Sign Language Translation Systems:
A Systematic Literature Review

Ankith Boggaram, PES University, India*
Aaptha Boggaram, PES University, India
Aryan Sharma, PES University, India
Ashwin Srinivasa Ramanujan, PES University, India
Bharathi R., PES University, India

ABSTRACT

Sign language, often termed "dactylology," is a mode of communication for those who are hard of hearing. With over 2.5 billion people projected to have hearing loss by 2050, there are very few efficient real-time sign language translation (SLT) applications present today despite extensive research in the domain. The main purpose of the systematic literature review is to analyze existing research in SLT systems and obtain results that will help in building an efficient and improved SLT system. A total of 125 different research articles within the time frame of 2015–2022 were identified. The study analyzes each paper against nine main research questions. The results obtained show the unique strengths and weaknesses of the different methods used, and while the reviewed papers showed significant results, there is still room for improvement in the implementations. This systematic literature review helps in identifying suitable methods to develop an efficient SLT application, identifies research gaps in this domain, and simultaneously indicates recent trends in the field of SLT systems.

KEYWORDS

Accuracy, Bidirectional, Conversion, Deep Learning, Machine Learning, Pre-Processing, Recognition, Sign Language, Translation

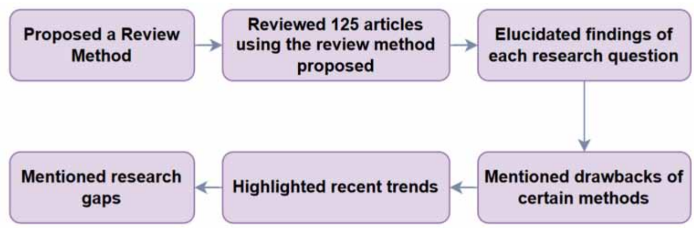
1. INTRODUCTION

Sign Language is one of the most sophisticated and structured means of communication for those who are hard of hearing and/or mute. Oftentimes, there are situations when they cannot have a conversation with people solely because many are unaware of the different taxonomies involved with Sign Language (SL) gestures. Bidirectional Sign Language Translation (SLT) is the conversion of SL gestures to their corresponding phrases and vice versa. It assists people who are hard of hearing to communicate in the wider world, which predominantly uses spoken language.

Spoken languages are sound-based, whereas sign languages, i.e., gesture-based languages, are concerned with appearance and hence use hand movements, signs in a particular order, and body language to create relevant words. SL communication can have different representations, one being phrase-based and another being character-based. The former representation, which depends on the movement of hands, would need dynamic input in the form of a video for translation, whereas the latter could be translated by taking static images.

DOI: 10.4018/IJSSCI.311448
*Corresponding Author

Figure 1. Flow chart of contributions



SLT systems present today use Machine Learning and Deep Learning methodologies integrated into working application software. These applications are available in the form of data-gloves, mobile and web applications for easy access to the general public.

A detailed overview of the research based on SLT, current developments in the field, and SL barrier concerns are addressed in the study. Different sign languages, modalities, and datasets in SL are addressed and given in a tabular form to make them easier to comprehend. The contributions made by this extensive SLT review paper are presented in the flow chart mentioned in Figure 1.

Firstly, the study proposes a review method that will be followed to assess articles; 125 articles were surveyed and the findings were presented in a proposed order in a laconic manner. Further, the drawbacks of specific methods used by some articles were highlighted; neoteric trends of recent papers were accentuated, and future scope and research gaps are identified.

The study analyses 125 articles against research questions to gain significant insights:

- Is the SLT system mentioned in the paper one-way or two-way?
- Dataset questions that provide more information about the dataset each article uses
- What different pre-processing methods are used in the articles?
- What different dimensionality reduction techniques are used?
- What feature extraction techniques are used in the different SLT systems?
- Informative questions about the different models used in each article
- What are the existing applications available to the general public and enabling technologies for Sign Language translation and recognition?
- What evaluation metrics are used in different articles?
- What validation techniques have been used in SLT systems?

These research questions were identified to explore the different methods that are proven to give good accuracy and can be applied in real-time SLT systems. The (World Health Organization, 2021) reports that nearly 2.5 billion people will have some form of hearing loss by 2050, with over a billion young adults who face the risk of permanent hearing loss. Yet, despite extensive study in this sector, SLT's potential applicability for real-time applications has to be fulfilled. Other possible reasons for the gap in theory and practice have been answered (Hao, 2019).

One of the objectives of the systematic review is to fill the gap in research and practice by helping to make future implementations easier and more efficient, and thus more people such as investors, researchers, or anyone with the capabilities and resources, can make a real-time application. The study also analyses unique aspects that have been not looked at sufficiently enough, by bringing into the picture the recognition of different angles of the same signs, as each person's gesticulations and body language differs from the others.

The structure of the study is as follows: Theoretical Background, Related Work, Methodology, Results, and Conclusion. Before beginning the literature review, the theoretical background of this topic is discussed. Followed by related articles addressing their approach and novelty (J. K. Appati et al., 2022). Next, the methodology being followed for conducting this review is elaborated upon; the results of the findings are explicated; and finally, the conclusion is presented.

2. THEORETICAL BACKGROUND

The (World Health Organization, 2021) estimates that around 300 million people worldwide have impaired hearing. The primary method used by special education institutions to educate and share information is SL, which is the everyday language used by the deaf and the hearing-impaired. SL is also the most convenient and natural form of communication between the two groups. In SL, the form, position, and movement of the hands are used to communicate meaning. However, very few people with normal hearing are fluent in SL, and theoretical research on SLT is still in its development phase in most countries.

SL gestures are identified by their complicated hand shapes, blurring movements, and overlap of the left and right hands. Since SL is made up of continuous motions, in addition to capturing images of signs, the SL translator also needs to be able to collect motion data over several subsequent video frames. Developing effective and appropriate SLT models is slowly becoming the focus of study for many researchers.

Methods for recognizing and translating SL are easily influenced by many factors, including hand movements, backgrounds, lighting, and more. Strategies for translating speech to SL are also affected by many factors, such as the pitch, amplitude, and speed of the speaker. For these reasons, pre-processing the input and extracting features play significant roles.

There are various approaches to translating SL to speech and speech to SL, ranging from the intricate techniques of Deep Learning to the construction of look-up tables. These methods have their advantages and drawbacks, but most importantly, they help with SLT, which can be used not only as a scientific and theoretical study, but its results can be used by business leaders to take decisions.

Additionally, researchers (Škare et al., 2022) provide an overview of how communication is linked to entrepreneurship. Whereas researchers (Aljohani et al., 2022; Nyagadza, 2022) provide an overview of several firms' future research directions as well as their implications.

3. METHOD

The methodology given by (Kitchenham & Charters, 2007) has been adopted for the systematic literature review. Researchers (Wadhawan & Kumar, 2021) have also opted for this research methodology in their SL systematic review.

3.1 Review Method

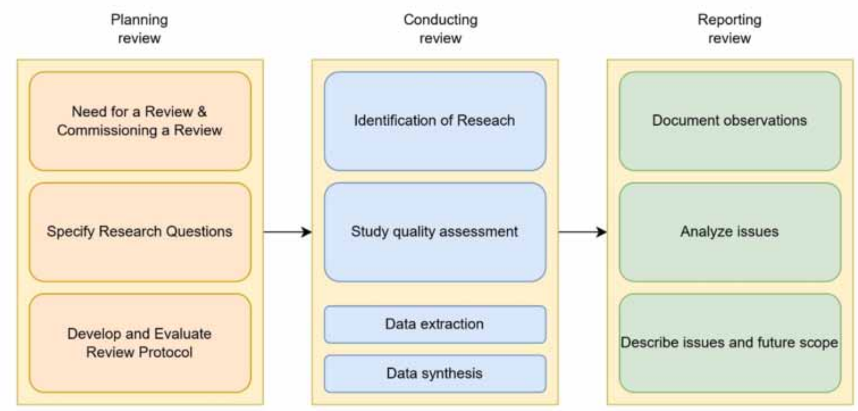
The review method is shown in Figure 2. It has three phases: planning, conducting, and reporting the review. These phases have been explained in detail below.

- Planning the review - This is the first phase in making a systematic literature review. It consists of three steps: determining the need for a review; defining research questions; and developing and evaluating the review protocol.
- Conducting the review - It constitutes the extraction and synthesis of data obtained. Additionally, answers to the mentioned research questions from the articles examined are elucidated.
- Reporting the review - This is the conclusion of the literature review. All the results and observations are mentioned. Future scope, issues regarding findings, and the research gaps in the articles are addressed.

3.2 Research Questions

The purpose of the research questions is to address the research gap in the field of SLT. A research gap is a topic or area in which reviewers' ability to conclude a specific issue is restrained by missing

Figure 2. Review method steps



or inadequate information. Furthermore, the research questions are identified and the most crucial ones are filtered for integration. Answering them would help in analyzing the trends in the domain of SLT systems and SSL translators. This systematic literature review looks into different Machine Learning and Deep Learning methodologies, datasets used, and aims to deepen the knowledge of the topic on hand.

Additionally, the study aims to bring to light the different methods of feature extraction, the types of pre-processing techniques employed, and the dimensionality reduction techniques used. A major reason for the study is to note the number of research articles that have implemented bidirectional SLT. Furthermore, the study looks into the evaluation metrics used and the different ways the Machine Learning or Deep Learning models have been validated. Table 1 contains the shortlisted research questions and the motivation behind them. Meanwhile, Figure 3 depicts the same in a mind map.

3.3 Search Strategy and Data Extraction

The search strategy involves four basic steps, namely the creation of search phrases, the identification of relevant databases, performing search operations, and evaluating the results obtained. The search strategy of the study was inspired by the text-mining approach (Martínez et al., 2022). To begin with, relevant keywords and phrases that will be utilized to retrieve research articles must be discovered. As a result, the following search terms were chosen.

- Sign language to speech conversion
- Speech to sign language conversion
- Bidirectional sign language translation
- Sign language recognition
- Sign language translation systems
- Sign language to text conversion
- Real-time sign language translation

Further, the extensive coverage of research articles necessitates a broad perspective. To ensure the same, specific databases were identified. These databases not only aided in the discovery of more information but also in the identification of highly relevant articles. The articles are chosen from the following electronic databases:

- IEEE Explore
- ACM Digital Library
- Springer

Table 1. Research Questions and their motivation

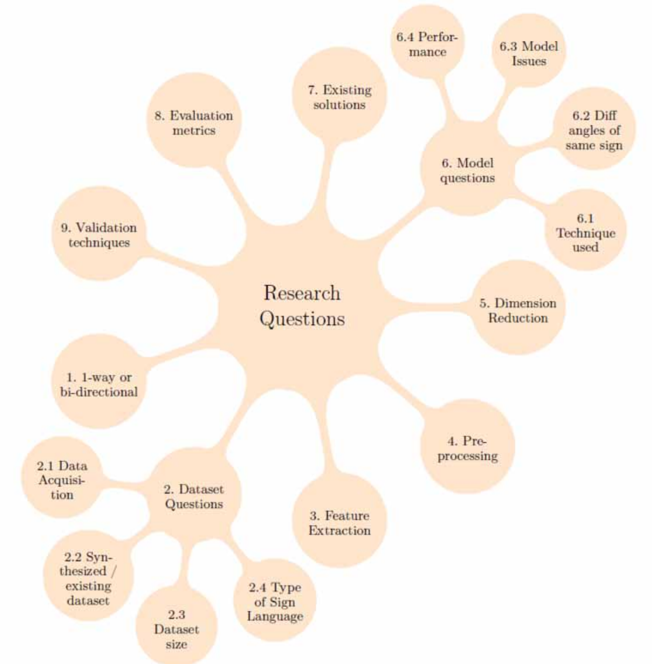
Research questions	Motivation
RQ1: Is the SLT performed one-way or bidirectional?	To categorize the implementations based on the type of SLT performed.
RQ2: Dataset questions:	To examine various datasets.
RQ2.1: What are the different methods used for the acquisition of data?	To determine the methods used for the collection of data.
RQ2.2: Is it a synthesized or an existing dataset?	To divide datasets based on data acquisition.
RQ2.3: What is the size of the dataset i.e., how many samples are present in them?	To classify datasets with the regard to the size of the dataset used.
RQ2.4: Which Sign Languages are used in SLT systems?	To group articles based on the SL used.
RQ3: What are the different pre-processing techniques used in SLT systems?	To identify different techniques used for pre-processing of input data.
RQ4: What are the dimensionality reduction techniques used?	To recognize different techniques used for reducing the dimensions of datasets.
RQ5: What are the feature extraction techniques used in SLT systems?	To pinpoint different methods used for extraction of features.
RQ6: Model questions:	To examine various models.
RQ6.1: What are the different Machine Learning / Deep Learning techniques used?	To discover different methods used for SLT.
RQ6.2: Can the model identify the same sign from different angles?	To infer if a model can perform seamless SLT despite capturing the same sign from different angles.
RQ6.3: What are the challenges in using these models for SLT systems?	To single out numerous challenges researchers face upon using certain models.
RQ6.4: What is the performance of prediction techniques used?	To analyze the performance of SLT systems proposed.
RQ7: What are the existing applications available to the general public and enabling technologies for Sign Language translation and recognition?	To examine different existing solutions, present for SLT.
RQ8: What are the evaluation metrics used in SLT systems?	To observe the trend in the use of different evaluation metrics
RQ9: What are the validation techniques used in SLT systems?	To determine methods used for validating the prediction techniques used in SLT systems.

- Elsevier
- Nature
- arXiv

Subsequently, the Google Scholar search engine was used for performing search operations and retrieving relevant results. These results were filtered to obtain those falling within the time frame of 2015 to 2022. Additionally, citation searching was performed on the retrieved articles. This facilitated the acquisition of comparable supplemental articles as well as the identification of relevant citations by other researchers.

Finally, to ensure that the assessment was thorough, a meticulous database search was conducted and articles were selected based on the criteria mentioned in the search strategy. Table 2 includes the inclusion and exclusion criteria used in the data extraction process.

Figure 3. Research questions in mind map form



3.4 Study Selection

Using the results acquired from implementing the search strategy and incorporating the theoretical framework (Saura, 2021), research articles were included in the review based on the inclusion and exclusion criteria mentioned in Table 2. (J. K. Appati et al., 2022) have incorporated a similar criterion in their systematic literature review.

Table 2. Inclusion and exclusion criteria

Inclusion criteria	Articles published between 2015 and 2022
	Articles from reputable publishers such as IEEE, Elsevier, Springer, the ACM digital library, and so on.
	Articles with at least one citation, except articles from 2022.
	Articles from selected journals and conferences mentioned in the search strategy are included.
	Articles that respond to at least one of the research questions posed.
Exclusion criteria	Publications in languages other than English.
	Articles that make no mention of SLT techniques.
	Articles that appear in multiple electronic resources to reduce redundancy.

3.5 Study Quality Assessment

Following filtering using the inclusion criteria, the remaining relevant articles were subjected to a quality assessment. Each study was evaluated for bias as well as the internal and external validity of the results. Additionally, a questionnaire proposed in Table 3 was used to evaluate the quality of the articles retrieved. Using the same, all of the included articles featured research on SLT or improving the reliability of the database selected.

Table 3. Form used for Quality Assessment of Articles selected

Assessment Questions	Considerations
I. First Screening Question	
Is the research article related to SLT? **Yes ☐ No ☐**	Check if SLT is either mentioned or performed in the study.
II. Secondary Screening Questions	
Is the goal of the research article to use machine translation techniques to perform SLT? **Yes ☐ No ☐**	The article should mention techniques used for SLT.
Is the goal of the research article to use machine translation techniques to perform SSL translation? **Yes ☐ No ☐**	
Is the goal of the research article to use machine translation techniques to perform both SLT and SSL translation? **Yes ☐ No ☐**	The article should mention techniques used for SSL translation.
Is the research relevant to any of the sub-categories? **Yes ☐ No ☐**	
If the answer to the previous question is affirmative, the next question is considered.	
III. Comprehensive Question	
Research Results	
Is there a succinct purpose for the findings? **Yes ☐ No ☐**	The article should give conclusive evidence regarding SLT performed. If it does, what is the method used for generating the relevant sign?

(Oxman & Guyatt, 1991) proposed a generic quality assessment questionnaire that can be used in the data extraction process to retrieve relevant articles. The question from their study, “Were the conclusions supported by reported data?”, corresponds to the “Comprehensive Question” mentioned in Table 3. (Kitchenham & Charters, 2007) used a quality assessment form similar to the one used in the study. The first and second screening questions in Table 3 correspond to the first four questions in the questionnaire used in their work.

3.6 Validity Threats

In this section, the internal and external validity threats that might be posed to the literature review are discussed. The subsequent sections will explain them in detail.

3.6.1. Publication Time Frame

Since the search strategy focused on articles published between 2015 and 2022, it is possible that a few articles were missed in the process. This lowers the number of articles that are relevant to the search queries used. As a result, other databases were employed to preserve a reasonable number of relevant articles.

3.6.2. Primary Study Selection

The proposed search queries in the search strategy are restrictive in nature. Therefore, to increase the number of relevant scholarly articles retrieved, Google Scholar was utilized. Having said that, some articles were not present in the search engine results. Thus, identical search queries were used in other databases and the results obtained were appended to the existing one.

4. RESULTS

4.1 Primary Studies

With the existence of Sign Language translators dating back to 1977, beginning with the RALPH project (Jaffe, 1994), the study in this field is extensive and diversified. Despite the research, however, there remains a gap in the literature reviews. Several articles review existing work in SLT systems. Some of them are mentioned below.

(Wadhawan & Kumar, 2021) reviewed research articles published between 2007 and 2017, and this systematic literature review continues their research by reviewing articles from the years 2015 to 2022. Their originality stems from their examination of works about the sign languages presented in the articles they reviewed and their large sampling frame. They examined work done on 13 different sign languages and highlighted key elements required to collaborate with them. Different data acquisition modes and classification techniques used were represented in a tabular and graphical manner. Furthermore, they discussed the benefits and drawbacks of various methodologies for various sign languages. For instance, the translation of American Sign Language (ASL) has relied heavily on neural networks (NN), followed by Support Vector Machines (SVM), Long-Short Term Memory (LSTM), k-Nearest Neighbor (KNN), etc. They concluded that approximately 90% of the research articles examined used single-handed SL gestures.

(Amin et al., 2022) reviewed articles involving hardware-based SLT. In each reviewed article, the researchers focused on the available glove types, techniques utilized for classification of SL, sensors used for capturing data, and details about the dataset and output devices of different recognition systems. Additionally, six distinct translation models were examined, including sensor-based, vision-based, hybrid recognition models, framework-based recognition models, and commercial and non-commercial data gloves. Next, a comparative analysis of all existing methodologies was provided. In conclusion, they inferred that increased accuracy was observed when multiple sensors were utilized, and that combining their sensory data would aid in the processing of different translation algorithms. However, they recommended the use of a minimal number of sensors to decrease system complexity and improve hardware performance. Not only did their findings aid in identifying the shortcomings of certain sensors and applications, but they also provided suggestions on SLT systems from the perspectives of researchers, governments, and developers. On a similar note, this systematic literature review provides suggestions and identifies several shortcomings that users can consider before developing an SLT system.

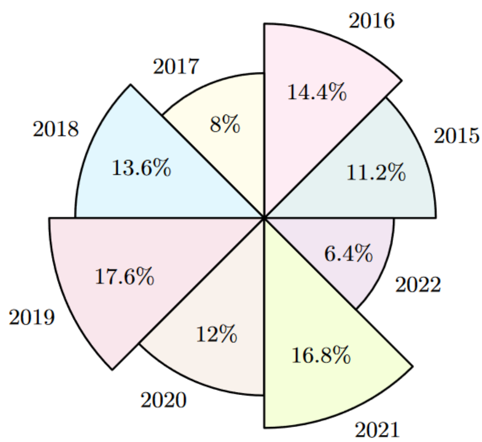
(M. A. Ahmed et al., 2021) conducted a review to provide a consistent taxonomy for describing recent research. The taxonomy was divided into four categories: development, framework, other hand gesture recognition, and reviews & surveys. Further, they investigated three sensor-based systems, namely the Inertial Measurement Unit (IMU) sensor, the touch sensor, and flex sensor systems. Unlike other systematic reviews, they proposed a recognition framework and conducted a comparative analysis of their execution with the articles they reviewed. The proposed system had more data channels and sensors than the summarized articles, allowing it to identify more gestures successfully. Their implementation involved a sensory-based data glove that captured wrist orientation, degree of bend, and hand motion. Their proposed system considers different orientations of the same SL gesture. However, different orientations of the same SL gestures are not considered in vision-based SLT systems. This is a research gap identified and the same is explained later in this review.

In the SLT review, the chosen articles are either journal entries or conference papers, as specified in the search strategy. Moreover, each of the articles presented in the review answers at least one of the research questions proposed. Additionally, the succeeding sections elaborate on the statistics of the primary studies evaluated. The study incorporates articles from the years 2015 to 2022 solely to compare recent works and their development. Comparing work before 2015 would include methods that might not be suitable for use with today's large-scale datasets and thus were not considered.

4.1.1. Publishing Year

Figure 4 shows the distribution of articles selected over the years, from 2015 up to 2022, and it can be observed that the selection is quite uniform, with an appreciable number of published articles chosen from each time frame.

Figure 4. Published articles count by year

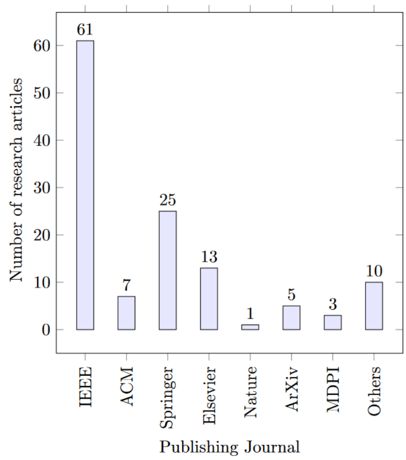


The majority of the work was presented between 2019 and 2021. Barring 2022 articles, only 6% of the articles reviewed were from this time frame because 8 months have passed and there have not been a significant number of relevant research articles. 75% of the pertinent articles published in 2022 used existing datasets, a 30% increase from prior years. However, only 12.5% of these articles utilized pre-trained models despite their accessibility. These observations are explained in detail in RQ2.2 and RQ6.1, respectively. Regardless of how few relevant publications from 2022 there are, the bulk of them achieved extremely high accuracy (approx. 99%) even on existing datasets. Therefore, their methodologies may be used to improve the performance of currently ongoing work. The methods that gave exceptionally high accuracy are mentioned in RQ6.4 in detail.

4.1.2. Publishing Source

Research articles have been collected from various sources as mentioned in the search strategy. Figure 5 illustrates the number of articles reviewed from each publishing database, and most of these were published by IEEE, followed by Springer and Elsevier.

Figure 5. Published article count by source

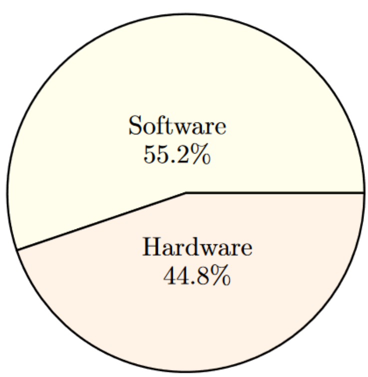


4.2 Mode of Implementation

The most striking difference among all these articles is their modes of implementation. For instance, if researchers used data gloves along with various sensors, it would be considered a hardware-based project. On the other hand, if researchers make an application software to run the SLT/SSL translation functionalities on a standalone device, it is classified as a software project. Figure 6 shows the statistics of both implementations.

It can be seen that the bulk of research articles used only software implementations in their studies. Hardware-enabled experiments necessitate the use of additional devices, sensors, and gadgets to perform SLT. A notable trend is that a majority of the research articles examined in the earlier years required additional hardware components, making them hardware-related research. Analogously, the sections that follow provide detailed answers to the research questions proposed in Table 1.

Figure 6. Mode of implementation



4.3 RQ1: Is the Sign Language Translation a One-Way/Bidirectional Implementation?

As recorded in Table 4, it can be observed that among the 125 articles reviewed, only twelve have implemented bidirectional SLT (Ahire et al., 2015; M. Ahmed et al., 2016; Bajpai et al., 2015; Camgoz et al., 2020; Cooper et al., 2011; Escudeiro et al., 2015; Fernandes et al., 2020; Hoque et al., 2016; Nath & Arun, 2017; Oliveira et al., 2019b, 2019a; Süzgün et al., 2015).

Table 4. Type of Sign Language Translation used

Type of Implementation	Number of articles
Only classification	71
Bidirectional	12
Sign Language to Speech	7
Speech/text to Sign Language	9
Speech to Sign Language	5
Sign Language to speech/text	17
Other	4

Further, bidirectional SLT involves two stages or phases, i.e., SLT as well as SSL translation combined into a single application. 71 papers textually classify SL gestures; 23 classify sign languages as speech/text (Abraham & Rohini, 2018; Adaloglou et al., 2020; Akoum & al Mawla, 2015; An et al., 2016; Arif et al., 2016; Bhagat et al., 2019; Cheok et al., 2019; Dutta et al., 2015; Hasan et al., 2016; Kamal et al., 2019; Liao et al., 2019; Mariappan & Gomathi, 2019; Pu et al., 2019; Rahman et al., 2019; Rajaganapathy et al., 2015; Tiku et al., 2020; Tornay et al., 2019; Truong et al., 2016; Vijayalakshmi & Aarthi, 2016; Warrier et al., 2016; Wen et al., 2021; Wu et al., 2016; Zhao et al., 2021); 15 convert speech/text to SL gestures (M. Ahmed et al., 2016; Aliwy & Ahmed, 2021; Bharti et al., 2019; das Chakladar et al., 2021; Dhanjal & Singh, 2021; Gago et al., 2019; Kang, 2019; Natarajan et al., 2022; Patel et al., 2020; Rayner et al., 2016; Saunders et al., 2020; Stoll, Camgoz, et al., 2020; Stoll, Hadfield, et al., 2020; Xiao et al., 2020; Zelinka et al., 2019), and 4 articles mention other forms of translation (SL numbers, other signs, etc.) (Amin et al., 2022; Chakraborty et al., 2018; Madhiarasan et al., 2022; Moryossef et al., 2021) The existence of bidirectional SLT systems is comparatively less than one-way implementations, which are either SLT or SSL translation systems.

4.4 RQ2: Dataset Questions

The objective of RQ2 is to obtain details about the datasets across the research articles concerning the modes of data acquisition; whether the dataset is synthesized or not, the size of the dataset, and the different types of SL used in SLT systems.

4.4.1. RQ2.1: What Are The Different Methods Used For The Acquisition Of Data?

The different methods for acquiring datasets are explained below and summarized in Table 5. Research articles (Akoum & al Mawla, 2015; Bheda & Radpour, 2017; Dutta et al., 2015; Hasan et al., 2016; M. M. Islam et al., 2017; Kamruzzaman, 2020; Mariappan & Gomathi, 2019; Nandi et al., 2022; Nikam & Ambekar, 2016; Qin et al., 2021; Rocha et al., 2020; Sahoo, 2021; Shivashankara & Srinath, 2018; Tripathi & Nandi, 2015; Truong et al., 2016; Yang & Zhu, 2017; Ye et al., 2018) have used a webcam for acquiring videos/images of SL gestures to build their datasets. (Warrier et al., 2016), (Bantupalli & Xie, 2018) and (Park et al., 2021) used a smartphone camera to build their dataset.

The researchers (M. Ahmed et al., 2016; Bhagat et al., 2019; Escudeiro et al., 2015; Kumar et al., 2017a, 2017b; Lee et al., 2016; Liu et al., 2016; Oliveira et al., 2019b, 2019a; Quesada et al., 2017; Rajaganapathy et al., 2015; Sharma & Singh, 2021; Stoll, Hadfield, et al., 2020; Süzgün et al., 2015; Wang et al., 2015; Xiao et al., 2019; Zhang et al., 2016) have taken advantage of the tool 'Kinect for Windows SDK beta' created by Microsoft Research. This tool allowed easy access to the Microsoft Kinect device's features for easy acquisition of data, research, and development.

Another mode of data collection includes sensory inputs. (Fok et al., 2015) used Multi-Sensor Data Fusion (MSDF) (Mitchell, 2007) and (Savur & Sahin, 2015) collected surface electromyography (sEMG) values, which are bio-electrical signals produced by the human body via the CleveMed BioRadio 150 device (a wireless data collection system for a subject's physiological signals), while articles (Abraham & Rohini, 2018; M. A. Ahmed et al., 2021; Arif et al., 2016; Bajpai et al., 2015; K. Li et al., 2016; Oliveira et al., 2019b, 2019a; Phing et al., 2019; Shukor et al., 2015; Vijayalakshmi & Aarthi, 2016; Wu et al., 2016) used a wearable glove as a mode acquiring data. Leap Motion sensors (Motion, 2015) were utilized by various researchers (Aliyu et al., 2017; Fok et al., 2015; Kumar et al., 2017a; Mittal et al., 2019; Quesada et al., 2017) to track the spatial features of hands.

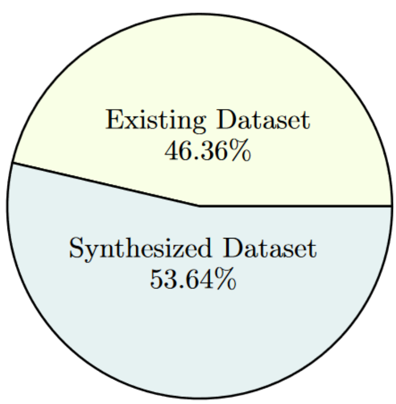
Table 5. Mode of data acquisition

Mode of acquisition	Number of Articles
Web camera	18
Smartphone camera	3
Microsoft Kinect camera	17
Data glove	11
Leap motion sensor	6
CleveMed BioRadio 150	1

4.4.2. RQ2.2: Is It a Synthesized Dataset or An Existing Dataset?

It is found that there are two categories of datasets used. Namely, synthesized datasets and existing datasets as shown in Figure 7. It is observed that a little over half the articles have synthesized their datasets, while the rest have collected data from existing ones.

Figure 7. Type of dataset



Synthesized datasets are the datasets generated by the authors of research articles, whereas existing datasets are the ones that are available as open-source databases. Existing datasets include ASLLVD, ATIS corpus, RWTH-PHOENIX-Weather, etc., and a detailed explanation regarding the utilization of the same is given in the following paragraphs.

The decision to choose between the two types depends on the model requirements. For example, if the system uses ASL, it is relatively easy to find existing open-source datasets. Whereas sieving through the internet for legitimate datasets for specific languages might not prove fruitful and researchers elect to go for synthesized datasets in these cases.

A few researchers used salient existing datasets, some of which are mentioned below.

- RWTH-PHOENIX-Weather 2014 dataset, made for Continuous SLT (Koller et al., 2015) contains approximately 45,670 samples. This dataset was used by (Adaloglou et al., 2020; Cui et al., 2017, 2019; Kamal et al., 2019; Pu et al., 2019; Xiao et al., 2020; Zhou et al., 2021). Another variant of the same dataset called RWTH-PHOENIX-Weather 2014 T was used by articles (Camgoz et al., 2020; Chen et al., 2022; Kamal et al., 2019; Moryossef et al., 2021; Natarajan et al., 2022; Saunders et al., 2020; Stoll, Camgoz, et al., 2020; Zhao et al., 2021; Zhou et al., 2021). Phoenix SD (Forster et al., 2014) and Phoenix SI (Camgoz et al., 2018) were used by (Adaloglou et al., 2020). These datasets are collected from the weather forecasts of a German TV station called Phoenix.
- Kaggle's Sign Language MNIST Dataset (tecperson, 2017), one of the most popular datasets, was used by (Tiku et al., 2020), (Chakraborty et al., 2018), (Abiyev et al., 2020), (Makarov et al., 2019) and (Fernandes et al., 2020). The dataset contains 27,455 training samples and 7,172 test samples.
- The WLASL dataset (D. Li et al., 2020) is the world's largest video-based dataset for ASL with over 21,000 samples. It is a collection of videos of approximately 2000 ASL words. This dataset was used by (Hosain et al., 2021) in their work.
- The SIGNUM corpus (von Agris et al., 2008), was widely used because of its availability and usability. It is a German Sign Language (Deutsche Gebärdensprache - DGS) dataset. (Adaloglou et al., 2020; Cui et al., 2019; Kamal et al., 2019) used the same for their work.
- RWTH-BOSTON-104 and RWTH-BOSTON-400 (Dreuw et al., 2008) were used by (Kamal et al., 2019). In their work, they analyze SLR works on numerous different existing datasets.
- ATIS (Bungeroth et al., 2008) and ASSLVD (Athitsos et al., 2008) SL Corpus are well-known datasets. ASSLVD has been used by (Kamal et al., 2019), (Bilge et al., 2022), and (Zheng et al., 2017). While ASLLVD consists of ASL gestures, the ATIS Sign Language corpus contains five different SL datasets - German Sign Language (DGS), South African Sign Language, Irish Sign Language, German and English. (Kamal et al., 2019) have also used the ATIS corpus in their work.
- Other notable datasets are DEVISIGN-L, USTC-SLT, and MSR Gesture3D; these are Chinese Sign Language (CSL) datasets. ISL-CSLTR, a Mendeley database for Indian Sign Language Recognition and translation was used by (Natarajan et al., 2022).
- Czech weather forecasting dataset was used by (Zelinka & Kanis, 2020). Bangladeshi Sign Language datasets BdSLHD-D1500 and 2300 were used by (Podder et al., 2022). Researchers (Pu et al., 2019), (C. Wei et al., 2019), and (Han et al., 2022) used the Continuous CSL and CSL-500 datasets.

The rest used synthesized datasets for their SLT and SSL implementations. Table 6 summarizes some of the different open-source SL datasets along with the SL they are based on.

Table 6. Open-source datasets used in research articles

Dataset	Sign Language
ATIS Corpus	German, Irish, and South African.
RWTH-PHOENIX-Weather 2014	German
RWTH-PHOENIX-Weather 2014 T	German
Phoenix SD	German
SIGNUM Corpus	German
SIGNUM SI	German
ISL-CSLTR	Indian
WLASL	American
ASLLVD	American
Sign Language MNIST	American
KETI dataset	Korean
DEVISIGN-L	Chinese
MSR Gesture3D	Chinese
CSL - 500	Chinese
Continuous CSL	Chinese
USTC-SLTR	Chinese
Czech weather forecasting dataset	Czech
BdSLHD-D1500	Bangladeshi
BdSLHD-D2300	Bangladeshi

4.4.3. RQ2.3: What Is the Size of The Dataset; I.E., How Many Samples Are Present In Them?

A recognition-based dataset would consist of images or videos as its data. They might be in the form of annotated images, pre-processed images that are converted into features, or video snippets. As mentioned in RQ2.2, a few datasets were video-based, i.e., datasets containing snippets of videos in which the SL gestures are captured in motion. Articles (An et al., 2016; Bheda & Radpour, 2017; Gupta et al., 2016; Masood, Thuwal, et al., 2018; Sahoo, 2021; Sruthi & Lijiya, 2019; Tiku et al., 2020) captured static images from a sequential input.

Table 7 includes the count of the number of articles using a particular range of samples in their datasets. A point to note is that the statistics mentioned in the table only account for the articles in which the authors have mentioned the number of samples present (Adaloglou et al., 2020; F. Ahmed

Table 7. Range of the number of samples used

Sample Size	Number of articles
>10,000	30
5,000 - 10,000	8
2,500 - 5,000	7
<2,500	15

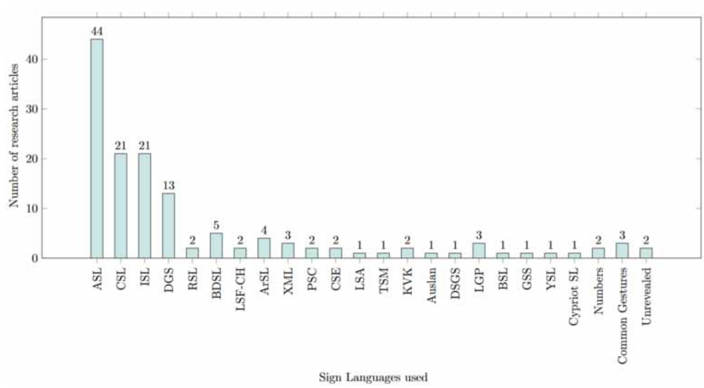
et al., 2016; Akoum & al Mawla, 2015; Aliyu et al., 2017; Ameen & Vadera, 2017; Amin et al., 2022; Barbhuiya et al., 2021; Bilge et al., 2022; Camgoz et al., 2020; Chen et al., 2022; Cui et al., 2017, 2019; Escudeiro et al., 2015; Fernandes et al., 2020; Fok et al., 2015; Han et al., 2022; Hasan et al., 2016; Hosain et al., 2021; Huang et al., 2015; M. Islam et al., 2018; Jiang et al., 2021; Kamal et al., 2019; Ko et al., 2018; Ku et al., 2019; Kumar et al., 2017a, 2017b; K. Li et al., 2016; Liao et al., 2019; Liu et al., 2016; Makarov et al., 2019; Mariappan & Gomathi, 2019; Mittal et al., 2019; Moryossef et al., 2021; Nandi et al., 2022; Natarajan et al., 2022; Oliveira et al., 2019b; Park et al., 2021; Patel et al., 2020; Podder et al., 2022; Pu et al., 2019; Quesada et al., 2017; Rahaman et al., 2018; Sahoo, 2021; Saunders et al., 2020; Savur & Sahin, 2015; Sridevi et al., 2018; Stoll, Camgoz, et al., 2020; Stoll, Hadfield, et al., 2020; Tiku et al., 2020; Truong et al., 2016; Vijayalakshmi & Aarthi, 2016; Wang et al., 2015; C. Wei et al., 2019; Xiao et al., 2019, 2020; Xu et al., 2021; Zhang et al., 2016; Zhao et al., 2021; Zhou et al., 2021).

4.4.4. RQ2.4: Which Sign Languages Are Used in SLT systems?

With over 300 sign languages in use and new ones recurrently emerging, SLT systems play a huge role in the day-to-day lives of many people. Some countries share sign languages, but under different names at times. For instance, Serbian Sign Language and Croatian Sign Language belong to the same SL family.

- It can be seen in Figure 8 that ASL was the most commonly used language (Abiyev et al., 2020; Abraham & Rohini, 2018; Akoum & al Mawla, 2015; Ameen & Vadera, 2017; Amin et al., 2022; Arif et al., 2016; Bajpai et al., 2015; Bantupalli & Xie, 2018; Barbhuiya et al., 2021; Bharti et al., 2019; Bheda & Radpour, 2017; Bilge et al., 2022; Bragg et al., 2021; Chakraborty et al., 2018; Cooper et al., 2011; Daroya et al., 2018; Fernandes et al., 2020; Fok et al., 2015; Hosain et al., 2021; M. M. Islam et al., 2017; Kang, 2019; Ku et al., 2019; Madhiarasan et al., 2022; Makarov et al., 2019; Masood, Thuwal, et al., 2018; Moryossef et al., 2021; Natarajan et al., 2022; Nath & Arun, 2017; Quesada et al., 2017; Rahman et al., 2019; Rastgoo et al., 2021; Savur & Sahin, 2015; Sharma & Singh, 2021; Shivashankara & Srinath, 2018; Sridevi et al., 2018; Tao et al., 2018; Taskiran et al., 2018; Tiku et al., 2020; Truong et al., 2016; Vijayalakshmi & Aarthi, 2016; Warrier et al., 2016; Wen et al., 2021; Wu et al., 2016; Ye et al., 2018). The motivation behind utilizing ASL for research is the abundance of existing datasets found for the same and their ease of accessibility, as mentioned in RQ2.2.
- CSL incorporates the use of both hands, making SLT tricky. And yet, it is the second most widely used SL in the field of research. There are 21 articles that have used CSL (Amin et al., 2022; An et al., 2016; Chen et al., 2022; Han et al., 2022; Kamal et al., 2019; Kang, 2019; K. Li et al., 2016; Liao et al., 2019; Liu et al., 2016; Pu et al., 2019; Qin et al., 2021; Rahaman et al., 2018; Wang et al., 2015; C. Wei et al., 2019; Xiao et al., 2019, 2020; Xu et al., 2021; Yang & Zhu, 2017; Zhang et al., 2016; Zhao et al., 2021; Zhou et al., 2021).

Figure 8. Sign languages used



- Indian Sign Language (ISL) (Ahire et al., 2015; Amin et al., 2022; Bhagat et al., 2019; das Chakladar et al., 2021; Dhanjal & Singh, 2021; Dutta et al., 2015; Gupta et al., 2016; Katoch et al., 2022; Kumar et al., 2017a, 2017b; Mariappan & Gomathi, 2019; Mittal et al., 2019; Nandi et al., 2022; Natarajan et al., 2022; Patel et al., 2020; Rao et al., 2018; Sahoo, 2021; Sharma & Singh, 2021; Sruthi & Lijiya, 2019; Tripathi & Nandi, 2015; Venugopalan & Reghunadhan, 2021) and German Sign Language (DGS) (Camgoz et al., 2020; Chen et al., 2022; Cui et al., 2017, 2019; Moryossef et al., 2021; Natarajan et al., 2022; Pu et al., 2019; Saunders et al., 2020; Stoll, Camgoz, et al., 2020; Stoll, Hadfield, et al., 2020; Tornay et al., 2019; Zhao et al., 2021; Zhou et al., 2021) are other widely used sign languages.
- Some researchers collected common gestures across all Sign Languages and made SLT systems. Researchers (Huang et al., 2015), (Rajaganapathy et al., 2015), and (Cheok et al., 2019) have used these common gestures for their implementations.

4.5 RQ3: What Are the Different Pre-Processing Techniques Used In SLT Systems?

Pre-processing techniques minimize the computational complexity of data processing. Different pre-processing approaches are used for different implementations of SLT.

Concerning hardware-enabled implementations of SLT, (Fok et al., 2015) have pre-processed their datasets by transforming them using data/shape registration so the data can be used by reference sensors. Their solution for transformation involved using the Kabsch algorithm, which is a method for determining the best rotation matrix between two paired sets of points that minimizes the Root Mean Squared Deviation (RMSD). After data alignment, they used the covariance matrix for the fusion of these matrices using covariance intersection. (Wu et al., 2016) synchronized the values from two sensors and converted them into one feature. (Savur & Sahin, 2015) pre-processed raw signal data using the band-pass and notch filters. (Kumar et al., 2017a, 2017b) normalized data from two sensors to synchronize them.

Researchers (Bajpai et al., 2015), (Escudeiro et al., 2015), (Fernandes et al., 2020), (Vijayalakshmi & Aarthi, 2016), (Shukor et al., 2015), (Aliyu et al., 2017), (Xiao et al., 2019), (Estrada Jiménez et al., 2017) , (Quesada et al., 2017), (K. Li et al., 2016), (Arif et al., 2016) and (Phing et al., 2019) chose not to pre-process their sensory inputs.

In the computer vision approach, however, there are a plethora of pre-processing techniques used. The various techniques used for pre-processing image and video data are shown below.

Normalization is the process of scaling individual samples to have a unit norm and was used in articles (Amin et al., 2022; Madhiarasan et al., 2022; Mittal et al., 2019; Nandi et al., 2022; Rahman et al., 2019). Morphological transformations such as erosion, dilation, opening, and closing were used for noise reduction by the researchers (Bhagat et al., 2019; Dutta et al., 2015; Gupta et al., 2016; Madhiarasan et al., 2022; Mariappan & Gomathi, 2019; Nikam & Ambekar, 2016; Truong et al., 2016; Warrier et al., 2016). (Cheok et al., 2019), (Bhagat et al., 2019), (Nath & Arun, 2017), and (Katoch et al., 2022) used Gaussian blurring, a mathematical function applied to an image to blur it and reduce the image background component. (Bheda & Radpour, 2017) used background-subtraction techniques on their image inputs.

The simple but effective pre-processing method of resizing images was utilized in numerous articles (Adaloglou et al., 2020; Akoum & al Mawla, 2015; Barbhuiya et al., 2021; Cui et al., 2017; Kamruzzaman, 2020; Kodandaram et al., 2021; Makarov et al., 2019; Nandi et al., 2022; Rahman et al., 2019; Rao et al., 2018; Rocha et al., 2020; Sahoo, 2021; Sharma & Singh, 2021; Truong et al., 2016), as using appropriately sized images helps in making the model perform efficiently. For better performance, many have chosen to convert images to grayscale (M. Islam et al., 2018; Makarov et al., 2019; Rahman et al., 2019; Rocha et al., 2020; Sahoo, 2021; Shivashankara & Srinath, 2018; Truong et al., 2016; Warrier et al., 2016).

(Nath & Arun, 2017), (Cooper et al., 2011) and (Warrier et al., 2016) used the image edge detection technique to determine the borders of objects within the images. Another variant of this technique called Canny Edge Detection was employed by (Tiku et al., 2020) and (Katoch et al., 2022). (Bantupalli & Xie, 2018) and (Ye et al., 2018) acquired input in the form of a video and converted it into multiple RGB frames. Researchers (Nikam & Ambekar, 2016), (Bhagat et al., 2019), (Nath & Arun, 2017), and (Katoch et al., 2022) used skin segmentation and color filtering, which divide a picture into sections depending on color, shape, or texture, for pre-processing their images.

Another widely used pre-processing technique was converting the signs in images to their skeletal form. This was done using OpenPose (Cao et al., 2017, 2019; Simon et al., 2017; S.E. Wei et al., 2016), a real-time multi-person system that detects key points on the human body and forms a 2D or 3D skeletal structure. Researchers (Gago et al., 2019; Hosain et al., 2021; Ku et al., 2019; Rastgoo et al., 2020; Stoll, Camgoz, et al., 2020; Tiku et al., 2020; Tornay et al., 2019) employed this technique. Another technique, thresholding, is used to further process a grayscale image by creating binary images. Here, a predetermined threshold is chosen and any pixels with an intensity greater than that are converted into black pixels, while the remaining pixels are converted to white pixels (Cooper et al., 2011; Gupta et al., 2016; Hasan et al., 2016; M. Islam et al., 2018; Nath & Arun, 2017; Nikam & Ambekar, 2016; Sahoo, 2021; Taskiran et al., 2018; Tripathi & Nandi, 2015; Warrier et al., 2016).

HOG (Histogram of Oriented Gradients), or histogram equalization, can be used to remove unwanted objects. This method, along with filters such as median and moving average, was used by (Sethi et al., 2012). (Saunders et al., 2020) used a novel pre-processing technique called "Progressive Transformers," a counter-decoding technique, to predict continuous sequences of variable lengths by tracking their production progress.

Considering SSL translation, the speech/text input must be pre-processed as a pre-requisite step, and multiple natural language processing techniques were used for this. Lemmatization, the process of reducing a word to its base form (known as a lemma), was the most popular pre-processing technique used. (Patel et al., 2020) and (Zelinka et al., 2019; Zelinka & Kanis, 2020) used this method to preprocess their data. (Patel et al., 2020) used stemming, the method of reducing a word to its word stem and, as a result, removing suffixes and prefixes. They also used stop-word removal, a method of removing commonly occurring words. Different Python libraries have been used for pre-processing text. The NLTK (Natural Language Toolkit) library (Loper & Bird, 2002) was used by (das Chakladar et al., 2021) while (Kang, 2019) used the CoreNLP library (Stanford, 2011).

4.6 RQ4: What Are the Dimensionality Reduction Techniques Used?

Dimensionality reduction techniques are used to reduce the size of the features and the datasets without losing key information present in them. Utilizing these methods helps in decreasing computational complexity. A considerable number of articles have more than 1000 image or video samples in their datasets. Principal Component Analysis (PCA) is the most prevalent dimensionality reduction technique used (Amin et al., 2022; Cheok et al., 2019; Kumar et al., 2017a; Lee et al., 2016; Madhiarasan et al., 2022; Nath & Arun, 2017; Sharma & Singh, 2021; Tiku et al., 2020; Tripathi & Nandi, 2015; Wang et al., 2015; Wen et al., 2021; Zhang et al., 2016).

Researchers (M. M. Islam et al., 2017) used pixel segmentation — highlighting the pixels in an image that represent required features. (Huang et al., 2015) used sub-sampling (select subsets and reduce the size of data); (Gupta et al., 2016) and (Tornay et al., 2019) used correlation features.

Many implementations involving the use of GPUs utilized very powerful algorithms and did not require performing dimensionality reduction. Thus, the usage of dimensionality reduction for one's implementation completely depends on the number of features, the size of the dataset, the model used, and the accuracy obtained.

However, articles (Abiyev et al., 2020; M. Ahmed et al., 2016; Camgoz et al., 2020; Cui et al., 2019; Deriche et al., 2019; Liao et al., 2019; Pu et al., 2019; Rahaman et al., 2018; Rajaganapathy et al., 2015; Rastgoo et al., 2021; Sridevi et al., 2018; Süzgün et al., 2015; Wang et al., 2015; C. Wei et al., 2019) chose not to pre-process their data because it was unnecessary or because their models were robust enough to do without. Additionally, articles that implemented Deep Learning models did not require additional dimensionality reduction, as convolutional layers tend to do it.

4.7 RQ5: What Are the Feature Extraction Techniques Used In SLT Systems?

SLT systems use an image or video input from which features are extracted. The extracted feature inputs can be fed into any algorithm for the classification and identification of SL gestures. SSL translators require the extraction of audio features before being used for translation. There are multiple methods used for feature extraction in both these systems.

Convolution seems to be the go-to feature extraction operation for Deep Learning and object detection models. In the convolution operation, the input feature arrays are fed into the convolutional operator, and this technique has been used in 25 articles. The HOG descriptor is often used in the extraction of features from images, especially in object detection algorithms. It decomposes the images into smaller cells inside which a HOG is calculated. These values are normalized and the feature descriptor for each cell is obtained. This descriptor was used by (Hasan et al., 2016), (Zhang et al., 2016), and (Tripathi & Nandi, 2015).

Long-Short Term Memory (LSTM) was used by (Rastgoo et al., 2021), Mean Absolute Value (MAV), Modified Mean Absolute Value (MMAV), Simple Square Integral (SSI), Root Mean Square (RMS), Log Detector, Average Amplitude Change (AAC), Maximum Fractal Length (MFL), Minimum, Maximum, and Standard Deviation were used for feature extraction by (Savur & Sahin, 2015). Support Vector Machine (SVM) was used as a method of extracting features by (Barbhuiya et al., 2021) and (Rastgoo et al., 2021).

Other methods of feature extraction include sensory data extraction (M. A. Ahmed et al., 2021), bag-of-features, a method of representing the features of images (Sridevi et al., 2018), and Enhanced Shape Context (Singh & Hazarika, 2007) used by (Zhang et al., 2016).

In SSL translation systems, feature extraction must be performed on speech. For the extraction of features from speech, formats, band energies, spectrum, pitch, and variation in pitch were the features used by (Bharti et al., 2019). Some used APIs for speech-to-text conversion and did not need to perform feature extraction (Ahire et al., 2015; M. Ahmed et al., 2016; Bharti et al., 2019). There are many more feature extraction methods for speech, such as Mel Frequency Cepstral Coefficients (MFCC), Deadweight Tonnage (DWT), Linear Prediction Cepstral Coefficients (LPCC), etc.

4.8 RQ6: Model Questions

The scope of this question is to find out more about the models or frameworks used for SLT by researchers with regard to the techniques used; the solutions to the difficulties of different angles of the same sign, the issues faced, and their performances. The systematic literature review conducted by (J. K. Appati et al., 2021) resembles this research question. In their study, findings are classified based on models utilized, accuracy, and performance metrics. Similarly, (Bharathi & Selvarani, 2019) utilized HMM models to carry out assessment of methods in a systematic approach.

4.8.1. RQ6.1: What are the Different Machine Learning / Deep Learning Techniques Used?

Various Machine Learning and Deep Learning algorithms that were used in SLT systems is mentioned in Figure 9 in the form of a waffle chart. HMMs and LSTM models were predominantly used for SLT that involved continuous video input. In addition, CNNs were heavily utilized, followed by SVM and k-NN (Amin et al., 2022; Cheok et al., 2019; Estrada Jiménez et al., 2017; Hasan et al., 2016; Sahoo, 2021; Shukor et al., 2015; Wu et al., 2016). Models such as Dynamic Time Warping algorithm (Escudeiro et al., 2015; Estrada Jiménez et al., 2017; Süzgün et al., 2015; Wang et al., 2015), Viterbi algorithm (Vijayalakshmi & Aarthi, 2016) and AdaBoost (M. Ahmed et al., 2016; Truong et al., 2016) were also used.

Table 8 lays out different pre-trained models used in SLT systems. These models are made specifically to handle real-time inputs.

Many have used hybrid models, as shown in Table 9, to make the system more robust and make SLT faster.

Figure 9. Waffle chart of models used

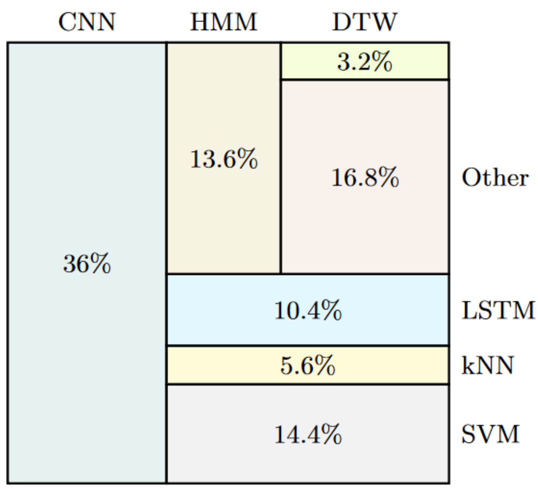


Table 8. Pre-trained models

Computer Vision Models	Research Articles
ResNet+LSTM	(Adaloglou et al., 2020)
GoogLeNet	(Abiyev et al., 2020; Adaloglou et al., 2020)
Single Shot detector	(Abiyev et al., 2020; Chen et al., 2022; Rastgoo et al., 2020, 2021)
ResNet	(C. Wei et al., 2019)
B3D ResNet	(Liao et al., 2019)
AlexNet	(Barbhuiya et al., 2021)
CNN VGG16 Architecture	(Barbhuiya et al., 2021; Gruber et al., 2018; Masood, Thuwal, et al., 2018)
DenseNet	(Daroya et al., 2018)
GoogLeNet + BiLSTM	(Venugopalan & Reghunadhan, 2021)

4.8.2. RQ6.2: Can the Model Identify Different Angles of The Same Sign?

There is no guarantee that each SL user gestures in the same way as the other, i.e., one person may use their hands at different angles to another while signing the same gesture. The SLT systems must recognize these differences between signers.

A Neural Network (NN) having an increased number of dimensions would help the model recognize different angles of the same sign. Microsoft's Kinect for Windows SDK provides users the ability to obtain the skeletal structure of the hand captured using the Kinect V2 sensor. This feature aids in identifying SL gestures from different angles. However, none of the reviewed articles considered this.

Table 9. Algorithms and their variants

Technique	Algorithmic Variants	Research articles
CNN	CNN + LSTM	(Bhagat et al., 2019; Mittal et al., 2019)
	CNN + LSTM + RNN	(Bantupalli & Xie, 2018)
	3D CNN + LSTM	(Adaloglou et al., 2020)
	VGG-S/ GoogLeNet+BiLSTM+CONV1D	(Cui et al., 2017)
	2DCNN	(Rastgoo et al., 2020)
	3DCNN	(Park et al., 2021; Rastgoo et al., 2020)
	CNN + BiLSTM	(Cui et al., 2019)
	Graph CNN (GCNN)	(Sharma & Singh, 2021)
	Segment CNN (Seg-CNN)	(Podder et al., 2022)
	Separable Spatial-Temporal Convolution Network (SSTCN)	(Jiang et al., 2021)
	ConvNet	(Hosain et al., 2021)
SVM	SVM	(Abiyev et al., 2020; Amin et al., 2022; An et al., 2016; Cheok et al., 2019; Cooper et al., 2011; Escudeiro et al., 2015; Hasan et al., 2016; Kamal et al., 2019; Katoch et al., 2022; Lee et al., 2016; Quesada et al., 2017; Savur & Sahin, 2015; Sridevi et al., 2018; Wu et al., 2016)
	SVM with HoG feature descriptor	(Chakraborty et al., 2018; Gupta et al., 2016; Tiku et al., 2020)
	SVM with RBF kernel	(Gruber et al., 2018)
LSTM	LSTM	(Gago et al., 2019; Kamal et al., 2019; Liu et al., 2016; Mittal et al., 2019; Rastgoo et al., 2020, 2021; Xu et al., 2021)
	Bidirectional LSTM (BiLSTM/ BLSTM)	(Bilge et al., 2022; Qin et al., 2021; C. Wei et al., 2019)
	I3D+LSTM	(Adaloglou et al., 2020)
	RNN + LSTM	(Xiao et al., 2020)
	BLSTM - NN	(Kumar et al., 2017a)
HMM	HMM	(Cheok et al., 2019; Fok et al., 2015; Kumar et al., 2017a; Lee et al., 2016; Liu et al., 2016; Tornay et al., 2019; Vijayalakshmi & Aarthi, 2016; Wang et al., 2015; Zhou et al., 2021)
	HMM+ BLSTM-NN	(Zheng et al., 2017)
	Hybrid CNN + HMM	(Zheng et al., 2017)
	Adaptive HMM	(Zhang et al., 2016)
	Parallel HMM	(Tornay et al., 2019)
	Coupled HMM	(Kumar et al., 2017b)
	Light HMM	(Wang et al., 2015)
	Gaussian Mixture Models - HMM (GMM-HMM)	(Kamal et al., 2019)

4.8.3. RQ6.3: What Are the Challenges In Using These Models For SLT Systems?

The main text format consists of flat left-right SLT systems that often require real-time conversion of sign languages. Therefore, Machine Learning and Deep Learning techniques used in developing SLT applications should be able to make predictions with very low latency. Moreover, the conditions related to the movement of application users, lighting, system capabilities, noise, etc. play an important role in SLT systems and SSL translation systems. Specific challenges faced by researchers are mentioned below.

- (Shukor et al., 2015) made a data glove that runs with the help of an Arduino board. In their work, the system they developed is strapped around the hand, which slightly restricts hand movement. Consequently, the researchers mention that the accuracy of their glove was reduced when there was a lot of hand movement.
- (Nandi et al., 2022) used a CNN model for ISL translation. But ISL contains both static and dynamic SL gestures. Although their model had a high accuracy of 99.52%, they stated that it performed better on static inputs than dynamic ones.
- (Bantupalli & Xie, 2018) proposed a combination network that included CNN, LSTM, and RNN for SLT. They observed a dip in performance when faces were included in the input and signers wore different clothing.
- (Cooper et al., 2011) stated that their multi-class SVM model was extremely sensitive to light and background cluttering.
- (M. Ahmed et al., 2016) claimed that their model performs better if the user is not Asian, as they inferred that their model had difficulty detecting edges. Lastly, some of the common challenges faced in SLT systems are mentioned in Table 10.

Table 10. Challenges observed in SLT systems

SLT Challenges	Observations
Computation speed and time	Complex layered models can take a lot of computation time.
Dynamic and non-uniform background environment	Noise, improper classification of hands, and faces affect the performance and mislead the SLT system.
Image orientation problem	Different angles of the same sign can be perceived differently by the same model.
Illumination of light	Most models employ the RGB model, whose performance varies with different lighting circumstances, hence making models extremely sensitive to lighting.

4.8.4. RQ6.4: What Is the Performance Of The Prediction Techniques?

This research question explores the recognition and translation accuracies of techniques covered in RQ6.1. Researchers who have employed hardware implementations or used very few classes and data samples in their work have observed a 100% accuracy for some of the SL gestures, such as (Fok et al., 2015) and (Vijayalakshmi & Aarthi, 2016). Similarly, (Quesada et al., 2017) attained an accuracy of 100% using the same approach with leap motion sensors and Microsoft's Kinect sensor.

Furthermore, several implementations of vision-based algorithms also attained 100% accuracy., who implemented the Dynamic Time Warping algorithms, HMM and SVM, respectively, attained 100% accuracies for numerous classes. The SL MNIST dataset gave respectable accuracies, ranging from 95% to 100%. and used this dataset. obtained a fairly high accuracy of 99% while training but saw a drop in the recognition rate (averaging around 91%) when different faces were involved in the testing dataset. Researchers who employed CNNs and their variants, as mentioned in Table 9, obtained accuracy ranging from 80% to 99%.

(Liu et al., 2016) used HMMs and LSTMs for CSL recognition but obtained very varied accuracies (ranging from 11 - 80%), having 125,000 samples in their dataset. They mentioned that their dataset used gestures involving the whole forearm, which was the cause of the resultant accuracy. (Kamruzzaman, 2020) compared model performance before and after image augmentation and saw a solid 5% increase in recognition accuracy. SVM also had varying performances, with accuracies ranging from 50% in some cases (Gruber et al., 2018; Lee et al., 2016) to 99% in others (Abiyev et al., 2020; An et al., 2016; Cheok et al., 2019; Cooper et al., 2011; Escudeiro et al., 2015; Gupta et al., 2016; Hasan et al., 2016; Katoch et al., 2022; Quesada et al., 2017; Savur & Sahin, 2015; Sridevi et al., 2018; Tiku et al., 2020; Wu et al., 2016).

Ultimately, it is observed that the recognition accuracy is affected by pre-processing techniques and the quality of the datasets used.

4.9 RQ7: What Are the Existing Applications Available to the General Public and Enabling Technologies For SLT and Recognition?

Of the articles reviewed, a few researchers made applications that can be used for real-time SLT. Among them, only one implementation called HOSPISIGN has been made available to the general public. This platform was created by (Süzgün et al., 2015). It is a Turkish SL translator, that implements bidirectional SLT, and is a real-time web-based platform made to help people who are hard of hearing during their hospital visits. Although it has limited functionality, it performs exceptionally well, with accuracy ranging from 83% to 100% for some classes.

Similarly, (Taskiran et al., 2018) have implemented a real-time system for SLT. However, the article does not talk about the system's availability to the general public. (Park et al., 2021) have implemented a SLT system on mobile platforms, but this application is not open-source.

(Rayner et al., 2016) made a web-based platform that converts spoken French to Swiss French SL (LSF-CH). The application is completely dependent on third-party software, making it very easy to use and implement, not just for LSF-CH but other sign languages as well.

4.10 RQ8: What Are the Evaluation Metrics Used in SLT Systems?

Once data is acquired and the model is trained, they must be evaluated and their performance needs to be assessed using evaluation metrics.

Accuracy - the measurement of the correctness of a particular prediction, was the most widely used evaluation metric with multiple researchers (Abiyev et al., 2020; Adaloglou et al., 2020; Akoum & al Mawla, 2015; An et al., 2016; Bajpai et al., 2015; Barbhuiya et al., 2021; Bhagat et al., 2019; Bharti et al., 2019; Cheok et al., 2019; Cooper et al., 2011; Fernandes et al., 2020; Fok et al., 2015; Gupta et al., 2016; Han et al., 2022; Hasan et al., 2016; Huang et al., 2015; Jiang et al., 2021; Kamruzzaman, 2020; Katoch et al., 2022; Ko et al., 2018; Kodandaram et al., 2021; Lee et al., 2016; Liao et al., 2019; Liu et al., 2016; Madhiarasan et al., 2022; Mariappan & Gomathi, 2019; Masood, Srivastava, et al., 2018; Masood, Thuwal, et al., 2018; Park et al., 2021; Patel et al., 2020; Podder et al., 2022; Pu et al., 2019; Rajaganapathy et al., 2015; Rao et al., 2018; Rastgoo et al., 2020, 2021; Sahoo, 2021; Savur & Sahin, 2015; Sharma & Singh, 2021; Shukor et al., 2015; Sruthi & Lijiya, 2019; Süzgün et al., 2015; Taskiran et al., 2018; Tiku et al., 2020; Tornay et al., 2019; Vijayalakshmi & Aarthi, 2016; Wen et al., 2021; Xiao et al., 2020; Xu et al., 2021; Yang & Zhu, 2017; Zhang et al., 2016) using it.

Other common evaluation metrics used were precision (Adaloglou et al., 2020; An et al., 2016; Escudeiro et al., 2015; Podder et al., 2022; Pu et al., 2019; Sahoo, 2021; Truong et al., 2016; Venugopalan & Reghunadhan, 2021; C. Wei et al., 2019), recall (Abiyev et al., 2020; Adaloglou et al., 2020; Truong et al., 2016; Venugopalan & Reghunadhan, 2021); sensitivity (Podder et al., 2022; Truong et al., 2016; Wu et al., 2016) and specificity (Podder et al., 2022; Truong et al., 2016).

(Podder et al., 2022) and (Truong et al., 2016) have presented a combination of these four metrics in the form of a confusion matrix. Another evaluation metric called the F1-score, a combination of the metrics precision and recall, was used by several researchers (Abiyev et al., 2020; Adaloglou et al., 2020; Podder et al., 2022; Sahoo, 2021; Sruthi & Lijiya, 2019; Truong et al., 2016; Venugopalan & Reghunadhan, 2021).

The model error, Mean Squared Error (MSE), and Root Mean Squared Error (RMSE) were other evaluation metrics used in articles (Abiyev et al., 2020; Adaloglou et al., 2020; An et al., 2016; Cheok et al., 2019; Rahman et al., 2019; Rastgoo et al., 2020; Stoll, Camgoz, et al., 2020; Zelinka et al., 2019; Zelinka & Kanis, 2020). (Bheda & Radpour, 2017) used categorical cross-entropy loss and (Chen et al., 2022) used Connectionist Temporal Classification (CTC) loss (Graves, 2012). Articles (Abiyev et al., 2020; Adaloglou et al., 2020; Chen et al., 2022; Gago et al., 2019; Gruber et al., 2018; Sharma & Singh, 2021; Zelinka et al., 2019) used translation loss as their evaluation metric.

(Tiku et al., 2020), (Qin et al., 2021) and (Venugopalan & Reghunadhan, 2021) have used temporal evaluation metrics, which include model speed and the average time taken to process an image (in ms). (Bharti et al., 2019) used the total time in delay, while (Sruthi & Lijiya, 2019) and (Madhiarasan et al., 2022) used the recognition rate of the model as their evaluation metric.

Newer evaluation metrics include Word Error Rate (WER) used in articles (M. A. Ahmed et al., 2021; Cheok et al., 2019; Kang, 2019; Madhiarasan et al., 2022; Pu et al., 2019; Qin et al., 2021; Zhou et al., 2021), Sentence Error Rate (SER) used by (Kang, 2019), Bilingual Evaluation Understudy (BLEU) used by (das Chakladar et al., 2021; Moryossef et al., 2021; Natarajan et al., 2022; Pu et al., 2019; Qin et al., 2021; C. Wei et al., 2019; Zhou et al., 2021) and Cross-lingual Optimized Metric for Evaluation of Translation (COMET) used by (Moryossef et al., 2021).

Researchers (Pu et al., 2019) and (C. Wei et al., 2019) used other metrics that included Consensus-based Image Description Evaluation (CIDEr), Recall-Oriented Understudy for Gisting Evaluation (ROUGE-L), and METEOR. (Aliwy & Ahmed, 2021) used their evaluation metric where the predicted Sign Language gestures were given to experts who rated them on a scale of 1 to 5.

4.11 RQ9: What Are the Validation Techniques Used in SLT Systems?

(Vijayalakshmi & Aarthi, 2016), (Bajpai et al., 2015), (Rayner et al., 2016), and (Phing et al., 2019) have used self-validation; the researchers confirm the correctness of the predicted SL gesture themselves, eliminating the need for test data.

Meanwhile, articles (Abiyev et al., 2020; M. A. Ahmed et al., 2021; Ameen & Vadera, 2017; Amin et al., 2022; Bilge et al., 2022; Chakraborty et al., 2018; Daroya et al., 2018; Dhanjal & Singh, 2021; Gruber et al., 2018; Han et al., 2022; Hasan et al., 2016; M. Islam et al., 2018; Jiang et al., 2021; Katoch et al., 2022; Madhiarasan et al., 2022; Masood, Srivastava, et al., 2018; Nandi et al., 2022; Park et al., 2021; Podder et al., 2022; Qin et al., 2021; Rao et al., 2018; Rocha et al., 2020; Savur & Sahin, 2015; Sharma & Singh, 2021; Sridevi et al., 2018; Süzgün et al., 2015; Tao et al., 2018; Taskiran et al., 2018; Wen et al., 2021; Yang & Zhu, 2017) have employed cross-validation, thereby dividing data into three parts - training, testing, and cross-validation datasets.

Both (Barbhuiya et al., 2021) and (Wang et al., 2015) utilized leave-one-out cross-validation (LOOCV), a form of cross-validation method in which each observation serves as the validation set while the remaining observations serve as the training set. They observed that it performed well on their dataset. (Das Chakladar et al., 2021) employed an approach called Absolute Category Rating. This classification is based on absolute values; for example, poor, excellent, good, and fair.

5. CONCLUSION

Although there is a vast number of systematic literature reviews when it comes to SLT and SSL translation systems, they do not comprehensively discuss bidirectional SLT systems, pre-trained models utilized, atypical observations, advantages and drawbacks of certain techniques, recent trends, and an in-depth evaluation of review articles. Taking this into account, the literature review provides explanations for the same and helps give disparate methods that can be beneficial in improving performance.

The systematic literature review was done in an effort to spare upcoming researchers time and effort by assisting them in understanding the introduction, requirements, uses, and procedures involved in these translation systems. The study provides important results that assist in creating better SLT systems.

During the review, it is noticed that most of the work is present in a one-way mode of communication, with 90.4% of articles implementing unidirectional SLT systems. In contrast, there is a dearth of bidirectional SLT systems, with only 9.6% of the research articles implementing them. RQ2 gave insights into the datasets utilized and modes of data acquisition that can be used. Despite the apparent increase in the number of benchmark datasets, synthesized datasets are used by 53.64% of the research articles because of their convenience of use. Articles published in 2022, however, majorly used existing datasets and achieved higher accuracies compared to past work with the same datasets. Additionally, this research question discusses the different sign languages in use for SLT systems, with ASL being the most widely used, with over 35% of the articles using it.

RQs 3-5 incorporate the pre-processing, feature extraction, and dimensionality techniques in use. Pre-processing techniques vary based on the type of implementation (hardware or software) of SLT systems. For instance, there are significantly fewer techniques used for pre-processing speech signals in SSL translation systems. With respect to dimensionality reduction, PCA was the most commonly used technique, with 0.1% of the articles using it.

Through the years, with the increase in computational capabilities of computers, there has been a notable shift from hardware-based implementations to the use of powerful Machine Learning and Deep Learning models such as CNN. Because of their efficacy in image classification applications, pre-trained models have been widely employed in recent years. Moreover, there has been a shift from using simple evaluation metrics like accuracy and precision to complex ones such as COMET, ROUGE-L, etc. in recent years.

Future research directions for this SLR are to focus on real-time and bidirectional SLT systems, consider the different ways SL gestures are held by different people, and take into account the environmental conditions such as lighting and background. Other notable directions are to improve object detection models to detect SL gestures in sign languages that involve the use of two hands and to improve the bounding boxes drawn by the object detection algorithms to accommodate both hands in the same bounding box.

In conclusion, the article is written in the hope of helping researchers broaden their knowledge and identify research gaps in accordance with existing work on SLT systems.

CONFLICT OF INTEREST

The authors of this publication declare there is no conflict of interest.

REFERENCES

Abiyev, R. H., Arslan, M., & Idoko, J. B. (2020). Sign language translation using deep convolutional neural networks. *Transactions on Internet and Information Systems (Seoul)*, *14*(2), 631–653.

Abraham, A., & Rohini, V. (2018). Real time conversion of sign language to speech and prediction of gestures using Artificial Neural Network. *Procedia Computer Science*, *143*, 587–594. doi:10.1016/j.procs.2018.10.435

Adaloglou, N., Chatzis, T., Papastratis, I., Stergioulas, A., Papadopoulos, G. T., Zacharopoulou, V., Xydopoulos, G. J., Atzakas, K., Papazachariou, D., & Daras, P. (2020). *A comprehensive study on sign language recognition methods.* ArXiv Preprint ArXiv:2007.12530

Ahire, P. G., Tilekar, K. B., Jawake, T. A., & Warale, P. B. (2015). Two way communicator between deaf and dumb people and normal people. *2015 International Conference on Computing Communication Control and Automation*, 641–644. doi:10.1109/ICCUBEA.2015.131

Ahmed, F., Bouillon, P., Destefano, C., Gerlach, J., Hooper, A., Rayner, M., Strasly, I., Tsourakis, N., & Weiss, C. (2016). Rapid construction of a web-enabled medical speech to sign language translator using recorded video. *International Workshop on Future and Emerging Trends in Language Technology*, 122–134.

Ahmed, M., & Idrees, M., ul Abideen, Z., Mumtaz, R., & Khalique, S. (2016). Deaf talk using 3D animated sign language: A sign language interpreter using Microsoft's kinect v2. *2016 SAI Computing Conference (SAI)*, 330–335. doi:10.1109/SAI.2016.7556002

Ahmed, M. A., Zaidan, B. B., Zaidan, A. A., Alamoodi, A. H., Albahri, O. S., Al-Qaysi, Z. T., Albahri, A. S., & Salih, M. M. (2021). Real-time sign language framework based on wearable device: Analysis of MSL, DataGlove, and gesture recognition. *Soft Computing*, *25*(16), 11101–11122. doi:10.1007/s00500-021-05855-6

Akoum, A., & al Mawla, N. (2015). Hand gesture recognition approach for asl language using hand extraction algorithm. *Journal of Software Engineering and Applications*, *8*(08), 419–430. doi:10.4236/jsea.2015.88041

Aliwy, A. H., & Ahmed, A. A. (2021). Development of arabic sign language dictionary using 3D avatar technologies. *Indonesian Journal of Electrical Engineering and Computer Science*, *21*(1), 609–616. doi:10.11591/ijeecs.v21.i1.pp609-616

Aliyu, S., Mohandes, M., & Deriche, M. (2017). Dual LMCs fusion for recognition of isolated Arabic sign language words. *2017 14th International Multi-Conference on Systems, Signals & Devices (SSD)*, 611–614.

Aljohani, N. R., Aslam, A., Khadidos, A. O., & Hassan, S.-U. (2022). Bridging the skill gap between the acquired university curriculum and the requirements of the job market: A data-driven analysis of scientific literature. *Journal of Innovation & Knowledge*, *7*(3), 100190. doi:10.1016/j.jik.2022.100190

Ameen, S., & Vadera, S. (2017). A convolutional neural network to classify American Sign Language fingerspelling from depth and colour images. *Expert Systems: International Journal of Knowledge Engineering and Neural Networks*, *34*(3), e12197. doi:10.1111/exsy.12197

Amin, M. S., Rizvi, S. T. H., & Hossain, M. M. (2022). A Comparative Review on Applications of Different Sensors for Sign Language Recognition. *Journal of Imaging*, *8*(4), 98. doi:10.3390/jimaging8040098 PMID:35448225

An, X., Yang, H., & Gan, Z. (2016). Towards realizing sign language-to-speech conversion by combining deep learning and statistical parametric speech synthesis. *International Conference of Pioneering Computer Scientists, Engineers and Educators*, 678–690. doi:10.1007/978-981-10-2053-7_61

Appati, J. K., Brown, G. A., Soli, M. A. T., & Denwar, I. W. (2021). A Review of Computational Intelligence Models for Brain Tumour Classification and Prediction. *International Journal of Software Science and Computational Intelligence*, *13*(4), 18–39. doi:10.4018/IJSSCI.2021100102

Appati, J. K., Nartey, P. K., Yaokumah, W., & Abdulai, J.-D. (2022). A Systematic Review of Fingerprint Recognition System Development. *International Journal of Software Science and Computational Intelligence*, *14*(1), 1–17. doi:10.4018/IJSSCI.300358

Arif, A., Rizvi, S. T. H., Jawaid, I., Waleed, M. A., & Shakeel, M. R. (2016). Techno-talk: An american sign language (asl) translator. *2016 International Conference on Control, Decision and Information Technologies (CoDIT)*, 665–670. doi:10.1109/CoDIT.2016.7593642

Athitsos, V., Neidle, C., Sclaroff, S., Nash, J., Stefan, A., Yuan, Q., & Thangali, A. (2008). American sign language Lexicon video dataset (ASLLVD). *Workshop on Human Communicative Behaviour Analysis*.

Bajpai, D., Porov, U., Srivastav, G., & Sachan, N. (2015). Two way wireless data communication and american sign language translator glove for images text and speech display on mobile phone. *2015 Fifth International Conference on Communication Systems and Network Technologies*, 578–585. doi:10.1109/CSNT.2015.121

Bantupalli, K., & Xie, Y. (2018). American sign language recognition using deep learning and computer vision. *2018 IEEE International Conference on Big Data (Big Data)*, 4896–4899. doi:10.1109/BigData.2018.8622141

Barbhuiya, A. A., Karsh, R. K., & Jain, R. (2021). CNN based feature extraction and classification for sign language. *Multimedia Tools and Applications*, *80*(2), 3051–3069. doi:10.1007/s11042-020-09829-y

Bhagat, N. K., Vishnusai, Y., & Rathna, G. N. (2019). Indian sign language gesture recognition using image processing and deep learning. *2019 Digital Image Computing: Techniques and Applications (DICTA)*, 1–8.

Bharathi, R., & Selvarani, R. (2019). Software reliability assessment of safety critical system using computational intelligence. *International Journal of Software Science and Computational Intelligence*, *11*(3), 1–25. doi:10.4018/IJSSCI.2019070101

Bharti, R., Yadav, S., & Gupta, S., & B, R. (2019). Automated Speech to Sign language Conversion using Google API and NLP. *Proceedings of the International Conference on Advances in Electronics, Electrical & Computational Intelligence (ICAEEC)*. doi:10.2139/ssrn.3575439

Bheda, V., & Radpour, D. (2017). *Using deep convolutional networks for gesture recognition in American sign language.* ArXiv Preprint ArXiv:1710.06836.

Bilge, Y. C., Cinbis, R. G., & Ikizler-Cinbis, N. (2022). Towards Zero-shot Sign Language Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1. doi:10.1109/TPAMI.2022.3143074 PMID:35025738

Bragg, D., Caselli, N., Gallagher, J. W., Goldberg, M., Oka, C. J., & Thies, W. (2021). ASL Sea Battle: Gamifying Sign Language Data Collection. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–13. doi:10.1145/3411764.3445416

Bungeroth, J., Stein, D., Dreuw, P., Ney, H., Morrissey, S., Way, A., & van Zijl, L. (2008). *The ATIS Sign Language Corpus*. LREC.

Camgoz, N. C., Hadfield, S., Koller, O., Ney, H., & Bowden, R. (2018). Neural sign language translation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 7784–7793.

Camgoz, N. C., Koller, O., Hadfield, S., & Bowden, R. (2020). Sign language transformers: Joint end-to-end sign language recognition and translation. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 10023–10033. doi:10.1109/CVPR42600.2020.01004

Cao, Z., Hidalgo Martinez, G., Simon, T., Wei, S., & Sheikh, Y. A. (2019). OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Cao, Z., Simon, T., Wei, S.-E., & Sheikh, Y. (2017). *Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields*. CVPR. doi:10.1109/CVPR.2017.143

Chakraborty, D., Garg, D., Ghosh, A., & Chan, J. H. (2018). Trigger detection system for american sign language using deep convolutional neural networks. *Proceedings of the 10th International Conference on Advances in Information Technology*, 1–6. doi:10.1145/3291280.3291783

Chen, Y., Wei, F., Sun, X., Wu, Z., & Lin, S. (2022). *A Simple Multi-Modality Transfer Learning Baseline for Sign Language Translation.* ArXiv Preprint ArXiv:2203. 04287.

Cheok, M. J., Omar, Z., & Jaward, M. H. (2019). A review of hand gesture and sign language recognition techniques. *International Journal of Machine Learning and Cybernetics*, *10*(1), 131–153. doi:10.1007/s13042-017-0705-5

Cooper, H., Holt, B., & Bowden, R. (2011). Sign language recognition. In *Visual analysis of humans* (pp. 539–562). Springer. doi:10.1007/978-0-85729-997-0_27

Cui, R., Liu, H., & Zhang, C. (2017). Recurrent convolutional neural networks for continuous sign language recognition by staged optimization. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 7361–7369. doi:10.1109/CVPR.2017.175

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. doi:10.1109/TMM.2018.2889563

Daroya, R., Peralta, D., & Naval, P. (2018). Alphabet sign language image classification using deep learning. *TENCON 2018-2018 IEEE Region 10 Conference*, 646–650.

das Chakladar, D., Kumar, P., Mandal, S., Roy, P. P., Iwamura, M., & Kim, B.-G. (2021). 3D Avatar Approach for Continuous Sign Movement Using Speech/Text. *Applied Sciences (Basel, Switzerland), 11*(8), 3439.

Deriche, M., Aliyu, S. O., & Mohandes, M. (2019). An intelligent arabic sign language recognition system using a pair of LMCs with GMM based classification. *IEEE Sensors Journal*, *19*(18), 8067–8078. doi:10.1109/JSEN.2019.2917525

Dhanjal, A. S., & Singh, W. (2021). An automatic machine translation system for multi-lingual speech to Indian sign language. *Multimedia Tools and Applications*, 1–39.

Dreuw, P., Forster, J., Deselaers, T., & Ney, H. (2008). Efficient approximations to model-based joint tracking and recognition of continuous sign language. *2008 8th IEEE International Conference on Automatic Face & Gesture Recognition*, 1–6.

Dutta, K. K., GS, A. K., SK, R. K., & Swamy B, S. (2015). Double handed Indian Sign Language to speech and text. *2015 Third International Conference on Image Information Processing (ICIIP)*, 374–377. doi:10.1109/ICIIP.2015.7414799

Escudeiro, P., Escudeiro, N., Reis, R., Lopes, J., Norberto, M., Baltasar, A. B., Barbosa, M., & Bidarra, J. (2015). Virtual sign–a real time bidirectional translator of portuguese sign language. *Procedia Computer Science*, *67*, 252–262. doi:10.1016/j.procs.2015.09.269

Estrada Jiménez, L. A., Benalcázar, M. E., & Sotomayor, N. (2017). Gesture recognition and machine learning applied to sign language translation. *VII Latin American Congress on Biomedical Engineering CLAIB 2016*, 233–236. doi:10.1007/978-981-10-4086-3_59

Fernandes, L., Dalvi, P., Junnarkar, A., & Bansode, M. (2020). Convolutional Neural Network based Bidirectional Sign Language Translation System. *2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT)*, 769–775. doi:10.1109/ICSSIT48917.2020.9214272

Fok, K.-Y., Ganganath, N., Cheng, C.-T., & Chi, K. T. (2015). A real-time asl recognition system using leap motion sensors. *2015 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery*, 411–414. doi:10.1109/CyberC.2015.81

Forster, J., Schmidt, C., Koller, O., Bellgardt, M., & Ney, H. (2014). Extensions of the sign language recognition and translation corpus RWTH-PHOENIX-Weather. *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, 1911–1916.

Gago, J. J., Vasco, V., Łukawski, B., Pattacini, U., Tikhanoff, V., Victores, J. G., & Balaguer, C. (2019). *Sequence-to-sequence natural language to humanoid robot sign language.* ArXiv Preprint ArXiv:1907.04198.

Graves, A. (2012). Connectionist temporal classification. In *Supervised sequence labelling with recurrent neural networks* (pp. 61–93). Springer. doi:10.1007/978-3-642-24797-2_7

Gruber, I., Ryumin, D., Hrúz, M., & Karpov, A. (2018). Sign language numeral gestures recognition using convolutional neural network. *International Conference on Interactive Collaborative Robotics*, 70–77. doi:10.1007/978-3-319-99582-3_8

Gupta, B., Shukla, P., & Mittal, A. (2016). K-nearest correlated neighbor classification for Indian sign language gesture recognition using feature fusion. *2016 International Conference on Computer Communication and Informatics (ICCCI)*, 1–5. doi:10.1109/ICCCI.2016.7479951

Han, X., Lu, F., & Tian, G. (2022). Efficient 3D CNNs with knowledge transfer for sign language recognition. *Multimedia Tools and Applications*, *81*(7), 10071–10090. doi:10.1007/s11042-022-12051-7

Hao, Y. (2019). Bridging the Gap between Translator Training and Practice: Can Theory Help Translators? *Flinders University Languages Group Online Review, 6*(1).

Hasan, M., Sajib, T. H., & Dey, M. (2016). A machine learning based approach for the detection and recognition of Bangla sign language. *2016 International Conference on Medical Engineering, Health Informatics and Technology (MediTec)*, 1–5. doi:10.1109/MEDITEC.2016.7835387

Hoque, M. T., Rifat-Ut-Tauwab, M., Kabir, M. F., Sarker, F., Huda, M. N., & Abdullah-Al-Mamun, K. (2016). Automated Bangla sign language translation system: Prospects, limitations and applications. *2016 5th International Conference on Informatics, Electronics and Vision (ICIEV)*, 856–862.

Hosain, A. A., Santhalingam, P. S., Pathak, P., Rangwala, H., & Kosecka, J. (2021). Hand pose guided 3d pooling for word-level sign language recognition. *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 3429–3439. doi:10.1109/WACV48630.2021.00347

Huang, J., Zhou, W., Li, H., & Li, W. (2015). Sign language recognition using 3d convolutional neural networks. *2015 IEEE International Conference on Multimedia and Expo (ICME)*, 1–6. doi:10.1109/ICME.2015.7177428

Islam, M., Mousumi, S. S. S., Rabby, A. K. M., & Hossain, S. A. (2018). A simple and mighty arrowhead detection technique of Bangla sign language characters with CNN. *International Conference on Recent Trends in Image Processing and Pattern Recognition*, 429–437.

Islam, M. M., Siddiqua, S., & Afnan, J. (2017). Real time hand gesture recognition using different algorithms based on American sign language. *2017 IEEE International Conference on Imaging, Vision & Pattern Recognition (IcIVPR)*, 1–6. doi:10.1109/ICIVPR.2017.7890854

Jaffe, D. L. (1994). Evolution of mechanical fingerspelling hands for people who are deaf-blind. *Journal of Rehabilitation Research and Development*, *31*(3), 236–244. PMID:7965881

Jiang, S., Sun, B., Wang, L., Bai, Y., Li, K., & Fu, Y. (2021). Skeleton aware multi-modal sign language recognition. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 3413–3423.

Kamal, S. M., Chen, Y., Li, S., Shi, X., & Zheng, J. (2019). Technical approaches to Chinese sign language processing: A review. *IEEE Access: Practical Innovations, Open Solutions*, *7*, 96926–96935. doi:10.1109/ACCESS.2019.2929174

Kamruzzaman, M. M. (2020). Arabic sign language recognition and generating Arabic speech using convolutional neural network. *Wireless Communications and Mobile Computing*, *2020*, 2020. doi:10.1155/2020/3685614

Kang, Z. (2019). Spoken language to sign language translation system based on HamNoSys. *Proceedings of the 2019 International Symposium on Signal Processing Systems*, 159–164. doi:10.1145/3364908.3365300

Katoch, S., Singh, V., & Tiwary, U. S. (2022). Indian Sign Language recognition system using SURF with SVM and CNN. *Array*, 100141.

Kitchenham, B., & Charters, S. (2007). *Guidelines for performing Systematic Literature Reviews in Software Engineering*. Academic Press.

Ko, S.-K., Son, J. G., & Jung, H. (2018). Sign language recognition with recurrent neural network using human keypoint detection. *Proceedings of the 2018 Conference on Research in Adaptive and Convergent Systems*, 326–328. doi:10.1145/3264746.3264805

Kodandaram, S. R., & Kumar, N. P., & GL, S. (2021). Sign Language Recognition. *Turkish Journal of Computer and Mathematics Education*, *12*(14), 994–1009.

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. doi:10.1016/j.cviu.2015.09.013

Ku, Y.-J., Chen, M.-J., & King, C.-T. (2019). A Virtual Sign Language Translator on Smartphones. *2019 Seventh International Symposium on Computing and Networking Workshops (CANDARW)*, 445–449. doi:10.1109/CANDARW.2019.00084

Kumar, P., Gauba, H., Roy, P. P., & Dogra, D. P. (2017a). A multimodal framework for sensor based sign language recognition. *Neurocomputing*, *259*, 21–38. doi:10.1016/j.neucom.2016.08.132

Kumar, P., Gauba, H., Roy, P. P., & Dogra, D. P. (2017b). Coupled HMM-based multi-sensor data fusion for sign language recognition. *Pattern Recognition Letters*, *86*, 1–8. doi:10.1016/j.patrec.2016.12.004

Lee, G. C., Yeh, F.-H., & Hsiao, Y.-H. (2016). Kinect-based Taiwanese sign-language recognition system. *Multimedia Tools and Applications*, *75*(1), 261–279. doi:10.1007/s11042-014-2290-x

Li, D., Rodriguez, C., Yu, X., & Li, H. (2020). Word-level deep sign language recognition from video: A new large-scale dataset and methods comparison. *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 1459–1469. doi:10.1109/WACV45572.2020.9093512

Li, K., Zhou, Z., & Lee, C.-H. (2016). Sign transition modeling and a scalable solution to continuous sign language recognition for real-world applications. *ACM Transactions on Accessible Computing*, *8*(2), 1–23. doi:10.1145/2850421

Liao, Y., Xiong, P., Min, W., Min, W., & Lu, J. (2019). Dynamic sign language recognition based on video sequence with BLSTM-3D residual networks. *IEEE Access: Practical Innovations, Open Solutions*, *7*, 38044–38054. doi:10.1109/ACCESS.2019.2904749

Liu, T., Zhou, W., & Li, H. (2016). Sign language recognition with long short-term memory. *2016 IEEE International Conference on Image Processing (ICIP)*, 2871–2875. doi:10.1109/ICIP.2016.7532884

Loper, E., & Bird, S. (2002). NLTK: The Natural Language Toolkit. *Clinical Orthopaedics and Related Research*. http://dblp.uni-trier.de/db/journals/corr/corr0205.html#cs-CL-0205028

Madhiarasan, D. M., Roy, P., & Pratim, P. (2022). *A Comprehensive Review of Sign Language Recognition: Different Types, Modalities, and Datasets.* ArXiv Preprint ArXiv:2204.03328.

Makarov, I., Veldyaykin, N., Chertkov, M., & Pokoev, A. (2019). American and russian sign language dactyl recognition. *Proceedings of the 12th ACM International Conference on PErvasive Technologies Related to Assistive Environments*, 204–210. doi:10.1145/3316782.3316786

Mariappan, H. M., & Gomathi, V. (2019). Real-time recognition of Indian sign language. *2019 International Conference on Computational Intelligence in Data Science (ICCIDS)*, 1–6.

Martínez, J. M. G., Carracedo, P., Comas, D. G., & Siemens, C. H. (2022). An analysis of the blockchain and COVID-19 research landscape using a bibliometric study. *Sustainable Technology and Entrepreneurship*, *1*(1), 100006. doi:10.1016/j.stae.2022.100006

Masood, S., Srivastava, A., Thuwal, H. C., & Ahmad, M. (2018). Real-time sign language gesture (word) recognition from video sequences using CNN and RNN. In *Intelligent Engineering Informatics* (pp. 623–632). Springer. doi:10.1007/978-981-10-7566-7_63

Masood, S., Thuwal, H. C., & Srivastava, A. (2018). American sign language character recognition using convolution neural network. In *Smart Computing and Informatics* (pp. 403–412). Springer. doi:10.1007/978-981-10-5547-8_42

Mitchell, H. B. (2007). *Multi-sensor data fusion: an introduction*. Springer Science & Business Media.

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. doi:10.1109/JSEN.2019.2909837

Moryossef, A., Yin, K., Neubig, G., & Goldberg, Y. (2021). *Data augmentation for sign language gloss translation.* ArXiv Preprint ArXiv:2105.07476.

Motion, L. (2015). Leap motion. Academic Press.

Nandi, U., Ghorai, A., Singh, M. M., Changdar, C., Bhakta, S., & Kumar Pal, R. (2022). Indian sign language alphabet recognition system using CNN with diffGrad optimizer and stochastic pooling. *Multimedia Tools and Applications*, 1–22. doi:10.1007/s11042-021-11595-4

Natarajan, B., Elakkiya, R., & Prasad, M. L. (2022). Sentence2SignGesture: A hybrid neural machine translation network for sign language video generation. *Journal of Ambient Intelligence and Humanized Computing*, 1–15. doi:10.1007/s12652-021-03640-9

Nath, G. G., & Arun, C. S. (2017). Real time sign language interpreter. *2017 IEEE International Conference on Electrical, Instrumentation and Communication Engineering (ICEICE)*, 1–5.

Nikam, A. S., & Ambekar, A. G. (2016). Sign language recognition using image based hand gesture recognition techniques. *2016 Online International Conference on Green Engineering and Technologies (IC-GET)*, 1–5. doi:10.1109/GET.2016.7916786

Nyagadza, B. (2022). Sustainable digital transformation for ambidextrous digital firms: a systematic literature review and future research directions. *Sustainable Technology and Entrepreneurship*, 100020.

Oliveira, T., Escudeiro, N., Escudeiro, P., Rocha, E., & Barbosa, F. M. (2019a). The virtualsign channel for the communication between deaf and hearing users. *IEEE Revista Iberoamericana de Tecnologias Del Aprendizaje*, *14*(4), 188–195. doi:10.1109/RITA.2019.2952270

Oliveira, T., Escudeiro, P., Escudeiro, N., Rocha, E., & Barbosa, F. M. (2019b). Automatic sign language translation to improve communication. *2019 IEEE Global Engineering Education Conference (EDUCON)*, 937–942. doi:10.1109/EDUCON.2019.8725244

Oxman, A. D., & Guyatt, G. H. (1991). Validation of an index of the quality of review articles. *Journal of Clinical Epidemiology*, *44*(11), 1271–1278. doi:10.1016/0895-4356(91)90160-B PMID:1834807

Park, H., Lee, Y., & Ko, J. (2021). Enabling Real-time Sign Language Translation on Mobile Platforms with On-board Depth Cameras. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *5*(2), 1–30. doi:10.1145/3463498

Patel, B. D., Patel, H. B., Khanvilkar, M. A., Patel, N. R., & Akilan, T. (2020). ES2ISL - An advancement in speech to sign language translation using 3D avatar animator. *2020 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*, 1–5. doi:10.1109/CCECE47787.2020.9255783

Phing, T. C., Ambar, R., Choon, C. C., & Wahab, M. H. A. (2019). Wireless Wearable for Sign Language Translator Device with Android-Based App. In Intelligent and Interactive Computing (pp. 411–422). Springer.

Podder, K. K., Chowdhury, M. E. H., Tahir, A. M., Mahbub, Z., Khandakar, A., Hossain, M. S., & Kadir, M. A. (2022). Bangla Sign Language (BdSL) Alphabets and Numerals Classification Using a Deep Learning Model. *Sensors (Basel)*, *22*(2), 574. doi:10.3390/s22020574 PMID:35062533

Pu, J., Zhou, W., & Li, H. (2019). Iterative alignment network for continuous sign language recognition. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4165–4174. doi:10.1109/CVPR.2019.00429

Qin, W., Mei, X., Chen, Y., Zhang, Q., Yao, Y., & Hu, S. (2021). Sign Language Recognition and Translation Method based on VTN. *2021 International Conference on Digital Society and Intelligent Systems (DSInS)*, 111–115. doi:10.1109/DSInS54396.2021.9670588

Quesada, L., López, G., & Guerrero, L. (2017). Automatic recognition of the American sign language fingerspelling alphabet to assist people living with speech or hearing impairments. *Journal of Ambient Intelligence and Humanized Computing*, *8*(4), 625–635. doi:10.1007/s12652-017-0475-7

Rahaman, M. A., Jasim, M., Ali, M. H., Zhang, T., & Hasanuzzaman, M. (2018). A real-time hand-signs segmentation and classification system using fuzzy rule based RGB model and grid-pattern analysis. *Frontiers of Computer Science*, *12*(6), 1258–1260. doi:10.1007/s11704-018-7082-4

Rahman, M. M., Islam, M. S., Rahman, M. H., Sassi, R., Rivolta, M. W., & Aktaruzzaman, M. (2019). A new benchmark on american sign language recognition using convolutional neural network. *2019 International Conference on Sustainable Technologies for Industry 4.0 (STI)*, 1–6. doi:10.1109/STI47673.2019.9067974

Rajaganapathy, S., Aravind, B., Keerthana, B., & Sivagami, M. (2015). Conversation of sign language to speech with human gestures. *Procedia Computer Science*, *50*, 10–15. doi:10.1016/j.procs.2015.04.004

Rao, G. A., Syamala, K., & Kishore, P. V. v, & Sastry, A. (2018). Deep convolutional neural networks for sign language recognition. *2018 Conference on Signal Processing And Communication Engineering Systems (SPACES)*, 194–197. doi:10.1109/SPACES.2018.8316344

Rastgoo, R., Kiani, K., & Escalera, S. (2020). Hand sign language recognition using multi-view hand skeleton. *Expert Systems with Applications*, *150*, 113336. doi:10.1016/j.eswa.2020.113336

Rastgoo, R., Kiani, K., & Escalera, S. (2021). Real-time isolated hand sign language recognition using deep networks and SVD. *Journal of Ambient Intelligence and Humanized Computing*, 1–21.

Rayner, M., Bouillon, P., Ebling, S., Gerlach, J., Strasly, I., & Tsourakis, N. (2016). An open web platform for rule-based speech-to-sign translation. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (*Volume 2*: Short Papers)*, 162–168. doi:10.18653/v1/P16-2027

Rocha, J., Lensk, J., Ferreira, T., & Ferreira, M. (2020). Towards a tool to translate brazilian sign language (libras) to brazilian portuguese and improve communication with deaf. *2020 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*, 1–4. doi:10.1109/VL/HCC50065.2020.9127257

Sahoo, A. K. (2021). Indian Sign Language Recognition Using Machine Learning Techniques. *Macromolecular Symposia*, *397*(1), 2000241. doi:10.1002/masy.202000241

Saunders, B., Camgoz, N. C., & Bowden, R. (2020). Progressive transformers for end-to-end sign language production. *European Conference on Computer Vision*, 687–705. doi:10.1007/978-3-030-58621-8_40

Saura, J. R. (2021). Using data sciences in digital marketing: Framework, methods, and performance metrics. *Journal of Innovation & Knowledge*, *6*(2), 92–102. doi:10.1016/j.jik.2020.08.001

Savur, C., & Sahin, F. (2015). Real-time american sign language recognition system using surface emg signal. *2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA)*, 497–502.

Sethi, A., Hemanth, S., Kumar, K., Bhaskara Rao, N., & Krishnan, R. (2012). Signpro-An application suite for deaf and dumb. *IJCSET*, *2*(5), 1203–1206.

Sharma, S., & Singh, S. (2021). Vision-based hand gesture recognition using deep learning for the interpretation of sign language. *Expert Systems with Applications*, *182*, 115657. doi:10.1016/j.eswa.2021.115657

Shivashankara, S., & Srinath, S. (2018). American sign language recognition system: An optimal approach. *International Journal of Image. Graphics and Signal Processing*, *11*(8), 18.

Shukor, A. Z., Miskon, M. F., & Jamaluddin, M. H., bin Ali, F., Asyraf, M. F., & bin Bahar, M. B. (2015). A new data glove approach for Malaysian sign language detection. *Procedia Computer Science*, *76*, 60–67. doi:10.1016/j.procs.2015.12.276

Simon, T., Joo, H., Matthews, I., & Sheikh, Y. (2017). *Hand Keypoint Detection in Single Images using Multiview Bootstrapping*. CVPR. doi:10.1109/CVPR.2017.494

Singh, L. B., & Hazarika, S. M. (2007). Enhanced shape context for object recognition. *15th International Conference on Advanced Computing and Communications (ADCOM 2007)*, 529–534. doi:10.1109/ADCOM.2007.125

Škare, M., Blanco-Gonzalez-Tejero, C., Crecente, F., & del Val, M. T. (2022). Scientometric analysis on entrepreneurial skills-creativity, communication, leadership: How strong is the association? *Technological Forecasting and Social Change*, *182*, 121851. doi:10.1016/j.techfore.2022.121851

Sridevi, P., Islam, T., Debnath, U., Nazia, N. A., Chakraborty, R., & Shahnaz, C. (2018). Sign Language recognition for Speech and Hearing Impaired by Image processing in matlab. *2018 IEEE Region 10 Humanitarian Technology Conference (R10-HTC)*, 1–4.

Sruthi, C. J., & Lijiya, A. (2019). Signet: A deep learning based indian sign language recognition system. *2019 International Conference on Communication and Signal Processing (ICCSP)*, 596–600.

Stanford. (2011). *CoreNLP*. http://nlp.stanford.edu:8080/corenlp/

Stoll, S., Camgoz, N. C., Hadfield, S., & Bowden, R. (2020). Text2Sign: Towards sign language production using neural machine translation and generative adversarial networks. *International Journal of Computer Vision*, *128*(4), 891–908. doi:10.1007/s11263-019-01281-2

Stoll, S., Hadfield, S., & Bowden, R. (2020). SignSynth: Data-Driven Sign Language Video Generation. *European Conference on Computer Vision*, 353–370.

Süzgün, M., Özdemir, H., Camgöz, N., Kindiroğlu, A., Başaran, D., Togay, C., & Akarun, L. (2015). Hospisign: An interactive sign language platform for hearing impaired. *Journal of Naval Sciences and Engineering*, *11*(3), 75–92.

Tao, W., Leu, M. C., & Yin, Z. (2018). American Sign Language alphabet recognition using Convolutional Neural Networks with multiview augmentation and inference fusion. *Engineering Applications of Artificial Intelligence*, *76*, 202–213. doi:10.1016/j.engappai.2018.09.006

Taskiran, M., Killioglu, M., & Kahraman, N. (2018). A real-time system for recognition of American sign language by using deep learning. In *2018 41st International Conference on Telecommunications and Signal Processing (TSP)*, 1–5. tecperson. (2017). *Sign Language MNIST*. Kaggle.

Tiku, K., Maloo, J., Ramesh, A., & Indra, R. (2020). Real-time conversion of sign language to text and speech. *2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA)*, 346–351. doi:10.1109/ICIRCA48905.2020.9182877

Tornay, S., Razavi, M., Camgoz, N. C., Bowden, R., & Doss, M. M. -. (2019). Hmm-based approaches to model multichannel information in sign language inspired from articulatory features-based speech processing. *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2817–2821.

Tripathi, K., & Nandi, N. B. G. C. (2015). Continuous Indian sign language gesture recognition and sentence formation. *Procedia Computer Science*, *54*, 523–531. doi:10.1016/j.procs.2015.06.060

Truong, V. N. T., Yang, C.-K., & Tran, Q.-V. (2016). A translator for American sign language to text and speech. *2016 IEEE 5th Global Conference on Consumer Electronics*, 1–2.

Venugopalan, A., & Reghunadhan, R. (2021). Applying deep neural networks for the automatic recognition of sign language words: A communication aid to deaf agriculturists. *Expert Systems with Applications*, *185*, 115601. doi:10.1016/j.eswa.2021.115601

Vijayalakshmi, P., & Aarthi, M. (2016). Sign language to speech conversion. *2016 International Conference on Recent Trends in Information Technology (ICRTIT)*, 1–6.

von Agris, U., Knorr, M., & Kraiss, K.-F. (2008). The significance of facial features for automatic sign language recognition. *2008 8th IEEE International Conference on Automatic Face & Gesture Recognition*, 1–6.

Wadhawan, A., & Kumar, P. (2021). Sign language recognition systems: A decade systematic literature review. *Archives of Computational Methods in Engineering*, *28*(3), 785–813. doi:10.1007/s11831-019-09384-2

Wang, H., Chai, X., Zhou, Y., & Chen, X. (2015). Fast sign language recognition benefited from low rank approximation. *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), 1*, 1–6.

Warrier, K. S., Sahu, J. K., Halder, H., Koradiya, R., & Raj, V. K. (2016). Software based sign language converter. *2016 International Conference on Communication and Signal Processing (ICCSP)*, 1777–1780. doi:10.1109/ICCSP.2016.7754472

Wei, C., Zhou, W., Pu, J., & Li, H. (2019). Deep grammatical multi-classifier for continuous sign language recognition. *2019 IEEE Fifth International Conference on Multimedia Big Data (BigMM)*, 435–442. doi:10.1109/BigMM.2019.00027

Wei, S.-E., Ramakrishna, V., Kanade, T., & Sheikh, Y. (2016). *Convolutional pose machines*. CVPR. doi:10.1109/CVPR.2016.511

Wen, F., Zhang, Z., He, T., & Lee, C. (2021). AI enabled sign language recognition and VR space bidirectional communication using triboelectric smart glove. *Nature Communications*, *12*(1), 1–13. doi:10.1038/s41467-021-25637-w PMID:34508076

World Health Organization. (2021). *Deafness and hearing loss*. https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss

Wu, J., Sun, L., & Jafari, R. (2016). A wearable system for recognizing American sign language in real-time using IMU and surface EMG sensors. *IEEE Journal of Biomedical and Health Informatics*, *20*(5), 1281–1290. doi:10.1109/JBHI.2016.2598302 PMID:27576269

Xiao, Q., Qin, M., & Yin, Y. (2020). Skeleton-based Chinese sign language recognition and generation for bidirectional communication between deaf and hearing people. *Neural Networks*, *125*, 41–55. doi:10.1016/j.neunet.2020.01.030 PMID:32070855

Xiao, Q., Zhao, Y., & Huan, W. (2019). Multi-sensor data fusion for sign language recognition based on dynamic Bayesin network and convolutional neural network. *Multimedia Tools and Applications*, *78*(11), 15335–15352. doi:10.1007/s11042-018-6939-8

Xu, B., Huang, S., & Ye, Z. (2021). Application of tensor train decomposition in S2VT model for sign language recognition. *IEEE Access: Practical Innovations, Open Solutions*, *9*, 35646–35653. doi:10.1109/ACCESS.2021.3059660

Yang, S., & Zhu, Q. (2017). Video-based Chinese sign language recognition using convolutional neural network. *2017 IEEE 9th International Conference on Communication Software and Networks (ICCSN)*, 929–934.

Ye, Y., Tian, Y., Huenerfauth, M., & Liu, J. (2018). Recognizing american sign language gestures from within continuous videos. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2064–2073. doi:10.1109/CVPRW.2018.00280

Zelinka, J., & Kanis, J. (2020). Neural sign language synthesis: Words are our glosses. *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 3395–3403. doi:10.1109/WACV45572.2020.9093516

Zelinka, J., Kanis, J., & Salajka, P. (2019). Nn-based czech sign language synthesis. *International Conference on Speech and Computer*, 559–568. doi:10.1007/978-3-030-26061-3_57

Zhang, J., Zhou, W., Xie, C., Pu, J., & Li, H. (2016). Chinese sign language recognition with adaptive HMM. *2016 IEEE International Conference on Multimedia and Expo (ICME)*, 1–6. doi:10.1109/ICME.2016.7552950

Zhao, J., Qi, W., Zhou, W., Nan, D., Zhou, M., & Li, H. (2021). Conditional Sentence Generation and Cross-modal Reranking for Sign Language Translation. *IEEE Transactions on Multimedia*.

Zheng, L., Liang, B., & Jiang, A. (2017). Recent advances of deep learning for sign language recognition. *2017 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, 1–7. doi:10.1109/DICTA.2017.8227483

Zhou, H., Zhou, W., Zhou, Y., & Li, H. (2021). Spatial-temporal multiw-cue network for sign language recognition and translation. *IEEE Transactions on Multimedia*.

APPENDIX 1

Acronyms

ArSL Arabic Sign Language
ASL American Sign Language
Auslan Australian Sign Language
BDSL Bangladeshi Sign Language
BLEU Bilingual Evaluation Understudy
BSL British Sign Language
CIDEr Consensus-based Image Description Evaluation
CNN Convolutional Neural Network
COMET Cross lingual Optimized Metric for Evaluation of Translation
CSE Czech Sign Language
CSL Chinese Sign Language
DGS Deutsche Gebärdensprache
DSGS Swiss German Sign Language
DTW Dynamic Time Warping
eCS Enhanced Shape Context
EM Expectation Maximization
GAN Generative Adversarial Network
GSS Greek Sign Language
HMM Hidden Markov Models
HOG Histogram of Oriented Gradients
ISL Indian Sign Language
kNN K-Nearest Neighbor
KVK Korean Sign Language
LGP Portuguese Sign Language. LSA Argentine Sign Language. LSE Spanish Sign Language
LSF-CH Swiss French & Chinese Sign Language
LSTM Long-Short Term Memory
NN neural Network
PCA Principal Component Analysis
PSC Persian Sign Language
RNN Recurrent Neural Network
ROUGE-L Recall-Oriented Understudy for Gisting Evaluation
RSL Russian Sign Language
SL Sign Language
SLT Sign Language Translation
SSL Speech and/or text to Sign Language
SVM Support Vector Machine
TSL Taiwanese Sign Language
TSM Turkish Sign Language
WER Word Error Rate
XML Malaysian Sign Language
YSL Yugoslavian Sign Language