounts of data at the frame level given only noisy video annotation. The iterative EM algorithm leverages the discriminative ability of the CNN to iteratively refine the frame level annotation and subsequent training of the CNN. The classifier achieves 62.8 % recognition accuracy on over 3000 manually labelled hand shape images. A model that is able to extract signs from videos, by processing the video frame by frame under minimally cluttered background was proposed by [98]. Signs are presented in a readable text, The system uses a Convolutional Neural Network (CNN) and fastai - a deep learning library, along with OpenCV for webcam input and displaying the predicted ASL sign. For high accurate image processing and classification tasks pre-trained ResNet-34 CNN classifier was adopted by the authors and datasets from Kaggle which consists of 26 alphabets along 3 special signs 'Space', 'Delete' and 'Nothing'. The datasets consist of 3000 images per character which comes to a total of 87000 images for the whole dataset. The model proposed achieved 78.5% accuracy on the testing set. In 2020 designed a mobile device-based sign language translation system using depth-only images [99]. The system performs image processing on a smartphone, collected depth images to emphasize the subject's hand and upper body gestures and exploits a convolutional neural network for feature extraction. The series of features gathered from word-representing videos are passed through a Long-Short Term Memory (LSTM) model for word-level sign lan-

---

guage translation. The authors trained and tested the system using a total of 2,200 samples collected from 26 people for 17 Korean Sign Language words. The classification accuracy of the proposed system using the self-collected data achieves 92% with an efficient image preprocessing phase.

In 2017 used temporal convolutions and recent advances in the deep learning like residual networks, batch normalization and exponential linear units (ELUs) to approach framewise classification problem [100]. The models were evaluated on three different datasets: the Dutch Sign Language Corpus (Corpus NGT), the Flemish Sign Language Corpus (Corpus VGT) and the ChaLearn LAP RGB-D Continuous Gesture Dataset (ConGD). The authors achieved a 73.5% top-10 accuracy for 100 signs with the Corpus NGT, 56.4% with the Corpus VGT and a mean Jaccard index of 0.316 with the ChaLearn LAP ConGD without the usage of depth maps.

A new feature extraction technique for hand pose recognition using depth and intensity images captured from a Microsoft KinectTM sensor was proposed by [101]. The technique was applied to American Sign Language fingerspelling classification using a Deep Belief Network for feature extraction. The authors evaluated results on a multi-user data set with two scenarios: one with all known users and the other with an unseen user and achieved 99 % recall and precision on the first, and 77 % recall and 79 % precision on the second.

Introduced a new Colombian sign language translation dataset (CoL-SLTD), that focuses on motion and structural information, which could be a significant resource to determine the contribution of several language components [102]. Encoder-decoder deep strategy was introduced to support automatic translation, including attention modules that capture short, long, and structural kinematic dependencies and their respective relationships with sign recognition. The evaluation in CoL-SLTD proves the relevance of the motion representation, allowing compact deep architectures to represent the translation. Also, the proposed strategy showed promising results in translation, achieving Bleu-4 scores of 35.81 and 4.65 in signer independent and unseen sentences tasks [103]. proposed a sign language recognition system based on wearable electronics and two different classification algorithms. The wearable electronics were made of a sensory glove and inertial measurement units to gather fingers, wrist, and arm/forearm movements. The classifiers were k-Nearest Neighbors with Dynamic Time Warping (that is a non-parametric method) and Convolutional Neural Networks (that is a parametric method). Ten sign-words were considered from the Italian Sign Language: cose, grazie, maestra, together with words with international meaning such as google, internet, jogging, pizza, television, twitter, and ciao. The adopted classifiers performed with an accuracy of  $96.6\% \pm 3.4$  (SD) for the k-Nearest Neighbors plus the Dynamic Time Warping and of  $98.0\% \pm 2.0$  (SD) for the Convolutional Neural Networks. Two-way communication was proposed by [104]. The objective of the authors was to develop a real time system for hand gesture recognition that is able to recognize hand gestures, features of hands

such as peak calculation and angle calculation and then convert gesture images into voice and vice versa. The ideas consisted of designing and implement a system using artificial intelligence, image processing and data mining concepts to take input as hand gestures and generate recognizable outputs in the form of text and voice with 91% accuracy. proposed a sign language translation system based solely on visual cues and deep learning for accurate translation of ASL [105]. The system applied Computer Vision and Neural Machine Translation for American Sign Language (ASL) gloss recognition and translation respectively. The authors were able to show that an end-to-end neural network system is not only capable of recognition of individual ASL glosses but also translation of continuous sign language videos into complete English sentences, making it an effective and practical tool for sign language communication.

The videos used to train Isolated Gloss Recognition System and the dataset used to train Gloss to Speech Neural Translator were obtained from The National Center for Sign Language and Gesture Resources (NCSLGR) Corpus [106]. proposed a real time sign language recognition system for ASL, convolutional neural network was trained by using dataset collected in 2011 by Massey University, Institute of Information and Mathematical Sciences, and 100% test accuracy was obtained. In the real-time system, the skin color is determined for a certain frame for hand use, and the hand gesture is determined using the convex hull algorithm, and the hand gesture is defined in real-time using the registered neural network model and network weights. The accuracy of the real-time system is 98.05% [107]. In a low-latency real-time sign language recognition application was developed to detect and process gestures performed from the Indian Sign Language (ISL) dictionary using a Convolutional Neural Network model, and to identify the words that were being communicated. The author's focused on of specific domain through custom made dataset containing 500 different images of the gesture corresponding to each word. The application detects both static and dynamic gestures performed by the user and generated a Python programming language like syntax for various constructs such as if, else. The major achievement of this research is the introduction of a novel method to taking programming knowledge to the hearing and speech impaired people and improving access to education for them. While most papers focused on developing systems for sign language recognition for various country using deep learning, proposed a layerwise optimized neural network architecture where batch normalization contributes to faster convergence of training, and introduction of dropout technique to mitigate data overfitting [108]. Batch normalization forces each training batch toward zero mean and unit variance, leading to improved flow of gradients through the model and convergence in shorter time. A constructed numerical hand gesture data set was used for validating the claims based on American Sign Language system and achieved a 98.50% accuracy. In the authors proposed method extracts upper body images directly from videos, and employs a pre-training convolutional network model to recognize the gesture in the image [109]. This method simplifies the hand-shape segmentation, and prevented information loss in feature extraction. The evaluation method on custom

self-built dataset includes 40 daily vocabularies, and showed that the proposed approach has good performance on sign language recognition task, with accuracy reaching to 99%. The size and quality of dataset used in training deep learning model determines to a great extent the quality of results [110]. trained MobileNet V1 convolutional neural network against the EgoHands dataset from Indiana University's UI Computer Vision Lab to determine if the dataset itself is sufficient to detect hands in LESCO (Costa Rican sign language) videos, from five different signers that wear short-sleeve shirts under complex backgrounds. Despite the high accuracy reported by the tests in the authors research, the hand detection module was unable to detect certain hand shapes such as closed fists and open hands pointing perpendicular to the camera lens. Therefore, complex egocentric views as captured in the EgoHands dataset might be insufficient for proper hand detection for Costa Rican sign language [111]. proposed a one-dimensional Convolutional Neural Network (CNN) array architecture for recognition of signs from the Indian sign language using signals recorded from a custom designed wearable IMU device. The array comprises of two individual CNNs in which one classify the general sentences and the other classify the interrogative sentence. The CNN array achieved a peak classification accuracy of 94.20% for general sentences and 95.00%.

In 2021 optimized a model for the recognition of Amharic Sign Language to Amharic characters [112]. A convolutional neural network model is trained on datasets gathered from a teacher of Amharic Sign Language. Two optimized algorithms namely the Faster R-CNN and SSD algorithm was evaluated with equal size of data sets to identify which model is better in terms of speed and accuracy. It is realized that Faster RCNN is better in accuracy for recognizing Amharic Sign Language while SSD is better in speed compared to Faster R-CNN but less accurate in recognizing Amharic Sign Language. Faster R-CNN model and SSD was able to detect and recognize the Sign Language with test different accuracy of 98. 25% and 96% respectively [113]. proposed a dynamic gesture recognition model based on CBAM-C3D. For a better network performance, key frame extraction technology, multimodal joint training, and network optimization with BN layer were used.

The experiments showed that the recognition accuracy of the proposed 3D convolutional neural network combined with attention mechanism reaches 72.4% on EgoGesture dataset [114]. built a symbiosis between a convolutional neural network (CNN) and a recurrent neural network (RNN) to recognize cultural/anthropological Italian sign language gestures from videos. The CNN extracts important features that are used by the RNN. RNNs is able to store temporal information inside the model to provide contextual information from previous frames to enhance the prediction accuracy. To avoid overfitting and provide small generalization error the authors used different data augmentation techniques and regularization methods from only RGB frames [115]. propose a fully convolutional network (FCN) for online SLR to concurrently learn spatial and temporal features from weakly annotated video sequences with only sentence-level annotations given. A gloss feature enhancement (GFE) module is introduced in the proposed network to enforce better sequence alignment learning [116]. In histograms of oriented gradients are used to extract the image features of hand sign. These features are then pass to the artificial neural network for training and recognition. The result showed that the proposed method is robust to detect the hand gestures in the complex background. It provides the accuracy recognition for the Thai fingerspelling of 84.05% [117]. Implement an algorithm for extracting Histogram of Gradient Orientation (HOG) features and these features are used to pass in neural network training for the gesture recognition purpose. The system is able to recognize alphabets characters (A-Z) and numerals (0-9) using Histograms of Oriented Gradients (HOG) features for Indian Sign language.

#### 4. Databases And Performance Metrics

There are different public databases available for gesture datasets that have been commonly adopted for evaluating deep learning-based methods. This study focuses on only four (7) popularly used to evaluate proposed methods. These include 20BN-jester dataset, Montalbano dataset(v2), ChaLearn LAP IsoGD, DvsGesture dataset, Sheffield kinect gesture dataset, EgoGesture dataset and Praxis gestures dataset [6,118,119]. The characteristics of each of these databases are briefly summarised in Table 6.

Dataset	Year	Acquisition device	Modality	#Classes	#Subjects	#Samples	#Scenes	Metric
20BN-jester dataset	2019	Laptop camera or webcam	RGB	27	1376	148092	-	Accuracy
Montalbano dataset(V2)	2014	Kinect v1	RGB, D S, UM	20	27	13858	-	Jaccard index
ChaLearn LAP IsoGD	2016	Kinect v1	RGB, D	249	21	47933	-	Accuracy
DVS128Gesture dataset	2017	DVS128and webcam	RGB	11	29	1342	1	Accuracy
SKIG	2013	Kinect v1	RGB, D	10	6	2160	3	Accuracy
EgoGesture dataset	2018	Intel RealSense	RGB, D	83	50	24161	6	Accuracy
Praxis gestures dataset	2017	Kinect v2	RGB-D	29	7	126	-	-

Table 6: 7 Publicly Available Gesture Databases

---

## 5. Key Knowledge Gaps

Despite all the progress and advancement in using gloves, embedded systems, machine learning and deep learning there is still no complete system for effective communication between the hearing and speech impaired and the general populace. Most of the research tends to focus on how to improve accuracy of recognition leaving other areas of speech and text for hearing impaired. Deep learning in this area of study is still in its infancy and has proved not sufficient due to prevailing problems like sign language recognition methods which are easily affected by human movement, change of gesture scale, small gesture area, complex background, illumination and so on. Compared with basic gestures, gestures in sign language are characterized by complex hand shape, blurred movement, low resolution of small target area, mutual occlusion of hands and faces, and overlapping of left and right hands. Therefore, how to build an efficient and suitable sign language recognition model has become a hot research area since they have some deficiencies [55,59]. Sign language recognition as a research direction with broad application and development space still has much room for improvement. Also, most methods of sign language recognition currently only consider the accuracy of the algorithm. However, for the application of sign language recognition in real scene, real-time performance is an important index. Therefore, it is also a direction worthy of further research to find how to improve the speed of locating hand and recognizing sign language words [55].

So far data acquisition can be done mainly with data gloves, video cameras, and new motion sensing devices. This is a big problem, research could be done to come up with standard and universal ways of capturing data and making such data available for other researchers to scrutinize its strength and weakness. Of all the research work reviewed only few and common datasets are used, there is no research with complete words or alphabet for any country sign language. There is no sign language translation from one country to other country's sign language as we have in spoken languages where NLP has been used to translate different languages (Multilingual). Getting clean, efficient and required dataset is a big gap that needs to be closed if deep learning must be used to solve the problem of sign recognition.

To the best of our knowledge, only proposed a model for the communication of hearing and speech impaired people in a small group [120]. There is no other research on how the speech/hearing impaired can communicate in either small or large group. This is because Automatic Speech Recognition (ASR) is still imperfect, and it contains errors in its output text in many real-world conversation settings. In Summary, a complete recognition system must be able to identify alphabets, numerals, static and dynamic words, contexts, emotions, coarticulation phase, facial expressions, eyebrow movement, body posture, and numerous other situations [62].

## 6. Trend Analysis, Challenges, And Future Directions

Sign language recognition problem is a broad research area which includes recognition problems like finger spelling dynamic alphabets, dynamic words, co-articulation detection and elimination for sentence identification. With advancement in technology and new researchers in this domain, architectures can be extended with additional modules and techniques to form a fully automated sign language recognition system in the future. Facial expression and context analysis are the other parts to be included in sign language recognition. An automated ISL recognition system with speech translator which can process videos in real time to produce voice output of the same can become a most effective assistive technology in near future [62].

presented a system that is capable of learning gestures by using the data from the Leap Motion device and the Hidden Markov classification (HMC) algorithm in a Virtual Reality (VR) environment using data produced by the Leap Motion device [121]. The authors achieved the gesture recognition accuracy (mean  $\pm$  SD) is  $86.1 \pm 8.2\%$  and gesture typing speed is  $3.09 \pm 0.53$  words per minute (WPM), when recognizing the gestures of the American Sign Language (ASL) [122]. proposed an approach, for sign language recognition, that makes use of a virtual reality headset to create an immersive environment using features from data acquired by the Leap Motion controller, using an egocentric view, can be used to automatically recognize a user signed gesture. The Leap features are used along with a random forest for real-time classification of the user's gesture. 26 letters of the alphabet in American Sign Language in a virtual environment with an application for learning was used to test the efficacy of the proposed approach. A classification accuracy of 98.33% and 97.1% was achieved using a random forest, and a deep feedforward neural network, respectively.

Another important application of gesture recognition is in the field of Augmented Reality and Virtual Reality. Using gesture recognition users can carry out tasks without the need of keyboards or voice. The user should only perform hand gestures and the computer will understand what action user is trying to perform and then translate the gestures into audio for a normal person to hear. While voice control is a good way of hands-free control, it brings with a long significant delay while gesture recognition can happen almost instantaneously. Also, voice control requires much more effort in AR/VR applications and gesture control feels much more intuitive and natural [63]. With augmented reality the barrier of communicating in different sign languages will be completely eliminated since the system is able to understand different sign languages and then translate it into any language using Natural Language Processing (NLP).

Author	Year	1D/2D/3D	Sign Language	Dataset	Technique used	Frame work	Gesture recognized	Pretrained	Data state	Microcontroller	Accuracy
[44]	2017	2D	-	-	HSV tracking	OpenCV	11 different hand gestures	-	Dynamic	Raspberry pi 2	-
[45]	2019	2D	-	-	-	OpenCV	-	-	Dynamic	Raspberry Pi	-
[46]	2017	2D	-	-	Gaussian blur, Convex Hull, Frame segmentation	OpenCV	-	-	Dynamic	Raspberry Pi	0.98
[47]	2015	2D	-	-	Color splitting, morphological processing and feature extraction & classification, template matching	MATLAB	-	-	Static	LPC2138	-
[48]	2017	2D	-	-	Tesseract OCR (Online Character Recognition), espeak, Speech to text (STT), (TTS)	OpenCV	-	-	Dynamic	Raspberry Pi	-
[49]	2022	3D	America	own dataset	CNNs, SSD Mobile net V2	TensorFlow, Object detection API, Open CV, LabelImg	2000 image of 5 symbols	Yes	Static	-	0.70-0.80
[50]	2019	2D	Indian	own dataset	CNN, OpenCV (find Contours)	Keras	96000 images of 48 gestures	-	Dynamic	-	0.99
[51]	2018	2D	-	LSA64 dataset	CNNs, hand and body skeletal features extracted from RGB, ImageNet VGG-19 network, linear dynamical system (LDS) histogram	-	-	Yes	Static	-	-
[53]	2020	2D	America	RWTH-Phoenix-Weather-2014, RWTH Phoenix-Weather-2014T, & CS	CNNs, gradient descent with momentum (SGDM) optimization function	-	10 different hands posture classes with 200 images & 24 letters of the ASL alphabet excluding the letters 'j' and 'z'	-	static	-	0.947 & 0.9996
[54]	2020	2D	-	RWTH-Phoenix-Weather-2014, RWTHPhoenix-Weather-2014T, & CS	CNN for spatial feature extraction, stacked 1D temporal convolution layers, bidirectional long short-term memory	-	-	-	Static	-	-

[55]	2019	3D	-	vocabulary	Faster R-CNN, long and short time memory	-	40 common words and 10,000 sign language images	Yes	Static	-	0.99
[56]	2019	3D	China	Custom Chinese Sign Language (CSL) dataset & ChaLearn14 benchmark	3D-CNN for capturing spatio-temporal features, spatial attention	-	500 categories	-	Static	-	-
[57]	2018	3D	America	Custom own dataset	3D-CNN	OpenCV	5 words, 20 images taken from different angles and total of 100 for each word	-	Static	Kinect sensor	0.923
[58]	2018	3D	America	Custom own dataset	video sequences and extracts temporal and spatial features, CNN, RNN, Inception V3	-	600 training samples of 300 frames each	Yes	Static	-	0.99
[59]	2019	2D	America	sign language MNIS	Capsule networks, LeNet	-	24 classes. 27455 examples in training set and 7172 in test se	-	Static	-	0.8893, 0.8219
[60]	2018	2D	America	Kaggle American Sign Language Letter	capsule-based deep neural network sign posture translator, adaptive pooling, CNNs	PyTorch	24 classes. 27455 examples in training set and 7172 in test se	No	Static	-	0.99
[61]	2019	3D	Bangla	Custom own dataset	customized Region of Interest (ROI) segmentation and Convolutional Neural Network (CNN)	Keras, OpenCV	100 signs from each class	-	Dynamic	Raspberry Pi	0.9754
[62]	2019	3D	India	binary hand region silhouette of the signer images	CNNs	-	500 images	-	Static	-	0.9864
[63]	2020	-	India	Custom own dataset	-	-	500 images for each alphabet	-	Dynamic	-	-
[64]	2018	3D	India	Custom own dataset	CNNs, skin segmentation to find the Region of Interest in the frame which shows the bounding box	Keras API with Tensor Flow as backend	300 images of each Indian Sign language numerals captured using RGB camera	-	Dynamic	-	.9956
[65]	2018	2D	India	Custom own dataset	CNNs, stochastic pooling	-	60000 images are used	-	Dynamic	-	0.9288
[66]	2020	3D	India	Custom own dataset	CNNs, YOLO, VGGNet	OpenCv	26 classes of 2000 plus images	Yes	Dynamic	-	-

[67]	2020	2D	America	ASL Finger Spelling benchmark	YCbCr Segmentation method and local binary pattern, SVM, VGG-1	-		total of 95,697 images and around for each user 4000 images are found in each alphabet, 24 static signs excluding the letters j and z	Yes	Static	-	0.9844
[68]	2018	2D	Bengali	Bengali custom dataset	CNNs, VGG16	-		37 classes of total 1147 images	Yes	Static	-	0.9633
[69]	2020	2D	Bengali	Bengali Public dataset	CNNs	Keras		4168 samples (basic characters: 18745 and numerals: 5423)	-	Static	-	0.9875
[70]	2020	2D	Bengali	Bengali Ishara-Lipi dataset	CNNs	OpenCv		3600 pre-processed images	-	Static	-	0.927
[71]	2020	2D	Bengali	Bengali Ishara-Lipi dataset	CNNs	Keras, TensorFlow		36,000 image samples for 36 alphabetical classes	-		-	0.9922
[72]	2017	2D	English	English Sign Language (ESL) is obtained from Kaggle	CNNs	Deep Learning Studio (DLS)		total of 810 images, of 26 English symbols and space	-	Static	-	0.82
[73]	2019	2D	America	Massey Dataset	CNNs, Histogram of Oriented Gradients (HOG) and Local Binary Pattern (LBP) features of each gesture will be extracted from training images. Then Multi-class Support Vector Machines (SVMs)	-		2524 images of static alphabetical hand gestures from a to z	-	Static	-	0.9708, 0.9830
[74]	2019	3D	India	Custom own dataset	CNNs modified LSTM model for continuous sequence	-		942 signed sentences, 35 different sign words	-	Static	Leap motion senso	0.895
[75]	2018	3D	America	Custom American Fingerspelling	Dense Net, CNN	-		50,000 images of letters	Yes	Dynamic	-	0.903
[76]	2018	2D	America	Custom own dataset	CNNs, YCbCr color space, AlexNet	MATLAB		four classes, 150 images per class	Yes	Dynamic	-	0.94
[77]	2019	3D	America	Custom own dataset	skin-color modeling technique, CNN	Keras and TensorFlow		1,200 images	-	Static	-	0.99
[78]	2019	3D	Russian	Custom own RSLdactyl dataset	CNN	-		33 gestures, only 26 gestures are static	-	Static	-	0.98
[79]	2020	2D	Arabic	Public ArASL	CNN, VGG16 and the Res-Net152	-		54,049 images distributed around 32 classes	Yes	Static	-	0.9945

[80]	2018	3D	Japan	Cohn-Kanade and Japanese Female Facial Expression	DNN, SVM	-	-	-	Static	-	0.946, 0.803
[81]	2018	2D	Hausa	Custom own dataset	Particle Swarm Optimization (PSO), Fourier descriptor, artificial neural network	MATLAB	21 classes with 10 sample searches for the static, manual and non-manual signs	-	Static	-	0.93.9
[82]	2020	2D	Indonesian	SIBI datasets from kaggle.com	CNN, model E, AlexNet	-	29 objects, 26 letters (a-z), nothing, delete, and space	Yes	Static	-	0.9682
[83]	2022	3D	India	Custom own dataset	CNN SSD MobileNet v2	TensorFlow Object Detection API, OpenCV	650 images in total, 25 images for each alphabet	Yes	Dynamic	-	0.8545
[84]	2016	2D	Bangladeshi	Custom own dataset	artificial neural network (ANN)	Matlab	518 images of 37 signs	-	static	-	0.99
[5]	2020	3D	Arabic	KSU-SSL dataset	3DCNN	openpose	40 dynamic hand gestures performed by 40 subjects	-	static	-	-
[85]	2020	3D	Arabic	KING SAUD UNIVERSITY SAUDI SIGN LANGUAGE (KSU-SSL), ARABIC SIGN LANGUAGE (ArSL), PURDUE RVL-SLLL AMERICAN SIGN LANGUAGE-DATASET	3DCNN, DenseImage Ne	-	0 classes and each class has 200 gesture, 3444 valid samples, 280 gesture sample	Yes	Static	-	0.9669,
[86]	2016	2D	Arabic	Custom won dataset	PCANet, support vector machine classifier	-	28 Arabic alphabetic 50 tunes for each alphabet.	-	Static	SOFTKI-NECT sensor	0.995
[87]	2020	3D	British	BSL dataset from kaggle.com	CNN, Leap Motion model VGG16	-	-	Yes	Dynamic	-	0.94444
[88]	2017	3D	Germany	RWTH-PHOENIX-Weather	CNN, recurrent convolutional neural network (LSTM) for spatio-temporal feature extraction and sequence learning. VGG-S	-	672 sentences in German sign language for training with 65,227 sign glosses and 799,006 frames	Yes	Static	-	-
[89]	2019	3D	-	RWTH-PHOENIX-Weather multi-signer 2014, SIGNUM signer-dependent	CNN, sequence learning, iterative training, multimodal fusion, sequence learning, iterative training, multimodal fusion (Bi-LSTM), VGG-S	-	-	Yes	Static	-	-

[90]	2018	3D	America	Custom own dataset	CNN, Inception v3	TensorFlow	24 labels of static gestures from letters A to Y, excluding J, average 100 images per class	Yes	Static	-	0.90
[91]	2018	3D	Japan	Custom own dataset	Sequence to sequence neural network model, CNN, LSTM	-	379 sentences (total of 812 sentences with a vocabulary of 195 words)	-	Dynamic	-	-
[92]	2019	3D	Brazil	LSA64 dataset, IBRAS-BSL dataset	CNN, multimodal information (RGB-D)	-	3200 videos; 10 subjects executed 5 repetitions of 64 different types of signs,	-	Dynamic	-	0.9692, 0.8702
[93]	2019	2D	Ghana	Custom own dataset	CNN ImageNet, VGG-16 and VGG-19	Keras	66000 images in the RGB colour space, with 33 classes of static gestures consisting of 24 alphabets and 9 digits (1-9)	Yes	Static	-	0.96
[94]	2019	3D	America	Custom own dataset	CNN	Keras TensorFlow	4300 images for each of the 29 classes	-	Dynamic	Raspberry Pi	0.93
[95]	2015	3D	-	Custom own dataset	deep neural network (DNN)	-	65,000 frame images	-	Static	Real-Sense and Kinect	0.989
[96]	2020	2D	America	-	CNN	-	-	-	Dynamic	-	-
[97]	2016	2D	-	RWTH-PHOENIX-Weather, Danish sign, New Zealand	CNN	-	-	-	Dynamic	-	0.628
[98]	2019	3D	America	dataset from Kaggle	CNN ResNet-34	Fastai, OpenCV, PyTorch	3000 images per character, which comes to a total of 87000 images	Yes	Dynamic	-	0.785
[99]	2020	2D	Korean	Custom own dataset	CNN Long-Short Term Memory (LSTM), MobileNet v2	Keras	total of 2,200 samples collected from 26 people for 17 words	Yes	-	-	0.92
[100]	2017	2D	-	Dutch Sign Language Corpus (Corpus NGT), the Flemish Sign Language Corpus (Corpus VGT) and the ChaLearn LAP RGB-D Continuous Gesture Dataset (ConGD)	temporal convolutions and recent advances in the deep learning field like residual networks, batch normalization and exponential linear units (ELUs)	-	-	-	Static	-	0.735

[101]	2014	2D	America	-	Deep BeliefNetwork	-	24 static letters of the fingerspelling alphabet, 500 images of each letter for each person, resulting in over 60000 total	-	Static	Microsoft Kinect sensor	-
[102]	n.d	3D	Colombia	CoL-SLTD	3D-CNN, bi-directional LSTM	-	-	-	Dynamic	-	-
[103]	2020	3D	Italian	-	k-Nearest Neighbors with Dynamic Time Warping, CNN	MATLAB	10 different gestures, repeated 100 times	-	Dynamic	sensory glove	0.98
[104]	2016	2D	-	-	-	MATLAB	-	-	Dynamic	-	0.91
[105]	2019	3D	America	National Center for Sign Language and Gesture Resources (NCSLGR) Corpus	-	-	-	-	Dynamic	-	-
[106]	2018	2D	America	Institute of Information and Mathematical Sciences, Massey University	CNN	Tensorflow and Keras	25 images of cropped images for each hand gesture, in total, 900 images	-	Dynamic	-	0.9805
[107]	2019	2D	India	Custom own dataset	CNN,	OpenCV	consists of 400 images	-	Static	-	-
[108]	2017	2D	America	Custom own dataset	Deep Convolutional Neural Networks (DCNN)	-	500 images for each class, totaling 5000 images in 10 numeral classes.	-	Static	-	0.9850
[109]	2017	3D	China	Custom own dataset	CNN	-	40 daily vocabularies	Yes	Dynamic	-	0.99
[110]	2019	2D,3D	Costa Rica	EgoHands dataset	CNN, MobileNet	-	four thousand eight hundred (4,800) labeled images (frames) taken from forty-eight (48) videos	Yes	Dynamic	-	0.961
[111]	2019	1D	Indian	Custom own dataset	CNN	-	total of 10 different subjects, out of which 5 were male and 5 are female	-	Dynamic	IMU device, Arduino	0.9420
[112]	2021	3D	Amharic	Custom own dataset	CNN, Faster R-CNN and SSD, VGG-16	KERAS	10 classes, 500 frames	Yes	Dynamic	-	0.9825, 0.96
[113]	2021	3D	China	EgoGesture datas	CNN	PyTorch, OpenCV	2081 RGB-D videos, 24161 gesturesamples, and 2953224 frames from six different themes	-	Dynamic	-	0.724

[114]	2021	3D	Italy	ChaLearn dataset	CNN, RNN	-	20 classes	-	Dynamic	Microsoft Kinect sensor,	0.772
[115]	2020	1D	-	Chinese Sign Language (CSL), RWTH-PHOENIX-Weather-2014 (RWTH) d	fully convolutional network (FCN)	-	6,841 different sentences signed by 9 different signers (around 80,000 glosses with a vocabulary of size 1,232), 00 sentences, each being signed for 5 times by 50 signers (in total 25,000 videos)	-	Static	-	-
[116]	2016	3D	Thai	Custom own dataset	artificial neural network	-	720 hand gestures images for training set, with 24 classes	-	Dynamic	Microsoft Kinect sensor	0.8405
[117]	2014	2D	India	Custom own dataset	CNN, Histograms of Oriented Gradients (HOG) features.	MATLAB	36 classes 26 alphanets and 0-9	-	Dynamic	-	-

**Table 7: The Summary of the Articles Reviewed for Vision Based Approach**

Author	Year	Algorithm	Sign Language	Dataset	Sensor	Microcontroller	Technique	Gesture recognized	Accuracy
[8]	2011	Hidden Markov Model classifier	America	RWTH-BOSTON-50 database	-	-	PCA as a global image descriptor, for describing hand shape and orientation	Static	-
[9]	2016	hidden Markov models (HMMs)	China	Custom own dataset	Kinect2.0	-	sign language recognition based on trajectory modeling	Dynamic	-
[10]	2016	K-means algorithm, Hidden Markov Model classifier	Taiwan	Custom own dataset	-	-	PCA as a global image descriptor	Static	0.913
[11]	1998	hidden Markov models (HMMs)	America	-	-	-	-	Dynamic	0.97
[12]	2006	hidden Markov models (HMMs)	-	Custom own dataset	-	-	Hidden Conditional Random Fields for Gesture Recognition	Dynamic	0.9775
[13]	2009	hidden Markov models (HMMs), Baum-Welch algorithm	-	Custom own dataset	-	-	-	Dynamic	0.932
[14]	2010	support vector machine	Irish	Jochen-Triesch static hand posture, ISL data set	-	-	eigenspace Size	Dynamic Function	0.973 and 0.935
[15]	2014	support vector machine	Arabic	Arabic sign language (ArSL) database	-	-	IHLS color space and random forest classifier	Static	0.995
[16]	2017	Mobile App	Sinhala	emoji	-	-	chatting	Dynamic	-
[17]	2021	Mobile App	English and Bahasa Malaysia	Vuforia Image Target Database	AR's Scene Generator,	-	Augmented reality (AR)	Dynamic	-
[18]	2014	DynamicTime Warping (DTW),	America	-	Microsoft Kinect camera	-	-	Dynamic	0.82
[19]	2013	Co-occurrences, efficient discriminative search	British	-	-	-	-	Dynamic	0.927
[20]	2010	SVM classifier	British	-	-	-	-	Dynamic	-
[21]	2002	time-delay neural network (TDNN)	America	Custom video database of 40 ASL signs	-	-	2D Motion Trajectories	-	-
[22]	2012	Camshift method and Hue, Saturation, Intensity (HSV) color model	India	-	-	-	-	Static	-
[23]	2015	-	-	SIGNUM database, WTH-PHOENIX-Weather	-	-	tracking, features, signer dependency, visual modelling and language modelling	Static	-

[24]	2016	feature covariance matrix based serial particle filter	America	RWTH-BOSTON-50	-	-	fusion of median and mode filtering is proposed for background modeling	Static	08733
[25]	2015	Boosting method, modal fusion, Feature pooling	America	ChaLearn-2014	Kinect sensor	-	-	Static	0.834
[26]	2013	SIFT (scale invariance Fourier transform)	India	Custom own dataset	-	-	Feature Extraction, Key point Matching	Static	0.95
[27]	2013	k - Nearest Neighbor (k-NN) and multi-class Support Vector Machine (SVM) classification	America	-	-	-	skin color as detection cue with RGB and YCbCr color spaces, and thresholding of gray level intensities.	Static	0.901
[28]	2018	Linear Discriminant Analysis (LDA)	India	Custom own dataset	-	-	features are extracted such as Eigen values and Eigen vectors	Static	-
[29]	2004	-	-	Custom own dataset	-	-	Wearable gloves	Dynamic	-
[30]	2019	-	-	-	Flex Sensors	Raspberry Pi,	Movable Device	Dynamic	-
[31]	2019	-	-	-	MEMS SENSOR	Arduino	Movable Device	Dynamic	-
[32]	2016	-	Arabic	-	Flex Sensor	Arduino	Wearable gloves	-	-
[33]–[35]	2014, 2017, 2013	-	-	-	Flex Sensor	PIC	Wearable gloves	Dynamic	0.989
[36]	2014	motion sensor.	America	-	Flex Sensor	PIC	Wearable gloves	Dynamic	-
[37]	2020	-	-	-	Flex Sensor	-	Wearable gloves	Dynamic	-
[38]	2016	-	India	-	Flex sensor	AVR	Wearable gloves	Dynamic	-
[39]	2020	-	America	-	Flex Sensor	NI USB6008 DAQcard	Wearable gloves	Dynamic	-
[40]	n.d.	OpenCv, Google API	-	-	Camera	Raspberry P	Wearable gloves	Dynamic	-
[41]	2007	fuzzy rule-based classification	Vietnam	-	MEMS accelerometers	BASIC Stamp micro-controller	Wearable gloves	Dynamic	-
[42]	2014	Hough Transform technique.	Tamil	-	-	-	Hand Segmentation Using Lab Color Space (HSL)	Static	-
[43]	2020	machine learning (ML)	America	-	RF sensor	-	Frequency warped cepstral coefficients (FWCC)	Static	0.95

**Table 8: The Summary of the Articles Reviewed for Non Vision Based Approach**

## 7. Conclusion

This paper carried out a comprehensive and critical analytic review of 110 published articles on gesture and sign language recognition approaches spanning from 1998 to 2022. It summarizes the non-vision and vision based techniques. It also presents trend analysis on recent literature, identifies gaps to be filled in future research and provides possible solutions to filling these gaps. This survey is very significant as it presents the strength and weakness of techniques that have been adapted from 1998 to 2022. Although vision-based approaches have made significant progress in recent years using deep learning and computer vision, there are still many prospects for improvement. Several insights, potential research

directions have been described in this survey and indicating the numerous opportunities in this field despite the advances achieved so far.

## References

1. World Health Organization. (2021).
2. Nwadinobi, V. (2019). Chapter eight hearing impairment. *Research Gate*.
3. Mohandes, M., Liu, J., & Deriche, M. (2014, February). A survey of image-based arabic sign language recognition.
4. Lahamy, H., & Lichti, D. D. (2012). Towards real-time and rotation-invariant American Sign Language alphabet

- recognition using a range camera. *Sensors*, 12(11), 14416-14441.
5. Al-Hammadi, M., Muhammad, G., Abdul, W., Alsulaiman, M., Bencherif, M. A., Alrayes, T. S., ... & Mekhtiche, M. A. (2020). Deep learning-based approach for sign language gesture recognition with efficient hand gesture representation. *Ieee Access*, 8, 192527-192542.
  6. Yuanyuan, S. H. I., Yunan, L. I., Xiaolong, F. U., Miao, K., & Miao, Q. (2021). Review of dynamic gesture recognition. *Virtual Reality & Intelligent Hardware*, 3(3), 183-206.
  7. Kieu, S. T. H., Bade, A., Hijazi, M. H. A., & Kolivand, H. (2020). A survey of deep learning for lung disease detection on medical images: state-of-the-art, taxonomy, issues and future directions. *Journal of imaging*, 6(12), 131.
  8. Zaki, M. M., & Shaheen, S. I. (2011). Sign language recognition using a combination of new vision based features. *Pattern recognition letters*, 32(4), 572-577.
  9. Pu, J., Zhou, W., Zhang, J., & Li, H. (2016, January). Sign language recognition based on trajectory modeling with HMMs. In *International conference on multimedia modeling* (pp. 686-697). Cham: Springer International Publishing.
  10. [10]Li, T. H. S., Kao, M. C., & Kuo, P. H. (2016). Recognition system for home-service-related sign language using entropy-based \$ K \$-means algorithm and ABC-based HMM. *IEEE transactions on systems, man, and Cybernetics: systems*, 46(1), 150-162.
  11. Starner, T., Weaver, J., & Pentland, A. (1998). Real-time american sign language recognition using desk and wearable computer based video. *IEEE Transactions on pattern analysis and machine intelligence*, 20(12), 1371-1375.
  12. Wang, S. B., Quattoni, A., Morency, L. P., Demirdjian, D., & Darrell, T. (2006, June). Hidden conditional random fields for gesture recognition.
  13. Kelly, D., Mc Donald, J., & Markham, C. (2009, September). Continuous recognition of motion based gestures in sign language.
  14. Kelly, D., McDonald, J., & Markham, C. (2010). A person independent system for recognition of hand postures used in sign language. *Pattern Recognition Letters*, 31(11), 1359-1368.
  15. Aly, S., & Mohammed, S. (2014, November). Arabic sign language recognition using spatio-temporal local binary patterns and support vector machine. In *International Conference on Advanced Machine Learning Technologies and Applications* (pp. 36-45). Cham: Springer International Publishing.
  16. Jayatilake, L., Darshana, C., Indrajith, G., Madhuwantha, A., & Ellepola, N. (2017). Communication between deaf-dumb people and normal people: Chat assist. *International Journal of Scientific and Research Publications*, 7, 12.
  17. Rum, S. N. M., Boilis, B. L. (2021 ).“Sign language communication through augmented reality and speech recognition (LEARNSIGN),” *International Journal of Engineering Trends and Technology*, vol. 69, no. 4, pp. 125–130.
  18. Jangyodsuk, P., Conly, C., & Athitsos, V. (2014, May). Sign language recognition using dynamic time warping and hand shape distance based on histogram of oriented gradient features. In *Proceedings of the 7th international conference on Pervasive technologies related to assistive environments* (pp. 1-6).
  19. Pfister, T., Charles, J., & Zisserman, A. (2013). Large-scale learning of sign language by watching TV.
  20. Buehler, P., Everingham, M., & Zisserman, A. (2010). Employing signed TV broadcasts for automated learning of British sign language.
  21. Yang, M. H., Ahuja, N., & Tabb, M. (2002). Extraction of 2d motion trajectories and its application to hand gesture recognition. *IEEE Transactions on pattern analysis and machine intelligence*, 24(8), 1061-1074.
  22. Ghotkar, A. S., Khatal, R., Khupase, S., Asati, S., & Hadap, M. (2012, January). Hand gesture recognition for indian sign language.
  23. Koller, O., Forster, J., & Ney, H. (2015). Continuous sign language recognition: Towards large vocabulary statistical recognition systems handling multiple signers. *Computer Vision and Image Understanding*, 141, 108-125.
  24. K. M. Lim, A. W. C. Tan, and S. C. Tan, “A feature covariance matrix with serial particle filter for isolated sign language recognition,” *Expert Syst Appl*, vol. 54, pp. 208–218, Jul. 2016,
  25. Camille, M., German, S., Andrey, O., (2015). “A Multi-scale Boosted Detector for Efficient and Robust Gesture Recognition,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 8927. Springer Verlag, p. VI.
  26. Goyal, S., Sharma, I., & Sharma, S. (2013). Sign language recognition system for deaf and dumb people. *International Journal of Engineering Research Technology*, 2(4), 382-387.
  27. Sharma, R., Nemani, Y., Kumar, S., Kane, L., & Khanna, P. (2013, July). Recognition of single handed sign language gestures using contour tracing descriptor. In *Proceedings of the world congress on engineering* (Vol. 2, pp. 3-5).
  28. Kumar, M. (2018). “Conversion of Sign Language into Text,”.
  29. Pradipa, R., Kavitha, M. S., Madurai, Nadu, T. (2004 ). “Hand Gesture Recognition-Analysis of various techniques, methods and algorithms”.
  30. Vigneshwaran, S., Fathima, M. S., Sagar, V. V., & Arshika, R. S. (2019, March). Hand gesture recognition and voice conversion system for dump people.
  31. Kiruthiga, B., Banu, R. N. B., Aparna, K. M., Kaiserea, J. D. K., Kumari, D. S. (2019). “Gesture Control Smart System for Deaf and Dumb People.”
  32. Abdulla, D., Abdulla, S., Manaf, R., & Jarndal, A. H. (2016, December). Design and implementation of a sign-to-speech/text system for deaf and dumb people.
  33. Lavanya, V., Akulapraavin, M. S., Mohan, M. (2014). “Hand Gesture Recognition And Voice Conversion System Using Sign Language Transcription System.”
  34. Prabakaran, S., Revathi, S., Soundarya, S., Tamilarasi, K.

- R., Rasu, R. (2017). "Hand Gesture Recognition and Voice Conversion System for Deaf and Dumb".
35. Verma, P., S. S. L., Priyadarshani, R. (2013). "Design of Communication Interpreter for Deaf and Dumb Person".
  36. Padmanabhan, V., & Sornalatha, M. (2014). Hand gesture recognition and voice conversion system for dumb people. *International Journal of Scientific & Engineering Research*, 5(5), 427.
  37. Shareef, S. R., Hussain, M. M., Gupta, A., & Aslam, H. A. (2020). Hand gesture recognition system for deaf and dumb. *International Journal of Multidisciplinary and Current Educational Research*, 2(4), 82-88.
  38. Jadhav, A. J., & Joshi, M. P. (2016, September). AVR based embedded system for speech impaired people.
  39. Kumuda, S., & Mane, P. K. (2020, February). Smart assistant for deaf and dumb using flexible resistive sensor: implemented on LabVIEW platform. In *2020 International Conference on Inventive Computation Technologies (ICICT)* (pp. 994-1000). IEEE.
  40. S. Raghul. M., Ramakrishna, S. "Raspberry-Pi Based Assistive Device For Deaf, Dumb And Blind People Assistant Professor Of Electronics and Communication."
  41. Bui, T. D., & Nguyen, L. T. (2007). Recognizing postures in Vietnamese sign language with MEMS accelerometers. *IEEE sensors journal*, 7(5), 707-712.
  42. Jayanthi, P., & Thyagarajan, K. K. (2013, December). Tamil alphabets sign language translator. In *2013 fifth international conference on advanced computing (ICoAC)* (pp. 383-388). IEEE.
  43. Gurbuz, S. Z., Gurbuz, A. C., Malaia, E. A., Griffin, D. J., Crawford, C. S., Rahman, M. M., ... & Mdrafi, R. (2020). American sign language recognition using rf sensing.
  44. Rishad, E. K., Vyshakh, C. B., & Shahul, S. U. (2017). Gesture controlled speaking assistance for dump and deaf.
  45. Karmel, A., Sharma, A., & Garg, D. (2019). IoT based assistive device for deaf, dumb and blind people. *Procedia Computer Science*, 165, 259-269.
  46. Abed, A. A., & Rahman, S. A. (2017). Python-based Raspberry Pi for hand gesture recognition. *International Journal of Computer Applications*, 173(4), 18-24.
  47. Agarwal, N., Hambarde, S. H. (2015). "Static Hand Gesture's Voice Conversion System Using Vision Based Approach For Mute People".
  48. Kumar, A., Raushan, R., Aditya, S., Jaiswal, V. K., & Divyashree, M. (2017). An innovative communication system for deaf, dumb and blind people.
  49. Pathak, A., Kumar, A., Priyam, P., Gupta, P., Chugh, G., Awasthi, K., ... & Jain, L. (2022). Real time sign language detection.
  50. Bohra, T., Sompura, S., Parekh, K., & Raut, P. (2019, November). Real-time two way communication system for speech and hearing impaired using computer vision and deep learning.
  51. Konstantinidis, D., Dimitropoulos, K., & Daras, P. (2018, June). Sign language recognition based on hand and body skeletal data.
  52. Dimitropoulos, K., Barmoutis, P., Grammalidis, N. (2017). "Higher Order Linear Dynamical Systems for Smoke Detection in Video Surveillance Applications," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 27, no. 5, pp. 1143-1154.
  53. Adithya, V., & Rajesh, R. (2020). A deep convolutional neural network approach for static hand gesture recognition. *Procedia computer science*, 171, 2353-2361.
  54. Papastratis, I., Dimitropoulos, K., Konstantinidis, D., & Daras, P. (2020). Continuous sign language recognition through cross-modal alignment of video and text embeddings in a joint-latent space. *IEEE Access*, 8, 91170-91180.
  55. He, S. (2019, October). Research of a sign language translation system based on deep learning. In *2019 International conference on artificial intelligence and advanced manufacturing (AIAM)* (pp. 392-396). IEEE.
  56. Huang, J., Zhou, W., Li, H., Li, W. (2019). "Attention-Based 3D-CNNs for Large-Vocabulary Sign Language Recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, no. 9, pp. 2822-2832.
  57. Soodtoetong, N., & Gedkhw, E. (2018, July). The efficiency of sign language recognition using 3D convolutional neural networks.
  58. Bantupalli, K., & Xie, Y. (2018, December). American sign language recognition using deep learning and computer vision.
  59. Bilgin, M., & Mutludoğan, K. (2019, October). American sign language character recognition with capsule networks.
  60. Jalal, M. A., Chen, R., Moore, R. K., & Mihaylova, L. (2018, July). American sign language posture understanding with deep neural networks.
  61. Khan, S. A., Joy, A. D., Asaduzzaman, S. M., & Hossain, M. (2019, April). An efficient sign language translator device using convolutional neural network and customized ROI segmentation.
  62. Sruthi, C. J., & Lijiya, A. (2019, April). Signet: A deep learning based indian sign language recognition system.
  63. Chhabria, K., Priya, V., & Thaseen, I. S. (2020, February). Gesture recognition using deep learning. In *2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE)* (pp. 1-4). IEEE.
  64. Sajanj, T. D., & Beena, M. V. (2018, April). Indian sign language numeral recognition using region of interest convolutional neural network.
  65. Rao, G. A., Syamala, K., Kishore, P. V. V., & Sastry, A. S. C. S. (2018, January). Deep convolutional neural networks for sign language recognition.
  66. Patil, U., Nagtilak, S., Rai, S., Patwari, S., Agarwal, P., & Sharma, R. (2020). Sign language translator using deep learning. *EasyChair*, (2541).
  67. Rajan, R. G., & Leo, M. J. (2020, February). American sign language alphabets recognition using hand crafted and deep learning features. In *2020 International conference on inventive computation technologies (ICICT)* (pp. 430-434). IEEE.
  68. Hossen, M. A., Govindaiah, A., Sultana, S., & Bhuiyan, A.

- (2018, June). Bengali sign language recognition using deep convolutional neural network.
69. Hossain, S., Sarma, D., Mitra, T., Alam, M. N., Saha, I., & Johora, F. T. (2020, July). Bengali hand sign gestures recognition using convolutional neural network.
  70. (2020). "A Deep Learning Approach for Recognizing Bengali Character Sign Language".
  71. Hasan, M. M., Srizon, A. Y., & Hasan, M. A. M. (2020, June). Classification of Bengali sign language characters by applying a novel deep convolutional neural network. In *2020 IEEE Region 10 Symposium (TENSYMP)* (pp. 1303-1306). IEEE.
  72. Krishnan, T., Palani, Balasubramanian, Parvathavarthini. (2017). "Detection of Alphabets for Machine Translation of Sign Language using Deep Neural Net".
  73. Nguyen, H. B., & Do, H. N. (2019, April). Deep learning for american sign language fingerspelling recognition system.
  74. Mittal, A., Kumar, P., Roy, P. P., Balasubramanian, R., & Chaudhuri, B. B. (2019). A modified LSTM model for continuous sign language recognition using leap motion. *IEEE Sensors Journal*, 19(16), 7056-7063.
  75. Daroya, R., Peralta, D., & Naval, P. (2018, October). Alphabet sign language image classification using deep learning.
  76. Shahriar, S., Siddiquee, A., Islam, T., Ghosh, A., Chakraborty, R., Khan, A. I., ... & Fattah, S. A. (2018, October). Real-time american sign language recognition using skin segmentation and image category classification with convolutional neural network and deep learning.
  77. Tolentino, L. K. S., Juan, R. S., Thio-ac, A. C., Pamahoy, M. A. B., Forteza, J. R. R., & Garcia, X. J. O. (2019). Static sign language recognition using deep learning. *International Journal of Machine Learning and Computing*, 9(6), 821-827.
  78. Makarov, I., Veldyaykin, N., Chertkov, M., & Pokoev, A. (2019, July). Russian sign language dactyl recognition.
  79. Saleh, Y., & Issa, G. (2020). Arabic sign language recognition through deep neural networks fine-tuning." *International journal of online and biomedical engineering*, vol. 16, no. 5, pp. 71-83.
  80. Song, N., Yang, H., & Zhi, P. (2018, November). A deep learning based framework for converting sign language to emotional speech.
  81. Hassan, S. T., Abolarinwa, J. A., Alenoghena, C. O., Bala, S. A., David, M., & Enenche, P. (2018). Intelligent sign language recognition using image processing techniques: a case of Hausa Sign language. *ATBU Journal of Science, Technology and Education*, 6(2), 127-134.
  82. Parapat, Y. (2020). Deep convolutional neural network for hand sign language recognition using model E. *Bulletin of Electrical Engineering and Informatics*.
  83. Srivastava, S., Gangwa, A., Mishra, R., Singh, S. (2022). "Sign Language Recognition System using TensorFlow Object Detection API".
  84. Ahmed, S. T., & Akhand, M. A. H. (2016, December). Bangladeshi sign language recognition using fingertip position.
  85. Al-Hammadi, M., Muhammad, G., Abdul, W., Alsulaiman, M., Bencherif, M. A., & Mekhtiche, M. A. (2020). Hand gesture recognition for sign language using 3DCNN. *IEEE access*, 8, 79491-79509.
  86. Aly, S., Osman, B., Aly, W., & Saber, M. (2016, December). Arabic sign language fingerspelling recognition from depth and intensity images.
  87. Bird, J. J., Ekárt, A., & Faria, D. R. (2020). British sign language recognition via late fusion of computer vision and leap motion with transfer learning to american sign language. *Sensors*, 20(18), 5151.
  88. Cui, R., Liu, H., & Zhang, C. (2017). Recurrent convolutional neural networks for continuous sign language recognition by staged optimization. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7361-7369).
  89. Cui, R., Liu, H., & Zhang, C. (2019). A deep neural framework for continuous sign language recognition by iterative training. *IEEE Transactions on Multimedia*, 21(7), 1880-1891.
  90. Das, A., Gawde, S., Suratwala, K., & Kalbande, D. (2018, January). Sign language recognition using deep learning on custom processed static gesture images.
  91. Balayn, A., Brock, H., & Nakadai, K. (2018, August). Data-driven development of virtual sign language communication agents.
  92. Escobedo, E., Ramirez, L., & Camara, G. (2019, October). Dynamic sign language recognition based on convolutional neural networks and texture maps. In *2019 32nd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)* (pp. 265-272). IEEE.
  93. Odaty, L. K., Huang, Y., Asantewaa, E. E., & Agbedanu, P. R. (2019, August). Ghanaian sign language recognition using deep learning. In *Proceedings of the 2019 the International Conference on Pattern Recognition and Artificial Intelligence* (pp. 81-86).
  94. Gupta, D., Mohanty, J. P., Swain, A. K., & Mahapatra, K. (2019, December). AutoGstr: Relatively Accurate Sign Language Interpreter. In *2019 IEEE International Symposium on Smart Electronic Systems (iSES)(Formerly iNiS)* (pp. 322-323). IEEE.
  95. Huang, J., Zhou, W., Li, H., & Li, W. (2015, July). Sign language recognition using real-sense.
  96. Bachani, S., Dixit, S., Chadha, R., & Bagul, A. (2020). Sign language recognition using neural network. *International Research Journal of Engineering and Technology*, 7(4), 583-586.
  97. Koller, O., Ney, H., & Bowden, R. (2016). Deep hand: How to train a cnn on 1 million hand images when your data is continuous and weakly labelled. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3793-3802).
  98. Kurhekar, P., Phadtare, J., Sinha, S., & Shirsat, K. P. (2019, April). Real time sign language estimation system.
  99. Park, H., Lee, J. S., & Ko, J. (2020, January). Achieving real-time sign language translation using a smartphone's true depth images.
  100. Pigou, L., Van Herreweghe, M., & Dambre, J. (2017). Gesture and sign language recognition with temporal

- residual networks. In *Proceedings of the IEEE International Conference on Computer Vision Workshops* (pp. 3086-3093).
101. Rioux-Maldague, L., & Giguere, P. (2014, May). Sign language fingerspelling classification from depth and color images using a deep belief network. In *2014 Canadian conference on computer and robot vision* (pp. 92-97).
  102. Rodríguez, J., Chacón, J., Rangel, E., Guayacán, L., Hernández, C., Hernández, L., & Martínez, F. (2020). Understanding motion in sign language: A new structured translation dataset.
  103. Saggio, G., Cavallo, P., Ricci, M., Errico, V., Zea, J., & Benalcázar, M. E. (2020). Sign language recognition using wearable electronics: implementing k-nearest neighbors with dynamic time warping and convolutional neural network algorithms. *Sensors*, 20(14), 3879.
  104. Shinde, S. S., Autee, R. M., & Bhosale, V. K. (2016, December). Real time two way communication approach for hearing impaired and dumb person based on image processing.
  105. Kumar, S. S., Wangyal, T., Saboo, V., & Srinath, R. (2019). Time series neural networks for real time sign language translation. In *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)* (pp. 243-248). IEEE.
  106. Taskiran, M., Killioglu, M., & Kahraman, N. (2018, July). 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). IEEE.
  107. Thanasekhar, B., Kumar, G. D., Akshay, V., & Ashfaq, A. M. (2019, December). Real Time Conversion of Sign Language using Deep Learning for Programming Basics. In *2019 11th International Conference on Advanced Computing (ICoAC)* (pp. 1-6). IEEE.
  108. Tushar, A. K., Ashiquzzaman, A., & Islam, M. R. (2017, December). Faster convergence and reduction of overfitting in numerical hand sign recognition using DCNN.
  109. Yang, S., & Zhu, Q. (2017, May). Video-based Chinese sign language recognition using convolutional neural network. In *2017 IEEE 9th international conference on communication software and networks (ICCSN)* (pp. 929-934).
  110. Zamora-Mora, J., & Chacón-Rivas, M. (2019, October). Real-time hand detection using convolutional neural networks for costa rican sign language recognition. In *2019 International Conference on Inclusive Technologies and Education (CONTIE)* (pp. 180-1806).
  111. Suri, K., & Gupta, R. (2019). Convolutional neural network array for sign language recognition using wearable IMUs.
  112. Feyera, I., & Seid, H. (2021). Developing Amharic Sign Language Recognition Model for Amharic Characters Using Deep Learning Approach.
  113. Liu, Y., Jiang, D., Duan, H., Sun, Y., Li, G., Tao, B., ... & Chen, B. (2021). [Retracted] Dynamic Gesture Recognition Algorithm Based on 3D Convolutional Neural Network. *Computational intelligence and neuroscience*, 2021(1), 4828102.
  114. Costanza, M. I., Jonas, B. (2021). "Dynamic Gesture Recognition ETH Zurich," in *ICMI 2013 - Proceedings of the 2013 ACM International Conference on Multimodal Interaction*, pp. 445-452.
  115. Cheng, K. L., Yang, Z., Chen, Q., & Tai, Y. W. (2020, August). Fully convolutional networks for continuous sign language recognition.
  116. Chansri, C., & Srinonchat, J. (2016). Hand gesture recognition for Thai sign language in complex background using fusion of depth and color video. *Procedia Computer Science*, 86, 257-260.
  117. Tavari, N. V., & Deorankar, A. V. (2014). Indian sign language recognition based on histograms of oriented gradient.
  118. Mazhar, O., Ramdani, S., & Cherubini, A. (2021). A deep learning framework for recognizing both static and dynamic gestures. *Sensors*, 21(6), 2227.
  119. Negin, F. *et al.* (2017). "PRAXIS: Towards Automatic Cognitive Assessment Using Gesture Recognition".
  120. Seita, M. (2020, April). Designing automatic speech recognition technologies to improve accessibility for deaf and hard-of-hearing people in small group meetings. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1-8).
  121. Vaitkevičius, A., Taroza, M., Blažauskas, T., Damaševičius, R., Maskeliūnas, R., & Woźniak, M. (2019). Recognition of American sign language gestures in a virtual reality using leap motion. *Applied Sciences*, 9(3), 445.
  122. Schioppo, J., Meyer, Z., Fabiano, D., & Canavan, S. (2019, May). Sign language recognition: Learning american sign language in a virtual environment. In *Extended abstracts of the 2019 CHI conference on human factors in computing systems* (pp. 1-6).

**Copyright:** ©2025 Okorie Emmanuel O, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.