DYNAMIC SIGN LANGUAGE RECOGNITION AND TRANSLATION THROUGH DEEP LEARNING: A SYSTEMATIC LITERATURE REVIEW

YESSANE SHRRIE NAGENDHRA RAO[1*], YUAN TING CHONG[2], REHMAN ULLAH KHAN[3], CHEE SIONG TEH[4], MOHAMAD HARDYMAN BARAWI[5], MOHD SHAHRIZAL SUNAR[6], JOAN JO JO SIM[7]

[1,2,3,4,5]Faculty of Cognitive Sciences and Human Development, Universiti Malaysia Sarawak, Sarawak, Malaysia
[6]School of Computing, Faculty of Engineering, Universiti Teknologi Malaysia, Johor Bahru, Malaysia
[7]Sarawak Society for the Deaf, Kuching, Malaysia

E-mail: [1*]yessaneshrrie@gmail.com, [2]yuantingc36@gmail.com, [3]krullah@unimas.my, [4]csteh@unimas.my, [5]bmhardyman@unimas.my, [6]shahrizal@utm.my, [7]jojosim_0411@yahoo.com

ABSTRACT

Sign language is the communication tool for deaf and hard-of-hearing (DHH) communities all around the world. But it is still difficult to establish proper communication between hearing and DHH individuals. As a result, numerous explorations and investigations that focused on sign language recognition and translation (SLRT) have garnered significant attention from researchers in related fields. This systematic literature review aims to provide a comprehensive study on current trends of state-of-the-art dynamic SLRT models proposed in 85 journal articles found in the Scopus database from 2020 to 2024. Based on the selected articles, this review produced an in-depth analyzation of dynamic SLRT models in terms of their frameworks, deep learning techniques, datasets, pre-processing techniques, and evaluation metrics used. Additionally, this review also highlights both the advancements and ongoing challenges in the domain. Notably, there have been considerable development in isolated and continuous SLRT models, particularly through the combinations of deep learning algorithms such as Convolutional Neural Network, Recurrent Neural Network and Transformer models, with suitable datasets. However, the complexities and challenges of developing robust continuous SLRT models for real-time SLRT persist. This systematic literature review was prepared to serve as a foundational reference that will assist future studies on dynamic SLRT.

Keywords: *Dynamic Sign Language, Sign Language Recognition, Sign Language Translation, Deep Learning, PICOC Criteria*

1. INTRODUCTION

Sign language is a standard language that is used by deaf people as their communication tool [1]. Based on the statistics, over 200 different SLs have been discovered and spoken around the world [2]. The deaf and hard-of-hearing (DHH) individuals and others can express sign language through two features, which are manual features (i.e., body and hand movements) and non-manual features (i.e., facial expressions) [3]. As the advancement of technology continues to improve, contributions towards sign language-based recognition and translation models are seemingly notable [4]. According to Al-Qurishi et al. [5], deep learning, also known as deep neural networks, has gained popularity in sign language recognition (SLR) and sign language translation (SLT) systems. In both the SLR and SLT processes, different deep learning models are commonly applied to achieve good performance [3]. There is said to be an advancement in this field of research, with papers on deep learning being increasingly published [6-7].

The deep learning models used in current studies include Convolutional Neural Networks (CNN) [8-11], Recurrent Neural Networks (RNN) [12-14], radial basis function neural networks (RBFNN) [15], 2D CNN [16], 3D CNN [17-19], Long Short Term Memory (LSTM) [20-24], Self-Organizing Map (SOM) [25], bidirectional LSTM (BiLSTM) [26], Residual Networks (ResNet) [27-29], Convolutional Block Attention Module (CBAM) [30-31], bidirectional encoder representations from Transformers (BERT) [27], Spatial-Temporal Transformer Network (STTN) [32], Gated Recurrent Unit (GRU) [33], and Transformers [34-35]. There are also studies that combine the deep learning models mentioned priorly such as a CNN-LSTM combination [4,36-38], CNN-RNN combination [39], and CNN-BiLSTM combination [30,40].

Another point that needs to be highlighted in current studies on SLR and SLT using deep learning models is the usage of either static or dynamic sign language gestures when conducting these processes. According to Johari et al. [41] sign language recognition and translation (SLRT) for static sign language was more commonly studied compared to dynamic sign language. SLR systems can process datasets of static gestures that are collected in image format directly and proceed with the SLR process after pre-processing methods, whereas dynamic gestures need to be collected in video format and split into key frames that represent each sign language movement [42-43]. Dynamic SLR and SLT, especially in real-time, is still a challenge [44-45]. Real-time systems face challenges in processing, detection and classification of sign language video datasets because each data has different durations, speeds and backgrounds [46].

There are two types of dynamic sign languages identified in past research, which are isolated and continuous sign languages. Isolated sign language includes different dynamic gestures of commands, alphabets, numbers or words [13,15,18,26,37,47-50] whereas continuous sign language is produced in sequences and divided into five frameworks, which are Sign2Gloss (S2G), Gloss2Text (G2T), Sign2Gloss2Text (S2G2T), Sign2Gloss+Text(S2(G+T)), and Sign2Text (S2T). Gloss refers to a label that represents the meaning of a sign gesture [51]. Models based on the S2G framework perform recognition from dynamic (continuous) signs to glosses that include information on the semantic and grammatical rules of the signs [20-21,52]. The G2T framework-based models translate from glosses to grammatically correct sentences in a target spoken language [53]. Next, models based on the S2G2T framework perform the recognition task from sequential dynamic (continuous) signs to gloss annotations, and then perform the translation task from gloss annotations into complete sentences [22,54]. The S2(G+T) framework-based models, on the other hand, output gloss and text simultaneously [53] while models based on the S2T framework conduct SLT tasks in an end-to-end manner from dynamic (continuous) sign language into spoken language sentences [24,27,32,39,56-57].

Researchers have utilized multiple varieties of models with modified functionalities to recognize and translate sign languages [4,33,59]. The datasets researchers use differ according to the needs of their models or systems. However, some of the most common datasets used are RWTH-PHOENIX-Weather-2014T [24,27,52,60], CSL dataset [24,27,52,54,60-62], and ASL dataset [14,48-49,61]. Many researchers have also created their own datasets according to the needs of their specific localities [22,17,19,31,63-66]. Current research on dynamic SLR and SLT has gained outstanding results such that the average accuracy of the models proposed by the researchers is approximately above 85% for alphabets or isolated words [8,11-13,23,35,48,50,59,67-71] and continuous sentences [10,40,56]. Researchers are able to compare the efficiency and reliability of their models by using datasets that are similarly used by other studies. Yet researchers still use different sign language datasets according to the specific sign language their research mainly focuses on.

The focus on real-time SLR and SLT depends on the development of dynamic SLRT. This is because merely focusing on dynamic isolated sign languages does not significantly contribute to the context of real-time sign language translation [16,25,38,48-49,72-73]. Besides that, although many studies [12,35-36,70-71,74-75] do claim to have developed dynamic SLRT systems using deep learning models, their study only relies on specific aspects of the sign language such as fingerspelling, numbers, alphabets, greetings, colors, action-based hand gestures, and common words. These studies that focused on isolated sign languages are still limited in producing efficient conversation in real-time [16,59,64,76-78], while continuous sign language models that include SLRT tasks on sentences-based output are more practical in real-time conversation [22,24,27-28,32,39,79-80].

Therefore, this systematic literature review aims to identify state-of-the-art dynamic SLRT deep learning models that can substantially contribute towards real-time SLRT in various frameworks. Previous literature reviews or surveys have typically focused on recognition or translation for dynamic sign languages in a separate manner [5,7,41,44-45,53]. Therefore, in order to bridge the gap found in previous reviews, this study provides a comprehensive examination of both aspects together in a systematic manner. Besides, it is found that previous reviews and surveys are yet to specifically focus on the SLRT processes that lead to real-time SLRT. Motivated by these limitations, this study undertakes a detailed and rigorous investigation, aiming to identify specific deep learning techniques used in dynamic sign language models and the aspects of dynamic sign language they address in recognition and translation processes which could direct SLRT research towards real-time SLRT. By specifically examining SLRT studies published from 2020 to 2024 in the Scopus database, this will offer a timely perspective on advancements and trends within this period. Scopus database was selected because it covers many articles under different journals worldwide and is commonly used in writing literature review [92].

Additionally, this systematic review also aims to understand the challenges and limitations that still exist using deep learning models in SLRT research, and what is the way forward in addressing them. This systematic literature review will also highlight the advancements of current dynamic SLRT systems to confer the processes needed for proper and comprehensive SLRT. Next, the study will analyze each paper retrieved to find out which datasets, pre-processing methods, and evaluation metrics are used by the researchers in their study. It will assess these elements on whether they enable or support both recognition and translation within dynamic SLRT models, rather than being limited to one aspect as seen in previous systematic reviews. This systematic literature review is significant in searching, investigating, and analyzing the potential tools and modifications of SLRT models in different frameworks, so that possible future development for advancements in the area of dynamic SLRT using deep learning models and techniques can be informed.

2. METHODOLOGY

2.1 Review Protocol

This study uses Kitchenham's methodology in conducting a systematic literature review as this methodology is commonly used as a referring point in software engineering or programming research [81]. A study by Usman et al. [82] explained that the Kitchenham methodology contributed to the growing number of systematic literature reviews as it published guidelines on writing a systematic literature review through identifying, evaluating, and integrating the evidence reported in scientific literature. To write a systematic literature review, the review should be prepared by answering specific research questions which are clearly formulated by collecting, learning, and analyzing the relevant studies found [83].

2.2 Research Questions

In this study, the formation of research questions will be based on the Population, Intervention, Comparison, Outcome, and Context (PICOC) criteria, should any of the criteria be deemed necessary by the researchers. Table 1 shows the descriptions of PICOC criteria.

Table 1: PICOC Criteria By Nishikawa-Pacher [84]

Criteria	Description
P (Population)	What population am I interested in?
I (Intervention)	What intervention exactly am I interested in reviewing? Is it one intervention, or a cluster of interventions?
C (Comparison)	With what is the intervention being compared?
O (Outcomes)	For many social interventions there is a wide range of outcomes, and the assessment of effectiveness involves collecting information on both positive and negative impacts and assessing the balance between them.
C (Context)	There is a further component, which needs to be considered – the *context* within which the intervention is delivered.

According to Kitchenham and Charters [81], the PICOC criteria was suggested by Petticrew and Roberts in generating research questions which was redefined as an extension from the original criteria named PICO (Population, Intervention, Comparison and Outcome) only without Context. PICO was a popular research question framing method, especially in clinical studies [85-86]. However, the study found that the PICO model was not only limited to clinical investigation studies, and it argues that the PICO model can be used universally for all study designs [84]. According to Nishikawa-Pacher [84], the researcher proved his argument by giving samples taken from Google Scholar Metrics that have used the universal PICO scheme in different research fields. One such sample is a study in the IEEE/CVF Conference on Computer Vision and Pattern Recognition which applied a residual learning framework (I) to neural networks (P) so as to generate substantially deeper neural nets despite lower complexity (O) compared to VGG nets (C) [87]. Table 2 shows the compiled list by Kitchenham and Charters [81] of the PICOC criteria that matches Computer Science related fields [81].

Table 2: PICOC Criteria In Software Engineering's Viewpoints By Kitchenham And Charters [81]

Criteria	Viewpoints in Software Engineering Fields
P (Population)	A question may refer to very specific population groups, e.g. novice testers, experienced software architects working on IT systems, or population area e.g. IT systems, command and control systems.
I (Intervention)	The intervention is the software methodology, tool, technology, or procedure that addresses a specific issue, e.g. technologies to perform specific tasks such as requirements specification, system testing, or software cost estimation.
C (Comparison)	This is the software engineering methodology, tool, technology, or procedure with which the intervention is being compared, e.g. compare people using a technique with people not using a technique.
O (Outcomes)	All relevant outcomes should be specified. For example, in some cases we require interventions that improve some aspect of software production without affecting another e.g. improved reliability with no increase in cost.
C (Context)	For Software Engineering, this is the context in which the comparison takes place (e.g. academia or industry), the participants taking part in the study (e.g. practitioners, academics, consultants, students), and the tasks being performed (e.g. small scale, large scale). Many software experiments take place in academia using student participants and tasks in small scale.

This study aims to analyze research gaps in published studies that focus on dynamic SLR, SLT, and SLRT models from 2020 to 2024. The research questions below have been formed in accordance with the PICOC criteria in Table 2 to achieve this aim.

1. What are the types of dynamic systems presented in the studies and the frameworks employed?
2. What are the deep learning techniques and models presented in the studies?
3. What are the datasets used to test the models in each published article?
    a. How was the data collected?
    b. What is the availability of the dataset?
    c. What is the size of the dataset?
    d. Which sign language is the dataset subjected to?
4. What pre-processing techniques are applied to the dynamic sign language datasets used in the models?
5. What are the evaluation metrics used to assess the models?

According to the research questions above, the types of dynamic sign language systems, the frameworks employed, deep learning techniques and models found in previous studies will first be identified. Next, the datasets used to test the models in each published article will be investigated and clarified by exploring the collection methods, availability, sizes and sign languages used of datasets. Then, the pre-processing techniques applied to the dynamic sign language datasets used in the models found in previous studies will be determined and explained. Lastly, the evaluation metrics that can be used to assess these models will be determined and assessed for suitability.

By obtaining these information from the reviewed studies, the major gap in dynamic SLR, SLT, and SLRT research can be identified to direct further research in this field. As current studies in this field differ in the aspects stated above, this study aims to identify the specific models, techniques, datasets, pre-processing techniques, and evaluation metrics that are required for a systematic dynamic SLRT system to be developed. This would inevitably guide researchers in choosing the appropriate models, datasets, and metrics for future research in the field of dynamic SLRT research directing them towards real-time SLRT. Nevertheless, it is important to note that this study does not include highly cited papers that are not available in the Scopus database which may exclude certain models, datasets, and metrics from being identified and investigated. However, future researchers are encouraged to replicate this systematic literature review with other databases such as Web of Science, ERIC, IEEE Xplore, and Science Direct to get a bigger picture of the current advancement in the field of dynamic SLRT research.

2.3 Search Strategy

The search strategy used by the researchers is in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 flow diagram. According to Trifu et al. [88], the recommendations of using PRISMA statements have been widely approved in assisting researchers in preparing review papers. The PRISMA flow diagram is published in the format of four divided stages, which are Identification, Screening, Eligibility and Included. The PRISMA flow diagram shows the stages of the literature search process that researchers use to finalize the articles that will be accounted for in their review [89]. The search process starts with the number of documents retrieved from different databases and resources using certain keywords or search strings (Identification Stage) [90]. Then, the documents are filtered manually or through the databases for specific reasons normally mentioned in the PRISMA 2020 flow diagram (Screening and Eligibility Stage). Afterwards, the PRISMA flow diagram will be completed with the finalized number of specific included articles (Included Stage), which will become the main discussants in the results section of the study [88]. The flow diagram can be modified

The publications retrieved range from the year 2020 till the year 2024. Studies found that the generation of the most recent systematic literature review is a challenging task, with one-third of review papers generally updated within two years, and publications within six years are considered as up-to-date references [95-96]. For this study, the time span of five years is selected. The PRISMA flow diagram for this review is shown in Figure 1.

according to whether it is a new and original review or an updated review [89].

For this review, the academic publications were retrieved from the Scopus database by Elsevier using the search string “TITLE-ABS-KEY (("sign language") AND ("recognition" OR "translation") AND ("deep learning" OR "neural network*" OR "artificial neural network*") AND ("dynamic" OR "continuous")) AND (PUBYEAR > 2019) AND (PUBYEAR < 2025) AND (LIMIT-TO (PUBSTAGE , "final")) AND (LIMIT-TO (DOC-TYPE , "ar")) AND (LIMIT-TO (SRCTYPE , "j"))”. Scopus is a citation and abstract database that was founded in 2004 [91]. It is a database that indexes peer-reviewed scientific literature that are highly curated [91-92]. Scopus contains publications indexed from all around the world, and it covers four main areas of publications, which are Physical Sciences, Social Sciences, Health Sciences, and Life Sciences [92]. One of the many reasons why Scopus is commonly used in systematic literature review studies is that it is a dependable search tool when searching for literature that ought to be reviewed through basic and advanced search queries [91-93]. Boolean operators, such as ‘AND’ and ‘OR’ are commonly used as the strategy in keywords combinations for search string purposes [94]. Moreover, it also offers easier forward and backward citation search which can assist researchers when conducting thorough literature reviews [91]. Hence, Scopus was chosen as the database in this study.

2.4 Study Selection

Inclusion and exclusion criteria are essential to set the boundaries of the systematic literature review and to determine the results of the articles acquired by implementing the search strategy. The researchers have determined the criteria beforehand to avoid any biases. Table 3 below shows the inclusion and exclusion criteria of this systematic review.

Table 3: Inclusion And Exclusion Criteria

No.	Inclusion Criteria	Exclusion Criteria
1.	Must be empirical studies	Review studies
2.	Must be final publication staged journal articles	In-press journal articles
3.	English articles	Other language journal articles
4.	Must be published between the year 2020 and 2024	Studies before the year 2020
5.	Relevant to dynamic or continuous sign language recognition or translation or both	No focus on dynamic or continuous sign language recognition or translation
6.	Relevant to the context of deep learning and neural networks	Irrelevant to the topic of deep learning and neural networks
7.	The SLR or SLT performed should be one-way, i.e. Sign language to text/speech	Bidirectional models

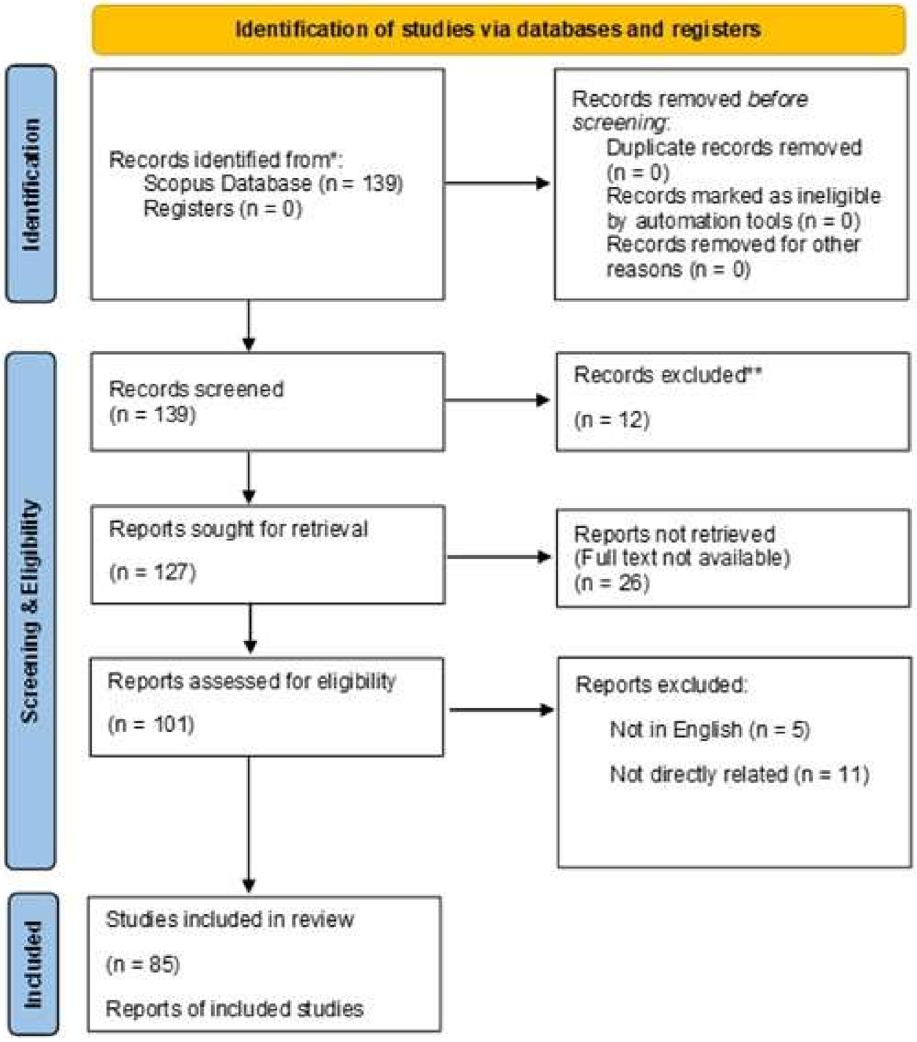
Identification of studies via databases and registers
Identification
Records identified from*:
Scopus Database (n = 139)
Registers (n = 0)
Records removed before screening:
Duplicate records removed (n = 0)
Records marked as ineligible by automation tools (n = 0)
Records removed for other reasons (n = 0)
Screening & Eligibility
Records screened (n = 139)
Records excluded** (n = 12)
Reports sought for retrieval (n = 127)
Reports not retrieved (Full text not available) (n = 26)
Reports assessed for eligibility (n = 101)
Reports excluded:
Not in English (n = 5)
Not directly related (n = 11)
Included
Studies included in review (n = 85)
Reports of included studies

Figure 1: PRISMA Flow Diagram (Source: Authors, Adopted From Page et al. [89])

2.5 Study Quality Assessment

The review was conducted only on journal articles that have been published or have reached the final publication stage. In order to maintain the quality of the study, duplicate records were thoroughly checked. In the screening stage, according to the PRISMA flow diagram, the abstracts of the 139 journal articles were read through, and only relevant articles were sought for retrieval which was 127 articles. 12 articles were found to be irrelevant. Next, from the records that were sought for retrieval, 26 records were not retrieved as the full-text articles were not available. Additionally, a Quality Assessment Form as shown in Table 4 was proposed to check the quality and eligibility of the full-text reports retrieved. This form was adapted from the proposed questionnaire by Boggaram et al. [44]. This questionnaire was created in accordance with the Kitchenham and Charters [81] quality assessment form and the Oxman and Guyatt [97] quality assessment questionnaire [44].

101 articles were assessed for eligibility, and five reports were removed due to being in a language other than English as stated in the inclusion and exclusion criteria in Table 3. Another 11 reports were removed after a full-text screening based on the proposed Quality Assessment Form due to irrelevancy to the subject matter of this systematic literature review. The studies were also checked for bias and internal and external validity. Finally, a total of 85 journal articles to be reviewed were included in this study.

Table 4: Quality Assessment Form

Assessment Questions	Considerations
A. Primary Screening Question	
Is the goal of the research article to use deep learning technique(s) to perform SLR, SLT or SLRT? Is the goal of the research article to use deep learning technique(s) to perform SLR, SLT or SLRT? Yes ( ) No ( )	The article should mention the deep learning technique used for SLR, SLT or SLRT.
B. Secondary Screening Question	
Is the goal of the research article to use deep learning technique(s) to develop dynamic sign language models? Yes ( ) No ( )	The article should mention or clearly show the type of sign language models (static or dynamic) being developed.
Is the goal of the research article to use deep learning technique(s) to develop dynamic sign language models? Yes ( ) No ( )	The article should mention or clearly show the type of sign language models (static or dynamic) being developed.
Is the type of dynamic sign language model(s) used in terms of isolated sign language model(s), S2G, G2T, S2(G+T), S2G2T, or S2T framework-based model(s) clearly shown in the research article? Yes ( ) No ( )	The article should clearly show the type of dynamic sign language models (isolated or continuous) and the frameworks (S2G, G2T, S2(G+T), S2G2T, or S2T) used to develop the model(s).
Is the research article testing the SLR, SLT or SLRT model(s) using dynamic sign language dataset(s)? Yes ( ) No ( )	The article should mention the dynamic sign language(s) dataset used to test their SLR, SLT or SLRT model(s).
Is the model in each article assessed using some kind of evaluation metric(s)? Yes ( ) No ( )	The article should mention the evaluation metric(s) used in assessing their model(s).
If the answer to the previous question is 'Yes', the next question is considered.	
C. Comprehensive Question	
Research Results	
Is there a succinct purpose for the findings? Yes ( ) No ( )	The article should give conclusive evidence regarding the SLR, SLT or SLRT performed.

3. RESULTS AND FINDINGS

3.1 Publications By Years

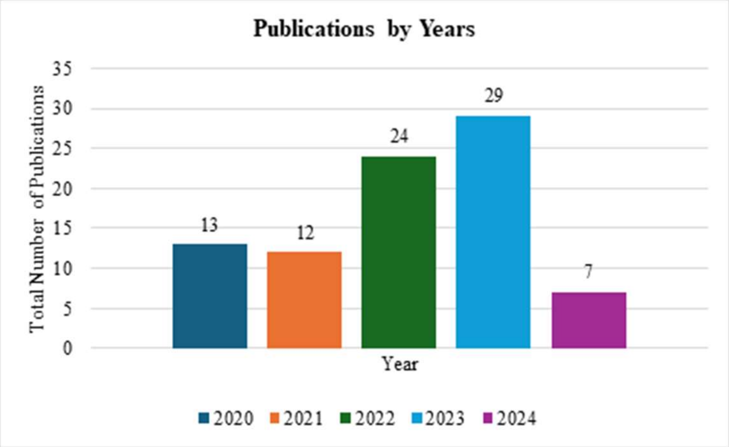

Figure 2: Total Number Of Publication By Years

The total number of publications for all five years chosen in this study are shown in Figure 2. The years 2020, 2021, 2022, 2023, and 2024 have 13, 12, 24, 29, and 7 publications respectively. There is a decrease in one publication from the year 2020 to 2021. However, the years 2021, 2022, and 2023 have recorded a continuous increase in publications, with a drastic increase in publications in the year 2022. This indicates that more researchers have gained an interest in the field of SLRT, especially in the years 2022 and 2023.

3.2 Findings Based On Research Questions

3.2.1 RQ1: What are the types of dynamic systems presented in the studies and the frameworks employed?

Table 5: Number Of Studies Identified Based On Different Dynamic System Types And Frameworks

Types/ Frameworks	Total Number of Studies
Isolated Signs	
Words, Phrases, Alphanumeric Characters or Simple Gestures	61
Continuous Signs (Sentences)	
Sign2Gloss	3
Gloss2Text	0
Sign2Gloss2Text	2
Sign2(Gloss+Text)	0
Sign2Text	19

Table 5 shows the number of studies found based on different dynamic system types and frameworks. The table above is divided into two types of dynamic systems, which are the systems that focus on continuous sign language (sentences) and isolated sign language only, such as alphabets, numbers, words, short phrases, and simple gestures for movement, direction or contexts that are not representative through words. Of the 85 studies reviewed, there are 61 studies [4,8-9,11-19,23,25-26,30-31,34-38,42,46-50,55,58-59,63-69,70-78, 98-111] that focused on isolated sign language recognition. In these studies, eight studies [12,18, 42,48-49,63,69,72] included static sign images. There are 19 studies [9,14-15,18-19,42,46,55, 58,67,69,72,75-77,99-100,106,111] that included action gesture recognition and eight of these studies [15,19,55,58,75,100,106,111] focused specifically only on action gesture recognition. The study conducted by Nadgeri and Kumar [38] converted the isolated signs into vectors only, while several other studies focused only on either alphabet [48-49,110] or fingerspelling [16,74] only.

For the continuous sign language systems, there are three studies [20-21,52] that used the Sign2Gloss framework. Next, there are two studies [22,54] that employed the Sign2Gloss2Text framework whereas the frameworks Gloss2Text and Sign2(Gloss+Text) were not employed by any of the studies. The Sign2Text framework has the highest number of studies with 19 studies [10,24,27-29,32-33,39-40,56-57,60-62,79-80,112-114] employing it. Out of all the continuous sign language-based studies, seven studies [20,22,29,32,79-80,112] included isolated signs as well. Additionally, the study conducted by Belissen et al. [39] recognized lexical signs but not gloss whereas Hu et al. [29] interestingly tested their model thrice separately with isolated signs, the Sign2Gloss framework for SLR, and the Sign2Text framework for SLT. Wang et al. [57] on the other hand worked on SLT, subsequently converting the spoken language text into voice output.

3.2.2 RQ2: What are the deep learning techniques and models presented in the studies?

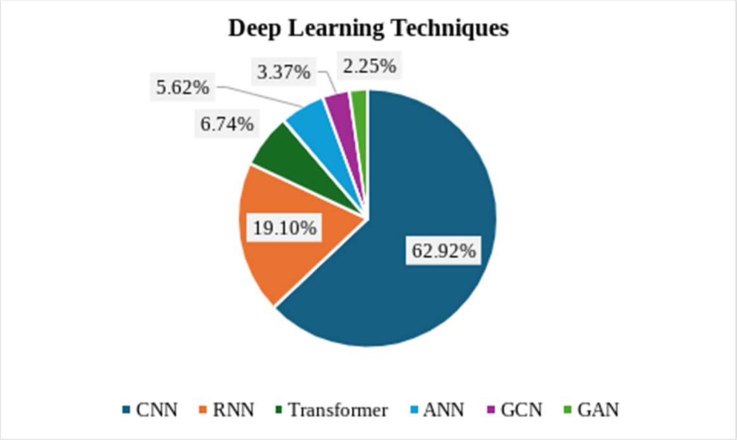

Figure 3: Deep Learning Techniques Presented In Reviewed Studies

Figure 3 shows the percentage of six deep learning baseline techniques found from 89 models from the reviewed studies. Convolution Neural Network (CNN) was the most common deep learning tool that covered 62.92%, of which are 56 CNN models. Then, Recurrent Neural Network (RNN) was the second most common deep learning technique used to develop 17 models (19.10%). Other than this, other deep learning techniques have published fewer than ten models. There are six Transformer models that have been proposed in the studies, which constitute 6.74%. In addition to that, Artificial Neural Network (ANN) techniques have also been applied to five models from the total (5.62%). While Graph Convolution Network (GCN) and Generative Adversarial Network (GAN) techniques only covered 3.37% and 2.25% respectively, which are three GCN and two GAN models in the reviewed studies.

Table 6: Deep Learning Models Identified In The Reviewed Studies

Techniques / Variants	Model	Framework
CNN (56)		
1DCNN	-	Isolated [69]
	Two-Stream Mixed-ResNet50	Isolated [49]
	1DCNN-BiLSTM	Isolated [104]
	VGG11-1DCNN-BiLSTM-Encoder (BiLSTM)-Decoder (SA-LSTM)	S2T [56]
CNN/ 2DCNN	-	Isolated [8,16,18,50,58,63-65,69,76,98,101,107]
		S2T [10]
	CNN-Window Sliding	S2T [79]
	CNN-FC	Isolated [26]
	CNN-LSTM	Isolated [36-38]
		S2G [20]
	DMN-AMN-SRN (CNN-LSTM)	Isolated [105]
	CNN-LSTM-CTC	S2G2T [22]
	CNN-LSTM-SelfMLP	Isolated [77]
	CNN-SLSTM	Isolated [4]
	CNN-BiLSTM	Isolated [46,66,109]
		S2T [28,39]
	CNN-BiLSTM-TCL	S2G [52]
	CNN-BiLSTM-GAN	S2T [40]
	2DCNN-Attentive Multi Feature-BiLSTM-CTC	S2G2T [54]
	CNN-Transformer-Reinforcement Learning	S2T [61]
	Meta Metric Learning with Triplet Loss Embeddings (MVDMML)	Isolated [108]
	CNN-Hybrid Arithmetic Hunger Games	Isolated [75]
	End-to-end Fourier-CNN	Isolated [9]
3DCNN/ C3D	-	Isolated [19,73]
	3DCNN-LSTM	Isolated [23]
	3DCNN-ConvLSTM	Isolated [11,18]
	C3D-MLP	Isolated [17]
	Deformable Convolution Sequence attention 3D Residual Network (DCSR3D) Encoder-GRU Decoder	S2T [33]
2D & 3DCNN	2D & 3DCNN- Batch-Normalization	Isolated [42]
Convolution Self-Organizing Map (CSOM)	CSOM-BiLSTM	Isolated [25]
Temporal Convolution Network (TCN)	CNN-TCN	Isolated [100]
		S2T [60]
Multi-channel CNN	Multi-channel CNN-LSTM	Isolated [106]
	Multi-channel CNN-Attention-based-Encoder-Decoder	S2T [57,113]
Convolutional block attention module (CBAM)	CBAM-3DResNet	Isolated [31]
	CBAM-BiLSTM	Isolated [30]
RNN (17)		
RNN	InceptionV3-RNN	Isolated [12]
Bidirectional RNN (BiRNN)	-	Isolated [14]
	BiRNN-LSTM-BiLSTM	Isolated [67]
Recurrent CNN (RCNN)	ResNet50-LSTM	Isolated [102]
Long Short-Term Memory (LSTM)	-	Isolated [13,48,55]
	LSTM-Attention Mechanism	Isolated [23]

	LSTM Encoder-Decoder	Isolated [26]
	VTCNN-LSTM Encoder-CTC and Attention-based Decoders	Isolated [74]
Convolution LSTM (ConvLSTM)	-	Isolated [78]
Bidirectional LSTM (BiLSTM)	BiLSTM-Fast Fisher Vector	Isolated [47,68]
	BiLSTM-Attention	Isolated [103]
	BiLSTM-ST-Attention	S2T [24]
	CNN Transfer Learning-BiLSTM	S2T [108]
Boundary Adaptive Encoder (BAE)	BAE-LSTM-Window Attention Encoder-Decoder	S2T [112]
Transformers (6)		
Transformers	-	Isolated [34-35]
Bidirectional Encoder Representations from Transformers (BERT)	ResNet-SignBERT	S2T [27]
	Cross-Attention-SignBERT	S2T [62]
SignBERT+	SignBert+ - Encoder-Decoder	S2T [29]
Spatial Temporal Transformer Network (STTN)	-	S2T [32]
GAN (2)		
GAN	GAN (Generator & Discriminator - CNN)	Isolated [71]
Hyperparameter based-GAN (H-GAN)	H-GAN (Generator-LSTM, Discriminator-3DCNN+LSTM)	S2G [21]
ANN (5)		
ANN	-	Isolated [110]
	SNN-ST-Back Propagating Training	Isolated [99]
	Radial Basis Function (RBF) Neural Network	Isolated [15]
Multilayer Perception (MLP)	-	Isolated [72]
	Neural Network	Isolated [22]
GCN (3)		
GCN	GCN-Attention-CNN	Isolated [59]
Spatial-Temporal Graph Convolution (STGC)	STGC-Transformer	S2T [80]
Dense GCN	Dynamic Dense ST GCN (DDSTGCN) - Dynamic ST CN Module (DSTCNM) -Encoder (BiLSTM)-Decoder (CTC)	S2T [56]

The baseline deep learning techniques identified in the studies in Table 6 are CNN, RNN, Transformers, GAN, ANN, and GCN. Researchers have used different variants of these baseline techniques and have developed their models according to those variants. For CNN, there are eight variants identified which are 1DCNN, 2DCNN (the original variant), 3DCNN, and 2D and 3D CNN, CSOM, TCN, Multi-channel CNN, and CBAM. 1DCNN has been found in developing isolated sign language models in an independent mode [69]. Moreover, it was also incorporated with other techniques to build Two-Stream Mixed-ResNet50 and 1DCNN-BiLSTM as two different isolated sign language models [49,104], and VGG11-1DCNN-BiLSTM in applying encoder-decoder tasks as a S2T model [56]. Next, 2DCNN was found to be highly used in developing models resulting in 34 models based on the studies reviewed. Independent 2DCNN was used to develop 13 isolated sign language models [8,16,18,50,58,63-65,69,76,98,101,107] and one S2T model [10]. There are also 13 isolated sign language models identified that were developed using 2DCNN combined with LSTM, LSTM-SelfMLP, SLSTM, BiLSTM, FC, meta metric learning, hybrid arithmetic hunger games, and end-to-end Fourier [4,9,26,36-38,46,66,75-77,105,108-109]. Besides this, 2DCNN was also used to develop six models using the S2T framework in an independent mode [10] by cooperating BiLSTM, BiLSTM-GAN, window learning, transformer-reinforcement learning [28,39-40,61,79].

Furthermore, there are also two S2G [20,52] and S2G2T [22,54] models found from the studies that have been developed using 2DCNN combined with LSTM, LSTM-CTC, BiLSTM-TCL, and attentive-multi feature-BiLSTM-CTC techniques. Next, 3DCNN was applied as an independent baseline model in developing two isolated sign language models [19,73] and contributed to the development of four hybrid isolated sign language models [11,17-18,23] by incorporating LSTM, ConvLSTM, and MLP, and one hybrid S2T DCSR3D encoder-GRU decoder model [33]. In addition to that, there is also an isolated sign language model that was developed by combining 2D and 3DCNN as a baseline with the Batch Normalization technique [42]. Other than this, CSOM and CBAM were used to develop several isolated sign language models by combining with BiLSTM [25,30] and 3DResNet [31] techniques in the reviewed studies. TCN also worked as a baseline to develop an isolated sign language [100] and a S2T

[60] hybrid model by cooperating CNN. An isolated sign language [106] and two S2T [57,113] hybrid models have been found under the baseline of multi-channel CNN in Multi-channel CNN-LSTM and attention-based Multi-channel CNN encoder-decoder models.

There are seven variants of RNN which are RNN (the original variant), BiRNN, RCNN, LSTM, ConvLSTM, BiLSTM, and BAE. RNN has been used to develop an isolated sign language model by incorporating InceptionV3 that is covered under the CNN technique [12]. Besides that, RNN has also been adapted into several extensions such as BiRNN and RCNN. BiRNN was used as an independent baseline model [14] and a com-bination with LSTM-BiLSTM [67] in isolated sign language model development, whereas RCNN was combined with ResNet50 and LSTM in developing an isolated sign language model as well [102]. LSTM is the variant of RNN that is often used by researchers in building isolated sign language models as identified in the reviewed studies. LSTM has been used as an independent tool in developing three isolated sign language models [13,22,48] and a baseline in three hybrid models [23,26,74] of LSTM-Attention Mechanism and two encoder-decoder models. LSTM has also been extended into ConvLSTM and BiLSTM to develop different models. For example, ConvLSTM was used to develop an isolated sign language model [78], whereas BiLSTM was used in developing five hybrid models, which are two models for S2T framework, combining spatial-temporal-attention and CNN Transfer Learning methods [24,114], and three models for isolated sign language models, incorporating attention mechanism and Fast Fisher Vector method [47,68,103]. Lastly, a BAE-based S2T was identified in the studies which combined the LSTM and window attention in encoder-decoder platform techniques [112].

Next, for Transformers, there are three variants identified which are Transformers (the original variant), BERT, BERT+ and STTN. Based on the studies reviewed, Trans-former was used as an independent baseline tool in developing two isolated sign language models [34-35]. Whereas, its variants, BERT, BERT+ and STTN were focused on proposing models based on the S2T framework. BERT and BERT+ have been found in the development of three hybrid models combined with ResNet, Cross-Attention and encoder-decoder task techniques [27,29,62], while STTN was only used in developing an independent S2T model without any additional tools [32]. GAN and ANN have two variants respectively which are the GAN technique itself, H-GAN for GAN, and the ANN technique itself, and MLP for ANN. According to the studies reviewed, the GAN technique was used to develop a Generator and Discriminator. It was found to be applied to an isolated sign language model with CNN using it as a Generator and Discriminator [71], whereas H-GAN was proposed in a model based on the S2G framework with a LSTM Generator and 3DCNN-LSTM Discriminator [21].

Apart from this, ANN was used most in isolated sign language models. From the 85 studies reviewed, ANN has been proposed as an independent baseline, combined with SNN-ST-Back Propagating Training and PCO-RBF neural network [15,99,110]. Its MLP variant was found in developing isolated sign language models in an independent mode with neural network extensions [22,72]. Finally, there are three variants identified for the GCN technique which are GCN (the original variant), STGC, and Dense GCN. GCN has been combined with attention mechanism and CNN to develop an isolated sign language model [59]. While STGC and Dense GCN are found in a couple of hybrid models based on the S2T framework, which are STGC-Transformer and DDSTGCN-DSTCNM Encoder-Decoder model [56,80]. Overall, there are a total of 89 models that have been identified from the 85 reviewed studies. There are some studies that have developed and proposed more than one model for different criteria and purposes. For example, several researchers have developed different models to process different formats of input data inserted [18,26,56,69]. In addition to that, Table 6 has only included the most efficient and finalized models found in the studies after ablation studies such as testing and comparing between various CNN layers [4,12-13,20,49,61,63,77,102,108-109] and different techniques, such as IDCNN and GRU were applied [76].

3.2.3 RQ3: What are the datasets used to test the models in each published article?

This question will focus on the detailed information of different datasets found in the research articles, what are the dataset data collection methods, availability, sizes, and sign languages.

3.2.3.1 RQ 3(a): How was the data collected?

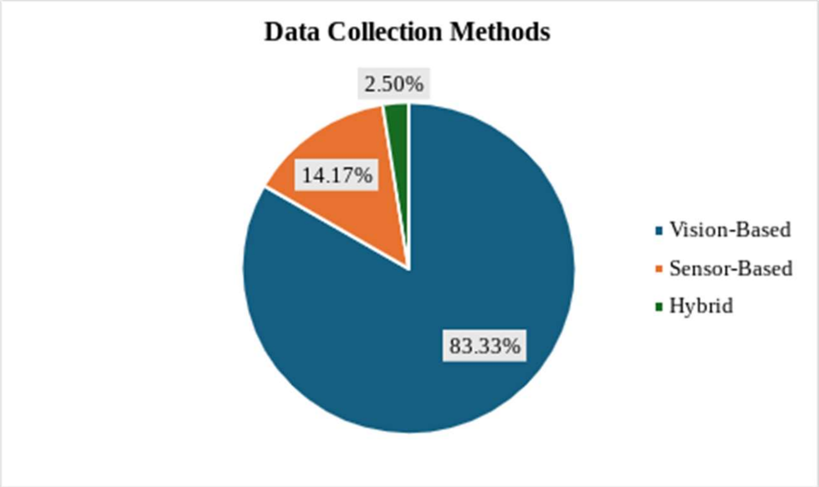

Figure 4: Data Collection Methods For Dynamic Sign Language Datasets

Based on Figure 4, there are three data collection methods found from the 85 studies reviewed. Data collection methods are divided into vision-based, sensor-based and hybrid methods. There is a total of 100 datasets collected according to the vision-based method, which achieved the highest percentage of 83.61% when compared to other methods. Sensor-based method achieved 13.93%, which was identified in 17 datasets from the total of 120 datasets. Lastly, there are only three datasets, which constitute 2.46% using hybrid mode of both methods found in the studies reviewed.

Table 7: Data Collection Tools For Dynamic Sign Language Datasets

Data Collection Tools			
Vision-based Method		**Sensor-based Method**	
Camera	71	Data Glove	8
Leap Motion Controller (LMC)	10	Armband	2
Microsoft Kinect	6	Millimeter Wave Radar	2
Smart Phone	5	Perception Neuron	2
Webcam	3	Electromyography (EMG)	1
Depth Camera	2	Position Tracker	1
Event Camera	1	Smart Watch	1
Open Computer Vision Camera	1	Total	17
Raspberry pi Camera	1		
Total	100		
Hybrid Method			
Camera + EMG		1	
Camera + Data Glove		1	
Smart Phone + Sign Ring		1	
Total		3	

Table 7 shows different types of data collection tools that have been found from 120 datasets in the 85 studies reviewed. There are nine tools that have been used in collecting vision-based data and seven tools for sensor-based data. For vision-based data collection, most of the datasets were collected by using camera as identified in 53 of the studies [4,8,10-11,16-18,20-22,23-25,27-29,32-35,37,39-40,42,46-47,49,52,54,56,59-61,63-66,70, 72-76,78,98-99,101,105-106,108-109,112,114]. Ten datasets were collected through LMC tool [9,14,48,55,58,67-68,106]. There are also six dataset data collected through Microsoft Kinect [9,30,76-77,102,112]. Smart phone cameras were also used as a data collection tool to capture videos and images for five of the datasets found [12,31,36,76,80]. There are three datasets collected through webcam [18,38,47] and two datasets using depth camera [46,58,75,106]. While there is only one dataset found from the 85 studies collected by event camera [99], Open Computer Vision [11], and Raspberry pi camera [76]. On the other hand, data glove was used as the popular tool in collecting sensor information for eight of the datasets studied [13,50,69,79,100, 103,110-111]. Next, there are two datasets developed through armband [55,113], millimeter wave radar system [19,71] and perception neuron [26]. Lastly, an EMG sensor system [15], a position tracker [111] and a smart watch [28] were used to capture sensor information for one dataset from the studies found. Apart from that, there are three datasets that have been developed through a combination of vision and sensor-based tools to collect hybrid data [22,57,104].

3.2.3.2 RQ 3(b): What is the availability of the dataset?

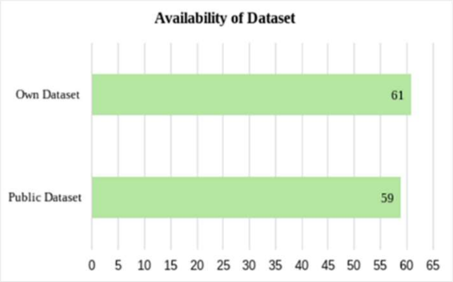

Figure 5: Availability of Datasets found from the Studies

According to Figure 5 above, there are two types of datasets that have been covered in the 85 studies reviewed, which are identified as own and public datasets. Both types of datasets have a similar total number in close proximity from the reviewed studies. There are 61 own datasets developed by the researchers in their studies, while some of these researchers used public datasets published in previous studies as well. There are 59 different public datasets that have been included in the studies reviewed.

Table 8: List Of Datasets

Availability	Datasets
Public Dataset (59)	26ASL Dataset, 50ASL Dataset, 500 CSL Dataset (7), American Sign Language (ASL) Dataset1, American Sign Language (ASL) Dataset2, American Sign Language Lexicon Video (ASLLVD) Dataset (3), American Sign Language (Triesch) Dataset, Ankara University Turkish Sign Language Dataset, Arabic Sign Language Dataset, Arabic Sign Language (Al-Jarrah) Dataset, ASL Alphabet Dataset, ASL Dataset (2), ASLLRP Dataset (2), Australian Sign Language Dataset, British Sign Language (BSL) Dataset, Cambridge Dataset, ChicagoFSWild Dataset, ChicagoFSWild+ Dataset, CSL-Daily Dataset, CSL Dataset (6), DEVISIGN-D Dataset, DJSLC Dataset, DHG Dataset (4), Dicta-Sign-LSF-v2 Dataset, First-Person Hand Action (FPHA) Dataset (2), Greek Sign Language (GSL) Dataset, GSL Dataset, HANDS17 Dataset, HKSL Dataset, How2Sign Dataset, ICSLD Dataset, INCLUDE Dataset, ISL-CSLTR Dataset (2), Isolate Words Arabic Sign Language Dataset, Jester Dataset, KArSL-190 Dataset, KArSL-502 Dataset (2), Kazakh-Rusian Sign Language (K-RSL) Dataset, KSU-SSL Dataset (2), Large KSL Dataset, LIBRAS-BSL Dataset, LMDHG Dataset (2), LSA64 Dataset (6), LSA40 Dataset, Marcel Dynamic Dataset, MNIST Dataset, MSASL Dataset, NVIDIA Gesture Dataset, RWTH-Boston-50 Dataset, RWTH-Boston-104 Dataset, RWTH-PHOENIX-Weather 2014 Dataset (10), RWTH-PHOENIX-Weather 2014T Dataset (5), Saudi Dictionary Dataset, Shanableh Dataset, Shape Retrieval Contest (SHREC) Dataset (4), SIGNUM Dataset, SLR-500 Dataset, SLR Dataset, WLASL Dataset (3)
Own Dataset (61)	48CSL Dataset, Isolated Hand Gestures Recognition of 26 letter in Alphabet Dataset, American Sign Language (ASL) Dataset, Armband New Dataset, ASL Alphabet Dataset, Arabic Dynamic Sign Language Dataset, ASL Alphabets Dataset, ASL Dataglove Dataset, ASL Dataset (Raspberry pi Camera), ASL Gestures Dataset, ASL IMU Dataset, Baby Sign Dataset, Channel State Information Dataset, Complex ASL Dataset, CSL Action (Dataglove) Dataset, CSL mmWave Dataset, CSL Standard Dataset, DSL-46 Dataset, DSL10 Dataset, DVS_Sign Dataset, DVS_Sign_v2e Dataset, Dynamic Arabic Sign Video Dataset, Dynamic ArSL Dataset, Dynamic Gesture Dataset, Dynamic Thai Fingerspelling Dataset, Gesture Dataset, Gesture Direction Dataset, Handicraft-Gesture Dataset, Hong Kong Continuous Sign Language (HKSL) Dataset, Hong Kong Sign Language (HKSL) Dataset, Indian Sign Language Custom Dataset, Indian Sign Language MIT-Anna University Dataset, Individual Alphabet Dataset, ISL Agricultural Words Dataset, ISL Dataset (Static & Dynamic), ISL Farmer Dataset, ISL Patient Dataset, Isolated Chinese Sign Language Dataset, Italian Sign Language Dataset, Japanese Sign Language Dataset, Japanese Video Dataset, Kazakh Sign Language (KSL) Dataset, KL_MV2DSL Dataset, KSL Lab Dataset, LMC Hand 3D Dataset, MuHAVi Dataset, MVDMML Dataset, Malaysian Sign Language (MSL) Videos Dataset, Myo Armband Sensor Dataset, New Dataset (Number, Alphabet, Phrase), NTU RGB D Dataset, NUMA Dataset, Number Dataset, Pakistan Sign Language Dataset, Portuguese Sign Language LGP Dataset, Sign-language Dataset, Sign Language Correctness Discrimination (SLCD) Dataset, Simple ASL Dataset, Talking Hands Dataset, Tunisian Sign Language (TunSigns) Dataset, WEIZMANN Dataset

Table 8 shows the list of 120 datasets that have been found in the 85 studies reviewed. Of the 59 public datasets, there are some datasets that have been used several times in different studies. The most popularly used public dataset found was RWTH-PHOENIX-Weather 2014 dataset which has been used in the training and testing for dynamic SLR models in ten studies [21,24,27,29, 32,52,54,60-62]. Then, its extension which is the RWTH-PHOENIX-Weather 2014T dataset was also found in five different studies [24,29,40, 52,60]. Other than these, the 500 CSL dataset was used in seven studies [24,27,29,52,61,98,112]. LSA64 [11,42,64,76,98,105] and the CSL dataset [32,54,57,60,62,112] is also often used by researchers in six studies found. In addition to that, four studies also applied the use of the DHG dataset [58,75,106] and SHREC [58,67-68,106] in testing their models' performances. Apart from this, the ASLLVD dataset [21,66,73] and the WLASL dataset [29,34-35] have been found in three different studies, whereas the ASL dataset [67-68], ASLLRP [22,61], FPHA dataset [75,106], ISL-CSLTR dataset [40,56], KArSL-502 [4,105], KSU-SLL [17,46] and LMDHG dataset [67-68] were found in two different studies. Besides that, the rest of public datasets have only been used in one study each from the total number of datasets [4,8-11,20-21,25-26,29-30,34-36,39-40,42,46,49,59-60,62,64,72,74,98,101,105, 108,111]. Whereas 61 own datasets that have been listed in Table 8 which were also newly developed datasets have only been used by the respective researchers only in their own studies [8,11-15,18-19,22-23,26-27,31,37-38,42,47-50,55,57-59,62-66, 69,70-71,76-80,99-104,107-113].

3.2.3.3 RQ 3(c): What is the size of the dataset?

Table 9: Sample Sizes Of Datasets

Sample Size	Number of Datasets	
	Own Datasets	Public Datasets
> 10000	12	13
5001-10000	5	11
2501-5000	3	13
1000-2500	12	6
< 1000	29	16

Through a detailed investigation, 120 datasets were found with different sizes of samples that have been developed by researchers. Table 9 shows all the datasets found arranged into different range of sample sizes, which are more than 10000, 5001-10000, 2501-5000, 1000-2500, and less than 1000 samples. According to the table, there are 28 datasets found in preparing over 10000 datasets, 12 for own datasets [8,13,18,33,42,57,80,108,110,112] and 13 for public datasets [4,9,17,24,27,29-30,34-35,46,52,56,61,64,98,101,105,112]. Next, there are five own datasets [50,71,76-77,102,107] and nine public datasets [4,9,16,21,24-27,29,32,39-40,52,54, 57,60-62,73-74,98,112] found for the sample sizes of 5001 to 10000. The 2501 to 5000 samples size range had the least number of datasets found, which are three own datasets [37,48,71] and 13 public datasets [4,8-9,11,20,25,28,36,42,58,60,64,67-68, 75,98,105-106,111]. 18 datasets were also found between the sizes of 1000 to 2500 samples, which are 12 own datasets [18,23,26-27,31,64,66,69,78, 100,104,108] and six public datasets [10,40,59,62, 75,101,106].

For the smallest size of datasets which is less than 1000 samples, there were 29 own datasets [11-12,14-15,18-19,22-23,28,38,47,55,59,63,65,79, 99,101,103,108-109,111,113] and 16 public datasets [11,22,26,29,40,46-47,49,58,61,67-68,72,101,106, 108] identified from the total of 120 datasets. Overall, the Isolated Chinese Sign Language Dataset was the largest own dataset that has been found from the studies reviewed, which contains 70000 vision-based samples obtained using Microsoft Kinect [112], while the Individual Alphabet Dataset has the highest number of own dataset samples (41600) for a sensor-based dataset that was developed using data glove [110]. Besides these, the largest public dataset that has been found from the studies reviewed is the SIGNUM dataset that has 33210 samples in video format [101]. However, the smallest sizes of own and public datasets were not able to be determined clearly in this study due to some of datasets found without the mention of an accurate number [15,19,22,28,49,55,59,62-63,65,72,98-99,103-104, 108,113-114]. The lowest number of samples that was obtained based on the accurate stated value was from the HANDS17 dataset which has 99 public data samples [29].

3.2.3.4 RQ 3(d): What is the size of the dataset?

Table 10: Sign Languages Used In Datasets

Sign Language	Number of Datasets	
	Own Datasets	Public Datasets
American Sign Language	13	15
Chinese Sign Language	6	6
Indian Sign Language	8	2
Arabic Sign Language	3	6
German Sign Language	0	6
Japanese Sign Language	2	2
Hong Kong Sign Language	2	1

Argentinian Sign Language	0	2
Kazakh Sign Language	1	1
Korean Sign Language	1	1
Saudi Sign Language	0	2
Australian Sign Language	0	1
Brazilian Sign Language	0	1
British Sign Language	0	1
French Sign Language	0	1
Greek Sign Language	0	1
Turkish Sign Language	0	1
Italian Sign Language	1	0
Malaysian Sign Language	1	0
Pakistan Sign Language	1	0
Portugues Sign Language	1	0
Thai Sign Language	1	0
Tunisian Sign Language	1	0
Others	19	9

Based on Table 10, there are 24 sign languages that have been found from the total of 120 datasets stated before. The most common sign language that were studied by the researchers is American Sign Language that had 13 own datasets [8,13-14,18,22,26,34,42,48,65,69,71,110] and 15 public datasets [4,16,21,26,29,35,40,47,49,62,72-75,102]. Follow by that, Chinese Sign Language had the same total number of datasets with six own datasets [4,33,79,103,112] and six public datasets [21-22,24,27,29,52,57,60-61,73,98], respectively. There was a total of ten datasets found for Indian Sign Language, where eight datasets were own datasets [12,23-24,30,63,66,108-109] and two were public datasets [20,40,56]. Nine datasets have been found using Arabic Sign Language with three own [11,77,102] and six public datasets [4,25,72,105, 108,113]. German Sign Language was also popular used in datasets, such as the RWTH's series and another two public datasets that are commonly used in the studies reviewed [21,24,27,29,32,40,52,54, 60-62,101].

Other than this, the rest of sign languages have less than five datasets respectively. Two own datasets [80,104] and public datasets [10,46] were developed using Japanese Sign language. Additionally, two own datasets [27-28] and one public dataset [62] were developed using Hong Kong Sign Language. There is no own dataset that has been developed by the researchers in their studies for Argentinian Sign Language [11], and Saudi Arabian Sign Language [17,46], but the researchers did train and test their models with two public datasets found for each sign language. According to Table 10, there are also two datasets found using Kazakh Sign Language [4,64] and Korean Sign Language [59]. Moreover, there are the total of six public datasets in a certain sign language that have been used in the studies directly without developing any of their own dataset, which are developed for Australian [14], Brazilian [76], British [9], French [39], Greek [9] and Turkish [64] sign languages. While there are also some researchers who developed their own unique dataset for Italian [50], Malaysian [31], Pakistan [107], Portuguese [78], Thai [37], and Tunisian [101] sign languages from the studies reviewed without using any other public datasets. Apart from that, 19 own datasets from 13 studies [14-15,18-19,33,38,47,55,99-100,108,111,113] and nine public datasets [29-30,46,58,67-68,72,75,106] found did not mention the sign languages used and obtained certain sign gestures for their model to recognize.

3.2.4 RQ4: What pre-processing techniques are applied to the dynamic sign language datasets used in the models?

There are many pre-processing techniques identified in the reviewed studies. Data pre-processing is usually conducted in SLRT studies in order to reduce the computational complexity of processing the data for recognition and translation [44]. Both the vision-based and sensor-based studies each have different data pre-processing methods. The most common pre-processing methods identified for vision-based studies are splitting the video data into consecutive frames, dimensionality reduction or resizing, normalization, frame cropping or segmentation, denoising or smoothing, RGB to grayscale conversion, and lighting elimination. 20 studies [4,11,16-18,22,30,31,38,42,46-47,62,64,73, 76-77,102,107-108] have pre-processed their video input by splitting it into a certain number of frames.

Another 19 studies [11-12,16,18,21,30-31, 38,42,46-47,49,54,61,66,73,77,99,102] conducted dimensionality reduction or resized the frames and images data. Next, there are 18 studies [9,14,16,17,23,25,31,34,35,47,49,57,63,68,78,106, 109,73] identified that conducted normalization to ensure data consistency. 12 studies [12,17,25, 31,33,54,56,66, 74,77-78,99] cropped or segmented the frames and images to ensure the effects of nonrelevant features are reduced.

Next, there are 15 studies [10,14,16,21,23,56-57,61,68,73,75,78,99,105,107] that conducted denoising or smoothing to remove outliers and improve the quality of the frames and images. Finally, there are five studies [16,18,25,73,107] that converted RGB videos into grayscale and two studies [16,73] that conducted lighting elimination using Histogram equalization. Besides these common pre-processing techniques, several studies also conducted different pre-processing methods according to the requirements of their data and system. For instance, Chen et al. [99] converted RGB video frames into Dynamic Vision Sensor data using the video-to-event method whereas Abdullah et al. [8] converted RGB sign images into Hue Saturation Value scale-based images. Abdallah et al. [76] and Adão et al. [78] flipped the video data to ensure full coverage of every sign. While Amrutha et al. [20] removed duplicates, deblurred frames, and fused frames to pre-process their data.

For sensor-based studies, the most common pre-processing techniques used are normalization, and denoising or smoothing. There are nine studies [13,15,50,55,57,69,79,110,113] identified conducting normalization and eight studies [15,26,28,50,55,57,79,113] identified conducting denoising or smoothing processes. Additionally, one study [22] was found to have up-sampled their data based on frame rates. Out of the studies reviewed, there are also 23 studies [9,24,27,29,32,36-37,39-40,48,52,58-60,65,71-72,80,100,104,108,111-112] that did not explicitly mention or explain any pre-processing techniques that were used in their studies while six studies [21,25,66,103,106,109] briefly mentioned the pre-processing techniques involved in the data processing or feature extraction section.

3.2.5 RQ5: What are the evaluation metrics used to assess the models?

Table 11: Types Of Evaluation Metrics Used In The Reviewed Studies

Evaluation Metrics	ISL	CSL		
		S2G	S2G2T	S2T
Accuracy	59	2	-	9
Word Error Rates (WER)	1	3	2	11
BiLingual Evaluation Understudy (BLEU) Score	-	-	-	2
Character Error Rate (CER) / Levenshtein distance	2	1	-	-
Others	4	1	-	2

Table 11 shows the types of evaluation metrics used in the reviewed studies. The findings show 59 studies [4,8-9,11-19,23,25-26,30-31,34-38,42,46-50,55,58-59,63-69,70-72,75-78,98-111] that have utilized Accuracy as the evaluation metric for isolated sign language systems. There are 11 studies that have used Accuracy for continuous sign language systems, namely Sign2Gloss studies and Sign2Text studies at two [20-21] and nine studies [10,24,29,33,39,40,56,79,108], respectively. Ten studies [23,31,36,49,59,64,66,69,75,100] included the findings for Precision, Recall, and F1-Score in order to find the Accuracy. Muthusamy & Murugan [56] only included Recall and Precision with Accuracy whereas Zhang et al. [33] used the Mean Average Precision metric with Accuracy. Interestingly, Sharma & Kumar [73] did not calculate the Accuracy of their system instead, they only used Precision, Recall, and F1-Score to evaluate their system.

Next, the WER evaluation metric was mostly used in continuous sign language system-based studies where there are three [20-21,52], two [22,54], and 11 studies [10,27-29,32,57,60,62,80,112-113] that used WER for the Sign2Gloss, Sign2Gloss2Text, and Sign2Text approaches respectively. Another evaluation metric identified in the studies is the BLEU Score. The BLEU Score was identified in two Sign2Text continuous sign language system studies which are Natarajan et al. [40] and Ananthanarayana et al. [61]. Next, there were three studies identified to have used the CER evaluation metric which are Kozhamkulova et al. [16] and Pannattee et al. [74] for isolated sign language systems and Elakkiya et al. [21] for a Sign2Gloss continuous sign language system. The CER evaluation metric is said to be derived from the Levenshtein Distance metric. Several studies have also included other types of evaluation metrics such as Computation Time [16], Validation Loss [16,31], Probability [20], Training Time [31], Fréchet Inception Distance for video (FID2vid) [40], Structural Similarity (SSIM) Index [40], Prediction Time Complexity [56], Specificity [75], and Time Response [110],

4. DISCUSSION

Based on the publications in the years 2020 till 2024, it can be procured that SLRT is a growing field of research. However, the majority of studies conducted throughout these past five years have been on isolated sign language, which does not effectively solve the real-time communication issues faced by DHH individuals. Nevertheless, the effort of researchers exploring continuous SLRT in 24 of the studies mentioned in the findings for the first research question must be acknowledged. A sign language recognition system is only efficient when the translation aspect is also covered. This is because users would not necessarily be able to comprehend the entire conversation with only glosses or lexical signs. Moreover, since conversations are usually sentence-based, the system developed should be able to recognize and translate continuous sign language. Unlike isolated signs, continuous signs are more complex to perform and process as sequential data in models.

Based on the studies reviewed, a couple of researchers have faced challenges in accurately detecting sign semantics and segmentations from the sequential signs performed by signers in continuous sign language systems [24,52]. Multiple deep learning techniques and additional tools, such as gloss annotations, vectors, and tokenization, are needed to develop and improve models for effective recognition and translation. As a result, some existing isolated sign language models identified in the studies have the potential to be extended and upgraded into powerful continuous sign language models by incorporating appropriate deep learning techniques. This approach has been implemented and documented by several researchers [24,79]. The suggested modifications of models, from isolated to continuous sign language systems, can be informed by the effective continuous sign language models discussed in the reviewed studies.

According to research questions one and two, researchers have focused on continuous sign language models that were developed using the S2T frameworks, which were a total of 19 studies. Most of these models are hybrid models with combinations of CNN with additional techniques such as RNN variants to better process temporal data of continuous signs. For example, there are eight CNN hybrid models found that were combined with LSTM, BiLSTM or GRU techniques to perform SLRT tasks on their datasets [22,28,33,39,40,54,56,108]. This proved that the combination of CNN-RNN and RNN variants in existing isolated sign language models has created opportunities for processing sentence-based data to upgrade them into SLRT models. In addition, Window Sliding Attention was found as a potential modification that may be able to produce better segmentation between signs in continuous presentations [79,112].

Other than this, Transformers and its variants were also highly recommended for developing continuous sign language models. This is because most of the techniques for Transformers, BERT, BERT+, and STTN were commonly tested for SLRT hybrid models using the S2T framework [27,29,32,60-62,80]. Thus, it can be derived that existing Transformers models can be modified to process SLRT tasks for sentence-based datasets. Nevertheless, some studies with existing isolated sign language models that achieved good results also suggested upgrading to SLRT models by combining Transformers techniques to perform sequence-to-sequence tasks from sign to text conversion. Apart from that, the encoder-decoder is another technique that is encouraged as it is a good additional tool in assisting models to perform recognition and translation tasks in an end-to-end manner. For instance, some existing encoder-decoder models found from the studies performed SLRT tasks concerning continuous sign language using the S2T framework by assigning certain algorithms to their encoder and decoder [29,33,56-57,112-113].

On the other hand, there are also some models that applied the methods of gloss annotations to perform recognition and translation using the S2G2T framework. These models are mostly developed with the combination of CNN and RNN techniques [22,54]. In both the S2G2T models found from the studies reviewed, there is a common ground in applying the CTC mechanism to improve the performance of processing long-term data into sequences. This mechanism was recommended as an efficient tool that may assist S2G or isolated sign language models with the G2T process. According to Papastratis et al. [52], recognition tasks for isolated sign language and S2G framework-based models follow a similar process of converting signs into corresponding words or glosses. It is highly recommended that S2G framework models be upgraded to S2G2T frameworks by incorporating robust attention mechanisms or leveraging existing G2T techniques from current models [21,52].

Next, according to the third research question, in order to develop a good continuous sign language model, the model must be trained and tested with suitable datasets. Based on the studies reviewed, researchers have often used RWTH-PHOENIX-Weather 2014 and its extension, the RWTH-PHOENIX-Weather 2014T datasets on German Sign Language to test their SLRT models [24,27,29,32,40,54,60,62]. Besides these, CSL and 500 CSL were also popular datasets that were highly used to train and test SLRT models [24,27,32,57,60-62]. The RWTH-PHOENIX datasets were prepared with around 7000 sentences by nine signers which resulted in 1232 words and 80000 glosses and are stored in video format, whereas the CSL datasets were also prepared on a large-scale, with around

25000 videos [54]. Because of their large-scale sizes, these public datasets are suitable for training and testing both recognition and translation performances for S2G2T and S2T framework-based models.

Furthermore, there is another public dataset called the CSLTR dataset that includes 18863 sentence-based frames for Indian Sign Language [56]. It has been found in testing two vision-based continuous sign language models in the reviewed studies [40,56]. On the other hand, there is a comprehensive sentence-level public dataset called SIGNUM, which includes a large vocabulary of German Sign Language in 33210 videos for recognition and translation tasks. However, this dataset was underutilized in the reviewed studies, with only eight gestures selected for testing in a single isolated sign language model [101]. Researchers should take the opportunity to utilize this large-scale dataset to train and test S2G2T or S2T framework-based models. Based on the public datasets introduced in this study, the training and testing of these datasets in novel SLRT models are highly encouraged in future studies to assist the researchers in assessing their model's performance.

Apart from this, the development of own datasets was also encouraged for researchers focusing on specific sign languages and requirements. Some researchers focused on developing their own sign language datasets for training and testing in order to determine whether their models were able to perform well for specific sign languages. Isolated Malaysian [31], Italian [50], Pakistan [107], Portuguese [78], and Thailand [37] sign languages' datasets were created in these studies, respectively. On another note, continuous sign language datasets have a more complex development process compared to isolated sign language datasets, where only three studies have developed continuous datasets which are the 48 CSL, Hong Kong Sign Language and Japanese Video dataset [28,79,80]. In the 48 CSL sensor-based dataset, there is a possibility of extending the isolated words-level dataset into a new sentence-level dataset by combining those vocabularies and their grammar rules [79]. Researchers could improve existing isolated sign language datasets by transitioning from word-level to sentence-level datasets, particularly for large vocabulary datasets such as the Isolated Chinese Sign Language dataset, which contains 70,000 samples [112].

Next, according to the findings for research question four, the studies reviewed have a diverse array of pre-processing techniques for both vision-based and sensor-based SLRT systems. The pre-processing techniques for isolated and continuous sign language videos are similar, which are frame splitting, dimensionality reduction, normalization, cropping or segmentation, and denoising or smoothing, RGB to grayscale, and lighting elimination. Most of the continuous sign language-based studies have specifically focused on frame splitting [62,108], dimensionality reduction [21,54], normalization [57,79,113], cropping or segmentation [33,54,56], and denoising or smoothing [10,21,28,56,61,79]. From these studies, it can be derived that continuous sign language videos introduce additional complexity due to their temporal nature. While the pre-processing techniques are similar to isolated sign language, their application must adapt to handle the longer continuous flow of data. The best pre-processing techniques for continuous sign language can be determined by the researcher depending on the dataset or data collection procedure. Several studies [27,29,32,40,60,80,112] that used public datasets did not mention any pre-processing techniques except for one study [62] that adopted a frame selection mechanism. Studies that have created their own dataset on the other hand, have pre-determined certain aspects of their data collection procedure. For example, Takayama et al. [80] collected video data at 30 frames per second hence, no pre-processing of the data frames splitting needed to be done after the data has been collected. Hence, the process of frame splitting to pre-processing the data was not required.

Therefore, researchers may need to conduct pre-processing techniques based on the nature of their data. The most common pre-processing that can be done to ensure the input data can be easily processed is denoising or smoothing. This is because dynamic data tends to have external noise or outliers which may affect the clarity of the data. Next, although some studies did not explicitly mention or explain the pre-processing techniques used, researchers are encouraged to explain the pre-processing techniques that will be conducted in future SLRT studies for research transparency. This will allow for reproducibility of the study so that improvements can be continuously made in this field of research. Besides, more research is also needed to evaluate the effectiveness of specialized pre-processing techniques like dynamic vision sensors or frame flipping as these techniques may offer significant benefits in specific contexts.

Finally, based on the findings for research question five, the WER evaluation metric is recommended for continuous sign language-based system as compared to accuracy. This is because accuracy works well with isolated data whereas WER is able to evaluate sequence-based data more efficiently [40]. On the other hand, the BLEU evaluation metric is also said to be popularly used for sentence-based data in machine translation [61]. Therefore, based on the studies reviewed, Accuracy is a popular evaluation metric that can be used for isolated sign language system evaluation for the recognition task whereas WER and BLEU evaluation metrics can be used for continuous sign language system evaluation for the translation task. Researchers are encouraged to use popular evaluation metrics to ensure that their studies can be compared with other studies as well. This will give a clearer picture on whether the model they have used is better in terms of performance, giving future researchers directions of the models that they can use, modify, and improvise in order to develop an efficient SLRT system.

Overall, with the wide range of deep learning techniques used in the dynamic field of SLRT that have been continually evolving through variants and combinations of namely CNN, RNN, Transformers deep learning models and techniques, key improvements in the field of SLRT have been observed and can be anticipated in future research. As the state-of-the-art research on SLRT has stepped into continuous sign language SLRT research, much future research on real-time continuous SLRT can be expected. Moreover, vision-based models have gained the interest of researchers as it is more likely to ensure every aspect of sign language such as the manual and non-manual cues have been covered for SLRT. This leads as a stepping-stone in real-time SLRT research as vision-based models are more convenient compared to hardware sensor-based models. Additionally, substantial advancement in additional techniques, such as attention mechanisms, are needed to ensure end-to-end translation can be conducted using the S2T framework while there is still a need for larger and more curated sign language datasets to be developed. Furthermore, heavy reliance on gloss-based models may not be entirely applicable for every sign language, researchers may work on developing larger datasets for a specific sign language or multi-lingual datasets with glosses as well to ensure both the gloss-based frameworks and the gloss-free frameworks can be tested for the best possible outcome. All in all, the state-of-the-art research on SLRT is one step closer to bridging the gap between hearing and DHH individuals.

5. CONCLUSION

In conclusion, a systematic literature review offers insights and meaningful findings regarding a specific field of study. This study has identified and summarized the key takeaways from 85 studies reviewed, which were obtained from the Scopus database from the year 2020 till 2024. This systematic literature review highlights the SLR and SLT frameworks, deep learning models and techniques, datasets, pre-processing techniques, and evaluation metrics used by researchers in SLRT research over the five years. Additionally, this research contributes to the body of systematic literature reviews in the field of SLRT systems, providing a detailed overview of the current leads in this research area. Based on the studies reviewed, it can be derived that researchers have begun to recognize the importance of SLRT research in aiding the DHH community.

As a summary of the findings, isolated SLR studies have outnumbered continuous SLRT studies (S2G, G2T, S2G2T, S2(G+T), S2T). However, continuous SLRT has gained the interest of researchers, especially using the S2T framework in recent years, indicating improvement and growth in this field. CNN, RNN, and Transformers are the most commonly used baseline deep learning models and techniques in a majority of the studies. Additionally, the reviewed studies suggest that existing isolated models have potential for adaptation to continuous SLRT, yet this remains underexplored. Techniques such as CNN-RNN hybrids, attention mechanisms, and Transformers have been shown to improve processing of sentence-level data, but their application to real-time tasks requires further validation.

The RWTH Phoenix Weather 2014 dataset and its extension (RWTH Phoenix Weather 2014T dataset) are the most commonly used public continuous sign language dataset in standardizing performance evaluation, while more diverse, multilingual, and real-time-appropriate datasets are still necessary. Pre-processing techniques differ with every study depending on the type of data that is being pre-processed and the data collection procedure, while Accuracy and WER are notably the most common evaluation metrics used in the studies. Nevertheless, the inconsistencies in evaluation metrics such as WER and BLEU hinder model comparability and the ability to draw broader conclusions. Future studies should prioritize transparency in choosing standardized sign language datasets and adopt suitable evaluation metrics based on their frameworks to enhance comparability across models.

This study acknowledges certain limitations, such as the exclusion of highly cited journal articles on SLRT that are not available in the Scopus database and another part of the SLRT process which is the feature extraction process that was not explored. It is acknowledged that the exact real-time SLRT process of the models is also difficult to determine due to the limitations of the current available models for dynamic sign language. Future researchers may consider reviewing SLRT literature from other databases, such as Web of Science, ERIC, IEEE Xplore, and Science Direct, in order to view this field of research from a much bigger perspective. SLRT researchers on the other hand may consider exploring the processes that can direct them towards real-time SLRT to ensure their SLRT models are more comprehensive and practical for daily usage. All in all, the research on SLRT is definitely advancing in many ways using deep learning and the current technological advantages such as the Internet of Things. While dynamic SLRT has seen notable progress, there remains a need for continued research to address and tackle the challenges and limitations of currently available studies in order to bridge the communication gap between DHH individuals and the wider society.

REFERENCES:

[1] Amin, M., Hefny, H., and Mohammed, A., "Sign Language Gloss Translation using Deep Learning Models", *International Journal of Advanced Computer Science and Application*, Vol. 12, No. 11, 2021, pp. 686-692.

[2] World Federation of the Deaf. Available online: https://wfdeaf.org/our-work/ (accessed on 20 May 2024).

[3] Papastratis, I., Dimitropoulos, K., and Daras, P., "Continuous Sign Language Recognition through a Context-Aware Generative Adversarial Network", *Sensors,* Vol.21, No. 7, 2021, pp. 1-20.

[4] Luqman, H., and El-Alfy, E., "Utilizing motion and spatial features for sign language gesture recognition using cascaded CNN and LSTM models", *Turkish Journal of Electrical Engineering and Computer Science*, Vol. 30, No. 7, 2022, pp. 2508–2525.

[5] Al-Qurishi, M., Khalid, T., and Souissi, R., "Deep learning for sign language recognition: current techniques, benchmarks, and open issues", *IEEE Access*, Vol.9, 2021, pp. 126917–126951.

[6] Samonte, C., Jane, M., Carl, J., Guingab, R., Relayo, J., Sheng, J., and Ray, D., "Using Deep Learning in Sign Language Translation to Text", *Proceedings of the International Conference on Industrial Engineering and Operations Management*, Istanbul, Turkey, March 7-10, 2022, pp. 4036-4044.

[7] Zhang, Y., and Jiang, X., "Recent advances on deep learning for sign language recognition", *Computer Modeling in Engineering and Science*, Vol. 139, No. 3, 2024, pp. 2399–2450.

[8] Abdullah, M. A., Wael, A., Marwan, A., Sara, A., and Mojtaba, N., "Intelligent gesture recognition system for deaf people by using CNN and IoT", *International Journal Advanced Science and Computer Applications*, Vol. 15, No. 1, 2023, pp. 145-158.

[9] Abdullahi, S. B., Chamnongthai, K., Bolon-Canedo, V., and Cancela, B., "Spatial–temporal feature-based end-to-end fourier network for 3D sign language recognition", *Expert System and Application*, Vol. 248, 2024, pp. 1-15.

[10] Brock, H., Farag, I., and Nakadai, K., "Recognition of non-manual content in continuous Japanese sign language", *Sensors*, Vol. 20, No. 19, 2020, pp. 1-21.

[11] Elsayed, E., and Fathy, D., "Semantic deep learning to translate dynamic sign language", *International Journal of Intelligent Engineering and Systems*, Vol. 14, No. 1, 2020, pp. 316–325.

[12] Bansal, N., and Jain, A., "Word recognition from Indian Sign Language using Transfer Learning Models and RNN Classifier". *International Journal of Intelligent System and Application in Engineering*, Vol. 12, No. 9s, 2024, pp. 182-189.

[13] Lee, B. G., Chong, T., and Chung, W., "Sensor Fusion of Motion-Based Sign Language Interpretation with Deep Learning", *Sensors*, Vol. 20, No. 21, 2020, pp. 1-17.

[14] Yang, L., Chen, J., and Zhu, W., "Dynamic hand gesture recognition based on a leap motion controller and Two-Layer bidirectional recurrent neural network", *Sensors,* Vol. 20, No.7, 2020, pp. 1-17.

[15] Yu, M., Li, G., Jiang, D., Jiang, G., Zeng, F., Zhao, H., and Chen, D., “Application of PSO-RBF neural network in gesture recognition of continuous surface EMG signals” *Journal of Intelligent and Fuzzy Systems*, Vol. 38, No. 3, 2020, pp. 2469–2480.

[16] Kozhamkulova, Z., Nurlybaeva, E., Kuntunova, L., Amanzholova, S., Vorogushina, M., Maikotov, M., and Kenzhekhan, K., “Two dimensional Deep CNN model for vision-based fingerspelling recognition system”, *International Journal of Advanced Computer Science and Applications*, Vol. 14, No. 9, 2023, pp. 1017-1024.

[17] Al-Hammadi, M., Muhammad, G., Abdul, W., Alsulaiman, M., Bencherif, M. A., Alrayes, T. S., Mathkour, H., and Mekhtiche, M. A., “Deep Learning-Based approach for sign language gesture recognition with efficient hand gesture representation”, *IEEE Access*, Vol. 8, 2020, pp. 192527–192542.

[18] Bendarkar, D. S., Somase, P. A., Rebari, P. K., Paturkar, R. R., and Khan, A. M., “Web Based Recognition and Translation of American Sign Language with CNN and RNN”, *International Journal of Online and Biomedical Engineering*, Vol. 17, No. 1, 2021, pp. 34-50.

[19] Dang, X., Wei, K., Hao, Z., and Ma, Z., “Cross-Scene sign Language gesture recognition based on Frequency-Modulated Con-tinuous Wave Radar”, *Signals,* Vol. 3, No. 4, 2022, pp. 875–894.

[20] Amrutha, K., Prabu, P., and Chandra, P. R., “LIST: A lightweight framework for continuous Indian Sign language translation”, *Information*, Vol. 14, No. 12, 2023, pp. 1-21.

[21] Elakkiya, R., Vijayakumar, P., and Kumar, N., “An optimized Generative Adversarial Network based continuous sign language classification”, *Expert System and Applications,* Vol. 182, 2021, pp. 1-12.

[22] Li, J., Zhong, J., and Wang, N., “A multimodal human-robot sign language interaction framework applied in social robots”, *Frontiers in Neuroscience*, Vol. 17, 2023, pp. 1-15.

[23] Joshi, J. M., and Patel, D. U., “Dynamic Indian Sign Language Recognition Based on Enhanced LSTM with Custom Attention Mechanism”, *International Journal of Electronics and Communication Engineering*, Vol. 11, No. 2, 2024, pp. 60–68.

[24] Xiao, Q., Chang, X., Zhang, X., and Liu, X., “Multi-Information Spatial–Temporal LSTM Fusion continuous sign language neural machine translation”, *IEEE Access*, Vol. 8, 2020, pp. 216718–216728.

[25] Aly, S., and Aly, W., “DeepARSLR: A novel Signer-Independent deep learning framework for isolated Arabic sign language gestures recognition”, *IEEE Access*, Vol. 8, 2020, pp. 83199–83212.

[26] Gu, Y., Sherrine, S., Wei, W., Li, X., Yuan, J., and Todoh, M., “American Sign Language Alphabet Recognition Using Inertial Motion Capture System with Deep Learning”, *Inventions,* Vol. 7, No. 4, 2022, pp. 1-15.

[27] Zhou, Z., Tam, V. W. L., and Lam, E. Y., “SignBERT: A BERT-Based deep learning framework for continuous sign language recognition”, *IEEE Access,* No. 9, 2021, pp. 161669–161682.

[28] Zhou, Z., Tam, V. W. L., and Lam, E. Y., “A Portable Sign Language Collection and Translation Platform with Smart Watches Using a BLSTM-Based Multi-Feature Framework”, *Micromachines*, Vol. 13, No. 2, 2022, pp. 1-15.

[29] Hu, H., Zhao, W., Zhou, W., and Li, H., “SignBERT+: Hand-Model-Aware Self-Supervised Pre-Training for Sign Language understanding", *IEEE Transactions on Pattern Analysis and Machine Intelligent*, No. 45, No. 9, 2023, pp. 11221–11239.

[30] Huang, Y., Huang, J., Wu, X., and Jia, Y., “Dynamic Sign Language Recognition Based on CBAM with Autoencoder Time Series Neural Network”, *Mobile Information Systems*, 2022, pp. 1–10.

[31] Khan, R. U., Wong, W. S., Ullah, I., Algarni, F., Haq, M. I. U., Barawi, M. H. B., and Khan, M. A., “Evaluating the efficiency of CBAM-Resnet using Malaysian sign language” *Computers, Materials and Continua*, Vol. 71, No. 2, 2022, pp. 2755–2772.

[32] Cui, Z., Zhang, W., Li, Z., and Wang, Z., “Spatial–temporal transformer for end-to-end sign language recognition”, *Complex and Intelligent Systems*, Vol. 9, 2023, pp. 4645–4656.

[33] Zhang, M., Yang, S., and Zhao, M., “Deep Learning-Based Standard sign Language Discrimination”, *IEEE Access,* Vol. 11, 2023, pp. 125822–125834.

[34] Eunice, J., J, A., Sei, Y., and Hemanth, D. J., “Sign2Pose: A Pose-Based approach for gloss

prediction using a transformer model", *Sensors*, Vol. 23, No. 5, 2023, pp. 1-23.

[35] Woods, L. T., and Rana, Z. A., "Modelling Sign Language with Encoder-Only Transformers and Human Pose Estimation Keypoint Data", *Mathematics,* Vol. 11, No. 9, 2023, pp. 1-28.

[36] Al-Mohimeed, B. A., Al-Harbi, H. O., Al-Dubayan, G. S., and Al-Shargabi, A. A., "Dynamic sign language recognition based on Real-Time videos", *International Journal of Online and Biomedical Engineering*, Vol. 18, No. 2, 2022, pp. 4–14.

[37] Lata, S., Phiphitphatphaisit, S., Gonwirat, S., Surinta, O., "Dynamic fingerspelling recognition from video using deep learning approach: from detection to recognition", *ICIC Express Letters Part B: Applications,* Vol. 13, No. 9, 2022, pp. 950-957.

[38] Nadgeri, S., and Kumar, D., "Deep learning based framework for dynamic baby sign language recognition system", *Indian Journal of Computer Science and Engineering,* Vol. 13, No. 2, 2022, pp. 550–563.

[39] Belissen, V., Braffort, A., and Gouiffès, M., "Experimenting the automatic recognition of non-conventionalized units in sign language", *Algorithms*, Vol. 13, No. 12, 2020, pp. 1-36.

[40] Natarajan, B., Rajalakshmi, E., Elakkiya, R., Kotecha, K., Abraham, A., Gabralla, L. A., and Subramaniyaswamy, V., "Development of an End-to-End Deep Learning Framework for sign language recognition, translation, and video generation", *IEEE Access,* Vol. 10, 2022, pp. 104358–104374.

[41] Johari, R. T., Ramli, R., Zulkoffli, Z., and Saibani, N., "A Systematic Literature Review on Vision-Based Hand Gesture for Sign Language Translation", *Jurnal Kejuruteraan*, Vol. 35, No. 2, 2023, pp. 287–302.

[42] Palraj, J., Bhama, P. R. K. S., Swetha, K., and Subash, S. A., "Real time static and dynamic sign language recognition using deep learning", *Journal of Scientific and Industial Research*, Vol. 81, No. 11, 2022, pp. 1186-1194.

[43] Suthar, J., Parikh, D., Sharma, T., Patel, A., "Sign language recognition for static and dynamic gestures", *Global Journal of Computer Science and Technology: D Neural and Artificial Intelligent*, Vol. 21, No. 2, 2021, pp. 1-7.

[44] Boggaram, A., Boggaram, A., Sharma, A., Ramanujan, A. S., and R, B., "Sign language translation systems", *International Journal of Software Science and Computer Intelligent*, Vol. 14, No. 1, 2022, pp. 1–33.

[45] Madhiarasan, D. M., and Roy, P. P. P., "A Comprehensive review of sign language recognition: different types, modalities, and datasets", *arXiv*, 2022, pp. 1-30.

[46] Abdul, W., Alsulaiman, M., Amin, S. U., Faisal, M., Muhammad, G., Albogamy, F. R., Bencherif, M. A., and Ghaleb, H., "Intelligent real-time Arabic sign language classification using attention-based inception and BiLSTM", *Computers and Electric Engineering*, Vol. 95, 2021, pp. 1-15.

[47] Buttar, A. M., Ahmad, U., Gumaei, A. H., Assiri, A., Akbar, M. A., and Alkhamees, B. F., "Deep Learning in Sign Language Recognition: A hybrid approach for the recognition of static and dynamic signs", *Mathematics*, Vol. 11, No. 17, 2023, pp. 1-20.

[48] Lee, C., Ng, K. K., Chen, C., Lau, H., Chung, S., and Tsoi, T., "American sign language recognition and training method with recurrent neural network", *Expert Systems with Applications*, Vol. 167, 2021, pp. 1-14.

[49] Ma, Y., Xu, T., and Kim, K., "Two-Stream Mixed Convolutional Neural Network for American sign language recognition", *Sensors,* Vol. 22, No. 16, 2022, pp. 1-17.

[50] Saggio, G., Cavallo, P., Ricci, M., Errico, V., Zea, J., and Benalcázar, M. E., "Sign Language Recognition Using Wearable Electronics: Implementing k-Nearest Neighbors with Dynamic Time Warping and Convolutional Neural Network Algorithms", *Sensors,* Vol. 20, No. 14, 2020, pp. 1-14.

[51] Yin, K., and Read, J., "Better Sign Language Translation with STMC-Transformer", *arXiv*, 2020, pp. 1-15.

[52] Papastratis, I., Dimitropoulos, K., Konstantinidis, D., and Daras, P., "Continuous sign language recognition through Cross-Modal alignment of video and text embeddings in a Joint-Latent space", *IEEE Access*, Vol. 8, 2020, pp. 91170–91180.

[53] Liang, Z., Li, H., and Chai, J., "Sign Language Translation: A survey of Approaches and techniques", *Electronics*, Vol. 12, No. 12, 2023, pp. 1-18.

[54] Aditya, W., Shih, T. K., Thaipisutikul, T., Fitriajie, A. S., Gochoo, M., Utaminingrum, F., and Lin, C., "Novel Spatio-Temporal continuous sign language recognition using an attentive Multi-Feature network", *Sensors*, Vol. 22, No. 17, 2022, pp. 1-18.

[55] Li, J., Huang, L., Shah, S., Jones, S. J., Jin, Y., Wang, D., Russell, A., Choi, S., Gao, Y., Yuan, J., and Jin, Z., "SignRing: Continuous American Sign Language Recognition Using IMU Rings and Virtual IMU Data", *Proceeding of the ACM Interactive, Mobile and Wearable Ubiquitous Technologies*, Vol. 7, No. 3, 2023, pp. 1–29.

[56] Muthusamy, P., and Murugan, G. P., "Recognition of Indian sign language using spatio-temporal hybrid cue network", *International Journal of Intelligent Engineering and Systems*, Vol. 16, No. 6, 2023, pp. 874-885.

[57] Wang, Z., Zhao, T., Ma, J., Chen, H., Liu, K., Shao, H., Wang, Q., and Ren, J., "Hear Sign Language: a real-time End-to-End sign language recognition system", *IEEE Transactions on Mobile Computer*, Vol. 21, No. 7, 2022, pp. 2398-2410.

[58] Zhang, X., Zeng, X., Sun, W., Ren, Y., and Xu, T., "Multimodal spatiotemporal feature MAP for dynamic gesture recognition", *Computer Systems Science and Engineering*, Vol. 46, No. 1, 2023, pp. 671–686.

[59] Shin, J., Miah, A. S. M., Suzuki, K., Hirooka, K., and Hasan, M. a. M., "Dynamic Korean sign language recognition using Pose Estimation based and Attention-Based neural network", *IEEE Access,* Vol. 11, 2023, pp. 143501–143513.

[60] Hu, H., Pu, J., Zhou, W., Fang, H., and Li, H., "Prior-aware cross modality augmentation learning for continuous sign language recognition", *IEEE Transaction on Multimedia*, Vol. 26, 2024, pp. 593–606.

[61] Ananthanarayana, T., Srivastava, P., Chintha, A., Santha, A., Landy, B., Panaro, J., Webster, A., Kotecha, N., Sah, S., Sarchet, T., Ptucha, R., and Nwogu, I., "Deep learning methods for sign language translation", *ACM Transactions Accessible Computing*, Vol. 14, No. 4, 2021, pp. 1–30.

[62] Zhou, Z., Tam, V. W., and Lam, E. Y., "A Cross-Attention BERT-Based framework for continuous sign language recognition", *IEEE Signal Processing Letters*, Vol. 29 2022, pp. 1818–1822.

[63] Ahmed, H. F. T., Ahmad, H., Narasingamurthi, K., Harkat, H., and Phang, S. K., "DF-WISLR: Device-Free Wi-Fi-based sign language recognition", *Pervasive and Mobile Computing*, Vol. 69, 2020, pp. 1-17.

[64] Amangeldy, N., Milosz, M., Kudubayeva, S., Kassymova, A., Kalakova, G., and Zhetkenbay, L., "A Real-Time dynamic gesture variability recognition method based on convolutional neural networks", *Applied Sciences,* Vol. 13, No. 19, 2023, pp. 1-18.

[65] Durdi, V. B., Kiran, A., Rao, A., Bhat, S. S., B, S., and N, M. K., "A novel method for recognizing hand gestured sign language using the stochastic gradient descent algorithm and convolutional neural network techniques", *International Journal of Intelligent Systems and Applied in Engineering,* Vol. 12, No. 7s, 2024, pp. 529-538.

[66] Venugopalan, A., and Reghunadhan, R., "Applying hybrid deep neural network for the recognition of sign language words used by the deaf COVID-19 patients", *Arabian Journal for Science and Engineering,* Vol. 48, 2023, pp. 1349–1362.

[67] Abdullahi, S. B., and Chamnongthai, K., "American Sign language words recognition using Spatio-Temporal Prosodic and Angle features: a sequential learning approach", *IEEE Access,* Vol. 10, 2022, pp. 15911–15923.

[68] Abdullahi, S. B., and Chamnongthai, K., "American sign language words Recognition of skeletal videos using processed Video driven Multi-Stacked Deep LSTM", *Sensors,* Vol. 22, No. 4, 2022, pp. 1-28.

[69] Faisal, M. a. A., Abir, F. F., Ahmed, M. U., and Ahad, M. a. R., "Exploiting domain transformation and deep learning for hand gesture recognition using a low-cost dataglove", *Science Report 2022,* Vol. 12, 2022, pp. 1-15.

[70] Palraj, J., Bharma, P. R. K. S., and Madhubalasi, B., "Sign Language Recognition using Deep CNN with Normalised Keyframe Extraction and Prediction using LSTM", *Journal of Scientific and Industrial Research,* Vol. 82, No. 7, 2023, pp. 745-755.

[71] Lan, S., Ye, L., and Zhang, K., "Applying MMWAVe Radar Sensors to Vocabulary-Level Dynamic Chinese Sign Language Recognition for the community with deafness and hearing loss", *IEEE Sensors Journal*, Vol. 23, No. 22, 2023, pp. 27273–27283.

[72] Agab, S. E., and Chelali, F. Z., "New combined DT-CWT and HOG descriptor for static and dynamic hand gesture recognition", *Multimedia Tools and Applications*, Vol. 82, 2023, pp. 26379–26409.

[73] Sharma, S., and Kumar, K., "ASL-3DCNN: American sign language recognition technique using 3-D convolutional neural networks", *Multimedia Tools and Applications,* Vol. 80, 2021, pp. 26319–26331.

[74] Pannattee, P., Kumwilaisak, W., Hansakunbuntheung, C., Thatphithakkul, N., and Kuo, C. J., "American Sign language finger-spelling recognition in the wild with spatio temporal feature extraction and multi-task learning", *Expert Systems and Applications,* Vol. 243, 2024, ppt. 1-17.

[75] Shanmugam, S., and Narayanan, R. S., "An accurate estimation of hand gestures using optimal modified convolutional neural network", *Expert Systems and Applications*, Vol. 249, 2024, pp. 1-11.

[76] Abdallah, M. S., Samaan, G. H., Wadie, A. R., Makhmudov, F., and Cho, Y., "Light-Weight Deep Learning Techniques with Advanced Processing for Real-Time Hand Gesture Recognition", *Sensors,* Vol. 23, No. 1, 2023, pp. 1-20.

[77] Podder, K. K., Ezeddin, M., Chowdhury, M. E. H., Sumon, M. S. I., Tahir, A. M., Ayari, M. A., Dutta, P., Khandakar, A., Mahbub, Z. B., and Kadir, M. A., "Signer-Independent Arabic Sign Language Recognition System using Deep learning model", *Sensors*, Vol. 23, No. 16, 2023, pp. 1-21.

[78] Adão, T., Oliveira, J., Shahrabadi, S., Jesus, H., Fernandes, M., Costa, Â., Ferreira, V., Gonçalves, M., Lopéz, M., Peres, E., and Magalhães, L., "Empowering Deaf-Hearing Communication: Exploring Synergies between Predictive and Generative AI-Based Strategies towards (Portuguese) Sign Language Interpretation", *Journal of Imaging*, Vol. 9, No. 11, 2023, pp. 1-30.

[79] Liu, Y., Jiang, X., Yu, X., Ye, H., Ma, C., Wang, W., and Hu, Y., "A wearable system for sign language recognition enabled by a convolutional neural network", *Nano Energy*, Vol. 116, 2023, pp. 1-10.

[80] Takayama, N., Benitez-Garcia, G., and Takahashi, H., "Sign language recognition based on Spatial-Temporal Graph Convolution-Transformer", *Journal of the Japan Society for Precision Engineering*, Vol. 87, 2021, pp. 1028–1035.

[81] Kitchenham, B., and Charters, S. M., "Guidelines for performing systematic literature reviews in software engineering", *EBSE Technical Report*, 2007, pp. 1-14.

[82] Usman, M., Ali, N. B., and Wohlin, C., "A quality assessment instrument for systematic literature reviews in software engineering", *e-Informatica Software Engineering Journal*, Vol. 17, No. 1, 2023, pp. 9-10.

[83] Mackenzie, H., Deway, A., Drahota, A., Kilburn, S., Kalra, P. R., Fogg, C., and Zachariah, D., "Systematic reviews: what they are, why they are important, and how to get involved", *Journal of Clinical Prevision Cardiology*, Vol. 1, No. 4, 2012, pp. 193-202.

[84] Nishikawa-Pacher, A., "Research Questions with PICO: A Universal Mnemonic". *Publications*, Vol. 10, No. 3, 2022, pp. 1-10.

[85] Aslam, S., and Emmanuel, P., "Formulating a researchable question: A critical step for facilitating good clinical research", *Indian Journal of Sexually Transmitted Diseases and AIDS,* Vol. 31, No. 1, 2010, pp. 47-50.

[86] Roever, L.," PICO: Model for Clinical Questions", *Evidence Based Medicine and Practice*, Vol. 3, No. 2, 2018, pp. 1-2.

[87] He, K., Zhang, X., Ren, S., and Sun, J., "Deep residual learning for image recognition", *Proceeding IEEE Conference Computer Vision and Pattern Recognition*, 2016, pp. 770-778. https://openaccess.thecvf.com/content_cvpr_2016/html/He_Deep_Residual_Learning_CVPR_2016_paper.html

[88] Trifu, A., Smîdu, E., Badea, D. O., Bulboacă, E., and Haralambie, V., "Applying the PRISMA method for obtaining systematic reviews of occupational safety issues in literature search", *MATEC Web of Conference*, Vol. 354, 2022, pp. 1-8.

[89] Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., . . . Moher, D., "The PRISMA 2020 statement: an updated guideline for reporting systematic reviews", *System Revision*, Vol. 10, No. 89, 2021, pp. 1-11.

[90] PRISMA Statement. Available online: https://www.prisma-statement.org/ (accessed on 22 July 2024)

[91] Burnham, J. F., "Scopus database: A review", *Biomedical Digital Libraries,* Vol. 3, No. 1, 2006, pp. 1-8.

[92] Baas, J., Schotten, M., Plume, A., Côté, G., and Karimi, R., "Scopus as a curated, high-quality bibliometric data source for academic research in quantitative science studies", *Quantitative Science Studies*, Vol. 1, No. 1, 2020, 377–386.

[93] Zhu, J., and Liu, W., "A tale of two databases: the use of Web of Science and Scopus in

academic papers", *Scientometrics*, Vol. 123, 2020, pp. 321–335.

[94] Ely, C., and Scott, I., "Essential study skills for nursing", *Mosby Elsevier*, 2007, pp. 66-72.

[95] Bashir, R., Surian, D., and Dunn, A. G., "Time-to-update of systematic reviews relative to the availability of new evidence", *Systematic Reviews,* Vol. 7, No. 195, 2018, pp. 1-8.

[96] Santini, A., "The importance of referencing", *The Journal of Critical Care Medicine*, Vol. 4, No. 1, 2018, pp. 3–4.

[97] Oxman, A. D., and Guyatt, G. H., "Validation of an index of the quality of review articles", *Journal Clinical Epidemiology,* Vol. 44, No. 11, 1991, pp. 1271–1278.

[98] Caliwag, A. C., Hwang, H., Kim, S., and Lim, W., "Movement-in-a-Video Detection Scheme for sign language gesture recognition using neural network", *Applied Sciences*, Vol. 12, No. 20, 2022, pp. 1-17.

[99] Chen, X., Su, L., Zhao, J., Qiu, K., Jiang, N., and Zhai, G., "Sign Language Gesture Recognition and Classification Based on Event Camera with Spiking Neural Networks", *Electronics*, Vol. 12, No. 4, 2023, pp. 1-14.

[100] Dong, Y., Liu, J., and Yan, W. Dynamic hand gesture recognition based on signals from specialized data glove and deep learning algorithms. *IEEE Transaction on Instrument and Measurement,* Vol. 70, 2021, pp. 1–14.

[101] Fakhfakh, S., and Jemaa, Y. B., "Deep learning shape trajectories for isolated word sign language recognition", The International Arab Journal of Information Technology, Vol. 19, No. 4, 2022, pp. 660-666.

[102] Ismail, M. H., Dawwd, S. A., and Ali, F. H., "Dynamic hand gesture recognition of Arabic sign language by using deep convolutional neural networks", *Indonesian Journal of Electrical Engineering and Computer Science,* Vol. 25, No. 2, 2022, pp. 952-962.

[103] Ji, A., Wang, Y., Miao, X., Fan, T., Ru, B., Liu, L., Nie, R., and Qiu, S., "Dataglove for Sign Language Recognition of People with Hearing and Speech Impairment via Wearable Inertial Sensors", *Sensors*, Vol. 23, No. 15, 2023, pp. 1-20.

[104] Lu, C., Kozakai, M., and Jing, L., "Sign Language Recognition with Multimodal Sensors and Deep Learning Methods", *Electronics,* Vol. 12, No. 23, 2023, pp. 1-15.

[105] Luqman, H., "An efficient Two-Stream network for isolated sign language recognition using accumulative video motion", *IEEE Access*, Vol. 10, 2022, pp. 93785–93798.

[106] Narayan, S., Mazumdar, A. P., and Vipparthi, S. K., "SBI-DHGR: Skeleton-based intelligent dynamic hand gestures recognition", *Expert Systems with Applications,* Vol. 232, 2023, pp. 161669- 161669.

[107] Saqib, S., Ditta, A., Khan, M. A., Kazmi, S. a. R., and Alquhayz, H., "Intelligent dynamic gesture recognition using CNN empowered by Edit Distance", *Computers, Materials and Continua,* Vol. 66, No. 2, 2020, pp. 2061–2076.

[108] Suneetha, M., Prasad, M. V. D., Venkata, V. K. P., and Anil, K. D., "Meta Triplet Learning for Multiview sign language recognition", *International Journal of Intelligent Engineering and Systems*, Vol. 16, No. 2, 2023, pp. 375–388.

[109] Venugopalan, A., and Reghunadhan, R., "Applying deep neural networks for the automatic recognition of sign language words: A communication aid to deaf agriculturists", *Expert Systems with Applications*, Vol. 185, 2021, pp. 1-9.

[110] Wu, R., Seo, S., Ma, L., Bae, J., and Kim, T., "Full-Fiber Auxetic-Interlaced yarn sensor for Sign-Language translation glove assisted by artificial neural network", *Nano-Micro Letters,* Vol. 14, No. 139, 2022, pp. 1-14.

[111] Pezzuoli, F., Corona, D., and Corradini, M. L. "Recognition and classification of dynamic hand gestures by a wearable Data-Glove", *SN Computer Science*, Vol. 2, No. 5, 2021, pp. 1-9.

[112] Huang, S., and Ye, Z., "Boundary-Adaptive encoder with attention method for Chinese sign language recognition", *IEEE Access*, Vol. 9, 2021, pp. 70948–70960.

[113] Padmanandam, K., M, R., V., Upadhyaya, A. N., K, R. C., B, C., and Sah, S., "Artificial intelligence biosensing system on hand gesture recognition for the hearing impaired", *International Journal of Operations Research and Information Systems*, Vol. 13, No. 2, 2022, pp. 1–13.

[114] Shanableh, T., "Two-Stage deep learning solution for continuous Arabic sign language recognition using word count prediction and motion images", *IEEE Access*, Vol. 11, 2023, pp. 126823–126833.