International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. 15, No.1 (Jan-24)

http://dx.doi.org/10.12785/ijcds/150116

A Detail Study of Sign Language Communication for Deaf-Mute to Normal Person

Chauhan Pareshbhai Mansangbhai[1] and Dr. Dineshkumar B. Vaghela[2]

[1]*Research Scholar,*
[1,2]*Gujarat Technological University, Ahmedabad, Gujarat, India*

Received 24 Jan. 2023, Revised 2 Jan. 2024, Accepted 6 Jan. 2024, Published 15 Jan. 2024

Abstract: Even though sign language recognition systems are crucial, there isn't a well-structured literature study to help Deaf-Mute. Approximately 160 research articles were found and reviewed in this survey of the sign recognition framework. From this pool, around 100 research articles were chosen, analyzed, and categorized. Each of these articles was grouped according to several criteria, such as datasets, kind of sign language, data collection methods, preprocessing, segmentation, and learning strategy. Notably, the largest study on sign language recognition utilized cameras and focused on steady, separated, and one-handed signs. Researchers are receiving increasing attention these days for creating cost-effective sign language recognition. According to the review, the main reason why commercialized sign language hasn't yet been developed as the current methods for identifying sign language either require expensive equipment or take too much computational time. The objective of this article is to provide an overview of recent developments in the field of computer vision, enabling other researchers to gain insights and potentially develop more efficient sign language systems.

Keywords: Hand-Sign, Sign Language Recognition, Deep Learning, Computer Vision

1. Introduction

Communication is a crucial part of sharing knowledge to live a joyful life. Ordinary people can freely and easily communicate with each other as they have a common communication language. But when ordinary people want to interact with a Deaf-Mute, the problem starts in communication due to a lack of shared language. Mostly, Deaf-Mute knows only Sign Language (SL) to interact with others. In the society of Deaf-Mute (D&M), SL is a basic necessity. D&M persons communicate with the help of gestures instead of oral language and audio patterns. The five primary elements of gesture language are hand appearance, angle, action, position, and non-manual gestures. To bridge the communication gap between normal individuals and deaf individuals, one possible solution is for the former to learn sign language. However, this may not always be feasible in practice. As per the report of WFD (World Federation of the Deaf), approximately seventy million people are Deaf-Mute who are currently facing communication problems [1]. Seventy million deaf people utilize more than three hundred sign languages globally, according to WFD [2].

The recognition of gestures holds the potential to alleviate social barriers faced by sign language users. Collaborative research across various fields, including template matching, machine learning, linguistics, language processing, and Sign Language Recognition (SLR), aims to develop techniques and algorithms for recognizing pre-defined gestures and their semantic meaning. Humanoid Machine Interaction (HMI) systems, which can recognize sign language, are designed to facilitate efficient and enjoyable communication. These systems employ a multidisciplinary approach that combines SL linguistics and gesture recognition technology. Implementing such techniques in public settings like hotels, trains, malls, businesses, and other venues can provide hearing-impaired individuals access to new ideas, information, and emotional support [3].

A translator that can translate sign language to voice or text enables the interaction between ordinary people and the Deaf-Mute. In this digital era, unlike a translator or device, a computer vision-based system can be a good alternative. It doesn't always make sense to carry a translator or gadget everywhere. Computer vision-based technique is essential because everyone should be given equal opportunities despite of where they are from or how they are physically. Numerous translators exist that can translate sign language into text or speech, but they have cost and computation restrictions. The main aim of this study is to provide an overview of recent developments in the field of sign language recognition. This study will help the researchers to gain more knowledge in this field. It may lead to improve the existing development in this field. The upcoming subsection is the research questions to direct the

E-mail address: pareshchauhan.ssec@gmail.com, dineshvaghela28@gmail.com

flow of this work in a structured manner.

A. Research Questions

Research Questions make it easier to figure out the depth of previous work. These questions help to direct the flow of review in the research. The below-listed questions also show the hierarchy of this research review. The answers to these questions commonly lead to achieving the objectives of findings. The following is a list of the research questions:
RQ1: How many papers have been published previously on Sign Language Recognition (SLR)?
RQ2: How many sign language recognition datasets are available?
RQ3: What are the available techniques for data acquisition in SLR?
RQ4: What are the available techniques for preprocessing in SLR?
RQ5: What are the available techniques for feature extraction in SLR?
RQ6: What are the available methods for classification in SLR?
RQ7: What is the overall accuracy of each existing SLR system?

The answers to the above questions guide this review's methodical and simplified format. The answers to the research questions provide the foundation for this paper's organizational structure. As this type of formation of hierarchical review is being conducted for the first time, it also differs and separates it from other SLR surveys. Table I and Table II shows the publisher-wise and year-wise summary table, respectively.

TABLE I. Referred papers (publisher-wise)

Sr. No.	Journals/Conferences Name	Number of Papers Referred
1	IEEE	31
2	Elsevier	13
3	ACM	07
4	Springer	18
5	Willey	02
6	Other	31

TABLE II. Referred papers (year-wise)

Sr. No.	Year	Number of Papers Referred
1	2017	10
2	2018	04
3	2019	20
4	2020	30
5	2021	27
6	2022	11

2. Different Sign Languages

Every location and nation has its unique sign language. Different sign languages can exist in various states or regions within the same nation. Around the world, there are over 300 sign languages. Among these, the mentioned sign languages are now receiving greater attention from research scholars.

A. American Sign Language

French SL gave rise to American SL (ASL), widely used in the US, French, and Canada. ASL users are in numbers between 2,50,000 and 5,000,000 [4]. West and Central Africa also utilize the ASL dialects. ASL signs include extensive phonemic elements like facial features and hand gestures. ASL has seen significant improvement in the modern period.

44% of the study on ASL reported cameras as data gathering, while 23% has been done using Kinect, 13% has been done using an armband, 8% has been done using gloves, Leap movements, electroencephalograms, and impulsive radio devices have all been used to measure 12%. Neural network (33%) has been used the most in ASL research. The researchers used SVM (21%) as the second number in ASL. After that hybrid approaches (21%) and CNN (13%), with AdaBoost, KNN, and DTW techniques receiving the least significant attention. 65% of SLR systems claimed accuracy levels above 90% on average. On the other hand, 23% of the approach's accuracy rates were in the 80%–89% range. Only 12% of techniques have a recognition rate below 80% [5].

B. Chinese Sign Language

In some regions of China, people communicate using a unique language called Chinese SL (CSL). Taiwanese and Malaysian speakers both speak it. Estimates for CSL users range from one million to twenty million [6].

64% of the CSL study was carried out using Kinect, the remaining 22% with cameras, 7% with hand gloves, and the final 7% with armbands. HMM and hybrid approaches both performed for 36% of the effort on CSL, with SVM contributing the least amount of work at 28%. Only 43% of gesture recognition frameworks have an overall recognition rate above 90%, while 50% of techniques have a recognition rate from 80% to 89%. Only 7% of frameworks have accuracy below 80% [5].

C. Indian Sign Language

In South Asia, Indian SL (ISL) is the most widely used common sign language. There are 2,700,000 ISL users worldwide, according to Lewis et al. [7]. ISL has three dialects: Mumbai-Delhi SL, Punjab-Sindh SL, and Bangalore-Chennai-Hyderabad SL.

Observations show that cameras have been used in 68% of ISL research projects, followed by Kinect, hand glove, and jump motion in 10%, 11%, and 11% of projects, respectively. Neural network (32%) and SVM (26%) have

been used the most in Indian SL research, with hybrid methods making up 16% of the total. KNN and DTW have had the least amount of work done on them. 68% of SL recognition frameworks for ISL have an overall recognition rate above 90%, according to research, whereas 24% of the systems are between 80% and 89% accurate. Frameworks with accuracy lower than 80% are accurate in just 8% of cases [5].

D. Arabic Sign Language

The Middle East and North Africa both use Arabic SL or ArSL. Arabs who are Deaf-Mute communicate their thoughts visually using ArSL. It is formed through the face, body, and hand motions. It is hard to find more details about ArSL.

A camera was used in 54% of the study on ArSL, followed by a Kinect, hand glove, and jump motion system in 13% of the study, and a Polhemus tracker in 7% of the study. Almost half (47%) of the research on ArSL has been completed using HMM, followed by hybrid approaches (20%), with Neural Network (NN) and SVM receiving the least amount of attention. Almost 70% of SL recognition frameworks have an overall recognition rate above 90%, whereas just 23% have an accuracy between 80% and 90%. There are only 7% of systems with accuracy lower than 80% [5].

E. Persian Sign Language

Iranians who are hard of hearing communicate in Persian Sign Language. There are around 1,50,000 Parsian SL users worldwide, according to the study [4]. Due to Iran's diverse geography, various forms of Persian SL are spoken around the country. Like ArSL, Persian SL has received less study.

The Persian SL review observes that cameras have been used as the acquisition instrument for all Persian SL studies. NN and hybrid approaches have been used to complete 80% of research on Persian SL, with HMM receiving the least attention (20%). Only 10% of systems had accurateness below 80%, while 90% of SL recognition frameworks recognized above 90% on average [5].

F. Brazilian Sign Language

Deaf persons in the area of Brazil use Brazilian SL. It's also referred to as Libras informally. Brazilian SL speakers are reported to number 3,000,000 [7]. Like ArSL, Brazilian SL has received less study.

The use of cameras (67%) and armbands (33%) have dominated research on Brazilian SL. NN has been used for 67% of the work on the Brazilian SL. The researchers used SVM for the remaining 33%. Only 17% of sign language recognition systems have an accuracy below 80%, while 83% of techniques have an average accuracy of better than 90% [5].

G. Greek Sign Language

Greece's official language is Greek, according to the law. There might be between 6000 and 60,000 speakers, according to estimates [8]. Greek SL has not received much attention due to fewer users.

A camera and an armband were used for most of the study on Greek SL at 34% and 33%, respectively. Jump motion (33%) was also used in this research. Distance metric has been used in 34% of the study with Greek SL, followed by SVM and HMM at 33% and 34%, respectively. Only 33% of systems have an accuracy between 80% and 89%, compared to 67% of systems that have an overall accuracy of higher than 90% [5].

H. Irish Sign Language

Irish SL is spoken in the Republic of Ireland and Northern Ireland, derived from French. There are 5000 deaf persons in Ireland [8]. Irish SL has not received much attention due to fewer users.

In all of the studies on Irish SL, cameras were used as the data-collecting tool. HMM has been used for 67% of the investigation with Irish SL, then with SVM, by 33%. The average accuracy of SL recognition frameworks is 100%, which is more than 90% [5].

I. Malaysian Sign Language

Malaysian SL is the primary gesture language used by the Deaf-Mute population in Malaysia. Like Greek SL, Malaysian SL has not received much attention.

In Malaysian SL, cameras have been used for the majority of research work (67%), followed by Kinect (33%). On Malaysian SL, Only NN has been used for all the research. Only 33% of systems have an accuracy between 80% and 89%, compared to 67% of systems that recognize signs with an overall accuracy of more than 90% [5].

J. Mexican Sign Language

Deaf individuals in Mexico's metropolitan areas communicate using Mexican SL. It is similar to French SL and has an estimated 130,000 native speakers [8].

In Mexican virtual worlds, Kinect has been used most frequently (67%), followed by cameras (33%). NN has been used to perform 67% of the development on Mexican SL, with DTW accounting for the remaining 33%. Approximately 33% of sign language recognition systems have accuracy levels below 80%, compared to 67% have average accuracy levels of better than 90% [5].

K. Taiwanese Sign Language

Taiwanese SL (TSL) is the name of the dialect used by deaf people in Taiwan. TSL originated from Japanese SL. Its user base is estimated to be over 20,000 [9]. Taiwanese SL has not received much attention due to fewer users.

The cameras have been used as an acquisition instrument for all of the research in this SL. NN has been used in

34% of the research on TSL, followed by SVM and hybrid approaches in 33% of the work, respectively. Around 66% of sign language recognition systems have accuracy levels between 80% and 89%, compared to 34% who have an overall accuracy of better than 90% [5].

L. Thai Sign Language

Thai is the deaf community's official gesture language. ASL and it both belong to the same linguistic family. According to estimates, 56,000 people in Thailand who are Deaf-Mute (20%) use Thai SL [9]. Thai SL has not received much attention due to fewer users.

Thai SL has been performed 67% with the cameras and 33% with gloves. NN has been used in 67% of the study with Thai gesture language, whereas SVM was used in 33% of the research work. 100% of SL recognition frameworks obtained an overall recognition rate of at least 90% [5].

By examining the variety of sign languages, ASL, CSL, and ISL are the sign languages that, in comparison to other SL, have the most direct impact on the huge Deaf-Mute community. The population of the individual nation is one of the main reasons. The above listed are the other gesture languages on which the very least amount of research work is done.

3. Sign Language Datasets

There are two widely used standard data sets for American Sign Language (ASL) at the letter level: ASL Finger-Spelling A and ASL FingerSpelling B [10]. Both datasets have 24 class labels as alphabets. Bostan ASLLVD [11] and MSR Gesture 3D [12] are two well-known standard datasets for ASL at the word level. There are over 3300 ASL signs in the ASLLVD dataset. The videos in this collection, which are all in sync, show the gestures from various angles. This dataset also contains the annotation of hand. Another word-level standard dataset of ASL is MSR Gesture 3D which includes 12 dynamic gesture signs. It has more information in the form of depth. It is publicly available for experiments.

In Chinese Sign Language (CSL), three well-known datasets are available: DEVISIGN-G, DEVISIGN-D, and DEVISIGN-L [13]. There are a total of 36 class labels in DEVISIGN-G. There are 26 letters and 10 numbers in this collection. The other two datasets contain daily used vocabulary words. DEVISIGN-D includes 500 daily used vocabulary (including the words in DEVISIGN-G). With 2000 Chinese SL vocabulary (including all the words in DEVISIGN-D), DEVISIGN-L is a large vocabulary collection. The sentence-level dataset of CSL is isoGD [14]. This dataset includes videos of gestures that have RGB and depth. The collection contains 22535 RGB-Depth videos with a total of 47933 gestures. There are 249 labels. For Indian Sign Language (ISL), the standard dataset for a word level is ISL-Emergency [15]. It contains eight emergency word gestures with a total sample of 416.

One of the most standard datasets for German Sign Language is RWTH-PHOENIX-Weather [16]. It is a big vocabulary collection of German Sign Language based on videos that may be used for statistical SL translation and recognition. The RWTH-PHOENIX-Weather database includes 600k frames, 45760 running glosses, 5356 phrases, and 1200 sign vocabulary. It has an annotation of hand and face. It is a publicly available dataset. The work of making the annotations better is still ongoing. Another standard dataset for German SL is SIGNUM [17]. It includes both isolated and continuous signing of several signers. There are 450 basic gestures in it. A total of 780 sentences were created using this set of words. The length of each phrase varies from two to eleven signs. Table III shows the well-known datasets used for different sign languages.

For the SL recognition framework field, this study includes the most relevant datasets, including videos, in Table III. Table III shows eight fields for each dataset. Those fields are the prime component to analyze the dataset. The contexts, characteristics, limitations, and complexity of these datasets vary. Although the gesture language recognition uses multiple languages as input information in frameworks, American SL (ASL) has received greater interest because of its increased use with popularity. The other language datasets, including Argentina, China, Germany, Greece, Poland, Turkey, India, and Netherlands are also used in many research studies today. As shown in Table III, it is preferable to use more gesture categories to build a generalized approach to provide solutions in the actual world. The majority of these data sources are for gesture categorization rather than detection, which is another crucial point to keep in mind. A few abbreviations for Table III are as follows: U-USA, G-Germany, Gr-Greek, P-Poland, C-China, Ity-Italy, Arg-Argentina, K-Korea, Ind-India, Irn-Iran.

4. Phase-wise Survey of Sign Language Processing

The traditional classification stages of image or video processing are as follows: A) Data Acquisition, B) Pre-processing, C) Segmentation, D) Feature Extraction, and E) Classification Method. As per the traditional image processing flow, it would be helpful for the researcher in this field to review each phase of the processing flow. The following subsections show the detailed survey for each phase of the general framework of sign language. At the end of this section, a detailed summary of the literature has been discussed.

A. Data Acquisition

Problems with classification are within the supervised learning category. For training and testing purposes, getting data along with its label is essential. Data collection must be consistent because the majority of classifiers demand a particular form and structure for the data. There are several techniques to get image data. It may be obtained manually or through online resources made available by someone. Table IV shows the study and comparison of a few typical methods for collecting data for SLR.

TABLE III. Sign language datasets

Year	Dataset Name	Country	Subject	Language Level	Class Number	Sample Number	Annotation
2011	Boston-ASL-LVD [11]	U.	Six	Word	3,300	9,800	Hand
2011	ASL-FingerSpelling-A [10]	U.	Five	Alphabets	24	1,31,000	***
2011	ASL-FingerSpelling-B [10]	U.	Nine	Alphabets	24	***	***
2012	DGS-Kinect-40 [18]	G.	Fifteen	Word	40	3,000	***
2012	RWTH-PHOENIX-Weather [16]	G.	Nine	Sentence	1,200	45,760	Face, Hand
2012	GSL-20 [19]	G.	Six	Word	20	840	***
2012	MSR-Gesture3D [12]	U.	Ten	Word	12	336	***
2013	PSL Kinect-30 [20]	P.	One	Word	30	300	***
2014	DEVISIGN-G [13]	C.	Eight	Word	36	432	***
2014	DEVISIGN-D [13]	C.	Eight	Word	500	6,000	***
2014	DEVISIGN-L [13]	C.	Eight	Word	2,000	24,000	***
2014	ChaLearn (Track 3) [21]	Ity.	***	Word	20	***	***
2014	CLAP14 [22]	Ity.	Twenty-Seven	Word	20	6,600	Hand
2015	SIGNUM [17]	G.	Twenty-Five	Sentence	450	33,210	***
2015	PSL-Fingerspelling-ToF [23]	P.	Three	Alphabets	16	960	***
2016	MSR [24]	U.	Ten	Word	12	336	***
2016	LSA64 [25]	Arg.	Ten	Word	64	3,200	Hand, Head
2016	TVC-hand-gesture [26]	K.	One	***	10	650	***
2016	LSA16-handshapes [27]	Arg.	Ten	***	16	800	***
2017	isoGD [14]	C.	Twenty-One	Sentence	249	47,933	Hand
2018	PHOENIX14T [28]	G.	Nine	Sentence	1,066	67,781	***
2018	KETI [29]	K.	Ten	Word, Sentences	4,19,105	14,672	Hand
2019	CMU [30]	G.	Ten	Word	70	700	***
2019	UTD-MHAD [30]	U.	Eight	Word	27	800	***
2020	ISL-Emergency [15]	Ind.	Twenty-Six	Word	08	416	***
2020	RKS-PERSIAN-SIGN [31]	Irn.	Ten	Word	100	10,000	Hand

The widely used technique is to use a normal camera to capture a two-dimensional video [32][33]. Since more individuals now own smartphones with integrated cameras, this approach is used more than the others. It is preferable to preprocess images first because the camera's quality is camera-dependent. A low-quality camera, for instance, can make it difficult to extract features from the acquired image because of excessive noise. Furthermore, since each camera's resolution differs, it is typically necessary to do preprocessing to make it consistent with the classifier.

Kinect is used in some studies. Kinect contains the sensors and a camera to take colorful photos and record object depth. These provide more detailed features which may assist in identification. With the discrimination between items in the foreground and background, the feature depth can help in segmentation. WiGest 11 is a novel technique designed by Abdelnasser et al. that used Wi-Fi signal strength to find floating hand gestures around the setup kit [41]. The variation in the signal helps to identify the input. This method is helpful because the user does not require anything to carry. It is multi-directional, but putting it into practice is challenging.

Additionally, information can be retrieved from ready-to-use data sources. Web datasets are developed priorly and made available online in huge quantities with high quality. Jalal et al. performed experiments on the Kaggle dataset that includes roughly 27,000 gesture images [43]. Researchers can concentrate on additional SLR elements if they use the

TABLE IV. Data acquisition

Source of Data	Authors	Pros.	Cons.
Standard Camera	[32] [33] [34] [35] [36] [37]	-Easy to use. -Easily understandable. -2D image data.	-Noise (Outlier).
Kinect	[38] [39] [40]	-Provides data free of errors. -Depth adds extra informational detail. -Easy to segment.	-Expensive to acquire. -Easily damaged by outside infrared supply.
WiGest	[41]	-Provides high-directional data.	-Execution is complex.
Online Web Dataset	[42] [43] [44]	-Lots of fully prepared components. -No input devices are needed. -Makes time and resource savings.	***
Data-Augmentation	[38] [45]	-Provides a large amount of data. -Prevents over-fitting.	***

TABLE V. Data preprocessing

Method	Authors	Pros.	Cons.
Gaussian Filter	[34] [46]	-Reduce noise. -Image smoothing.	-Omit some details.
Median Filter	[40] [47]	-Reduce background noise. -Preserve accurate data.	-Too simple. -Only good for a limited type of noise.
Image Cropping	[40] [44]	-Helps to provide consistent input.	-Complex implementation.
Bootstrapping	[48]	-Mostly preserves the useful information. -Simple kind of implementation.	-Time consuming.

currently available high-resolution datasets.

Data augmentation can obtain more details through insufficient data through augmenting it [38][45]. A piece of the current set-of information is enhanced to provide new, distinct data. It may avoid the classifier overfitting in some sign languages as it gives more stable data. It also helps to save time and increase the classifier's accuracy. Hand images acquired from different angles and distances in SLR, turn the results into invariant in size and rotation. A typical data augmentation example is rotation, where the image is rotated at a particular angle.

B. *Preprocessing*

Preprocessing is the stage immediately following data acquisition. Additionally, it provides better results for classification after removing any noisy data that may have been there. Table V compares a few of the typical preprocessing methods. Two well-famous techniques for reducing undesirable noise in 2D images to improve curve recognition in the process of segmentation are Gaussian filters [34][46] and median filters [40][47]. To save computation time and provide a uniform data format for the classifier, it is also possible to crop or scale the image size. A 28x28 image converted from an HD image might serve as an example.

C. *Segmentation*

The hand sign is the most crucial component of SLR. It is recognized through the motions. So, image/video segmentation is generally used to eliminate extra information like the background and other objects. Then it will interact with the classification model. The classifier will only consider the interested region (ROI) by reducing the area of data. At this point, the researcher must determine aspects of the image portion the classifier needs. Table VI provides an analysis of the segmentation techniques used in the research.

Gray scaling is one segmentation technique [32][33][37]. It converts a colored or RGB photo into a gray-scale one. Generally, studies frequently use gray scaling before continuing because gesture language does not consider people's skin tone. As a result, it forced a classifier to ignore color. The crucial aspect of gray scaling is to make it easy to separate the foreground from the background. The thresholding approach is the most often used segmentation technique [37][49][47][40][46][50]. It converts an image to binary form. It is used after the image is already gray-scaled. Gray scale conversion makes a photo black and white. Therefore, the thresholding technique is used to separate the back-end from the front-end, with black standing for the background and white for the foreground. The researcher will select a

TABLE VI. Segmentation

Method	Authors	Pros.	Cons.
Gray Scaling	[32] [33] [37]	-Less calculation. -Simple steps for implementation.	-Not much effective.
Thresholding	[40] [37] [49] [47] [46] [50]	-Less calculation. -Fast performance.	***
Otsu Algorithm	[32]	-Find automatic threshold value.	***
Morphological Filter	[32] [34] [49] [46]	-Good for finding the Region of Interest (ROI).	-Features might be less precise.
Canny Edge Detection	[32] [33] [41] [49]	-Impressive against the diverse and noisy environment.	-Extreme calculation. -Execution takes a lot of time.
Seeded-Region-Growing	[33]	-Quickly segment images. -Strong and efficient for expanding ROI.	-Issues with noise.
Sobel Edge	[34]	-Simple steps for implementation.	-Increase noise in information.
Skin Segmentation	[36] [49] [51] [50]	-Simplicity in the implementation.	-Issues with illumination.
Viola And Jones	[49]	-Do segmentation of facial components effectively.	-Vulnerable to the brightness and invariant rotation.
Background Subtraction	[49] [46]	-Low calculation.	-Depend on object's moving speed and frame rate. -Issues with illumination.

threshold value, and based on that value, the back-end and front-end colors can be identified.

Joshi et al. [32] found the Otsu method that automatically chooses the appropriate threshold value. The researcher needs not to identify the suitable threshold value [32]. Another approach is skin segmentation [36][49][51][50]. The popularity of this approach is due to its ease of usage. It uses a human skin color histogram or predefined color range to ensure that the image only shows the specified color. By doing this, any extra information, such as background or objects, won't be transmitted to the classifier. First, the researcher must choose the color space. The drawback is that other body parts or the face will also be recognized if they have a color similar to the skin sometimes. The segmentation of skin is light-sensitive.

Furthermore, the binary image often undergoes a morphological filter or operation [32][34][49][46]. By extending the Region of Interest (ROI) to match the image, morphological filters efficiently reduce errors from the foreground or background. Additionally, a canny edge identification approach can retrieve the boundary of the object in the photo [32][33][41][49]. This technique works well for locating the ROI and mapping out the hand's boundaries. The authors [49][46] also used Background Subtraction for the research. Steady objects are identified and subtracted from the input video. It can be used for a quick segmentation procedure since the hand movement will not be considered as background. However, it also sees the foreground in a video to find steady objects.

D. Feature Extraction

Feature extraction is the process of taking useful information from the data and enhancing it. It starts to gather attributes for the classifier after removing unnecessary information from the ROI. SLR properties may alter based on the researcher's thinking to gesture recognition. For example, Kumar et al. [48] used the location and orientation of the fingertips whereas Ahmed et al. [50] used the center of mass to determine the motions.

Table VII consists of a list of the feature extraction techniques used in the SLR research. The Discrete Cosine Transform (DCT) summarizes sinusoids with different amplitudes and frequencies to describe the data. The 2D Discrete Cosine Transform (DCT) is calculated of the data through its dct2 function. The DCT has the feature that, for a typical image, the majority of the image's visually important information is contained in just a few of the DCT coefficients [36]. A few researchers applied the DCT for image compression. Because the energy of Discrete Wavelet Transformation (DWT) focuses more on the time domain and still it has wave-like (periodic) qualities, wavelets enable simultaneous time and frequency analysis of data. Abdelnasser et al. [41] used the characteristics of DWT to extract the feature in their research work.

Principal Component Analysis (PCA) is a different technique that can reduce the information and make the attributes non-dependent. Rao et al. [34] used PCA to improve classification performance and minimize overfitting. Before using PCA, DCT is used to compress the informa-

TABLE VII. Feature extraction

Method	Authors	Pros.	Cons.
Discrete Cosine Transform	[34]	-Quick calculation.	-Additional memory required.
Principle Component Analysis	[34]	-Compress data. -Prevent over-fitting.	-Require data standardization.
Speeded Up Robust Features (SURF)	[33]	-Effective against invariant data. -Faster and more efficient than SIFT.	-Sometimes lead to false matching.
PCA-Net	[39]	-Effective computation.	-Occupy large space.
Discrete Wavelet Transform	[41]	-Simple for filtering noise.	-Complex implementation.

TABLE VIII. Classification method

Method	Authors	Pros.	Cons.
Cross-Correlation Coefficient	[32]	-Minimum calculation.	-Too basic. -Fails to learn.
SVM	[33] [39]	-Memory effective. -Efficient for multi-classification.	-Sensitive to noise.
ANN	[37] [52]	-Learning ability. -Resilient to faults and intelligent.	-Convergence speed is low.
HMM	[44] [38] [48]	-Good learning algorithm.	-Additional calculation. -Large training samples required.
CNN	[52] [43] [42] [47] [51] [46] [40]	-Highly precise for classification. -Effective even without pre-processing or segmentation.	-Additional calculation. -Require strong hardware.
CNN With Transfer Learning	[45] [35] [36] [44]	-Exceptional classification accuracy. -Pre-trained. -Reduces time.	-High calculation. -Require strong device. -Preprocessing must be compatible with the system.

tion. PCANet [39], a PCA network, is another option for feature extraction that is computationally effective. Jin et al. [33] use SURF or Speeded Up Robust Features. The approach was built around the SIFT (Scale Invariant Feature Transform). SURF (Speeded Up Robust Features) is more computationally efficient than SIFT and, like SIFT, can find an interesting point in a dataset by locating nearby features. For discovering features, this approach is invariant to scale, rotation, occlusion, and variance. Because the classifier can accommodate various sign rotations, this approach is effective for classifying images.

E. Classification Method

Table VIII displays the various classifying techniques used in various research. The convolution layers of Convolutional Neural Network (CNN), such as those described in the papers by Jalal et al. [43], Huang et al. [53], and other authors, are the most often used method for feature extraction. The convolution layer can extract the key details of the image. The Max-Pooling layer uses these features. It also improves it to speed up the computation. CNN is the most widely utilized source because of its consistency. In general, CNN is well-known for the extraction of features automatically by setting a different number of kernels. As per the architecture, it includes both the feature extraction and learning parts as a single unit. The study of Cayamcela et al. [35], in which the mentioned recognition rate is more than 99%, is an example of CNN. CNN uses additional convolutional layers to extract features not unlike Artificial Neural Network (ANN). Following that, it establishes the node where the data belongs and continues to learn based on the training results. With more layers, accuracy rises, but so does the cost of computing. CNN requires a lot of experimentation to get the layers perfect.

Some people used CNN with Transfer Learning. They had taken care of training and prediction components and ignored the layer concerns. The study of Cayamcela et al. [35] used AlexNet [36], which performs well at identifying hand motions. The pre-trained architecture of CNN reduces the training time precondition to fit the input data in the available architecture. Human hand detection, face detection, body detection, etc. pre-trained architectures are available nowadays. So, as per the suitable input, it can be used directly to process further to recognize the gestures.

Support Vector Machine (SVM) [39], a technique widely used to resolve margin maximization and optimization issues, is used by Jin et al. [33]. The result generates a selection boundary, which can be very helpful to increase the use of linear classifiers. It also outperforms compared to nonlinear and linear classifiers. For multi-class recognition, it can perform well with the limitation of the nature of the dataset and the availability of training/testing samples.

The Hidden Markov Model (HMM) is another approach that maximizes posterior probability while minimizing prior probability error. GMM-HMM, an HMM based on the Gaussian Mixture Model (GMM), is used by Guo et al. [38]. It performs well when there is sequential data. Image/video frames can be treated as a piece of sequential information in many of the techniques. Their analysis demonstrates that the model can increase accuracy compared to conventional ones. In the study of Joshi et al. [32], they used the cross-correlation coefficient. By utilizing several time-shifted functions, this approach compares two signals for similarity. In sign language, dynamic motions are crucial. This technique works well for them. It is not performance-intensive. So, researchers may implement it on standard hardware. A Minimum Distance Classifier (MDC) is also present for classification. Rao et al. [34] combined the MDC and Mahalanobis distance for effective classification. Although it struggles with handling hands of various sizes or forms, it is the point-to-point matching in the features. It can be used on a variety of devices because it is the least expensive method compared to the others.

F. Combined Summary

This section provides a summary of the approaches taken and the findings from earlier SLR research. The five steps of gesture recognition frameworks are 1) Collection of Data, 2) Preprocessing, 3) Segmentation, 4) Extraction of Features, and 5) Categorization. Every step is crucial to the effectiveness of succeeding stages. The five phases with the recognition rate from the earlier gesture recognition frameworks are compiled in Table IX. Even though each study may employ various versions of sign identification, such as steady or dynamic signs, this study combines them all in a single table and also focuses on the methodologies.

The most common way to collect data is by utilizing a camera. As per Table IV, it has been found that the camera is the most frequently used technique due to its accessibility and low price. Other researchers did experiments using Kinect or pre-existing datasets from various origins to gather extra information like depth to process and produce good results. Some studies use preprocessing techniques like cropping, gray-scaling, and Gaussian smoothing. It has been observed that researchers widely used gray-scale and median filtering as preprocessing. Most researchers used canny edge detection, skin color, and thresholding method to perform the segmentation in their research. The results of each study were satisfactory. Convolution Neural Network (CNN) [43][35][36][53][47][40][46][45] is the most often used feature extraction and classification technique among the variety of various research chosen in Table IX. Other researchers also used ANN [37], HMM [36], and other classification methods. Some researchers, in addition to CNN, also use SURF [33], Principal Component Analysis [39][34], etc., for feature extraction. One of the researchers also suggested the combination of SURF as feature extraction and SVM as a classification to recognize signs with a significant accuracy level [33]. Overall, it comes to know that with the help of CNN, the recognition of static or steady gestures is almost above 95%. Table IX includes other significant combinations of data acquisition, preprocessing, segmentation, and feature extraction followed by classification. In the entire process of the SLR system, a few researchers didn't mention the technique in the attribute of either preprocessing or segmentation, or both. For such a field, the symbol *** is written in Table IX.

By analyzing the combined summary table with accuracy and other attributes, the initial phase of identifying alphabets, numbers (finger-spelling), or specifically static gestures in large datasets for SL has shown that machine learning approaches can do well. Additionally, deep learning techniques with a hybrid approach are also one dimension that can be explored to recognize dynamic gestures. It can be recognized at a higher-level language to recognize continuous gesture recognition to word or sentence-level. The upcoming section shows a comparative summary to recognize higher-level sign language using hybrid techniques.

5. Why Deep Learning?

This section presents a quick overview of Deep Learning (DL). It has recently performed better than the framework based on machine learning techniques in various areas, including computer vision and Natural Language Processing (NLP) [54]. CNN with automated extracted features can handle large image datasets [55][56]. In addition, CNN with automatic feature extraction can perform well to process the video to identify action with a suitable number of layers [57]. CNN [58], Generative Adversarial Network (GAN) [59], Auto Encoder (AE) [60], Variational Auto Encoder (VAE) [61], Deep Belief Network (DBN) [62], and Deep Boltzmann Machine (DBM) [63] are few notable DL methods which can be applied in the area of computer vision. Avoiding the need to create or extract features is one of the prime objectives of deep learning models with the limitation to computation power with powerful hardware. With the help of DL, a calculative approach with multiple executive layers can study from input and store it at various levels of abstraction, simulating the operations of the human mind and indirectly maintaining its patterns and structures.

McCulloch and Pitts (1943) made the first attempt to simulate a human brain to know the functionality of the mind and how it creates extremely complicated formations by core cells linked together or neurons. The pattern of significant role persisted, and Hinton et al. [62] proposed

TABLE IX. Combined summary

Author	Data-Acquisition	Preprocessing	Segmentation	Feature-Extraction	Classifier	Accuracy (%)
[32]	Normal-Camera	Gray-scale, Morphological Operators	Otsu	Boundary-Identification	Cross-Correlation Coefficient	94
[33]	Normal-Camera	Gray-scale	Canny-Edge	SURF	K-Means, Bag-of-Features, SVM	97.13
[34]	Normal-Camera	Kernel-Gaussian	Sobel-Edge, Morphological-Gradient	DCT, PCA	Minimum-Distance	90.58
[43]	Kaggle-ASL	***	***	CNN	CNN	99
[35]	Normal-Camera	***	***	CNN	CNN-Transfer-Learning	99.39
[36]	Normal-Camera	***	Skin-Color	CNN	CNN-Transfer-Learning	94.70
[37]	Normal-Camera	Gray-scale	Threshold	Contour-Identification	ANN	95
[39]	Kinect	***	***	PCA-Net	SVM	88.70
[41]	WiGest	***	Special-Preamble	DCT, Edge-Detection	String-Matching	96
[49]	Normal-Camera	***	Background-Exclusion, Threshold, Canny-Edge, Skin-Color	Contour-Identification	Principle-of-Heuristics	88.66
[53]	Normal-Camera	***	***	CNN	CNN	82.70
[48]	Normal-Camera	Resampling-Bootstrap	***	Fingertip-Direction	HMM-Coupled	90.80
[47]	Normal-Camera	Filter-Median	Threshold	CNN	CNN	96.20
[52]	ISL	***	***	***	ANN, GA, PSO	99.96
[43]	Massey-University-Dataset-216, Normal-Camera	***	Skin-Color, Convex-Hull	CNN	CNN	98.05
[40]	Kinect	Cropping-Image, Filter-Median	Threshold	CNN	CNN	91.70
[46]	Normal-Camera	Kernel-Gaussian	Background-Exclusion, Threshold, Morphological-Operator	Contour-Identification, CNN	CNN	99.80
[50]	Normal-Camera	***	Skin-Color, Threshold	Center-Mass	DTW	90
[45]	ASL, Data-Augmentation	***	***	CNN	CNN-Transfer-Learning	98

DBN as one of the notable deep learning approaches. The wide range of techniques used in deep learning includes NN, hierarchical probabilistic frameworks, classification, and clustering algorithms. The power of parallel GPU processing and the availability of large, high-resolution, and fully accessible annotated datasets are two crucial factors that have aided in the rapid development of deep learning. The expansion of some powerful frameworks like TensorFlow [64], Theano [65], and MXNET [66], as well as the reduction of a vanishing gradient and the proposal of several novel regularization methods, including batch adjustment, dropout, and data augmentation, all played a crucial part in the advancement to depth kind of learning. The focus of this review is on DL-based models for gesture recognition in the area of computer vision.

A. Hybrid Models Using Deep Learning Methods

The mixture of depth-based approaches and traditional classifiers, this survey highlights current results in gesture identification systems and relevant fields in this subsection. The details of these models are shown in Table X.

Although the researcher Cheron et al. [104] combined conventional descriptors and classifiers for learning, these fields are outside the scope of this review. Rastgoo et al. [73] developed a depth-based technique for hand gesture identification with a Single Shot Detector (SSD), CNN, and Long Short Term Memory (LSTM) by taking advantage of hand posture data when recognizing hand gestures using an RGB-based mixture method. They improved the SSD model's hand detection rate on five online sign dictionaries. Additionally, they integrated the retrieved attributes from the CNN model with other handcrafted features. Koller et al. [44] used CNN nested within an HMM to recognize continuous sign language. They constructed the model from beginning to finish as a hybrid CNN-HMM approach. In contrast to advanced models, participated in a study on the RWTHPHOENIX-Weather-2014, Multi-Signer dataset revealed that this framework decreased the error margins on development and testing from 51.6/50.2 to 38.3/38.8 [44].

Chen et al. [77] suggested the hybrid architecture for hand sign identification using CNN with automated keypoint detection in combination with SVM for final prediction using raw EMG image input in RGB-based hybrid gesture recognition. Results from experiments using their dataset showed that the accuracy was better by the margins of error of 2.5% and 9.7% when compared to situations when just CNN or conventional approaches were used [77]. In addition, the other way of conversation from text to sign language is also the research area in which hybrid models with CNN can be effective [105].

Cardenas and Chavez [75] suggested the hybrid architecture for hand sign identification in multi-modal hybrid gesture recognition by combining CNN with a Histogram Cumulative Magnitude (HCM). Their three input modalities included RGB, skeleton, and depth. Two techniques, skeleton estimation and sampling, were applied to the input video to include a predetermined number of keyframes. They combined the acquired spatiotemporal characteristics and put them into SVM for classification for the final decision. This model's efficiency was proven on UTD-MHAD, ChaLearn-LAP-isoGD, and UFOP-LIBRAS datasets, which showed an accuracy of 94.81%, 67.36%, and 64.33%, respectively. On the isoGD and UFOP-LIBRAS datasets, this model performed similarly to state-of-the-art (latest) approaches, but to the UTD-MHAD dataset, it exceeded the most recent method with a relative improvement of 0.16% [75]. Ma et al. [76] applied a CNN-based model to recognize hand-sign from RGB with depth data, two different modalities. With the help of segmentation technique based on depth information, the sign area was retrieved. The SVM approach is used for final recognition and follows feature extraction using a CNN. The suggested model, which had the benefit of a mixture of CNN with SVM, attained a recognition rate above 95%, according to experimental results on their dataset [76].

In depth-based mixture arm pose estimation, Chen et al. [74] suggested a visual type methodology for 3-Dimension arm pose prediction by integrating deep CNN and a Spherical Part Model (SPM). In this approach, precise hand position detection by depth maps was done through historical knowledge of the human hand. Utilizing spherical description and the hand-centric coordinate method, SPM was used to derive skeletal structures starting with the steadiest joints. Results using datasets from NYU and NTU showed improvements in the marginal derivation of 0.063 mm and 3.358 mm in contrast with cutting-edge techniques [74].

Adithya V. and Rajesh R. [15] suggested a method to identify emergency words in ISL without relying on RGB-D data. To assess the performance of their generated dataset, they merged the CNN framework, namely GoogleNet-LSTM, and claimed a 96.25% recognition rate. LSTM has the characteristics to store the sequential data effectively in combination with CNN, providing better results for recognizing dynamic gestures. There is also a need to have a benchmark dataset of the different counties that are missing for most of the country.

Uchil et al. [90] suggested a novel dimension for recognizing the dynamic gesture by OpenPose and the hybrid approach with CNN. They eliminated the use of cameras and sensors by providing a way in Colab to identify the dynamic sign. With the help of ISLRTC YouTube and RKMVERI - the campus recognized a few healthcare-related keywords and prepared the dataset. They claimed 85% accuracy with their hybrid approach of CNN-RNN [90]. RNN (Recurrent Neural Network) provides more memory with the limitation of computational complexity.

Kinjal et al. [101] attempted to recognize ISL sentence-level recognition. They converted the continuous gesture language into the text sentence of the English language.

TABLE X. Hybrid models for deep learning methods

Author	Dataset	Method	Objective	Year	Accuracy (%)
[67]	RWTH-PHOENIX-Weather-2014	3D-CNN	Sign Language Recognition	2018	62.7
[68]	NGT	Heuristic Approach, LSTM	Continuous Dynamic Sign Language Recognition	2017	80.70
[69]	Chinese Dataset	3D-CNN, Bi-LSTM	Continuous Dynamic Sign Language Recognition	2019	94.90
[70]	One-Million Hands	CNN	Continuous Dynamic Sign Language Recognition	2017	59.20
[71]	RWTH-PHOENIX-Weather-2014, SIGNUM	3D-CNN, Bi-LSTM	Sign Language Recognition	2019	77.14, 97.20
[72]	Own Dataset	3D-RCNN	Continuous Dynamic Human Pose Estimation	2018	69.2
[73]	RKS-PERSIANSIGN, isoGD	SSD, CNN, LSTM, Hand-Crafted Features	Sign Language Recognition	2020	98.42, 86.32
[44]	RWTHPHOENIX-Weather-2014, Multi-Signer dataset	CNN, HMM	Sign Language Recognition	2016	61.7, 61.2
[74]	Own Dataset	CNN, SVM	Gesture Recognition	2018	49.88
[75]	UTD-MHAD, isoGD, UFOP-LIBRAS	CNN, HCM	Gesture Recognition	2020	94.81, 67.36, 64.33
[76]	ASL	CNN, SVM	Gesture Recognition	2016	96.1
[77]	ICVL, NYU, NTU	CNN, SPM	Pose Estimation	2018	91.36, 84.1, 87.19
[78]	Own Dataset	3D-CNN, LSTM, HOG	Gesture Recognition	2019	87.7
[79]	Bengali Character Sign Language	CNN	Gesture Recognition	2020	92.7
[80]	RGB-D Own Dataset	CNN	Gesture Recognition	2017	86
[81]	Bangla Alphabets-Numbers	CNN	Gesture Recognition	2018	92.85
[82]	DEVISIGN_D, SLR_Dataset	BLSTM-3D	Sign Language Recognition	2019	89.8, 86.9
[29]	KETI	CNN, LSTM, SVM	Pose Estimation	2019	89.5, 82.0
[83]	CSL (Chinese Sign Language)	CBAM-ResNet	Gesture Recognition	2019	83.3
[84]	Japanese SL (Own Dataset)	RNN	Continuous Dynamic Sign Language Recognition	2020	***
[85]	South African Sign Language (Own Dataset)	CNN	Gesture Recognition	2016	67
[86]	ASL Fingerspelling	CNN	Gesture Recognition	2017	82
[87]	Own Dataset	CNN	Gesture Recognition	2018	99
[88]	ASLLVD	3-SU (Subunit Sign Model)	Continuous Dynamic Sign Language Recognition	2019	88.7
[15]	ISL (Emergency)	CNN, LSTM	Gesture Recognition	2020	96.25
[89]	Arabic Sign Language (Own Dataset)	SVMRTS	Gesture Recognition	2019	83

TABLE X. Hybrid models for deep learning methods (continue)

Author	Dataset	Method	Objective	Year	Accuracy (%)
[90]	ISL (Own Dataset)	CNN, RNN	Gesture Recognition	2019	85
[91]	Marathi Sign Language (MSL)	CNN	Gesture Recognition	2020	99.28
[92]	ASL	CNN	Gesture Recognition	2021	98.67
[93]	ISL	CNN	Gesture Recognition	2020	99.72
[73]	isoGD	CNN, LSTM	Gesture Recognition	2020	86.32
[94]	Own Dataset	KELM (Kernel-Based Extreme Learning Machine)	Gesture Recognition	2019	97.81
[95]	LSA	CNN, RNN	Gesture Recognition	2018	95.21
[96]	SmartDeaf	CNN, LSTM	Gesture Recognition	2019	59.6 to 72.3
[97]	HDM05, CMU, NTU RGBD, ISL3D	JDTD, JATD, CNN	Gesture Recognition	2019	93.42, 92.67, 94.42, 93.01
[98]	ASL	LSTM, KNN	Gesture Recognition	2021	99.44
[31]	RKS-PERSIANSIGN, First-Person, NYU	CNN, Multi-View-Skeleton, Heat-Map, 3D-CNN, LSTM	Gesture Recognition, Pose Estimation	2020	99.80, 91.12, 95.36
[99]	BVCSL3D	CNN Multi-Stream	Gesture Recognition	2019	86.66
[100]	Bhutanese Sign Language (Own Dataset)	CNN	Gesture Recognition	2020	97.62
[101]	ISL (Own Dataset)	CNN, LSTM	Continuous Dynamic Sign Language Recognition	2021	93.89
[102]	Chinese Sign Language (Own Dataset)	Bi-LSTM	Gesture Recognition	2020	82.55
[103]	IISL2020 (Own Dataset)	RNN, LSTM, GRU	Gesture Recognition	2022	97

In their approach, they applied four steps. 1) Video to frame conversion, 2) Horizontal flipping, 3) Frame sequence generator with image augmentation, and 4) Training with MobileNetV2 + RNN. In the first step, they converted their video into frames. They performed horizontal flipping in the second step to include both the left-handed and right-handed signs on the batch of frames. In the third step, they performed a frame sequence generator with image augmentation. Frames are chosen in batches to get a collection of shifted frames for each unique video. For example, the first batch may contain frames 1, 7, 13, 19, 25, and so on, while the second batch may contain frames 2, 8, 14, 20, 26, and so on. This technique for image augmentation is supported by this unique generator. In the last step, they passed the output to the architecture CNN. With CNN, they used the pre-trained model MobileNetV2 [106]. Furthermore, MobileNetV2 is a compact model that uses deep neural network, which is the most effective for mobile and embedded vision applications [106]. Moreover, they used LSTM to store the output in a long sequence to generate the English text sentence. They compared their work with different pre-trained models like MobileNet, ResNet50, VGG16, and MobileNetV2. Comparing the model size, MobileNet took 16 MB, ResNet50 took 98 MB, VGG16 took 528 MB, and MobileNetV2 took only 14 MB. Comparing the time taken to get accuracy with all these pre-trained models, MobileNet took 23 hours, ResNet50 took 35 hours, VGG16 took 12.7 hours, and MobileNetV2 took only 12.1 hours. Their dataset includes a total of 1289 videos of ISL sentences. They considered 55 English text sentences as class labels. They conclude that by performing certain preprocessing and post-processing to the hybrid approach, it is feasible to develop a low-cost camera-based sign language recognition system that can translate the gestures into text.

The researchers [103] proposed a deep learning-based architecture combining RNN, LSTM, and GRU. They provided video as input, consisting of ISL sign-language ges-

ture sequences. Their primary goal was to identify the works from the gestures used in real life. For that, they divided their video files from a video into sub-sections to represent different words. They separated their videos by identifying the beginning and ending of each gesture. They split the separated videos into frames in the next step. The frames were passed to the architecture InceptionResNetV2 [107]. An architecture inceptionResNetV2 extracts the features from frames. These features are passed to an RNN, LSTM, and GRU to preserve the storage of past information for final prediction. They achieved an accuracy of 97% in sign identification from the ISL custom dataset (IISL2020).

By analyzing Table X, it has been observed that many researchers proposed hybrid deep learning approaches using CNN with LSTM or Bi-LSTM to recognize dynamic gesture recognition at the word level. It is noticed by analyzing the performance measure parameter as accuracy that the average accuracy of recognition is around 85%. A brief survey is shown in Table X. This detailed survey includes the dataset, deep learning method, objectives, publication year, and accuracy. In the observation, the hybrid approach, CNN with HMM, SVM, SSD, LSTM, handcraft features, 3-SU, KELM, etc., is integrated to recognize the higher-level language to recognize dynamic gestures or continuous sentences. It comes to know that dynamic sign language recognition remains a challenging issue. There is a significant scope to improve the result with the low-cost constraint.

6. Conclusion

In the older days when sign language was first invented very few static signs and alphabets were existed. But, nowadays many effective SLRs are available for the different sign languages. The system must be improved to correctly recognize the dynamic motions that appear in continuous visual sequences. In addition, researchers are currently focusing more on developing a broad vocabulary to recognize gesture language. Many researchers created their databases and limited vocabularies for sign language recognition systems. The large database has yet to be made public for a few countries. There are various categorization techniques to classify sign language. The comparison of one approach to another method for the gesture recognition framework still depends on the individual's preferences and constraints. Because sign language varies across nations, researchers have to set their limitations. Thus, it creates difficulty in comparing approaches fairly and directly. Almost all gesture languages used in the country vary in their syntax and the way to present words or sentences using gestures. Sign recognition accuracy is increased recently with the emergence of deep learning techniques. This survey also focuses on the current deep learning methods to recognize sign language. Recently, researchers put out several models. The CNN architecture was used in most of the research to extract features from input images due to excellent capabilities in this regard. Processing on the video, sequential approaches like GRU, LSTM, and RNN were used in most of the research. Several models have been merged into two or more techniques (hybrid techniques) to improve recognition accuracy. This brief review helps to touch on the subject of high-level sign language recognition to identify words or sentences to minimize the issues still faced by Deaf-Mute.

7. Future Work

There is a significant scope in the gesture recognition system to recognize the dynamic sign used to represent words and sentences. The interactive AI-based system can be developed for sign language to voice using computer vision which will minimize the overall cost as compared to the existing translator approach. It is a challenging problem in computer vision because of its dynamic nature. There should be a cost-effective system to recognize higher-level sign language. It will be a good contribution if there will be publicly available benchmark dataset representing words and sentences to perform the experiments. There is also a significant scope as it is a societal need for an effective interactive communication system between Deaf-Mute and ordinary people in a higher level of language form.

References

[1] A. Harshada, "Smart communication assistant for deaf and dumb people," *International Journal for Research in Applied Science and Engineering Technology*, vol. 9, pp. 1358–1360, 07 2021.

[2] J. Murray, "World Federation of the deaf," http://wfdeaf.org/our-work/, 2018, [Online; accessed 19-June-2022].

[3] P. Garg, N. Aggarwal, and S. Sofat, "Vision based hand gesture recognition," *World Academy of Science, Engineering and Technology, International Journal of Computer, Electrical, Automation, Control and Information Engineering*, vol. 3, pp. 186–191, 2009.

[4] R. E. Mitchell, T. A. Young, B. Bachleda, and M. A. Karchmer, "How many people use asl in the united states? why estimates need updating," *Sign Language Studies*, vol. 6, pp. 306 – 335, 2006.

[5] P. Bhatia and A. Wadhawan, "Sign language recognition systems: A decade systematic literature review," *Archives of Computational Methods in Engineering*, pp. 785–813, 01 2019.

[6] "Chinese Sign Language," https://en.wikipedia.org/wiki/Chinese_Sign_Language, [Online; accessed 19-June-2022].

[7] S. G. Lewis, M.P. and C. E. Fennig, "Ethnologue: Languages of the World. 17th Edition," http://www.ethnologue.com, 2014.

[8] C. E. Fennig and S. G. Lewis, M.P., "Ethnologue: Languages of the World. 18th Edition," http://www.ethnologue.com, 2015.

[9] J. Sri-on, "A history of the education of deaf people in thailand /," 01 2001.

[10] N. Pugeault and R. Bowden, "Spelling it out: Real-time asl fingerspelling recognition," *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, pp. 1114–1119, 2011.

[11] A. Thangali, J. P. Nash, S. Sclaroff, and C. Neidle, "Exploiting phonological constraints for handshape inference in asl video," *CVPR 2011*, pp. 521–528, 2011.

[12] C. Chen, B. Zhang, Z. Hou, J. Jiang, M. Liu, and Y. Yang, "Action recognition from depth sequences using weighted fusion of 2d and 3d auto-correlation of gradients features," *Multimedia Tools and Applications*, vol. 76, pp. 4651 – 4669, 2016.

[13] X. Chai, G. S. Li, Y. Lin, Z. Xu, Y. B. Tang, and X. Chen, "Sign language recognition and translation with kinect," 2013.

[14] J. Wan, S. Li, Y. Zhao, S. Zhou, I. R. Subramanian, and S. Escalera, "Chalearn looking at people rgb-d isolated and continuous datasets for gesture recognition," *2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 761–769, 2016.

[15] V. Adithya and R. Rajesh, "Hand gestures for emergency situations: A video dataset based on words from indian sign language," *Data in Brief*, vol. 31, p. 106016, 07 2020.

[16] J. Forster, C. Schmidt, T. Hoyoux, O. Koller, U. Zelle, J. H. Piater, and H. Ney, "Rwth-phoenix-weather: A large vocabulary sign language recognition and translation corpus," in *International Conference on Language Resources and Evaluation*, 2012.

[17] O. Koller, J. Forster, and H. Ney, "Continuous sign language recognition: Towards large vocabulary statistical recognition systems handling multiple signers," *Computer Vision and Image Understanding*, vol. 141, pp. 108–125, 12 2015.

[18] H. Cooper, E.-J. Ong, N. Pugeault, and R. Bowden, "Sign language recognition using sub-units," *Journal of Machine Learning Research*, vol. 13, no. 72, pp. 2205–2231, 2012. [Online]. Available: http://jmlr.org/papers/v13/cooper12a.html

[19] N. Adaloglou, T. Chatzis, I. Papastratis, A. Stergioulas, G. T. Papadopoulos, V. Zacharopoulou, G. J. Xydopoulos, K. Atzakas, D. Papazachariou, and P. Daras, "A comprehensive study on sign language recognition methods," *ArXiv*, vol. abs/2007.12530, 2020.

[20] M. Oszust and M. Wysocki, "Polish sign language words recognition with kinect," *2013 6th International Conference on Human System Interactions (HSI)*, pp. 219–226, 2013.

[21] X. Baró, J. Gonzàlez, J. Fabian, M. Bautista, M. Oliu, H. J. Escalante, I. Guyon, and S. Escalera, "Chalearn looking at people 2015 challenges: Action spotting and cultural event recognition," 06 2015, pp. 1–9.

[22] S. Escalera, X. Baró, J. Gonzàlez, M. Á. Bautista, M. Madadi, M. Reyes, V. Ponce-López, H. J. Escalante, J. Shotton, and I. R. Subramanian, "Chalearn looking at people challenge 2014: Dataset and results," in *ECCV Workshops*, 2014.

[23] T. Kapuscinski, M. Oszust, M. Wysocki, and D. Warchoł, "Recognition of hand gestures observed by depth cameras," *International Journal of Advanced Robotic Systems*, vol. 12, 04 2015.

[24] C. Chen, B. Zhang, Z. Hou, J. Jiang, M. Liu, and Y. Yang, "Action recognition from depth sequences using weighted fusion of 2d and 3d auto-correlation of gradients features," *Multimedia Tools and Applications*, vol. 76, pp. 4651 – 4669, 2016.

[25] F. Ronchetti, F. M. Quiroga, C. Estrebou, L. Lanzarini, and A. Rosete, "Lsa64: An argentinian sign language dataset," 2016.

[26] S. Kim, Y. Ban, and S. Lee, "Tracking and classification of in-air hand gesture based on thermal guided joint filter," *Sensors (Basel, Switzerland)*, vol. 17, 2017.

[27] F. Ronchetti, F. M. Quiroga, C. Estrebou, and L. Lanzarini, "Handshape recognition for argentinian sign language using probsom," 2016.

[28] N. C. Camgoz, S. Hadfield, O. Koller, H. Ney, and R. Bowden, "Neural sign language translation," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7784–7793.

[29] S.-K. Ko, J. Son, and H. Jung, "Sign language recognition with recurrent neural network using human keypoint detection," 10 2018, pp. 326–328.

[30] D. A. Kumar, A. Sastry, P. Kishore, and E. K. Kumar, "3d sign language recognition using spatio temporal graph kernels," *J. King Saud Univ. Comput. Inf. Sci.*, vol. 34, no. 2, p. 143–152, feb 2022. [Online]. Available: https://doi.org/10.1016/j.jksuci.2018.11.008

[31] R. Rastgoo, K. Kiani, and S. Escalera, "Hand sign language recognition using multi-view hand skeleton," *Expert Syst. Appl.*, vol. 150, p. 113336, 2020.

[32] A. Joshi, H. Sierra, and E. Arzuaga, "American sign language translation using edge detection and cross correlation," *2017 IEEE Colombian Conference on Communications and Computing (COLCOM)*, pp. 1–6, 2017.

[33] C. M. Jin, Z. Omar, and M. H. Jaward, "A mobile application of american sign language translation via image processing algorithms," *2016 IEEE Region 10 Symposium (TENSYMP)*, pp. 104–109, 2016.

[34] G. A. Rao and P. V. V. Kishore, "Sign language recognition system simulated for video captured with smart phone front camera," *International Journal of Electrical and Computer Engineering*, vol. 6, pp. 2176–2187, 2016.

[35] M. E. Morocho-Cayamcela and W. Lim, "Fine-tuning a pre-trained convolutional neural network model to translate american sign language in real-time," *2019 International Conference on Computing, Networking and Communications (ICNC)*, pp. 100–104, 2019.

[36] S. Shahriar, A. Siddiquee, T. Islam, A. Ghosh, R. Chakraborty, A. I. Khan, C. Shahnaz, and S. A. Fattah, "Real-time american sign language recognition using skin segmentation and image category classification with convolutional neural network and deep learning," *TENCON 2018 - 2018 IEEE Region 10 Conference*, pp. 1168–1171, 2018.

[37] A. Thongtawee, O. Pinsanoh, and Y. Kitjaidure, "A novel feature extraction for american sign language recognition using webcam," in *2018 11th Biomedical Engineering International Conference (BMEiCON)*, 2018, pp. 1–5.

[38] D. Guo, W. gang Zhou, M. Wang, and H. Li, "Sign language recognition based on adaptive hmms with data augmentation," *2016 IEEE International Conference on Image Processing (ICIP)*, pp. 2876–2880, 2016.

[39] W. Aly, S. K. H. Aly, and S. Almotairi, "User-independent american sign language alphabet recognition based on depth image and pcanet features," *IEEE Access*, vol. 7, pp. 123 138–123 150, 2019.

[40] Pigou, Lionel and Dieleman, Sander and Kindermans, Pieter-Jan and Schrauwen, Benjamin, "Sign language recognition using

convolutional neural networks," in *Lecture Notes in Computer Science*. Springer, 2015, pp. 572–578.

[41] H. Abdelnasser, K. A. Harras, and M. Youssef, "Wigest demo: A ubiquitous wifi-based gesture recognition system," *2015 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, pp. 17–18, 2015.

[42] R. Cui, H. Liu, and C. Zhang, "Recurrent convolutional neural networks for continuous sign language recognition by staged optimization," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1610–1618, 2017.

[43] M. A. Jalal, R. Chen, R. K. Moore, and L. S. Mihaylova, "American sign language posture understanding with deep neural networks," *2018 21st International Conference on Information Fusion (FUSION)*, pp. 573–579, 2018.

[44] O. Koller, S. Zargaran, H. Ney, and R. Bowden, "Deep sign: Hybrid cnn-hmm for continuous sign language recognition," in *British Machine Vision Conference*, 2016.

[45] B. Garcia, "Real-time american sign language recognition with convolutional neural networks."

[46] P. Xu, "A real-time hand gesture recognition and human-computer interaction system," 04 2017.

[47] C. J. L. Flores, A. E. G. Cutipa, and R. L. Enciso, "Application of convolutional neural networks for static hand gestures recognition under different invariant features," *2017 IEEE XXIV International Conference on Electronics, Electrical Engineering and Computing (INTERCON)*, pp. 1–4, 2017.

[48] P. Kumar, H. Gauba, P. P. Roy, and D. P. Dogra, "Coupled hmm-based multi-sensor data fusion for sign language recognition," *Pattern Recognit. Lett.*, vol. 86, pp. 1–8, 2017.

[49] H.-S. Yeo, B.-G. Lee, and H. Lim, "Hand tracking and gesture recognition system for human-computer interaction using low-cost hardware," *Multimedia Tools and Applications*, vol. 74, pp. 2687–2715, 2015.

[50] W. Ahmed, K. Chanda, and S. Mitra, "Vision based hand gesture recognition using dynamic time warping for indian sign language," *2016 International Conference on Information Science (ICIS)*, pp. 120–125, 2016.

[51] M. Taskiran, M. Killioglu, and N. Kahraman, "A real-time system for recognition of american sign language by using deep learning," *2018 41st International Conference on Telecommunications and Signal Processing (TSP)*, pp. 1–5, 2018.

[52] S. Hore, S. Chatterjee, V. Santhi, N. Dey, A. S. Ashour, V. E. Balas, and F. Shi, "Indian sign language recognition using optimized neural networks," pp. 553–563, 2017.

[53] J. Huang, W. gang Zhou, Q. Zhang, H. Li, and W. Li, "Video-based sign language recognition without temporal segmentation," in *AAAI Conference on Artificial Intelligence*, 2018.

[54] A. Voulodimos, N. D. Doulamis, A. D. Doulamis, and E. E. Protopapadakis, "Deep learning for computer vision: A brief review," *Computational Intelligence and Neuroscience*, vol. 2018, 2018.

[55] P. Patel and D. Vaghela, "CROP DISEASES AND PESTS DETECTION USING CONVOLUTIONAL NEURAL NETWORK TO INCREASE AGRICULTURAL PRODUCTIVITY," Master's thesis, GUJARAT TECHNOLOGICAL UNIVERSITY, GANDHINAGAR, 2018.

[56] P. Pruthvi and V. Dineshkumar, "Crop diseases and pests detection using convolutional neural network," 02 2019, pp. 1–4.

[57] S. Dave and P. Chauhan, "REAL-TIME DRIVER'S DROWSINESS DETECTION BY CONVOLUTION NEURAL NETWORK(CNN) OF DEEP LEARNING APPROACH," Master's thesis, GUJARAT TECHNOLOGICAL UNIVERSITY, GANDHINAGAR, 2018.

[58] J. Wu, "Introduction to convolutional neural networks," 2017.

[59] T. Wang, "Recurrent neural network," https://www.cs.toronto.edu/~tingwuwang/rnn_tutorial.pdf, 2016, machine Learning Group, Univer- sity of Toronto, for CSC 2541, Sport Analytics.

[60] R. Grosse, "Csc321 lecture 20: Autoencoders," https://www.coursehero.com/file/139888740/lec20pdf/, 2017, toronto University.

[61] C. Doersch, "Tutorial on variational autoencoders," 06 2016.

[62] G. Hinton, "Deep belief nets," https://www.slideshare.net/zukun/nips2007-deep-belief-nets/, 2007, nIPS. Vancouver, B.C., Canada.

[63] A. Fischer and C. Igel, "An introduction to restricted boltzmann machines," in *Iberoamerican Congress on Pattern Recognition*, 2012.

[64] Tensorflow, "Tensorflow," https://www.tensorflow.org/, [Accessed on June 2022].

[65] F. Bastien, P. Lamblin, R. Pascanu, J. Bergstra, I. Goodfellow, A. Bergeron, N. Bouchard, D. Warde-Farley, and Y. Bengio, "Theano: new features and speed improvements," *arXiv preprint arXiv:1211.5590*, 2012.

[66] MXNET, "Apache mxnet," https://mxnet.apache.org/versions/1.9.1/, [Accessed on June 2022].

[67] J. Pu, W. gang Zhou, and H. Li, "Dilated convolutional network with iterative optimization for continuous sign language recognition," in *International Joint Conference on Artificial Intelligence*, 2018.

[68] B. Mocialov, G. Turner, K. Lohan, and H. Hastie, "Towards continuous sign language recognition with deep learning," in *Proc. of the Workshop on the Creating Meaning With Robot Assistants: The Gap Left by Smart Devices*, 2017.

[69] C. Wei, W. gang Zhou, J. Pu, and H. Li, "Deep grammatical multi-classifier for continuous sign language recognition," *2019 IEEE Fifth International Conference on Multimedia Big Data (BigMM)*, pp. 435–442, 2019.

[70] N. C. Camgöz, S. Hadfield, O. Koller, and R. Bowden, "Subunets: End-to-end hand shape and continuous sign language recognition," *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 3075–3084, 2017.

[71] R. Cui, H. Liu, and C. Zhang, "A deep neural framework for continuous sign language recognition by iterative training," *IEEE Transactions on Multimedia*, vol. 21, pp. 1880–1891, 2019.

[72] Y. Ye, Y. Tian, M. Huenerfauth, and J. Liu, "Recognizing american sign language gestures from within continuous videos," *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 2145–214 509, 2018.

[73] R. Rastgoo, K. Kiani, and S. Escalera, "Video-based isolated hand sign language recognition using a deep cascaded model," *Multimedia Tools and Applications*, vol. 79, no. 31, pp. 22 965–22 987, 2020.

[74] T.-Y. Chen, P.-W. Ting, M.-Y. Wu, and L.-C. Fu, "Learning a deep network with spherical part model for 3d hand pose estimation," *Pattern Recognition*, vol. 80, 02 2018.

[75] E. J. E. Cardenas and G. C. Chávez, "Multimodal hand gesture recognition combining temporal and pose information based on cnn descriptors and histogram of cumulative magnitudes," *J. Vis. Commun. Image Represent.*, vol. 71, p. 102772, 2020.

[76] M. Ma, Z. Chen, and J. Wu, "A recognition method of hand gesture with cnn-svm model," in *International Conference on Bio-Inspired Computing: Theories and Applications*, 2016.

[77] H. Chen, R. Tong, M. Chen, Y. Fang, and H. Liu, "A hybrid cnn-svm classifier for hand gesture recognition with surface emg signals," *2018 International Conference on Machine Learning and Cybernetics (ICMLC)*, vol. 2, pp. 619–624, 2018.

[78] S. He, "Research of a sign language translation system based on deep learning," *2019 International Conference on Artificial Intelligence and Advanced Manufacturing (AIAM)*, pp. 392–396, 2019.

[79] D. Aich, A. Al Zubair, K. M. Zubair Hasan, A. D. Nath, and Z. Hasan, "A deep learning approach for recognizing bengali character sign langauage," in *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, 2020, pp. 1–5.

[80] Y. Ji, S. Kim, and K.-B. Lee, "Sign language learning system with image sampling and convolutional neural network," *2017 First IEEE International Conference on Robotic Computing (IRC)*, pp. 371–375, 2017.

[81] A. J. Rony, K. H. Saikat, M. Tanzeem, and F. M. R. H. Robi, "An effective approach to communicate with the deaf and mute people by recognizing characters of one-hand bangla sign language using convolutional neural-network," *2018 4th International Conference on Electrical Engineering and Information & Communication Technology (iCEEiCT)*, pp. 74–79, 2018.

[82] Y. Liao, P. Xiong, W. Min, W. Min, and J. Lu, "Dynamic sign language recognition based on video sequence with blstm-3d residual networks," *IEEE Access*, vol. 7, pp. 38 044–38 054, 2019.

[83] H. Chao, W. Fenhua, and Z. Ran, "Sign language recognition based on cbam-resnet," in *Proceedings of the 2019 International Conference on Artificial Intelligence and Advanced Manufacturing*, ser. AIAM 2019. New York, NY, USA: Association for Computing Machinery, 2019.

[84] H. Brock, F. Law, K. Nakadai, and Y. Nagashima, "Learning three-dimensional skeleton data from sign language video," *ACM Transactions on Intelligent Systems and Technology*, vol. 11, p. 30, 04 2020.

[85] K. Jacobs, M. Ghasiazgar, I. Venter, and R. Dodds, "Hand gesture recognition of hand shapes in varied orientations using deep learning," in *Proceedings of the Annual Conference of the South African Institute of Computer Scientists and Information Technologists*, ser. SAICSIT '16. New York, NY, USA: Association for Computing Machinery, 2016.

[86] S. Ameen and S. Vadera, "A convolutional neural network to classify american sign language fingerspelling from depth and colour images," *Expert Systems*, vol. 34, no. 3, p. e12197, 2017.

[87] Y. Ji, S. Kim, Y.-J. Kim, and K.-B. Lee, "Human-like sign-language learning method using deep learning," *ETRI Journal*, vol. 40, pp. 435–445, 08 2018.

[88] E. R. and S. K., "Subunit sign modeling framework for continuous sign language recognition," *Comput. Electr. Eng.*, vol. 74, no. C, p. 379–390, mar 2019. [Online]. Available: https://doi.org/10.1016/j.compeleceng.2019.02.012

[89] M. Almasre and H. Al-Nuaim, "A comparison of arabic sign language dynamic gesture recognition models," *Heliyon*, vol. 6, p. e03554, 03 2020.

[90] A. P. Uchil, S. Jha, and B. Sudha, "Vision based deep learning approach for dynamic indian sign language recognition in healthcare," in *International conference on computational vision and bio inspired computing*. Springer, 2020, pp. 371–383.

[91] A. M. Deshpande and S. R. Kalbhor, "Video-based marathi sign language recognition and text conversion using convolutional neural network," in *Emerging Trends in Electrical, Communications, and Information Technologies*. Springer, 2020, pp. 761–773.

[92] P. Kartik, K. B. Sumanth, V. Ram, and P. Prakash, "Sign language to text conversion using deep learning," in *Inventive Communication and Computational Technologies*. Springer, 2021, pp. 219–227.

[93] A. Wadhawan and P. Kumar, "Deep learning-based sign language recognition system for static signs," *Neural computing and applications*, vol. 32, no. 12, pp. 7957–7968, 2020.

[94] J. Imran and B. Raman, "Deep motion templates and extreme learning machine for sign language recognition," *The Visual Computer*, vol. 36, no. 6, pp. 1233–1246, 2020.

[95] S. Masood, A. Srivastava, H. C. Thuwal, and M. Ahmad, "Real-time sign language gesture (word) recognition from video sequences using cnn and rnn," in *Intelligent Engineering Informatics*. Springer, 2018, pp. 623–632.

[96] K. Sakamoto, E. Ota, T. Ozawa, H. Nishimura, and H. Tanaka, "Feasibility study on deep learning scheme for sign language motion recognition," in *Conference on Complex, Intelligent, and Software Intensive Systems*. Springer, 2018, pp. 1106–1115.

[97] E. K. Kumar, P. Kishore, M. T. K. Kumar, and D. A. Kumar, "3d sign language recognition with joint distance and angular coded color topographical descriptor on a 2–stream cnn," *Neurocomputing*, vol. 372, pp. 40–54, 2020.

[98] C. K. Lee, K. K. Ng, C.-H. Chen, H. C. Lau, S. Chung, and T. Tsoi, "American sign language recognition and training method with recurrent neural network," *Expert Systems with Applications*, vol. 167, p. 114403, 2021.

[99] S. Ravi, M. Suman, P. Kishore, K. Kumar, A. Kumar *et al.*, "Multi

modal spatio temporal co-trained cnns with single modal testing on rgb–d based sign language gesture recognition," *Journal of Computer Languages*, vol. 52, pp. 88–102, 2019.

[100] K. Wangchuk, P. Riyamongkol, and R. Waranusast, "Real-time bhutanese sign language digits recognition system using convolutional neural network," *Ict Express*, vol. 7, no. 2, pp. 215–220, 2021.

[101] K. Mistree, D. Thakor, and B. Bhatt, "Towards indian sign language sentence recognition using insignvid: Indian sign language video dataset," *International Journal of Advanced Computer Science and Applications*, vol. 12, 01 2021.

[102] Q. Xiao, M. Qin, and Y. Yin, "Skeleton-based chinese sign language recognition and generation for bidirectional communication between deaf and hearing people," *Neural networks*, vol. 125, pp. 41–55, 2020.

[103] D. Kothadiya, C. Bhatt, K. Sapariya, K. Patel, A.-B. Gil-González, and J. M. Corchado, "Deepsign: Sign language detection and recognition using deep learning," *Electronics*, vol. 11, no. 11, 2022. [Online]. Available: https://www.mdpi.com/2079-9292/11/11/1780

[104] G. Chéron, I. Laptev, and C. Schmid, "P-cnn: Pose-based cnn features for action recognition," *2015 IEEE International Conference on Computer Vision (ICCV)*, pp. 3218–3226, 2015.

[105] S. Lakhara, S. Chauhan, and D. Vaghela, "Voice/text to sign: A survey on communication system for normal and "deaf or mute" people based on natural language programming," vol. 40-ISSUENO.-9-2020, 2020, [Online available link: https://tpnsindia.org/index.php/sipn/article/view/9044/8651].

[106] A. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," 04 2017.

[107] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," *AAAI Conference on Artificial Intelligence*, vol. 31, 02 2016.

Chauhan Pareshbhai Mansangbhai
Chauhan Pareshbhai Mansangbhai is a Ph.D. research scholar from Gujarat Technological University, Ahmedabad, Gujarat, India. He completed his M.Tech. in Information Technology from Dharmsinh Desai University, Nadiad, Gujarat, India, in 2014. He completed his B.E. in Information Technology from Gujarat Technological University, Ahmedabad, Gujarat, India, in 2012. He is currently working as an Assistant Professor at the Department of Information Technology, Shantilal Shah Engineering College, Bhavnagar, Gujarat, India. He has 18 years of teaching experience. His research interest spans computer vision, image processing, and machine learning. He has published research papers in prestigious conferences concerning computer vision, image processing, and machine learning.

Dr. Dineshkumar B. Vaghela
Dr. Dineshkumar B. Vaghela completed his Ph.D. in Computer Engineering from Gujarat Technological University, Chandkheda, Gujarat, India, in 2017. He completed his M.E. in Computer Engineering from Sardar Patel University, V.V. Nagar, Gujarat, India, in 2010. He completed his B.E. in Information Technology from Gujarat University, Ahmedabad, Gujarat, India, in 2004. He is currently working as an Assistant Professor at the Department of Information Technology, Shantilal Shah Engineering College, Bhavnagar, Gujarat, India. He has 18 years of teaching experience. He guided many students in the field of computer vision, image processing, machine learning, cloud computing, big data analytics, and IoT. His research interest spans computer vision, image processing, machine learning, cloud computing, big data analytics, and IoT. He has published research papers in prestigious conferences and journals concerning computer vision, image processing, machine learning, and cloud computing.