IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 9 August 2024, accepted 2 September 2024, date of publication 10 September 2024, date of current version 20 September 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3457692

RESEARCH ARTICLE

Advancements in Sign Language Recognition: A Comprehensive Review and Future Prospects

BASHAER A. AL ABDULLAH, GHADA A. AMOUDI, AND HANAN S. ALGHAMDI
Department of Information Systems, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah 21589, Saudi Arabia
Corresponding author: Bashaer A. Al Abdullah (bmuhammadalabdullah@stu.kau.edu.sa)

This Project was funded by the Deanship of Scientific Research (DSR) at King Abdulaziz University, Jeddah, under grant no. (GPIP-624-612-2024). The authors, therefore, acknowledge with thanks DSR for technical and financial support.

ABSTRACT Sign language (SL) is a vital mode of communication, bridging the gap between the hearing impaired and hearing communities. However, SL, despite its paramount importance, has received relatively limited attention from researchers. Its unique structural characteristics, distinct from those of natural languages, present novel challenges that require innovative solutions. Remarkable technological advancements, notably in Artificial Intelligence (AI) and machine learning, offer promising avenues for automated Sign Language translation Systems (SLTS). This review study addresses the crucial need for a comprehensive synthesis of existing research by systematically examining and evaluating the progress made in SLTS. By analyzing 58 research papers, with a particular emphasis on the most frequently cited papers from each year up to 2023, we shed light on the field's current state, identifying key advancements and challenges. This review followed a systematic approach based on clear guidelines. The methodology involved defining research questions, formulating queries, selecting studies based on clear criteria, and extracting pertinent information to address the research objectives. This review found that deep learning techniques, such as convolutional and recurrent neural networks, have shown high accuracy in sign language recognition, and their performance in recognizing the variety of signs has steadily improved over time. Additionally, integrating non-manual features has proven pivotal in enhancing recognition accuracy. Future research should refine advanced deep learning models and integrate non-manual features to improve system accuracy and applicability. These ongoing advancements hold the potential to revolutionize communication and break down barriers for individuals who rely on sign language as their primary mode of communication.

INDEX TERMS Computer vision-based recognition systems, deep learning, manual features, non-manual features, sensor-based recognition systems, sign language, sign language recognition systems.

I. INTRODUCTION

Sign language (SL) is an essential parallel to all audible languages. It is considered the only language that connects vocal people with the hearing impaired community, which globally numbers around 430 billion with total deafness, and 1.5 billion are partially hearing impaired, according to the World Health Organization (WHO) [1]. During the past few years, awareness of the importance of sign language has increased worldwide [2]. All the hearing impaired-related associations motivate the countries to celebrate this community and empower them socially [3].

The associate editor coordinating the review of this manuscript and approving it for publication was Sudhakar Radhakrishnan.

Undoubtedly, after the emergence of the computer and the possibility of processing natural languages, proposing an automated translation system for audible speech became most common over a range of generations [4]. However, sign language has received less attention from scholars since it is structurally different from natural languages and requires more sophisticated techniques [5]. Recently, Artificial Intelligence (AI) has slowly but steadily become a more significant part of our daily lives. The expanding usage of this technological revolution has exciting implications for the hearing impaired community [6]. Subsequently, developing SL translation systems has become a realistic and applicable idea for researchers [7]. Machine translation and image recognition have interested many researchers for several decades. The surveys indicate that initial research in this field was done in the 1940s [8]. Recognizing sign language is deeply related to the machine translation field; instead, it can be considered a sub-domain with many exciting challenges to solve [8].

There were massive worldwide contributions dealing with SL translation, intending to create a communication environment for the hearing impaired and hard of hearing (DHH). An enormous number of scholars have introduced SL translation systems for different sign languages based on multiple computer technologies across the years. This review article aims to summarise previous literature's main findings in the Automated Sign Language Translation Systems (SLTS) and discusses them in the light of the review questions.

This paper is structured into six key sections. The initial section serves as the introduction, whereas the second provides a brief background on sign language systems. The third section outlines the methodology and discusses the search strategy employed. In the fourth section, we delve into a comprehensive review of related works, categorizing them into three facets: alphabet-level translation, word-level translation, and sentence-level translation. Subsequently, the fifth section critically examines these associated works. Finally, the sixth section offers a concise conclusion summarizing the essential findings and insights.

II. BACKGROUND

Sign language is a system of communication that conveys meaning through gestures, facial expressions, and body language. Sign languages worldwide are primarily spatial-gestural; thus, they can not be compared to spoken languages [9]. Unlike spoken languages, sign languages do not have well-defined word order, grammatical rules, and sentence structure. Furthermore, it has no predefined formatting standards [9].

There are many ways in which a sign language can differ from another, including syntax, semantics, grammar, morphology, and phonology [8]. Syntax refers to the rules that control the arrangement of gestures to construct a phrase, where semantics determines the meaning of sentences. Grammar defines the structure of the language. Similarly, morphology accounts for the various ways to build a sentence [8]. Sign language differs based on languages and dialects [10].

There is no global sign language is used everywhere; instead, there are nearly 200 different sign languages, as each nation has its own sign language [11]. Table 1 includes the most common sign languages along with their acronyms [8].

For years, sign language recognition (SLR) has been an active topic in the sign language processing field [12]. It refers to detecting and recognizing signs performed by signers to provide readable text or audible speech; these signs could express letters, numbers, isolated words, or phrases. Moreover, the input types of the SLR systems could be either static or dynamic signs. The first helps express the language's alphabet and numbers by fingerspelling. In contrast, the second type is much more efficient for real-world conversation [13], as it could represent isolated words and sentences by a sequence of frames. SLR includes two main phases: Feature extraction and classification. Features could be manual features, which indicate the signer's hands' shape, motion, location, and orientation [14]. Otherwise, it could be non-manual features that represent the signer's emotions, such as facial expressions, mouth gestures, and body postures [13]. Features extraction is a critical phase in creating dynamic sign language recognition systems [15], as these systems require a complete base of knowledge about the characteristics of the signs to deliver the desired recognition performance. The characteristics of the dynamic signs require a sequential modeling technique for the extraction process. Each sign is constructed of a series of frames influenced by the temporal and spatial sequence. The second main phase in sign language recognition is classifying the extracted features. Wherein this process, the classifier takes the extracted features as inputs. It identifies them to a set of classes and assigns probabilities for these inputs to indicate their particular classes to identify and convey their meaning for the recognition process [16]. Fig. 1 shows the Arabic SLR high-level architecture.

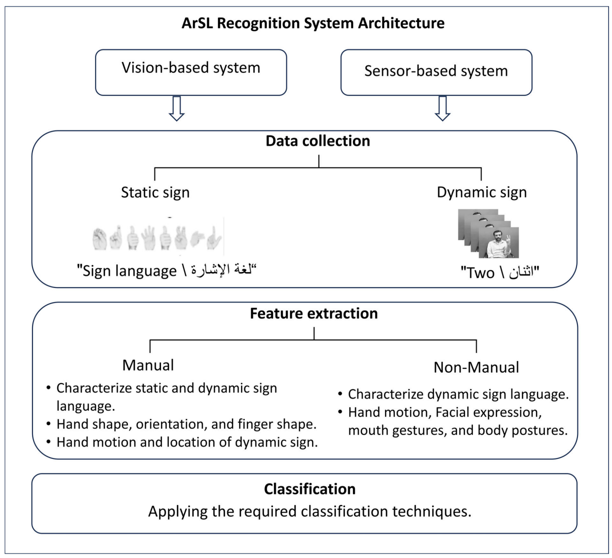

FIGURE 1. **Arabic SLR high-level architecture.**

In general, sign language recognition still lacks an appropriate system that can be practically applied in the real world [17]. Many challenges still face these systems and limit their capabilities, such as the extracted features of the signs, the complexity of the dynamic signs, and the translation of sign language sentences. Scholars have proposed various techniques to automate sign language recognition in recent decades. Most of these approaches focused on manual features [18]. However, few studies targeted the non-manual features alone or in conjunction with manual features. Dynamic signs are much more effective in daily-life

FIGURE 2. Example of similarity in dynamic arabic sign language.

communication [19]. However, some dynamic signs begin with a similar frame sequence until a specific time, which may confuse the classifier. For example, as shown in Figure 2, the words "brother" and "sister" in Arabic sign language begin with similar frame sequences. Fig. 2 shows an example of similarity in dynamic Arabic sign language.

III. METHODOLOGY

This review was carried out systematically, following Farooq et al. strategy [8], who conducted the review following Altman's [20] and Kitchenham's [21] guidelines. Whereby different research questions are defined to meet the research's objectives. A protocol is followed for literature selection to ensure that all relevant studies are included. This protocol includes: (a) Establishing the research questions to be addressed. (b)Defining appropriate queries and applying them to an appropriate platform. (c)Choosing appropriate studies based on well-defined inclusion and exclusion criteria. (d)Selecting, analyzing, and extracting useful information from selected studies in the light of the review questions.

A. RESEARCH QUESTIONS

An initial step in a systematic review is to define the research questions. Fig. 3 clarifies the questions the paper intends to answer.

The question	The objective
What deep learning techniques have been successfully applied to sign language?	Highlight the deep learning models that have a positive influence on the sign language recognition improvement.
How significant are non-manual features in the recognition of sign language?	To determine the non-manual features' level of significance in the overall process of understanding and interpreting sign language.
What are the future directions for sign language translation systems?	Discuss the main challenges and the required investigations to improve the solutions in this field.

FIGURE 3. Research questions.

B. SEARCH STRATEGY

The articles were searched and collected from the Semantic Scholar database, a free AI-powered research tool for scientific literature [22], using the search string below: (sign language recognition repository) OR (sign language translation) OR (gesture recognition using deep learning) AND (sign language recognition using deep learning) AND (computer vision). Fig. 4 shows the PRISMA diagram describing the article selection process.

The search went through 4 main phases. Firstly, define the related articles. In this phase, the database provided about 1090 results related to the search string. Then, the results were filtered based on the relatability to the computer science field, which narrowed the results to 997. After that, the results screened depending on five criteria: (1) choosing the studies most relevant to the field, (2) choosing the most influential studies, (3) focusing on recent studies, (4) including the fundamental studies, (5) and focusing extensively on the Arabic works. Consequently, the results narrowed to 400. Finally, 58 research papers were conducted, placing particular emphasis on the most extensively cited papers from each year, spanning up to 2023.

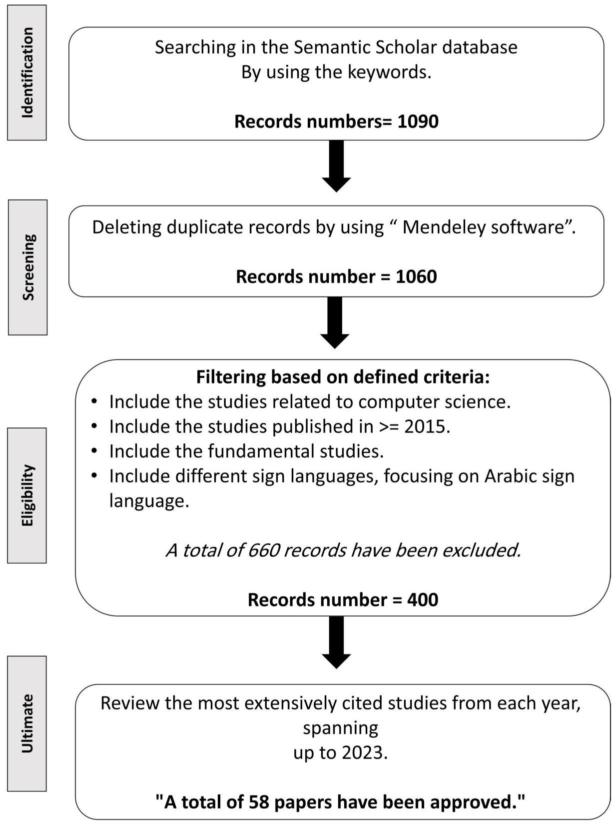

FIGURE 4. PRISMA diagram of the articles' selection process.

IV. RELATED WORKS REVIEW

The journey of recognizing sign language began with the first research paper published in the mid-90s, which was oriented toward recognizing American Sign Language (ASL) [8]. Since then, research has been confined to ASL for years until researchers from different poles of the world began applying the proposed solutions to their languages and dialects [23]. Thus, these systems became diverse and applicable to different dialects, especially after the emergence of the concept of Transfer Learning [23]. This sub-section reviewed fundamental research worldwide, primarily Arabic, focusing mainly on deep learning techniques. The studies are addressed in three recognition levels: alphabet, word, and sentence.

TABLE 1. Most common sign languages with their acronyms.

The language	Acronym
American Sign Language	ASL
Argentinian Sign Language	ArgSL
British Sign Language	BSL
Chinese Sign Language	CSL
German Sign Language	DGS
Irish Sign Language	IrSL
Malaysian Sign Language	MSL
New Zealand Sign Language	NzSL
Portuguese Sign Language	PorSL
Spanish Sign Language	LSE
Arabic Sign Language	ArSL
Australian Sign Language	AusLan
Brazilian Sign Language	LSB
Greek Sign Language	GSL
Indian Sign Language	ISL
Japanese Sign Language	JSL
Mexican Sign Language	MxSL
Pakistan Sign Language	PSL
Russian Sign Language	RSL
Turkish Sign Language	TSL
Arabic Sign Language	ArSL

A. RESEARCH IN ALPHABET SIGN LANGUAGE RECOGNITION WORLDWIDE

The sign language alphabet could be represented by either a static gesture or by drawing the shape of the letter with fingerspelling, which refers to writing letters and sometimes numbers only with the hands [24].

Artificial Neural Network (ANN) has recently been a hot topic in recognizing sign language gestures. Several authors applied it in their works to recognize signed alphabet and letters, such as [25], the earliest paper in this field, the authors aimed to present a posture and gesture recognition system to translate the 42 symbols of the Japanese sign language alphabet into text using a Recurrent Neural Network (RNN). The recognition rate improved by using a new encoding method, reaching 96%.

In the same context, [10] used ANN beside data gloves containing a Polhemus sensor to recognize American Sign Language (ASL) based on two stages: the first stage deals with recognizing the sign phonology (i.e., handshape, movement, position, and orientation). After that, the actual signs are recognized in the second stage. However, it was assessed based on a small lexicon. The results gave an overall accuracy of 86%.

Likewise, both [26] and [27] adopted the skin color segmentation technique, besides ANN, to represent the hand shape and region to translate the Thai sign language alphabet and the ASL. The authors in [26] depended on the time-delay neural network to represent 40 words in ASL. Skin color was employed for the hand region segmentation. The time delay neural network was then used to model the segmented hand motion trajectory to recognize the signed words. The experimental results have shown an accuracy of 99%, using all trajectories in the hand area. Where [27] Presented a method to translate the Thai alphabet, relying on the ANN classifier and the skin color segmentation, using the HOG to represent the segmented hand shape. The approach offered a rate of 84.05% recognizing Thai fingerspelling.

In pursuit of real-time sign language recognition, the study in [28] collected a dataset of American Sign Language (ASL) alphabet signs. A HandDetector module captured these images via a PC webcam. A convolutional neural network (CNN) with three convolutional layers and a SoftMax output layer was trained using the Adam optimizer and categorical cross-entropy loss function. Impressively, the system achieved outstanding results: training accuracy of 99.86%, validation accuracy of 99.94%, and test accuracy of 94.68%. These findings represent a significant leap forward in sign language recognition compared to previous studies.

On the other hand, many machine learning algorithms are well-applied for SL recognition. For example, [29] Discussed the applicability of a PCA-based model for detecting fingerspelling alphabets. The suggested model was tested on a vast and diverse real-time dataset, including many fingerspelling alphabet images captured from 20 persons in the real world. The authors used a pre-processing procedure during training and recognition to improve the performance of a PCA-based model. The extensive experiment shows that the proposed pre-processing significantly influences the performance of a PCA-based model.

In addition, [30] showed 96.23% accuracy in detecting six bespoke signs using a SIFT-based bag-of-features and an SVM classifier. However, these techniques seldom attain good accuracies when identifying ASL signals, which are complicated and have a lot of inter-person variabilities. Along, [31] used a Gabor Filter-based approach to detect 24 static ASL alphabet signs with only 75% average accuracy. Where [32] used a random forest regression model to estimate the 3D coordinates of the hand joints. The model is trained on a large dataset of hand poses and tested on several benchmark datasets to evaluate its accuracy. The evaluation results showed that the proposed system achieves state-of-the-art accuracy with an average error of 6.8mm for joints in 3D space. The proposed system can also highly accurately classify hand poses, achieving an average classification accuracy of 96.5% on the NYU hand pose dataset.

Likewise, [33] presented a static gesture recognition system of the sign language alphabet, depending on the magnetic positioning system with wearable transmitting nodes, to measure the hand's and fingers' three-dimensional position and orientation. After that, the collected data was processed and classified based on three different Support Vector Machine (SVM) models, trained in a vast sign language-alphabet image dataset that is publicly available. The system achieved an accuracy of about 97% on all English sign language alphabet letters.

Arab researchers also made many contributions to the field of ArSL. Some notable studies that interpret alphabet signs are summarised below. Reference [34] developed an automatic translation system for the manual gestures of the Arabic sign language alphabet; created a sequence of the Adaptive Neuro-Fuzzy Inference System (ANFIS) networks. The system could translate the Arabic alphabet with a recognition rate of 93.55%. Similarly, [35] proposed an adaptive neuro-fuzzy interference system for alphabet recognition using a colored glove for the hands' region segmentation. The system achieved an accuracy of 95.5%.

In the same context, [36] aimed to develop, implement, and test a non-visual-based smart glove to enhance performance accuracy and simplify implementation. Specifically, it utilized five flex sensors and an accelerometer to facilitate sign language recognition and its subsequent conversion into speech and text. It also employed well-known Machine Learning classifiers, including LR, SVM, MLP, and RF, to recognize American Sign Language (ASL) and Arabic Sign Language (ArSL). It ultimately achieved a classification accuracy of 99.7% for ASL and 99.8% for ArSL using the Random Forests (RF) classifier. Feature Importance analysis highlighted the dominance of accelerometer features in sign language recognition compared to flex sensor features. The research could compare the implementation and performance aspects of non-vision and vision-based sign language recognition methods as a potential next step.

Moreover, [37] developed an Arabic alphabet signs translator with an accuracy of up to 91.3%. The proposed systems extract the rotation, scale, and translation invariant from the input video, which makes the system more flexible. The authors depended on the Multilayer Perceptron (MLP) neural network and a Minimum Distance classifier (MDC) for recognition.

Furthermore, [38] Worked on recognizing the Arabic alphabet using a dataset of 1,500 photos for each class, where each class indicates a distinct meaning through its hand gestures or signs. The pre-processing and segmentation stages are done using different techniques. Furthermore, the experiments were performed using pre-trained models. The EfficientNetB4 model has been deemed the best match for the case, given that it obtained 98% training and 95% testing accuracy.

In addition, [39] worked on detecting, segmenting, and classifying the continuous sequence of gestures in real-time, with a 70% recognition rate, by suggesting a novel graph-matching approach and solving the difficulties of standard production of image identifiers and their standardization for picture and static alphabet recognition using Pulse Coupled Neural Networks (PCNN).

Furthermore, [40] developed a unique method for coping with posture changes in 3D object identification by employing PCNN to produce picture characteristics from two perspectives. They suggested a 90% identification rate approach that leverages two 2D picture characteristics to generate optimum 3D quality features. The authors then suggested a new strategy to improve feature quality by employing PCNN, Continuity Factor, which achieved a 90% recognition rate.

Moreover, [41] showed a machine learning-based alphabet recognition system for Arabic sign language utilizing a database of 2,800 images. To improve the image's value, the authors improved the color and used optimization and image enhancement algorithms, classification, and morphological straining to the color photograph in several ways; this allowed them to extract the best features afterward and achieve the most efficiency. The feature extraction process depended on a hand-shape-based description. Each hand picture is characterized by a vector of 15 values, where the values correspond to the positions of the critical points. The classification process was implemented using different methods, such as the KNN classifier, Naive Bayes, and MLB. The authors tested and had a 97.548% accuracy rate. Moreover, [42] developed a vision-based technique using CNNs to identify Arabic characters based on hand signs and convert them into Arabic speech with a recognition rate of 90%.

In addition, [43] proposes using the CNN technique to recognize 28 Arabic sign letters and numbers based on a 7-layer model, where the classification layer is attached to each network's last layer. Two approaches for categorizing dynamic gestures were proposed. Because dynamic gestures comprise a series of frames, the first technique categorized each frame independently using the SVM, KNN, and NN classifiers, with the result chosen by a simple majority. The DTW approach was employed in the second method. The second technique outperformed the first, with an accuracy of 97.45% for the palm characteristics set and 96.97% for the bone features set, respectively.

Furthermore, [44] Proposed a sensor-based system depending on a leap motion controller and latte panda. The recommended method is expanded to 30 hand motions, 20 of which are single-handed and 10 of which are double-handed. AdaBoost's accuracy results are 92.3% for single-hand gestures and 93% for double-hand motions. Besides, [11] developed a clever technique for recognizing words in ArSL that uses a Leap Motion device to build a 3D model of the human hand utilizing infrared and achieves classifying rates of 89% for one-hand moves and 96% for two-hand movements.

Additionally, [45] presented the first automated ArSLRS based on Hidden Markov models and proved how to detect ArSL utilizing a single video camera (HMMs). On ran data in offline mode, test data in offline mode, and test data in online mode, the device, in signer-dependent situations, achieves word recognition rates of 98.13%, 96.74%, and 93.8%, respectively. However, the gadget achieves a word recognition score of 90.6% in the offline mode and 94.2% in the signer-independent instances.

Moreover, [46] presented two unique techniques for feature extraction for the video-based ArSLR: motion representation through motion estimation and residuals. Comparisons with recent work show that up to 39% of misclassifications have been rectified.

B. RESEARCH IN WORD-LEVEL SIGN LANGUAGE RECOGNITION WORLDWIDE

Word-level translation refers to systems capable of recognizing simple words such as (''happy'' and ''teacher'') or complex words such as (''fireman'' and ''post-man'') and converting them into readable text or speech [47].

Several researchers introduced SVM-based systems to improve the accuracy of classification and recognition of signed words. Such as, [48] depended on the Hu moments technique, combined with the weight eigenspace functions to extract features from the camera data. The colored gloves were used to gather the hand blob's shape from the signs. This technique gathered two distinct datasets. The system had an accuracy of 0.973 and 0.935 when evaluated on the Irish Sign Language (ISL) and the Treisch datasets.

Also, [49] used the SVM beside the skin color model for hand segmentation. The authors employed RGB-D data and a Sparse Observation description to identify sign language from hand movements and postures. However, these sensor-based systems have drawbacks regarding user comfort and practicality. Additionally, these methods would become quite complex if the recognized item had a large sign language vocabulary.

The initial system [50] aimed to identify ten ASL words by gathering motion data from wrist sensors and wearable ring sensors on the index and middle fingers. Support Vector Machine and Discriminant analysis classifiers were employed, along with selected features. After training, the system achieved an 82% classification accuracy.

Moreover, both [51] and [52] introduced a gesture recognition methodology to translate the Indian sign language based on the ANN as a gesture classifier. Reference [51] developed an Artificial Neural Network-based method for the automated recognition of Indian sign language fingerspelling. The article suggested a way to distinguish between single-handed and double-handed signs. The authors photographed 720 Indian signs in total, including alphabets and digits. Skin color was used to partition the gathered signs in the YCbCr color space.

Likewise, [52] proposed a vision-based system distinguished from others using RGB images instead of leap motion or depth images. The system used the multi-class neural network classifier that provides ten different output classes. The system obtained a high accuracy rate compared with other research in Indian sign language, as high as 98

Many researchers Built their recognition systems depending on the Hidden Markov Models (HMMs) as a classifier, such as [53] and [54].

In [53], the ABC-based HMM with the Entropy-based K-Means algorithm was used to recognize and model Taiwan sign language words. The authors offered a recognition method for deciphering words in sign language used in the home services industry. The number of states in the HMM model is evaluated using an entropy diagram and an entropy-based K-means technique. In addition, the artificial bee colony technique and the Baum-Welch algorithm are combined in a data-driven method to discover the structure of HMM. Finally, 11 HMM models are used to create the recognition system, and cross-validation results show an average recognition rate of 91.3%.

While [54] employed HMM to model 100 isolated Chinese sign words for recognition, depending on partitioning the trajectory data into several segments. The curve feature descriptor is applied to represent each segment to be modeled after by the HMM. The system performed better with this approach than with the standard coordinate feature.

Likewise, [55] applied the Gaussian Mixture Model Hidden Markov Model (GMM-HMM) on a recognition system consisting of a camera and glove-based armbands. The GMM-HMM provided the system with a classification rate of up to 96% for ASL gestures.

Moreover, Neural Networks are applied by other researchers, such as [56]. The authors combined the Convolutional Neural Network (CNN) with Long short-term memory (LSTM), intending to act as a Recurrent Neural Network (RNN) architecture to recognize sign video frames of 40 Chinese sign language words. The system was able to achieve a recognition rate of 95%.

Also, [57] proposed a hand posture recognition system for ASL based on two stages: hand tracking and feature detection. In the first stage, a hybrid tracking algorithm was proposed that combined Skeletal and Kalman tracking methods to track and detect hand posture. The algorithm can handle hard-to-detect situations, such as when the hand is near the body, the face, or the other hand. After that, the deep neural network (DNN) is applied in the second stage to detect and learn the hand features from 3D images (i.e., hand rotation and movement). The experimental results achieved a recognition rate of about 98.12%.

Moreover, Jebali et al. proposed a computer vision system for recognizing 33 isolated signs in continuous sign language videos [58]. The system employed a novel algorithm for accurate word boundary detection. This algorithm leverages hand shape and motion features to isolate signs within the video. A Hidden Markov Model (HMM) has been utilized for sign recognition. The proposed framework achieves high accuracy (95.18% for single gestures and 93.87% for two-handed gestures). The performance was improved after including head pose and eye gaze features, with 2.24% and 2.9% for one- and two-hand gestures, respectively.

The field of sign language recognition has recently become a matter of interest for Arab researchers [8]. The researchers in this field are still few and limited, but it is on the way to flourishing.

Some Arabic scholars have utilized the neural network approaches to support the system's performance, such as [59]. The authors concentrated on NN's ability to assist with ArSL hand gesture recognition. This paper aimed to show how different types of NN may be used to identify human hand gestures in static and dynamic images. The authors demonstrated the use of Feed Forward Neural Networks (FFNN) and RNNs and their many designs, including entirely and partly recurrent networks. They studied their offered framework then, and the results showed that the proposed framework with the whole recurrent architecture had a static gesture detection execution accuracy of 95%.

Moreover, [60] Demonstrated a unique hand gesture detection approach for recognizing the ArSL alphabet and converting it to voice communication, allowing hearing impaired people to converse with hearing people. The current technique took color images of the hand sign. It converted them to the YCbCr color model, which gave a precise and effective way to extract skin patches from colored images after brightening alterations. The segmented hand sign edges were extracted using the Prewitt edge identifier. Using the PCA technique, the extricated edges were used to frame the specified feature vector for the motions library and signs. The Euclidean distance was used to calculate the similarity between the signs and feature vectors. The nearest sign was picked, and the corresponding sound portion was performed. The approach given here was used to recognize the ArSL alphabet and certain well-known Arabic signs. They used the approach to recognize over 150 signs and gestures in a practical test for three different signers, with an accuracy of over 97 percent. However, this research looked only at the current approach and its results.

Both [61] and [62] applied the digital sensor leap motion to extract the hands' features. In [62], the sensor addressed essential issues in vision-based systems, such as skin color and illumination. In the 3D model, Leap motion-captured finger and hand motions. Every action frame produced 3D data from the sensor. An MLP NN was used to deploy the spatial and temporal characteristics. The model was tested on 50 different dynamic signs, and the identification accuracy was 88 percent for two people. Even though Leap motion correctly follows two hands, it regrettably does not follow non-manual aspects. This model might be enhanced by adding more sensors to track non-manual aspects such as body shape and facial expressions. The suggested sensor could observe every feature of a sign while simultaneously working in leap motion.

Therefore, [61] proposed the system using image- and sensor-based approaches. The system discussed three main categories of recognition: Alphabet, Isolated Words, and Continuous Sign Language Recognition. The results have only been satisfactory for alphabet recognition, with accuracy passing 98%. Also, Amiri et al. demonstrated an ArSL hand gesture detection method that combined the Leap Motion Controller (LMC) with an SVM classifier [63]. For static hand movements, the system had a 91.3 percent accuracy rate.

Additionally, using a machine-learning approach, [64] developed a sign language-to-speech system to recognize and convert BISINDO's sign language into speech. With a dataset created in the study and Mediapipe for feature extraction, the model achieved an impressive accuracy of 98% using the Support Vector Machine method. However, during user trials, the model's accuracy dropped significantly to 78% as it exceeded the system's effective range. Despite this, the Sign Language-to-Speech implementation successfully produced audio speech output without requiring an internet connection. The system could detect dynamic and static user gestures in real-time.

On the other hand, the LMC had several system limitations, including a restricted detection range, the inability to recognize hands in the presence of impediments, and the capacity only to detect one hand at a time. Deriche et al. [65] presented a dual LMCs-based paradigm for ArSL identification. Hand occlusion and missing data issues were addressed using a front LMC and a side LMC. There were 100 dynamic indicators in the database that were utilized in daily talks. Two scenarios were evaluated, with 16 characteristics acquired from both LMCs. The first occurred when one LMC detected a gesture, and the other happened when both LMCs sensed a gesture. Only the data from the discovered LMC was provided to the classifier in the first case. The collected data from both LMCs were merged using the Dempster–Shafer

(DS) evidence-based fusion technique and then provided to the classifier in the second scenario. A Bayesian classifier based on the Gaussian mixture model GMM and linear discriminant analysis was employed for classification, with a 92%.

Also, sensor-based gloves were introduced by Assaleh et al. [66] to recognize Arabic sign language. In this system, accumulated differences (ADs) were used as the basis for extraction. The proposed system yielded 92.5% and 95.1% recognition rates based on user-dependent and user-independent models.

Hisham and Hamouda suggested an LMC-based model for ArSL recognition that can recognize static and dynamic gestures [67]. The model involved five phases: Pre-processing, tracking, feature extraction, classification, and sign recognition. Therefore, The four classifiers SVM, KNN, NN, and dynamic time wrapping (DTW) were employed using two feature sets: a palm features set and a bone feature set. Researchers utilized SVM, KNN, and NN to categorize static gestures, but the KNN classifier outperformed the others, with accuracies of 99% and 98% for the palm and bone data sets, respectively.

Furthermore, [68] introduced an Arabic sign language recognition system depending on fine-tuning deep convolutional neural networks (CNN) to improve the accuracy rate. The system applied 32 Arabic hand gestures in 2D images based on multiple models similar to the ResNet152 and VGG16 structures. The experimental result acquired an accuracy as high as 99%.

Additionally, [12] introduced a computer application designed for the machine translation of Iraqi sign language in two directions: from sign language to Arabic language (text/speech) and from Arabic language (text) to Iraqi sign language. The system utilized a Convolutional Neural Network (CNN) to classify sign language based on its features and predict the meanings of signs. Notably, the section translating sign language to Arabic text/speech achieved an impressive accuracy rate of 99.3% for letters.

In addition, [69] used the Kinect Sensor to create a Real-Time System for automatic ArSL recognition based on a Dynamic Time Warping coordination method for sign comparison without using power or data gloves in the system. Many tests were used to find custom-made signals from the Standard ArSL for a vocabulary of 30 separated terms. The framework could operate in three modes: online, signer-independent, and dependent. According to the test findings, the current model has a high detection value for each mode. The framework achieved a detection rate of 97.58 percent and a rate of error of 2.42 percent for signer-dependent. The model then achieved a detection rate of 95.25 percent and a rate of error of 4.75 percent for signer-independent.

Moreover, [70] presented an automatic visual recognition system for ArSL that translates Arabic word signs into Arabic text, depending on four primary stages: hand segmentation, hand tracking, hand feature extraction, and classification. The dynamic skin detector segments the hand based on the face color. Then, hand tracking was implemented using a proposed skin-blob tracking method with the aid of the head. The geometric features were also employed to extract the hand features, and the Euclidean distance classifier was applied as the recognition technique. According to the experimental results, the proposed system has a recognition rate of 97

C. RESEARCH IN SENTENCE-LEVEL SIGN LANGUAGE RECOGNITION WORLDWIDE

Sign language sentences are composed of gestures separated by movement epenthesis [70]. Sentence recognition using fingerspelling or isolated words is challenging and time-consuming. Studies in this aspect are few and recent, given that scholars still face problems translating continuous sign language at the word level; thus, a complete sentence is a more significant challenge [17].

In general, [71] introduced the first sentence-level sign language recognition system based on Radio-frequency Identification (RFID). The essential function of the system is to gather the phase sequence of signals obtained by RFID devices. The system has reasonably pure phase properties and can perform sign language segmentation and recognition using a new SOS feature. Extensive experiments have been carried out in various scenarios to assess the system's performance from many perspectives. The approach has a high recognition accuracy and resilience, with data indicating that the method's average accuracy is between 96 and 98.11 percent in various multipath circumstances.

Moreover, [72] proposed a method for identifying Indian sign language signs made by individuals with hearing impairments at the sentence level. The signs are recorded as a video. Each frame is processed to quickly extract sign information to model the sign and detect instances of new test signs—low-dimensional global "gist" descriptors record sign information from each video frame. K-means clustering is used to select a specific number of frames that are discriminatory enough to differentiate between signs. Furthermore, using a fixed number of frames allowed the authors to cope with an uneven number of frames between instances of the same sign caused by different signers, reducing the complexity of subsequent processing. In addition, the authors used symbolic data analysis to express a sign effectively. A fuzzy trapezoidal membership function determines the similarity between the test and a reference sign. The nearest neighbor classification algorithm is used to detect the supplied test sign. A comprehensive sign database (UoM-ISL) is developed, and significant testing is carried out on this database to investigate the effectiveness of the suggested technique. Experiments have yielded promising outcomes.

Moreover, the authors conducted the UoM-ISL dataset in a second attempt to recognize signs at the sentence level [73]. The paper explored the concept of Sign Energy Images (SEIs) and proposed a technique for extracting

Fuzzy-Gaussian Local Binary Patterns (FzGLBPs) from SEIs to identify signs. The study examined the effectiveness of interval-valued symbolic data for efficiently representing signs in a knowledge base. The matching between the reference and test indications is established using a Chi-square proximity metric. A straightforward nearest-neighbor classification method is employed to classify signs. Several tests are carried out to investigate the proposed system's effectiveness. With only 324 representatives, the suggested technique achieved a 79.13 percent recognition accuracy. The findings demonstrate unequivocally that the suggested feature extraction and representation strategies outperform alternative combinations.

The presented approach by [74] also used symbolic data analysis to deal with changes that resulted in the same sign being produced by various signers or by the same signer at different times for various practical reasons. A database of signs produced by communication-impaired individuals in the Mysore region is constructed, and comprehensive tests are performed on this database to illustrate the effectiveness of the proposed technique. Compared to standard crisp representation strategies, the suggested symbolic representation methodology effectively lowers the number of reference sign samples required to train the system. With 243 sign representations, the system achieved an average recognition rate of 92.73%.

In addition, [75] presented another method for detecting Indian sign language at the sentence level. The authors extracted the local contrast information from the video frames of the signs by utilizing the LBPV. For a variety of practical reasons, the use of symbolic data analysis is being investigated to handle differences among signs. Experiments were carried out using the UoM-ISL dataset with the assistance of hearing impaired persons from the Mysore region. The experimental results were more promising for the dataset under consideration. The suggested technique was compared to various classifier combinations and found to be more effective regarding recognition accuracy and storage requirements. The experimental findings revealed that the suggested technique performed well in terms of F-measure rates. However, using an extensive dataset, the proposed method's scalability must be investigated.

proposed an approach to learning sign language at the sentence level [76] by combining 3D-CNNs with bidirectional recurrent neural networks (Bi-RNNs). Even though 3D CNN networks are computationally intensive, they are large and bulky. To achieve a lightweight 3D CNN network, the 3D convolution was broken down into a "2+1D" convolution. The research relied on 30 sentences collected from the SIBI video dataset. The authors concluded that the depth of the network does not necessarily guarantee a successful outcome. In addition, the volume of datasets affects the performance of the system.

Furthermore, [17] introduced DeepASL. This transformational deep learning-based sign language translation solution allows ubiquitous and non-intrusive American sign language translation at both the word and sentence levels.

DeepASL captures ASL signals non-intrusively by using infrared light as a sensor technique. Infrared light is used as a sensor to capture ASL signs non-intrusively. The authors employed a newly developed hierarchical bidirectional deep recurrent neural network (HB-RNN) and a probabilistic framework based on Connectionist Temporal Classification (CTC) for word-level and sentence-level ASL translation. The authors collected 7,306 samples from 11 people to test the system's effectiveness, which included 56 regularly used ASL terms and 100 ASL sentences [17].

DeepASL system achieved an average of 94.5% on the word level. Moreover, it observed an average 8.2% word error rate while translating unseen ASL sentences. Furthermore, it obtained an average word error rate of 16.1% while translating ASL sentences produced by anonymous users across 100 often-used ASL sentences [17].

Assaleh et al. [19] made the first effort to recognize continuous Arabic sign language for a database of 40 phrases comprising 80 regularly used terms. The authors demonstrated a user-dependent continuous Arabic sign language recognition system. The dataset involved 19 repetitions of each of the 40 sentences accomplished by one person.

Also, this work has no grammatical or sentence length limits as it is based on limited vocabulary data. The authors published only word recognition and accuracy rates, with no sentence recognition rates for comparison. Spatiotemporal features were used with an HMM classifier based on the Discrete Cosine Transformation (DCT) as a feature extraction method. The number of features, states, and Gaussian mixtures was all tuned in the classifier. The adjusted settings resulted in a word recognition rate of 94% [19].

Furthermore, [77]. DG5-VHand data gloves were used to collect the same set of 40 sentences; each repeated ten times. Moreover, a camera coordinates hand gestures with matching sign language words during data labeling. Low-complexity pre-processing and feature extraction approaches to capture and accentuate the data's temporal dependency.

The classification is then done using a Modified k-Nearest Neighbor (MKNN) technique. For accurate classification, the proposed MKNN takes advantage of the context of feature vectors. The proposed approach obtained a 98.9% sentence recognition rate. The findings are compared to those of a previous vision-based method employing an identical text collection. The suggested method outperforms vision-based systems in classification rates while removing their limitations. Their technique, however, requires extensive training due to the human tagging of word boundaries [78].

Also, the same dataset was used besides a dataset collected using the Polhemus G4 tracker in [46] to propose a sentence-level sign language recognition system based on motion data. The study presented the system as a continuous sensor-based ArSLR that depends on Hidden Markov Models (HMM) and a modified k-nearest neighbor (KNN) version.

The system was tested on two datasets with 40 Arabic sentences with a perplexity of 80 words. The results showed that the system performs well, with a sentence recognition rate of 97 percent. Moreover, it showed that augmenting raw sensor data with statistical features enhanced the classification accuracy, and using a motion tracker alone leads to good accuracy and promising results. The study also showed that MKNN outperformed HMM in terms of sentence classification accuracy. It is almost the opposite for word recognition rate, where HMM outperformed MKNN [46].

Tables 2-5 summarize previous research on sign language recognition, categorized by the translation level achieved. Tables 2 and 3 focus on alphabet-level translation, with Table 2 specifically addressing studies that translate to Arabic and Table 3 encompassing studies for other languages. Table 4 presents research on word-level translation, while Table 5 summarizes studies on sentence-level translation.

Each table includes columns for ''The Paper'' (referencing the research work), ''Target Language,'' ''Signs No.'' (number of signs studied), ''Signer'' (specificity of the signer), ''Isolated / Continuous'' (presentation format of the signs), ''Input Data'' (modality of the sign language input), ''Manual'' (extraction of manual features), ''Non-Manual'' (extraction of non-manual features like facial expressions), ''Standard Dataset'' (use of a standardized dataset for evaluation), ''Model'' (the machine learning model employed), and ''Accuracy'' (the reported performance of the translation system).

V. DISCUSSION

This section summarizes the key advances in SLR systems and answers the three research questions.

A. KEY ADVANCEMENTS IN SIGN LANGUAGE RECOGNITION

SLR systems can be categorized into two broader types: hardware-based and software-based. Hardware-based approaches involve specific wearable hardware, such as Gloves, Kinect, or any other sensor-based device [35]. On the other hand, the software-based approaches depend on computer vision and image processing and use probabilistic models or machine/deep learning models for gesture recognition [35].

- Hardware-based systems advancement

Hardware-based sign language recognition has seen some advancements in improving sensors by embedding them in wearable devices, such as data gloves, watches, and bands [36]. Data gloves have been working extensively to capture hand movements, orientation, and location positively [77]. On the other hand, electromyography (EMG) sensors recognize signs by detecting electrical muscle movement while the signs are being signed [8]. Researchers have worked on miniaturizing these sensors over the years, helping to create more comfortable wearable devices [8]. However, wearing these devices for a long while could be uncomfortable. Also, the cost associated with manufacturing, maintaining, and developing these sensors is expensive. Fig 5 shows different examples of wearable devices, cyber-gloves [79], EMG sensors [80], and data-gloves [81].

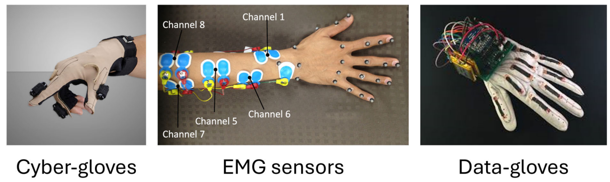

FIGURE 5. Examples of wearable devices.

- Software-based systems advancement

While hardware has made some strides, the field is currently dominated by vision-based systems utilizing cameras [8]. The researchers have been working on Improving software that enables real-time sign translation to spoken or text. They have also been developing software that could recognize signs regardless of the signers' identity, signing style, clothes, or backgrounds [18]. Moreover, they are working on providing systems able to incorporate other modalities, such as facial expressions, lip movements, and body language, that contribute to the meaning of sign language [18]. Deep learning has started a new era in improving these systems, opening new solutions and valuable techniques for this task, such as using different neural Networks spirited or combined with Transformer models [8]. Neural networks have shown competitive performance since they were adopted, as they can analyze hand queries beside other modalities, such as artificial expression in static or dynamic scenarios [18]. This technology is affected by the amount of data and the training time, like human brains. The more data provided and the more the neural networks are trained, the better the performance [82]. This is crucial for real-world applications. Fig 6 shows the positive correlation between the model performance and the amount of data in the context of deep learning and machine learning [82]. However, the limited dataset and the environmentally affected factors such as camera singles, image quality, background noise, and lighting still affect the production of the ideal practical system [18].

B. WHAT DEEP LEARNING TECHNIQUES HAVE BEEN SUCCESSFULLY APPLIED TO RECOGNIZE SIGN LANGUAGE?

Deep learning is an emerging machine-learning area focusing on learning data representations [11]. Deep algorithms learn from large quantities of data and process it at a deeper level, like human neurons. Therefore, deep learning could eliminate some pre-processing required while using machine learning. Moreover, it could increase the prediction accuracy for various applications [78]. Deep learning techniques have been used effectively in diverse applications, including computer vision, speech recognition, natural language processing,

TABLE 2. Arabic alphabet sing language recognition works summary.

The Paper	Target language	Signs No.	Signer	Isolated \Continuous	Input Data	Manual	Non-Manual	Standard dataset	Model	Accuracy
(Al-Jarrah and Halawani, 2001) [34]	Arabic sign language	30	1	Isolated	Vision-based system. (Gray-scale Images)	Hand location, shape, and orientation	-	-	Adaptive Neuro-Fuzzy Inference System (ANFIS) networks	93.55%
(Al-Rousan and Hussain, 2001) [35]	Arabic sign language	30	-	Isolated	Vision-based system. (RGB images)	Hand shape and orientation	-	-	Adaptive Neuro-Fuzzy Inference System (ANFIS) networks	95.5%
((Al-Rousan, Assaleh, and Tala'a, 2009) [45]	Arabic sign language	30	18	Isolated	Vision-based system. (RGB images)	Hand location, shape, and movement	-	-	Hidden Markov Models(HMM)	98.13%
(El-Bendary et al. 2010) [37]	Arabic sign language	15	-	Isolated	Vision-based system. (RGB images)	Hand location, shape, and orientation	-	-	Multilayer Perceptron (MLP) neural network and a Minimum Distance classifier (MDC)	91.3%
(Elons, Aboull-Ela, and Tolba, 2013) [40]	Arabic sign language	50	_	Isolated	Vision-based system. (RGB and depth images)	Hand shape	-	-	Pulse Coupled Neural Networks (PCNN) And Continuity Factor	90%
(Tolba, Samir, and Aboull-Ela, 2013) [39]	Arabic sign language	158 static postures and 50 dynamic gestures	-	Isolated \Continuous	Vision-based system. (RGB images)	Hand location, shape, and movement	-	-	Pulse Coupled Neural Networks (PCNN).	Up to 70%

TABLE 2. *(Continued.)* Arabic alphabet sing language recognition works summary.

(Tubaiz et al., 2015) [77]	Arabic sign language	80	1	Continuous	Sensor-based system. (Electronic glove and Mocap (Motion capture))	Hand location and shape	-	-	Modified k-Nearest Neighbor (MKNN) approach	98.9%
(Khelil et al., 2016) [63]	Arabic sign language	10	10		Vision-based system. (RGB images) Leap Motion Controller (LMC) and the Microsoft Kinect system used	Hand location and shape	-	-	Pattern recognition methods	91.3%
(Hayani et al., 2019) [43]	Arabic sign language	39	-	Isolated	Vision-based system. (RGB images)	Hand shape and orientation	-	-	Convolutional Neural Networks (CNN)	97.45 % for the palm characteristics set and 96.97 % for the bone features set
(Jesi et al., 2020) [11]	Arabic sign language	10	-	Continuous	Vision-based system. (RGB and depth images) Leap motion sensor used	Hand location, shape, and orientation	-	-	Weighted Bias Mean Convolutional Neural Networks (WBM-CNN)	89% for one-hand moves and 96% for two-hand movements

and other areas. Furthermore, the domain advancements have significant implications and applications for neural network-based sign language interpretation [11]. Many deep learning techniques have been applied successfully in recent sign language recognition experiments, as these approaches have been proven in pattern and speech recognition [47], including Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Hybrid CNN-RNN models, Long Short-Term Memory (LSTM) networks, and Transfer Learning.

TABLE 2. *(Continued.)* **Arabic alphabet sing language recognition works summary.**

(Hisham and Hamouda, 2021) [44]	Arabic sign language	20 single-hand gestures 10 double-hand gestures	3	Isolated	Vision-based system. (RGB images) Leap motion controller and latte panda used	Hand location, shape, and orientation	-	-	k-Nearest Neighbor (KNN) & Support Vector Machine (SVM) & an Ada-Boosting technique	92.3% for single-hand gestures and 93% for double-hand motions
(Kamruzzaman, 2020) [42]	Arabic sign language	31	1	Isolated	Vision-based system. (RGB images)	Hand location, shape, and orientation	-	-	Convolutional neural networks (CNN)	90%
(Tharwat, Ahmed, and Boual-legue, 2021) [41]	Arabic sign language	28	10	Isolated	Vision-based system. (RGB images)	Hand shape and orientation	-	-	KNN classifier, Naive Bayes, and MLB	97.55%
(Zakariah et al., 2022) [38]	Arabic sign language	32	40	Isolated	Vision-based system. (Grayscale Images)	Hand shape and orientation	-	ArSL2018	Different neural networks techniques	The best match is The Efficient-NetB4 with 95%
(Alosail et al., 2023) [36]	Arabic sign language / American sign language	29 letters of ArSL and 24 letters of ASL.	-	Isolated	Sensor-based system. (Smart glove using flex sensors and accelerometer)	Hand shape and orientation	-	-	well-known Machine Learning classifiers, including LR, SVM, MLP, and RF	99.7% for ASL and 99.8% for ArSL

CNNs have been extensively used for sign language recognition, as they can capture intricate local patterns in image data. These networks apply a series of convolutional operations to the input image, followed by pooling and non-linearity layers, to extract features that are then used for classification. CNN-based approaches have achieved state-of-the-art performance on several benchmark datasets [83].

RNNs, on the other hand, are used to model temporal dynamics in sign language recognition. They are well-suited for capturing the sequential nature of sign language gestures and have been shown to outperform other traditional machine learning methods. Long Short-Term Memory (LSTM) networks are RNNs capable of learning long-term dependencies. They have been applied to sign language recognition tasks with promising results [59]. LSTM transforms the input sign

TABLE 3. Worldwide alphabet sign language recognition works summary.

The Paper	Target language	Signs No.	Signer	Isolated \Continuous	Input Data	Manual	Non-Manual	Standard dataset	Model	Accuracy
(Murakami & Taguchi, 1991) [25]	Japanese sign language	42	1	Isolated	Sensor-based system. (Electronic glove)	Hand shape and orientation	-	-	Recurrent Neural Networks (RNN)	96%
(Waldron & Kim, 1995) [10]	American sign language	14	-	Isolated	Sensor-based system. (Electronic glove) Mocap (Motion capture) used	Hand location, movement, shape, and orientation	-	-	Artificial Neural Networks (ANN)	86%
(M.-H. Yang et al., 2002) [26]	American sign language	40	1	Isolated	Vision-based system. (RGB images)	Hand shape and movement	-	-	Time-delay neural network (TDNN)	99%
(Suraj and Guru, 2007) [29]	American sign language	24	20	Isolated	Vision-based system. (Gray-scale Images)	Hand location, shape, and orientation	-	-	Diagonal PCA-based model	90.17% with Histogram equalization
(Dardas and Georganas, 2011) [30]	American sign language	10	-	Isolated	Vision-based system. (RGB images)	Hand location, shape, and orientation	-	-	SVM classifier	96.23%
(Pugeault et al., 2011) [31]	American sign language	24	4	Isolated	Vision-based system. (RGB and depth images)	Hand shape	-	-	Random Forest Algorithm	75%
(Keskin et al., 2011) [32]	American Sign Language	10	10	Isolated	Vision-based system. (RGB and depth images)	Hand shape and orientation	-	-	Random forest regression model	96.5%
(Chansri & Srinonchat, 2016) [27]	Thai sign language	24	1	Isolated	Vision-based system. (RGB and depth images)	Hand location, shape, and orientation	-	-	Back-propagation of neural network	84.05%

TABLE 3. ***(Continued.)*** **Worldwide alphabet sign language recognition works summary.**

(Rinalduzzi et al., 2021) [33]	American sign language	24	-	Isolated	Sensor-based system. (Magnetic glove with transmitting nodes)	Hand location, shape, and orientation	-	-	Magnetic Positioning System (MPS) and Support Vector Machine (SVM)	97%
(Orovwode et al., 2023) [28]	American sign language	24	-	Isolated	Vision-based system. (RGB images)	Hand shape and orientation	-	-	Convolutional Neural Network (CNN)	99.94%

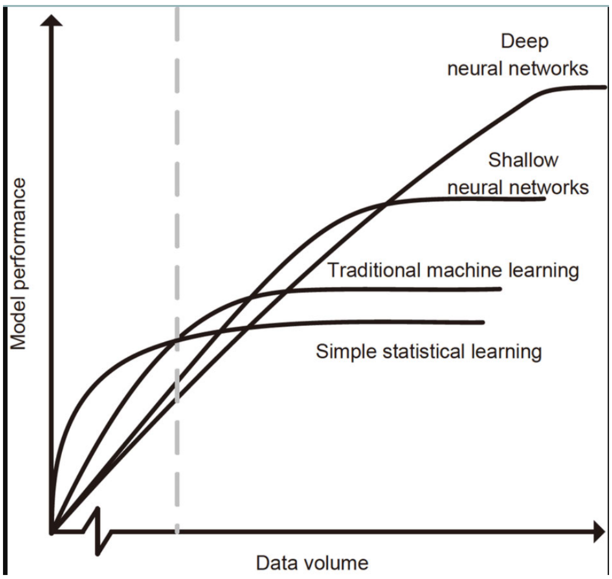

FIGURE 6. **The correlation between model performance and the amount of data.**

language into a time-series signal. Then, these networks process the signal to identify the sign language. LSTM networks are generally preferred over RNNs as they can maintain long-term memory [84].

CNN-RNN models have been successfully applied to recognize sign language. In general, CNNs are good at extracting spatial features from images or videos, while RNNs are better at capturing the temporal dependencies in a sequence of events. By combining these two types of networks, the hybrid CNN-RNN models can achieve better accuracy in recognizing sign language gestures [76]. These models first pass the input (a video of sign language gestures) through a CNN, which extracts spatial features such as the shape and position of the hand, fingers, and other relevant body parts. These features are then fed into an RNN, which captures the temporal dependencies in the sequence of gestures [76]. Some studies further incorporate attention mechanisms to improve these models' accuracy and interpretability. These mechanisms allow the model to selectively focus on certain input parts, making it easier to understand which features are most important for recognizing the sign language gesture [13].

Transfer learning is another technique that has been applied to sign language recognition. It involves taking a pre-trained model on a large dataset and fine-tuning it for the specific sign language recognition task. This approach has been shown to improve the recognition accuracy of models trained on smaller sign language datasets by leveraging the knowledge learned from the larger dataset [38].

In summary, CNNs, RNNs, and Transfer Learning are some of the successful deep-learning techniques used for sign language recognition. Each method provides unique advantages in capturing different aspects of sign language. CNNs have been used for feature extraction and classification, while RNNs have been used for sequence modeling and temporal dependencies. Hybrid CNN-RNN models combine the strengths of both CNNs and RNNs to improve performance. Transfer learning and data augmentation techniques have also been employed to improve the accuracy and robustness of sign language recognition models.

C. HOW SIGNIFICANT ARE NON-MANUAL FEATURES IN RECOGNIZING SIGN LANGUAGE?

Sign language recognition is a critical task that has been studied extensively in recent years. Sign language is a visual language that uses manual and non-manual features to convey meaning [18]. Manual features refer to hand and finger movements and shapes when signing. In contrast, non-manual features refer to body movements, facial expressions, and other actions that complement the manual signs. Non-manual features, such as manual features, are essential to convey grammatical information and add context to signs. Therefore, ignoring it can lead to inaccuracies in

TABLE 4. Word-level sing language recognition works summary.

The Paper	Target language	Signs No.	Signer	Isolated \Continuous	Input Data	Manual	Non-Manual	Standard dataset	Model	Accuracy
(Maraqa & Abu-Zaiter, 2008) [59]	Arabic sign language	30	2	Isolated	Sensor-based system. (Colored gloves)	Hand shape	-	-	Feedforward Neural Networks with back-propagation & RNNs	95%
(Hemayed & Hassanien, 2010) [60]	Arabic sign language	30	35	Isolated	Vision-based system. (RGB images)	Hand shape	-	-	The Euclidean distance and Principal Component Analysis algorithm	97%
(Kelly, Mcdonald, and Markham 2010) [48]	Irish sign language	23	16	Isolated	Sensor-based system. (Colored gloves)	Hand shape and orientation	-	-	Hu moments technique, combined with the weight eigenspace functions	97.3%
(K. T. Assaleh et al., 2012) [66]	Arabic sign language	100	-	Continuous	Sensor-based system. (CyberGlove)	Hand shape, Movement, and orientation	-	-	Accumulated Differences (ADs)	Yielded 92.5% and 95.1%
(Adithya et al., 2013) [51]	Indian sign language	26 letters and 10 numercal	-	Isolated	Vision-based system. (RGB images)	Hand shape	-	-	Digital image processing techniques and ANN	91.11%
(Mohandes et al., 2014) [61]	Arabic sign language	38	10	Isolated	Vision-based system. (Depth images)	Hand shape and orientation	-	-	Multilayer Perceptron (MLP) & Nave Bayes classifier	98%

TABLE 4. *(Continued.)* **Word-level sing language recognition works summary.**

(Elons et al., 2014) [62]	Arabic sign language	50	-	Continuous	Vision-baed system. (RGB images) The Microsoft Kinect sensor and Leap Motion used	Hand location and movement	-	-	Multilayer perceptron Neural Network (MLP)	Up to 88%
(Wang, Chai, and Chen, 2016) [49]	Chinese sign language	1000	1	Isolated	Vision-based system. (RGB and Depth images)	Hand shape and movement	-	-	SVM beside the skin color model	98.4%
(T. H. S. Li et al., 2016) [53]	Taiwan Sign Language	11	-	Isolated	Vision-based system. (RGB images)	Hand shape and location	-	-	Entropy-Based K-Means Algorithm and ABC-Based HMM	91.3%
(Junfu et al., 2016) [54]	Chinese sign language	100	-	Isolated	Vision-based system. (RGB images)	Hand shape	-	-	Trajectory Modeling with HMMs	82.7&
(Fatmi et al., 2017) [55]	American sign language	13	3	Isolated	Sendor-based system. (Wearable motion sensor (Myo armbands))	Hand location and orientation	-	-	HMM models	96%

recognizing the intended meaning of signed phrases [18]. For example, facial expressions distinguish between a question and a statement, and body movements indicate tense, aspect, or mood. Recognizing and incorporating non-manual features is essential for accurately understanding and translating signed languages. However, its influence on recognizing sign language's accuracy remains an open question. Therefore, recent articles attempt to demonstrate the effectiveness of incorporating these features by combining them with hand gestures. Accordingly, some researchers have been able to achieve higher accuracy rates in sign language recognition systems, such as:

In [15], the researchers examined the performance of a recognition system that included both manual and non-manual features compared to a system that only used manual features. They found that including non-manual features improved recognition accuracy by up to 10%. Moreover, the study revealed that certain non-manual

TABLE 4. *(Continued.)* Word-level sing language recognition works summary.

(Hisham & Hamouda, 2017) [67]	Arabic sign language	28 static letters, 10 numbers, and 28 dynamic words	2	Isolated \Continuous	Vision-based system. (RGB images) Leap motion sensor used.	Hand shape, orientation, and movement	-	-	Support Vector Machine (SVM), K-Nearest Neighbour (KNN), Artificial Neural Network (ANN), and Dynamic Time Wrapping (DTW)	99% and 98% for the palm and bone data sets
(S. Yang & Zhu, 2017) [56]	Chinese sign language	40	-	Continuous	Vision-based system. (Long Short-Term Memory (LSTM) network, RGB, and optical flow data)	Hand shape, orientation, and movement	-	-	Convolutional Neural Networks	95%
(J. Li et al., 2019) [13]	American sign language	500	50	Isolated	Vision-based system. (RGB and Depth images)	Hand location, movement, and shape	-	-	Deep Neural Networks (DNNs)	98.12%
(Abdel-Samie et al., 2018) [69]	Arabic sign language	30	1	Isolated	Vision-based system. (RGB images) Kinect sensor used	Hand location, movement, and shape	-	-	Translator based on Dynamic Time Warping	95.25%

features, such as eyebrow movement and mouth opening, were particularly informative for recognizing specific signs. For example, the researchers found that eyebrow movement was crucial for distinguishing signs that differ only in handshape, while mouth opening was critical for identifying signs that vary in location. These findings highlight the importance of incorporating non-manual features to produce more accurate and practical models for various applications, such as communication aids for people with hearing impairments.

Also, [85], the authors conducted experiments using a dataset of American Sign Language (ASL) videos. They compared the results of recognition systems that used only manual features with those that incorporated non-manual

TABLE 4. *(Continued.)* **Word-level sing language recognition works summary.**

(Ibrahim et al., 2016) [70]	Arabic sign language	30	-	Isolated	Vision-based system. (RGB images) Dynamic skin detector used	Hand location and movement	The head	-	Skin-blob tracking technique	97%
(Deriche et al., 2019) [65]	Arabic sign language	100	2	Isolated	Vision-based system. (RGB images) Dual Leap Motion controllers (LMC) used	Hand location and movement	-	-	Bayesian approach with a Gaussian mixture model (GMM) and a simple linear discriminant analysis (LDA) approach	92%
(Kulkarni & Chaitanya Badhe, 2020) [52]	Indian sign language	10	-	Isolated	Vision-based system. (RGB images)	Hand location and movement	-	-	Artificial Neural Networks (ANN)	98%
(Saleh & Issa, 2020) [68]	Arabic sign language	32	-	Isolated	Vision-based system. (RGB images)	Hand location and orientation	-	-	Transfer learning and fine-tuning deep Convolutional Neural Networks	99%
(Jemni et al., 2021) [58]	Different sign language	33	3	Isolated \Continuous	Vision-based system. (RGB images) Kinect sensor and Lap motion used	Hand shape and movement	Head pose and eye gaze	-	SqueezeNet and Hidden Markov Models(HMM)	95.18
(Raja'a et al., 2023) [12]	Arabic sign language	28	-	Isolated	Vision-based system. (Gray-scale images)	Hand shape and orientation	-	-	Convolutional Neural Network (CNN)	99.3%

TABLE 4. ***(Continued.)*** **Word-level sing language recognition works summary.**

(Fauzi et al., 2023) [64]	Indonesian Sign Language	72	-	Isolated	Vision-based system. (RGB images) Webcam used.	Hand shape	-	-	Support Vector Machine method (SVM)	78%
(Dibba & Min, 2023) [50]	American sign language	10	-	Isolated	Sensor-based system. (Wrist sensors and wearable ring sensors))	Hand location and movement	-	-	Support Vector Machine method (SVM)	82%

features. The results showed that recognition accuracy increased by up to 11.7% when non-manual features were used. Furthermore, the paper explored different feature extraction techniques to capture non-manual features such as optical flow, 3D facial landmarks, and head pose estimation. The results showed that combining these techniques led to an even more significant improvement in recognition accuracy, reaching up to 17.4% compared to models using only manual features.

In conclusion, the authors suggest that incorporating non-manual features into sign language recognition systems can significantly improve the accuracy and help overcome the challenges posed by the diversity and complexity of sign languages.

D. WHAT ARE THE FUTURE DIRECTIONS FOR SIGN LANGUAGE TRANSLATION SYSTEMS?

Since multiple obstacles are inherent in the SLR task, it is currently not viable to construct SLR tools with near-perfect accuracy [8], [13]. As a result, it is critical to keep creating new methodologies and evaluating their respective advantages to arrive at increasingly effective solutions. Several challenges may serve as excellent starting points for researchers interested in developing these systems. Among the most significant of these challenges are the following:

1) DATASET

• Lack of benchmark dataset: A significant portion of artificial intelligence system learning is based on datasets. There are no datasets for Arabic Sign Language at the sentence level. Furthermore, Arabic datasets at the word level are often considered limited compared to other languages. Similarly, corpora for gestures in sign language are limited and must be expanded by incorporating more and more hearing impaired people and sign language specialists [8].

• Quality of datasets: Until the present time, the Arabic sign language datasets are considered limited in terms of the number of signs, the number of individuals involved in producing the signs, and the number of samples from each sign. Moreover, the included vocabulary is confined to specific use and not varied to include aspects of essential communication words, such as computer science terminologies, medical, financial, and so on [8]. That could improve the communication between vocal and hearing impaired employees or customers in different work environments. Furthermore, each individual has a unique manner of making gestures, which may influence a person's ability in sign language, the speed with which motions are performed, the size of the hands, and other factors that make it challenging.

• Crowdsourcing could be a viable option. This method can aid in the collection of signs for gesture recognition with multiple signers, the more accessible collection of data for regional sign languages, the involvement of the hearing impaired community in the design of gestures for upcoming words, and the participation of the hearing impaired community in evaluating the acceptability of the developed translation and avatar-based system.

2) HARDWARE

• Sensor-based approaches have long been famous for collecting actions for various purposes. These devices were also used in sign language recognition to improve sign visualization. However, these techniques are not without their own set of difficulties. For example, there is a need to reduce hardware costs so that the public can benefit from hardware-based technology to identify gestures produced by a hearing impaired person. Therefore, most current research has focused on vision-based techniques that use images, video, and depth data to determine the semantic content of hand signs. It allows users to fully utilize their smartphones without needing an external device, significantly lowering costs [8].

3) CONTINUOUS SIGN LANGUAGE

Continuous sign language recognition is still a significant challenge, and even the most advanced automated systems

TABLE 5. Sentence-level sing language recognition works summary.

The Paper	Target language	Signs No.	Signer	Isolated \Continuous	Input Data	Manual	Non-Manual	Standard dataset	Model	Accuracy
(Assaleh et al., 2010) [19]	Arabic sign language	80	1	Continuous	Vision-based system. (RGB images)	Hand location	-	-	Spatiotemporal feature extraction and Hidden Markov Models(HMM)	94%
(Nagendraswamy, Kumara, and Chinmayi, 2015) [74]	Indian sign language	15	4	Isolated	Vision-based system. (RGB images)	Hand location	-	-	The nearest neighbor classification algorithm	88.75%
(Tubaiz et al., 2015) [77]	Arabic sign language	80	1	Continuous	Sensor-based system. (Electronic gloves) Mocap used	Hand location and movement	-	-	Modified k-Nearest Neighbor (MKNN) approach	98.9%
(Nagendraswamy, Kumara, Guru, et al. 2015) [72]	Indian sign language	17	4	Isolated	Vision-based system. (RGB images) Symbolic data analysis implemented	Hand location and movement	-	-	Symbolic data analysis-based approach	92.73.
(Kumara and Nagendraswamy, 2016) [73]	Indian sign language	26	4	Continuous	Vision-based system. (RGB images)	Fullframe	Fullframe	-	Fuzzy-Gaussian Local Binary Patterns (FzGLBPs)	79.13%
(Hassan et al., 2016) [46]	Arabic sign language	80	1	Continuous	Sensor-based system. (DG5-VHand data gloves)	Hand location	-	-	Hidden Markov Models (HMM) and a modified version of k-nearest neighbor (KNN)	97%

TABLE 5. *(Continued.)* Sentence-level sing language recognition works summary.

(Fang et al., 2017) [17]	American sign language	16	11	Continuous	Vision-based system. (Depth images) Leap Motion sensor used	Hand location, movement, and shape	-	-	Long-Short Term Memory, Hierarchical Bidirectional Recurrent Neural Network, and Connectionist Temporal Classification	Average 16.1% word error rate
(Nagendraswamy and Kumara, 2017) [75]	Indian sign language	26	4	Continuous	Vision-based system. (RGB images)	Fullframe	Fullframe	-	Utilizing the LBPV	84.81%
(Ariesta et al., 2018) [76]	Indonesian Sign Language	30	10	Isolated	Vision-based system. (RGB images)	Hand location, movement, and shape	9	8	3D Convolutional Neural Network and Bidirectional Recurrent Neural Network	97.5%
(Meng et al., 2019) [71]	Different sign languages	-	-	Continuous	Vision-based system. (RGB images) commercial RFID device used	Hand shape and movement	-	-	Random Forest (RF) classifier	Between 96% and 98.11%

struggle with the linguistic subtleties reflected in sign language phrases. This challenge is partly due to the limited vocabulary and the small number of phrases in most available datasets. In contrast, training models for sophisticated linguistic tasks require significantly more vast libraries with various examples [23]. This topic should continue to be investigated in multiple ways to find a configuration that can overcome the limitations that hinder the formation of highly effective solutions. Continuous SLR is more effective for real-world communication [23].

VI. LIMITATIONS

This study has potential limitations. The study's ability to comprehensively evaluate sign language recognition techniques may be impeded due to limited access to well-annotated datasets, particularly for less common sign languages or those with complex non-manual features. Employing restrictions in systematic reviews, such as language, publication date, or study type, could potentially impact the generalizability of the findings. For Future Research, we recommend including sign languages that have not been studied extensively and exploring a broader range of aspects of sign language development. It would also be relevant to discuss ethical considerations, such as privacy, accessibility, and cultural sensitivity, associated with developing and deploying sign language translation systems.

Advancements in Natural Language Processing (NLP) and Human-Computer Interaction (HCI) can help overcome limitations in sign language recognition. NLP techniques like transfer learning and data augmentation can improve model performance with limited datasets [86]. In addition, sentiment analysis can be used to recognize facial expressions in sign language videos. Also, anonymization techniques can protect user privacy. HCI offers user-centered design approaches and participatory methods, ensuring systems are developed with community input [86]. Additionally, HCI research on user interface design for people with disabilities can make sign language recognition systems accessible to diverse users [86]. Cultural sensitivity can be maintained by developing systems that respect the cultural nuances in sign languages.

VII. FUTURE WORKS

The future of sign language recognition holds immense potential that could contribute to breaking communication barriers for the hearing impaired community. We can expect advancements in developing deep learning models to overcome the existing challenges. Embedding non-manual features could open new opportunities for providing systems with intricate context and sentence structure. Moreover, we expect to provide more friendly, lightweight architecture software that can run on mobile, which could minimize the costs associated with hardware, increase the acceptability, and provide the users with more personalization options.

VIII. CONCLUSION

In conclusion, several deep learning techniques have been successfully applied to recognize sign language, including convolutional neural networks, recurrent neural networks, and hybrid models. These models have demonstrated high accuracy rates and have been used in various applications, such as interpreting sign language in real time and enhancing communication between hearing and hearing impaired communities. Non-manual features are essential for recognizing sign language. They provide valuable contextual information that improves recognition accuracy and helps disambiguate signs with similar manual features. Regional variations in sign language also rely heavily on non-manual features, making them critical for accurate recognition.

Generally, sign language translation systems will continue to evolve, driven by technological advancements and the growing demand for accurate and accessible communication. We believe that datasets are crucial elements for these systems. We suggest directing efforts toward generating a comprehensive dataset that includes major sign languages worldwide. This dataset could be collected using a crowdsourcing approach. Also, the focus should be on empowering the systems to recognize different modalities, such as lip and cheek movements. This would lead to more efficient and practical systems for real-world communication. Additionally, it creates exciting opportunities for further development of deep learning technology. The combination of these advancements could prove useful for various other tasks beyond sign language translation.

ACKNOWLEDGMENT

This Project was funded by the Deanship of Scientific Research (DSR) at King Abdulaziz University, Jeddah, under grant no. (GPIP-624-612-2024). The authors, therefore, acknowledge with thanks DSR for technical and financial support.

REFERENCES

[1] *World Health Organization (WHO) Deafness*. Accessed: Oct. 18, 2021. [Online]. Available: https://www.who.int/news-room/facts-in-pictures/detail/deafness

[2] The United Nations. *We Sign for Human Rights*. Accessed: Nov. 2, 2021. [Online]. Available: https://www.un.org/en/observances/sign-languages-day

[3] World Federation of the Deaf. *International Week of Deaf People and International Day of Sign Languages*. Accessed: Nov. 2, 2021. [Online]. Available: http://wfdeaf.org/iwdeaf2021/

[4] Translator Studio. *The History of Machine Translation*. Accessed: Nov. 2, 2021. [Online]. Available: https://translatorstudio.co.uk/machine-translation-history/

[5] J. M. Power, G. W. Grimm, and J.-M. List, "Evolutionary dynamics in the dispersal of sign languages," *Roy. Soc. Open Sci.*, vol. 7, no. 1, Jan. 2020, Art. no. 191100.

[6] J. Rowe, "Artificial intelligence and the future of sign language," LinkedIn, 2017. Accessed: Nov. 3, 2021. [Online]. Available: https://www.linkedin.com/pulse/artificial-intelligence-future-sign-language-james-rowe

[7] B. S. Parton, "Sign language recognition and translation: A multidisciplined approach from the field of artificial intelligence," *J. Deaf Stud. Deaf Educ.*, vol. 11, no. 1, pp. 94–101, Oct. 2005.

[8] U. Farooq, M. S. M. Rahim, N. Sabir, A. Hussain, and A. Abid, "Advances in machine translation for sign language: Approaches, limitations, and challenges," *Neural Comput. Appl.*, vol. 33, no. 21, pp. 14357–14399, Nov. 2021.

[9] M. A. Abdel-Fattah, "Arabic sign language: A perspective," *J. Deaf Stud. Deaf Educ.*, vol. 10, no. 2, pp. 212–221, Apr. 2005.

[10] M. B. Waldron and S. Kim, "Isolated ASL sign recognition system for deaf persons," *IEEE Trans. Rehabil. Eng.*, vol. 3, no. 3, pp. 261–271, Sep. 1995.

[11] V. E. Jesi, S. M. Aslam, and R. Fatima, "Deep learning based dynamic hand gesture recognition with leap motion controller," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 9, no. 5, pp. 7309–7315, 2020.

[12] M. Raja'a, M. Mohammed, and S. M. Kadhem, "Automatic translation from Iraqi sign language to Arabic text or speech using CNN," *Iraqi J. Comput., Commun., Control Syst. Eng.*, vol. 23, pp. 112–124, Jun. 2023.

[13] J. Huang, W. Zhou, H. Li, and W. Li, "Attention-based 3D-CNNs for large-vocabulary sign language recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 9, pp. 2822–2832, Sep. 2019.

[14] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andretti, and H. Adam, "MobileNets: Efficient convolutional neural networks for mobile vision applications," *CoRR*, vol. abs/1704.04861, Apr. 2017. [Online]. Available: https://arxiv.org/abs/1704.04861

[15] M. Mukushev, A. Sabyrov, A. Imashev, K. Koishybay, V. Kimmelman, and A. Sandygulova, "Evaluation of manual and non-manual components for sign language recognition," in *Proc. Int. Conf. Lang. Resour. Eval.*, 2020, pp. 6073–6078.
[16] K. Wong, R. Dornberger, and T. Hanne, "An analysis of weight initialization methods in connection with different activation functions for feedforward neural networks," *Evol. Intell.*, vol. 17, no. 3, pp. 2081–2089, Jun. 2024.
[17] B. Fang, J. Co, and M. Zhang, "DeepASL: Enabling ubiquitous and non-intrusive word and sentence-level sign language translation," in *Proc. 15th ACM Conf. Embedded Netw. Sensor Syst.*, Nov. 2017, pp. 1–13.
[18] H. Luqman and E.-S.-M. El-Alfy, "Towards hybrid multimodal manual and non-manual Arabic sign language recognition: MArSL database and pilot study," *Electronics*, vol. 10, no. 14, p. 1739, Jul. 2021.
[19] K. Assaleh, T. Shanableh, M. Fanaswala, F. Amin, and H. Bajaj, "Continuous Arabic sign language recognition in user dependent mode," *J. Intell. Learn. Syst. Appl.*, vol. 2, no. 1, pp. 19–27, 2010.
[20] A. Liberati, D. G. Altman, J. M. Tetzlaff, C. D. Mulrow, P. C. Gøtzsche, J. P. A. Ioannidis, M. Clarke, P. J. Devereaux, J. Kleijnen, and D. Moher, "The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration," *PLoS Med.*, vol. 6, no. 7, 2009, Art. no. e1000100.
[21] P. Brereton, B. A. Kitchenham, D. Budgen, M. Turner, and M. Khalil, "Lessons from applying the systematic literature review process within the software engineering domain," *J. Syst. Softw.*, vol. 80, no. 4, pp. 571–583, Apr. 2007.
[22] *Semantic Scholar | AI-Powered Research Tool*. Accessed: Nov. 8, 2022. [Online]. Available: https://www.semanticscholar.org/
[23] N. B. Ibrahim, H. H. Zayed, and M. M. Selim, "Advances, challenges and opportunities in continuous sign language recognition," *J. Eng. Appl. Sci.*, vol. 15, no. 5, pp. 1205–1227, Dec. 2019.
[24] A. Puente, J. M. Alvarado, and V. Herrera, "Fingerspelling and sign language as alternative codes for reading and writing words for Chilean deaf signers," *Amer. Ann. Deaf*, vol. 151, no. 3, pp. 299–310, Jun. 2006.
[25] K. Murakami and H. Taguchi, "Gesture recognition using recurrent neural networks," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, 1991, pp. 237–242.
[26] M.-H. Yang, N. Ahuja, and M. Tabb, "Extraction of 2D motion trajectories and its application to hand gesture recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 8, pp. 1061–1074, Aug. 2002.
[27] C. Chansri and J. Srinonchat, "Hand gesture recognition for Thai sign language in complex background using fusion of depth and color video," *Proc. Comput. Sci.*, vol. 86, pp. 257–260, Jan. 2016.
[28] H. Orovwode, I. D. Oduntan, and J. Abubakar, "Development of a sign language recognition system using machine learning," in *Proc. Int. Conf. Artif. Intell., Big Data, Comput. Data Commun. Syst. (icABCD)*, vol. 2, Aug. 2023, pp. 1–8.
[29] M. G. Suraj and D. S. Guru, "Appearance-based recognition methodology for recognizing fingerspelling alphabets," in *Proc. Int. Joint Conf. Artif. Intell.*, 2007, pp. 605–610.
[30] N. H. Dardas and N. D. Georganas, "Real-time hand gesture detection and recognition using bag-of-features and support vector machine techniques," *IEEE Trans. Instrum. Meas.*, vol. 60, no. 11, pp. 3592–3607, Nov. 2011.
[31] N. Pugeault and R. Bowden, "Spelling it out: Real-time ASL fingerspelling recognition," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCV Workshops)*, Nov. 2011, pp. 1114–1119.
[32] C. Keskin, F. Kiraç, Y. E. Kara, and L. Akarun, "Real time hand pose estimation using depth sensors," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCV Workshops)*, Nov. 2011, pp. 1228–1234.
[33] M. Rinalduzzi, A. De Angelis, F. Santoni, E. Buchicchio, A. Moschitta, P. Carbone, P. Bellitti, and M. Serpelloni, "Gesture recognition of sign language alphabet using a magnetic positioning system," *Appl. Sci.*, vol. 11, no. 12, p. 5594, Jun. 2021.
[34] O. Al-Jarrah and A. Halawani, "Recognition of gestures in Arabic sign language using neuro-fuzzy systems," *Artif. Intell.*, vol. 133, nos. 1–2, pp. 117–138, Dec. 2001.
[35] M. A. Al-Rousan and M. Hussain, "Automatic recognition of Arabic sign language finger spelling," *Int. J. Comput. Appl.*, vol. 8, pp. 80–88, Jan. 2001.
[36] D. Alosail, H. Aldolah, L. Alabdulwahab, A. Bashar, and M. Khan, "Smart glove for bi-lingual sign language recognition using machine learning," in *Proc. Int. Conf. Intell. Data Commun. Technol. Internet Things (IDCIoT)*, Jan. 2023, pp. 409–415.
[37] N. El-Bendary, H. M. Zawbaa, M. S. Daoud, A. E. Hassanien, and K. Nakamatsu, "ArSLAT: Arabic sign language alphabets translator," in *Proc. Int. Conf. Comput. Inf. Syst. Ind. Manage. Appl. (CISIM)*, Oct. 2010, pp. 590–595.
[38] M. Zakariah, Y. A. Alotaibi, D. Koundal, Y. Guo, and M. M. Elahi, "Sign language recognition for Arabic alphabets using transfer learning technique," *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–15, Apr. 2022.
[39] M. F. Tolba, A. Samir, and M. Aboul-Ela, "Arabic sign language continuous sentences recognition using PCNN and graph matching," *Neural Comput. Appl.*, vol. 23, nos. 3–4, pp. 999–1010, Sep. 2013.
[40] A. S. Elons, M. Abull-Ela, and M. F. Tolba, "A proposed PCNN features quality optimization technique for pose-invariant 3D Arabic sign language recognition," *Appl. Soft Comput.*, vol. 13, no. 4, pp. 1646–1660, Apr. 2013.
[41] G. Tharwat, A. M. Ahmed, and B. Bouallegue, "Arabic sign language recognition system for alphabets using machine learning techniques," *J. Electr. Comput. Eng.*, vol. 2021, pp. 2995851:1–2995851:17, Oct. 2021.
[42] M. M. Kamruzzaman, "Arabic sign language recognition and generating Arabic speech using convolutional neural network," *Wireless Commun. Mobile Comput.*, vol. 2020, pp. 3685614:1–3685614:9, May 2020.
[43] S. Hayani, M. Benaddy, O. El Meslouhi, and M. Kardouchi, "Arab sign language recognition with convolutional neural networks," in *Proc. Int. Conf. Comput. Sci. Renew. Energies (ICCSRE)*, Jul. 2019, pp. 1–4.
[44] B. Hisham and A. Hamouda, "Arabic sign language recognition using ada-boosting based on a leap motion controller," *Int. J. Inf. Technol.*, vol. 13, no. 3, pp. 1221–1234, Jun. 2021.
[45] M. Al-Rousan, K. Assaleh, and A. Tala'a, "Video-based signer-independent Arabic sign language recognition using hidden Markov models," *Appl. Soft Comput.*, vol. 9, no. 3, pp. 990–999, Jun. 2009.
[46] M. Hassan, K. Assaleh, and T. Shanableh, "User-dependent sign language recognition using motion detection," in *Proc. Int. Conf. Comput. Sci. Comput. Intell. (CSCI)*, Dec. 2016, pp. 852–856.
[47] I. Papastratis, C. Chatzikonstantinou, D. Konstantinidis, K. Dimitropoulos, and P. Daras, "Artificial intelligence technologies for sign language," *Sensors*, vol. 21, no. 17, p. 5843, Aug. 2021.
[48] D. Kelly, J. McDonald, and C. Markham, "A person independent system for recognition of hand postures used in sign language," *Pattern Recognit. Lett.*, vol. 31, no. 11, pp. 1359–1368, Aug. 2010.
[49] H. Wang, X. Chai, and X. Chen, "Sparse observation (SO) alignment for sign language recognition," *Neurocomputing*, vol. 175, pp. 674–685, Jan. 2016.
[50] M. Dibba and C.-H. Min, "American sign language recognition system using wearable sensors and machine learning," in *Proc. 21st IEEE Interregional NEWCAS Conf. (NEWCAS)*, Jun. 2023, pp. 1–5.
[51] V. Adithya, P. R. Vinod, and U. Gopalakrishnan, "Artificial neural network based method for Indian sign language recognition," in *Proc. IEEE Conf. Inf. Commun. Technol.*, Apr. 2013, pp. 1080–1085.
[52] P. C. Badhe and V. Kulkarni, "Artificial neural network based Indian sign language recognition using hand crafted features," in *Proc. 11th Int. Conf. Comput., Commun. Netw. Technol. (ICCCNT)*, Jul. 2020, pp. 1–6.
[53] T. S. Li, M.-C. Kao, and P.-H. Kuo, "Recognition system for home-service-related sign language using entropy-based K-means algorithm and ABC-based HMM," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 46, no. 1, pp. 150–162, Jan. 2016.
[54] J. Pu, W.-G. Zhou, J. Zhang, and H. Li, "Sign language recognition based on trajectory modeling with HMMs," in *Proc. Conf. Multimedia Modeling*, 2016, pp. 686–697.
[55] R. Fatmi, S. Rashad, R. Integlia, and G. Hutchison, "American sign language recognition using hidden Markov models and wearable motion sensors," *Trans. Mach. Learn. Data Mining*, vol. 10, no. 2, pp. 41–55, 2017.
[56] S. Yang and Q. Zhu, "Continuous Chinese sign language recognition with CNN-LSTM," *Proc. SPIE*, vol. 10420, pp. 83–89, Jul. 2017, Art. no. 104200F.
[57] J. Li, Z.-L. Wang, H. Zhao, R. Gravina, G. Fortino, Y. Jiang, and K. Tang, "Networked human motion capture system based on quaternion navigation," in *Proc. 11th EAI Int. Conf. Body Area Netw.*, vol. 5, 2016, pp. 38–44.
[58] M. Jebali, A. Dakhli, and M. Jemni, "Vision-based continuous sign language recognition using multimodal sensor fusion," *Evolving Syst.*, vol. 12, no. 4, pp. 1031–1044, Dec. 2021.
[59] M. Maraqa and R. Abu-Zaiter, "Recognition of Arabic sign language (ArSL) using recurrent neural networks," in *Proc. 1st Int. Conf. Appl. Digit. Inf. Web Technol. (ICADIWT)*, Aug. 2008, pp. 478–481.

[60] E. E. Hemayed and A. S. Hassanien, "Edge-based recognizer for Arabic sign language alphabet (ArS2V-Arabic sign to voice)," in *Proc. Int. Comput. Eng. Conf. (ICENCO)*, Dec. 2010, pp. 121–127.

[61] M. Mohandes, S. Aliyu, and M. Deriche, "Arabic sign language recognition using the leap motion controller," in *Proc. IEEE 23rd Int. Symp. Ind. Electron. (ISIE)*, Jun. 2014, pp. 960–965.

[62] A. S. Elons, M. Ahmed, H. Shedid, and M. F. Tolba, "Arabic sign language recognition using leap motion sensor," in *Proc. 9th Int. Conf. Comput. Eng. Syst. (ICCES)*, Dec. 2014, pp. 368–373.

[63] B. Khelil and H. Amiri, "Hand gesture recognition using leap motion controller for recognition of Arabic sign language," in *Proc. 3rd Int. Conf. ACECS*, 2016, pp. 233–238.

[64] M. Z. Fauzi, R. Sarno, and S. C. Hidayati, "Recognition of real-time BISINDO sign language-to-speech using machine learning methods," in *Proc. Int. Conf. Comput. Sci., Inf. Technol. Eng. (ICCoSITE)*, Feb. 2023, pp. 986–991.

[65] M. Deriche, S. O. Aliyu, and M. Mohandes, "An intelligent Arabic sign language recognition system using a pair of LMCs with GMM based classification," *IEEE Sensors J.*, vol. 19, no. 18, pp. 8067–8078, Sep. 2019.

[66] K. T. Assaleh, T. Shanableh, and M. M. Zourob, "Low complexity classification system for glove-based Arabic sign language recognition," in *Proc. Int. Conf. Neural Inf. Process.*, 2012, pp. 262–268.

[67] B. Hisham and A. Hamouda, "Arabic static and dynamic gestures recognition using leap motion," *J. Comput. Sci.*, vol. 13, no. 8, pp. 337–354, Aug. 2017.

[68] Y. Saleh and G. F. Issa, "Arabic sign language recognition through deep neural networks fine-tuning," *Int. J. Online Biomed. Eng.*, vol. 16, no. 5, pp. 71–83, May 2020.

[69] A.-G.-A.-R. Abdel-Samie, F. A. Elmisery, A. M. Brisha, and A. H. Khalil, "Arabic sign language recognition using Kinect sensor," *Res. J. Appl. Sci., Eng. Technol.*, vol. 15, no. 2, pp. 57–67, Feb. 2018.

[70] N. B. Ibrahim, M. M. Selim, and H. H. Zayed, "An automatic Arabic sign language recognition system (ArSLRS)," *J. King Saud Univ., Comput. Inf. Sci.*, vol. 30, no. 4, pp. 470–477, Oct. 2018.

[71] X. Meng, L. Feng, X. Yin, H. Zhou, C. Sheng, C. Wang, A. Du, and L. Xu, "Sentence-level sign language recognition using RF signals," in *Proc. 6th Int. Conf. Behav., Econ. Socio-Cultural Comput. (BESC)*, Oct. 2019, pp. 1–6.

[72] H. S. Nagendraswamy, B. M. C. Kumara, and R. L. Chinmayi, "GIST descriptors for sign language recognition: An approach based on symbolic representation," in *Proc. Int. Conf. Mining Intell. Knowl. Explor.*, 2015, pp. 103–114.

[73] C. Kumara and H. S. Nagendraswamy, "Sign energy images for recognition of sign language at sentence level," *Int. J. Comput. Appl.*, vol. 139, no. 2, pp. 44–51, Apr. 2016.

[74] H. S. Nagendraswamy, B. M. C. Kumara, D. S. Guru, and Y. G. Naresh, "Symbolic representation of sign language at sentence level," *Int. J. Image, Graph. Signal Process.*, vol. 7, no. 9, pp. 49–60, Aug. 2015.

[75] H. S. Nagendraswamy and B. M. C. Kumara, "LBPV for recognition of sign language at sentence level: An approach based on symbolic representation," *J. Intell. Syst.*, vol. 26, no. 2, pp. 371–385, Apr. 2017.

[76] M. C. Ariesta, F. Wiryana, S. Suharjito, and A. Zahra, "Sentence level Indonesian sign language recognition using 3D convolutional neural network and bidirectional recurrent neural network," in *Proc. Indonesian Assoc. Pattern Recognit. Int. Conf. (INAPR)*, Sep. 2018, pp. 16–22.

[77] N. Tubaiz, T. Shanableh, and K. Assaleh, "Glove-based continuous Arabic sign language recognition in user-dependent mode," *IEEE Trans. Human-Mach. Syst.*, vol. 45, no. 4, pp. 526–533, Aug. 2015.

[78] P. V. V. Kishore, M. V. D. Prasad, D. A. Kumar, and A. S. C. S. Sastry, "Optical flow hand tracking and active contour hand shape features for continuous sign language recognition with artificial neural networks," in *Proc. IEEE 6th Int. Conf. Adv. Comput. (IACC)*, Feb. 2016, pp. 346–351.

[79] *CyberGlove Systems*. Accessed: Jun. 20, 2024. [Online]. Available: https://www.cyberglovesystems.com/

[80] J. G. Ngeo, T. Tamei, and T. Shibata, "Continuous and simultaneous estimation of finger kinematics using inputs from an EMG-to-muscle activation model," *J. Neuroeng. Rehabil.*, vol. 11, no. 1, pp. 1–14, Dec. 2014.

[81] E. J. Holden and R. Owens, "Visual sign language recognition," in *Multi-Image Analysis* (Lecture Notes in Computer Science), vol. 2032, R. Klette, G. Gimel'farb, and T. Huang, Eds. Berlin, Germany: Springer, 2001, doi: 10.1007/3-540-45134-X_20.

[82] Q. Guo, S. Jin, M. Li, Q. Yang, K. Xu, Y. Ju, J. Zhang, J. Xuan, J. Liu, Y. Su, Q. Xu, and Y. Liu, "Application of deep learning in ecological resource research: Theories, methods, and challenges," *Sci. China Earth Sci.*, vol. 63, no. 10, pp. 1457–1474, Oct. 2020.

[83] A. A. Barbhuiya, R. K. Karsh, and R. Jain, "CNN based feature extraction and classification for sign language," *Multimedia Tools Appl.*, vol. 80, no. 2, pp. 3051–3069, Jan. 2021.

[84] C. Olah. *Understanding LSTM Networks*. Colah Blog. Accessed: Oct. 18, 2020. [Online]. Available: http://colah.github.io/posts/2015-08-Understanding-LSTMs/

[85] U. Canzler and T. Dziurzyk, "Extraction of non-manual features for video-based sign language recognition," in *Proc. IAPR Int. Workshop Mach. Vis. Appl.*, 2002, pp. 318–321.

[86] *Special Issue: Advances in Computer Vision and Machine Learning for Human-Centric AI*. Accessed: Jun. 20, 2024. [Online]. Available: https://onlinelibrary.wiley.com/doi/toc/10.1155/8483.si.612913

BASHAER A. AL ABDULLAH received the B.S. degree in computer science from King Khalid University, in 2019. He is currently pursuing the master's degree with the Department of Computer Information Systems, King Abdulaziz University, Jeddah.

GHADA A. AMOUDI received the Ph.D. degree in computer science from Dalhousie University, Canada, in 2016. She is currently an Associate Professor with the Faculty of Computing and Information Technology, King Abdulaziz University. She has been involved in academic activities, since 2005, and worked as an Educator, a Lecturer, and a Researcher. She has co-authored an article in the imaging deep learning domain titled deep learning approaches for detecting COVID-19 from chest X-ray images: A survey and another recent one in the deep learning NLP domain titled "Arabic Rumor Detection: A Comparative Study." Her research interests include applying AI to Arabic text mining, health domains, and social media. She is the Co-Founder and a member of the Director Board of the Saudi Artificial Intelligent Association.

HANAN S. ALGHAMDI received the Ph.D. degree in computer science from the University of Surrey, Guildford, U.K., in 2018. Her current research interests include medical image analysis, explainable artificial intelligence, deep learning, reinforcement learning, generative adversarial networks, evolutionary algorithms, and optimization. She has published several papers in the field of medical image analysis; examples include "Measurement of Optical Cup to Disk Ratio in Fundus Images for Glaucoma Screening," "Ensemble Learning Optimization for Diabetic Retinopathy Image Analysis," "Automatic Optic Disc Abnormality Detection in Fundus Images: A Deep Learning Approach," and "Deep Learning Approaches for Detecting COVID-19 From Chest X-Ray Images: A Survey." Currently, she is an Assistant Professor with King Abdulaziz University, Jeddah, Saudi Arabia. She is involved in several research activities and has participated as a reviewer for multiple journals and conferences locally and internationally. Examples include *Scientific Reports* journal, *Expert Systems with Applications* journal, IEEE ACCESS journal, *ACM Transactions on Asian and Low-Resource Language Information Processing* journal, the International Conference on Medical Image Computing and Computer-Assisted Intervention, MICCAI, International Conference on Computational Intelligence and Intelligent Systems, CIIS, International Conference on Machine Learning and Human-Computer Interaction, MLHMI and Asia Conference on Algorithms, Computing and Machine Learning, and CACML. Regarding professional experience, she was the Strategic Communication Manager with the Savola Group, Saudi Arabia, and as a Demonstrator with the University of Surrey, U.K.