جامعة شقراء
Shaqra University

Department of Computer Science
SHAQRA UNIVERSITY
May 2025

A SYSTEMATIC REVIEW OF ARABIC SIGN LANGUAGE TRANSLATION: METHODS, CHALLENGES, AND TECHNOLOGICAL ADVANCES

A Systematic Review Submitted in Partial Fulfillment of the Requirements for the Bachelor's Degree in Computer Science

by

Huda Obaid Almutiri

442690020

Batul Nasser ALHAFI

443690183

ALjohrah saud ALOTAIBI

443690176

Amani Owaidh Aldammas

443690172

Amirah Ayed Alotaibi

443690173

Supervised by: Dr.Fayha Al-mutairy

Abstract

Arabic Sign Language (ArSL) presents unique linguistic and technological challenges due to its rich dialectal diversity and lack of standardization. This systematic review explores the current landscape of ArSL translation technologies, including vision-based and sensor-based recognition methods, text-to-sign and speech-to-sign translation, and the integration of deep learning frameworks. Vision-based systems, such as those using OpenPose and MediaPipe, show strong potential but remain limited by environmental factors like lighting variations and occlusion. Sensor-based approaches (e.g., Myo, Kinect, Leap Motion) offer higher precision but face challenges in scalability and accessibility. Text-to-sign translation suffers from the lack of standardized gloss systems, while speech-to-sign systems are hindered by dialectal variability and real-time processing limitations. Deep learning models—including CNNs, RNNs, and Transformers—have notably advanced the field, yet progress is constrained by the scarcity and inconsistency of available datasets. This review emphasizes the urgent need for comprehensive ArSL datasets, standardized glossing protocols, and active community involvement to develop culturally adaptive, scalable, and ethically responsible translation systems. Future directions include leveraging GANs for data augmentation, employing multimodal fusion techniques, and deploying edge-AI solutions for real-time translation to enhance communication accessibility across the MENA region.

Table of Contents

1. Introduction

Sign languages are mature, native languages that express meaning in terms of a combination of hand gestures, blinked faces, and body movements. These visual –manual languages are the mother tongue for millions of deaf and hard of hearing around the world. Among those, Arabic Sign Language (ArSL), which can be viewed as a set of highly diverse sign languages which are used by Arabic speakers in the MENA region. Unlike spoken Arabic, the differences between ArSL and language are drastically different from one nation to another and there is no common standard which makes it linguistically diverse and difficult technologically. With increased momentum delivered by digital accessibility and inclusive technologies across the globe, the need for robust Arabic Sign Language translation system becomes more important. Such systems are intended to close communication gaps between the deaf and being able to hear, as well as help integrate deaf and able to hear people and to secure their better education and access to basic facilities.

Notwithstanding the growing advancement of sign language translation technologies in other languages, ArSL continues to be rather under-explored because of linguistic intricacy, regional dialectal variation, as well as unavailable standardized resources. Several contributions have been made to vision-based recognition, motion tracking with sensors, and natural language processes, although application to ArSL is still finding itself. Furthermore, advents of artificial intelligence and deep learning in the past few years hold transformational promise for constructing scalable and context aware ArSL translation systems. These technologies can facilitate the identification of subtle gestures, and the identification of facial expressions will become more precise, and there will be more accurate animated or textual sign outputs. But other limitations such as scarcity of dataset, variable signers, and lack of standard benchmarks, still have prevented development of the field.

This interdisciplinary area and the growing importance of ArSL translation technologies in computer science justify a detailed and organized review synthesis of the current state of the art. The purpose of this systematic review is to bring together the available research on computational aspects of Arabic Sign Language translation, to determine the most promising methods, and to point out the major challenges that require attention. Precisely, the review covers progress in important areas including vision-based and sensor-based recognition, text-to-sign and speech-to-sign translation, deep learning methods, availability of the datasets, practices of the evaluation, and real-world applications.

The chief research questions that inform this review are: “And What are the main technological methods of the translation of the Arabic Sign Language ?” (1) 2. How efficient are existing models of recognition and translation with respect to the problem of linguistic and regional diversity of ArSL? What problems and prospects arise for future study and application? In responding to these questions, the review not only maps out the territory of ArSL translation technologies but also provides the basis for building diverse, culturally responsive communication systems that are technologically robust and universally accessible.

The methodology for reviewing consisted in the structured search performed in several databases such as IEEE Xplore, Science Direct, Scopus, and Google Scholar searching the following keywords: "Arabic Sign Language," "gesture recognition," "speech-to-sign" and "sign language translation." Peer reviewed articles from 2013 to 2025 were linked that discussed ArSL technologies

2. Vision-Based Recognition in ArSL

Vision based recognition represents a core component in the creation of Arabic Sign Language (ArSL) translation systems, providing a non-intrusive and non-invasive approach for the identification and interpretation of gestures conducted at affordable cost. Hand shapes, facial expressions and body movements (which are core to sign language communication) are determined based on camera-based input in most of these systems. As cameras are ubiquitous in cell phones, laptops, and surveillance devices, vision-based ArSL recognition systems promise great scalability potential in personal, educational, and public service platforms.

2.1 2D vs. 3D Pose Estimation

Poise Estimation is one of the most basic processes of vision-based sign recognition, which is focused on the detection of some major points of the human body; for example, hands, joints and facial landmarks to determine the pose and movement of a signer. In 2D pose estimation, these key points are rebalanced on a flat plane that allows tracking of gestures on the basic level with video input (Abbas, Al-Barhamtoshy, & Alotaibi, 2021). Gestures that involve depth, overlapping limbs, or rotation are outside the domain of models such as the 2D models constructed with OpenPose and MediaPipe, where they have shown reasonable success in body and hand movement detection.

On the other hand, 3D poses estimation offers richer spatial information since it calculates depth and orientation in the three-dimensional space. This will make it easier to model complex gestures accurately particularly with dynamic signing that also contains motion in the z-axis. The added 3D spatial awareness is especially useful in differentiating what might seem like similar signs, in a 2D projection, but can be difficult to tell apart depending on the hand orientation or placement relative to the body. Methods of 3D pose estimation tend to leverage stereo camera systems or a monocular depth inference model using convolutional neural networks (CNNs). Even as they empirically approximate more closely to the reality, 3D systems are computationally more demanding, and therefore need higher quality input, which may not always be available in low-resource situations.

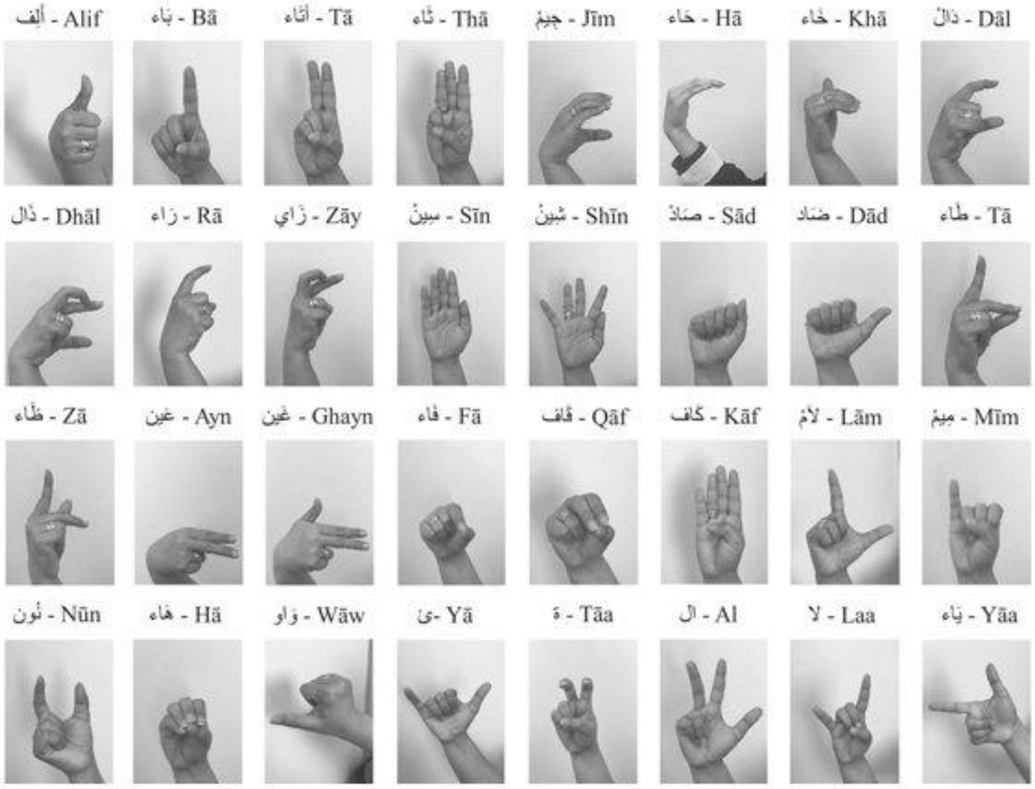

Figure 1: Vision-based Arabic Sign Language recognition using camera input and deep learning techniques

(Source: researchgate.net, 2025)

2.2 OpenPose, MediaPipe and YOLO on ArSL Recognition

Among the many existing vision-based frameworks for gesture recognition, OpenPose, MediaPipe, and YOLO (You Only Look Once) are identified for their performance and popularity in real-time applications. Open pose developed by the Carnegie Mellon Perceptual Computing Lab provides accurate multi-person pose estimations by detecting body, hand and face landmarks. This gives an opportunity to track fingers and upper body movement, which are important for identifying similar signs (Alani & Cosma, 2021). However, the device's resource usage has been the case for OpenPose, and it could be inefficient on devices with limited processing ability.

MediaPipe is an efficient and lightweight framework that codes for real-time hand and face tracking created by Google. Factoring in its cross platform-ability and low-latency PLAYABILITY, it makes an excellent candidate for mobile and web ArSL translation tools. Though MediaPipe's articulation of fingers may not be as detailed as OpenPose, the convenience of MediaPipe in live environments has resulted in its popularity among applications that use gesture recognition technology.

YOLO that has previously been utilized for object detection, has been simplified to detect and follow the movement of hands and body parts in real-time. Its main advantage is its speed and it's also capable of detecting various pieces of objects within a single sweep of an image (Mohammdi & Elbourhamy, 2023). In ArSL systems, YOLO is commonly applied with pose estimation frameworks for ensuring detection consistency, even if signs have rapid movements or occlusions. However, YOLO needs massive training data and may fall short of accuracy in low light condition or during detecting the fine gestures.

2.3 Difficulties concerning Lighting, Background and Occlusion.

Despite the progress of technology, vision-based ArSL recognition systems are burdened with significant issues with the environment. This type of lighting inconsistency is the most prevalent problem as poor or uneven lighting can obscure hand shapes and expressions making for poor recognition accuracy. Shadows and overexposure can distort the form of key areas creating a difficulty for models to accurately decipher signs. This is especially challenging in the uncontrolled environment such as outdoor, or home environment, in which illumination conditions would fluctuate greatly.

Background clutter is another problem which influences the robustness of vision-based models. In the case of complicated or dynamic background pictures, models have a tendency to be ineffective in distinguishing between the signer's hand or face – if the amount of contrast provided is insufficient (Luqman, 2023). Though some of the more sophisticated frameworks do contain segmentation algorithms to extract the foreground, their performance is otherwise hampered by the complexity of the scene.

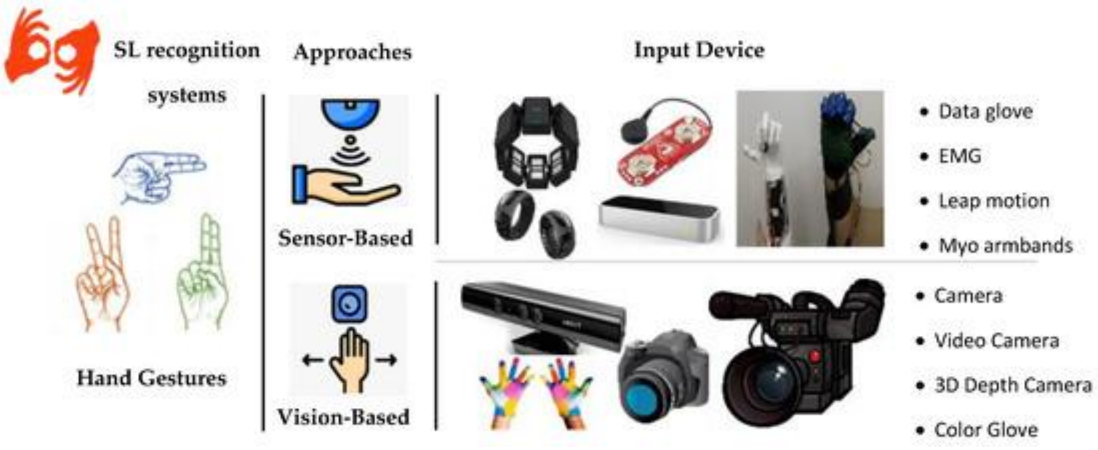

Figure 2: Architecture of Vision-Based Arabic Sign Language Recognition Using Pose Estimation and Deep Learning

(Source: mdpi.com, 2025)

Occlusion, where parts of the body are overlapped by one another, is a difficult obstacle for both two and three-dimensional models (Dabwan, Jadhav, Ali, & Olayah, 2023). For example, signaling or gestures that visibly involve both hands moving close to one's torso may partially or fully obscure important features. To overcome these challenges, advanced modeling techniques such as temporal smoothing should be applied, the predictive motion tracking technique should be used, and multi-view camera inputs should be used.

2.4 Future perspective of vision-based ArSL Recognition.

In terms of future, the vision based ArSL recognition will require more complex but more sophisticated machine learning models, which will involve multimodal data fusion with edge-computing deployment (Mohamed, Mustafa, & Jomhari, 2021). Transformer-based models with prospects in natural language processing and image analysis can be instructed to systematically analyze the spatial-temporal dynamics of sign languages. In addition, the combination of facial emotion recognition with hand gesture tracking may be able to provide richer contextual information and improve the correct deciphering of expressive or emotive signs.

3. Sensor-Based Recognition in ArSL

Sensor-based recognition has become an important paradigm in Arabic Sign Language (ArSL) translation systems because of its ability to record very accurate motion and orientation data from users. Whereas vision-based systems which use external cameras to interpret signs from visual cues are not sensor-based systems, which involve a direct acquisition of motion data from physical sensors placed on the body or in woven wearables. Such systems are especially touchy to minute finger movements and spatial hand positions thus appropriate for complex sign recognition jobs of high precision and resilience.

3.1 ArSL recognition the types of sensors used in.

Multiple sensors' types are used in the development of ArSL recognition systems based on sensors. These are accelerometers, gyroscopes, magnetometers and flex sensors, most often built into data gloves, or wearable armbands (Guo, Lu, & Yao, 2021). Linear acceleration and angular velocity are measured with use of accelerometers and gyroscopes, respectively, while conducting the dynamic movement of hand gestures. Flex sensors are to be sewn into gloves to determine the extent of finger bending, which is essential for discerning between handshapes. In certain more sophisticated systems sensor fusion is employed, in which data streams from different types of sensors are combined to improve the accuracy and resolution of movement recording.

Another class of sensor devices is composed of depth sensors and infrared sensors like the kind used in Microsoft Kinect. These sensors deliver spatial and skeletal data covering the whole upper body and thus allow full gesture tracking in a to-be space (Nahar et al., 2023). Another popular device is the Leap Motion Controller that uses infrared cameras and LEDs to detect motion of both hands and fingers, without any contact to the user. Similarly, the Myo armband measures both electrical activity in muscle (EMG) and motion data to provide a multimodal view of the execution of a gesture.

3.2 Comparative Analysis: Myo, Kinect, and Leap Motion

The Myo armband provides the compact wearable as a combo of motion sensing and muscle activity detection. It can be trained to identify gestures from the orientation of arms, acceleration and through electromyographic (EMG) signals. Its strength lies in its ability to capture muscle activation patterns that prove especially useful in the recognition of subtle or occluded signs (Qi et al., 2024). However, its major disadvantage is its poor ability to separate finger-level gestures because it mainly registers forearm muscle activity. This makes it more appropriate for general gesture categories than for fine tuning of ArSL fingers used in the sophisticated interpretation of ArSL.

Walking in and out of a doorframe wearing the Microsoft Kinect's original or even modern variants is likely an unpleasant experience if one hopes to use them for full body skeletal tracking. Its method of operation relies on structured light and depth sensors. This allows it to register trajectories of hands, body posture, and direction of facial orientation to reasonable accuracy. Kinect systems are beneficial in situations involving signers' independence and free space interaction without attachment.

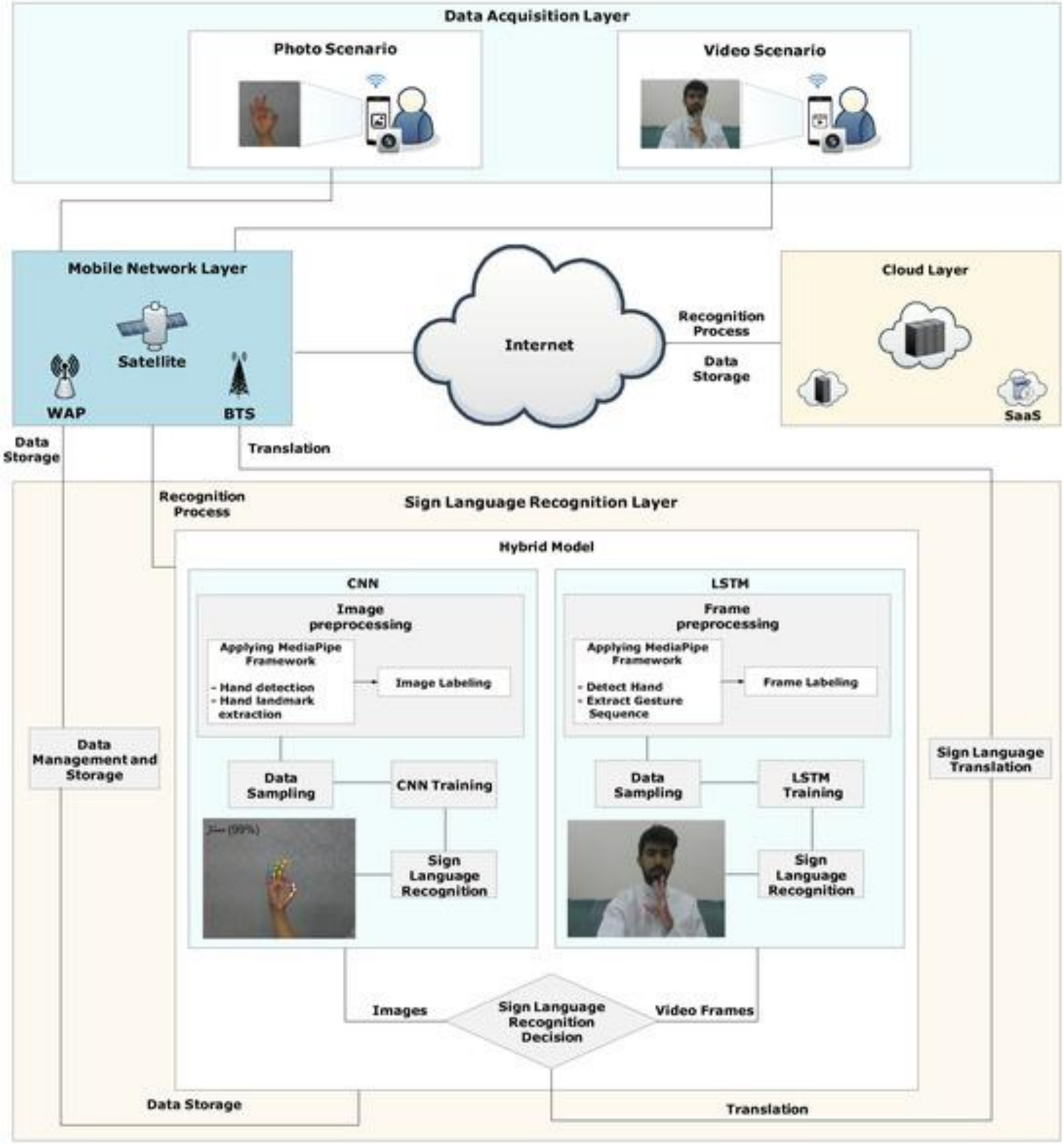

Figure 3: Comparative Overview of Sensor Devices Used in ArSL Recognition (Myo Armband, Leap Motion, Kinect)

(Source: mdpi.com, 2025)

Leap Motion Controller has been developed for specific tracking of hands and fingers at high resolution in a limited area of view. It is able to capture data of fine-grained hand gestures made with stereo infrared cameras and can determine individual finger positions with millimeter level accuracy. The controller shows good performance on the static and dynamic sign recognition task and can be used for real-time interaction (Chen, Wei, Sun, Wu, & Lin, 2022). Some of its limitations include sensitivity to occlusion and performance under the harsh condition of strong ambient lighting or dark tones where skin color can distort the infrared mechanism for reliable movement tracking.

3.3 Strengths and Weaknesses of Sensor-based vs. Vision based Methods

Sensor-based systems have a number of advantages over vision-based recognition, in particular, the precision of recognition, ability to tolerate changes in brightness and background, and ability to work if the target is occluded. They do not require visual inputs as such and are therefore immune to degradations associated with factors such as bad illumination or background clutter on the performance of camera dependent systems. Furthermore, sensor data is computationally week even to process and allows real-time feedback with lower latency when they are embedded in closed loop systems.

From a developmental perspective, sensor-based systems usually require intricate signal processing and calibration routines to decode raw sensor signals in meaningful gesture classifications (Zhou et al., 2021). This entails mapping output of sensors to the positions of hands or classes of signs through the use of machine learning models such as the support vector machines (SVM) or decision trees or neural networks. Although these models may have high accuracy in controlled environments, their performance may deteriorate in real-life situation that are variable in terms of users and also noise containing environment.

3.4 Scalability and Real-World Application Issues

Major challenges still lie ahead of the sensor-based ArSL recognition and this is scalability. The implementation of such systems over many people necessitates the standardization of the sensor hardware, the availability of strong user adaptation mechanisms, and the low-cost manufacture. The fact that there is a diversity of ways to perform signs shaped by regional dialects, signing styles and physical characteristics of users is one of major limitations. Sensor-based systems require training on large and diverse corpora to generalize well, a task made difficult due to the lack of public ArSL gesture corpora.

4. Text-to-Sign Translation for ArSL

Text-to-sign translation is no doubt a vital breakthrough towards the connectivity of the gap in communication between hearing citizens and the deaf community whereupon written or typed Arabic text can be translated automatically to Arabic Sign Language (ArSL). This process is composed of a number of linguistic and computational stages occurring between syntactic / semantic analysis up to intermediate representation through to the generation of sign output, which is usually in the form of animated avatars or video segments (Farooq et al., 2021). In spite of the promising application of text-to-sign translation for ArSL in educational platforms, public information systems and inclusive technology interfaces, this translation remains challenging due to the structural complexity of Arabic language and lack of standardization within ArSL itself.

4.1 Gloss Systems in Text-to-Sign Translation

One of the elemental building-blocks of any text-to-sign system is the presence of glosses i.e. a simplified form of the source language without any inflects act as a genism layer between the source text and the sign output. Glossing is used to simplify the target languages, so as to make it easy to map the Arabic words to their respective signs. Nevertheless, in ArSL the lack of a standardized glossing framework is a major barrier. Some researchers use gloss systems based on Modern standard Arabic (MSA), which do not, however, capture regional changes in signing and spoken dialects.

4.2 NLP Approaches: Rule-Based vs. Neural Methods

Standardized translation of Arab text into ArSL glosses or signs involves the overlying with natural language processing (NLP) methods which examine the grammatical properties and semantic aspect of the source text. In the past, a large number of rule-based systems have been applied to sign language translation tasks. These systems use handcrafted grammatical rules and dictionaries to translate text to glosses which give plenty of interpretability and control. Rearrangement of word order, negation, finding verbs Arabic conjugation and so forth may be examples of the rules that can be defined (Zhang et al., 2024). Here rule-based systems are fruitful in very constrained domains with low vocabulary but fail to generalize and scale up as rule construction is time consuming and the language of natural language is inconsistent.

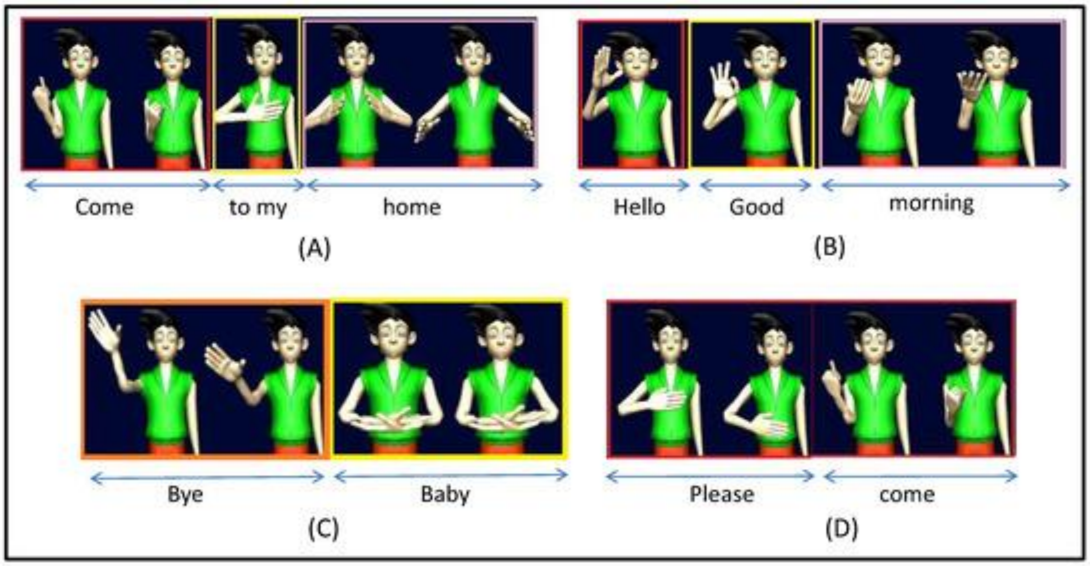

Figure 4: Workflow of Text-to-Arabic Sign Language Translation Using Gloss and Avatar Animation (e.g., Sign3D)

(Source: mdpi.com, 2025)

Neural NLP models, particularly transformer models, however, seem to be successful in capturing intricate syntactic patterns and semantic relations in Arabic text. These models are trained on huge input data sets and are able to learn a set of contextual mappings from a text to given glosses or direct sign representations. Systole to systole models with attention mechanisms have facilitated smoother and more context relevant sign language outputs enhancing the naturalness of the translated sign language outputs. However, neural model requires a significant amount of high-quality training data, which is unfortunately (for Arabic Sign Language) absent at this time (and particularly in the form of parallel text-gloss datasets).

4.3 Avatar Animation and Sign Rendering

After a gloss sequence has been produced, the second step is rendering of signs through animated avatars / visual displays. Virtual signers are used for what can be seen literally, but in the form of a human and help the deaf to understand the translation easily. One of the key tools in this domain is Sign3D, a 3D avatar driven animation platform, that takes gloss as an input to render the gestures in a set of animations. Such systems normally depend on pre-recorded motion-capture-data or synthetic animation-engines to generate realistic motions of hands, faces and body postures associated with ArSL signs.

Although they are staringly beautiful (or rather relays of effect-motif), avatar-based systems are plagued by a number of technical and perceptual limitations. Varied movements of many avatars as well as failure to capture the details of regional ArSL signs (including variations in facial grammar and hand orientation) are common (Xie, Qin, & Li, 2021). Moreover lip- syncing, emotion expressions, and non-manual markers (critical to sign language), are often scarcely managed, thereby limiting comprehension and user participation. The challenge increases further when trying to animate + continuous signs/idioms which do not have any one-to-one mapping from Arabic text.

4.4 Limitations and Future Considerations

The area of text-to-sign translation remains in infancy for Arabic Sign Language due to a lack of standardized glossing systems, the scarcity of linguistic resources and the absence of annotated corpora. Syntactic flexibility of Arabic and morphological richness of Arabic only aggravates the parsing and the translation procedures. Though both rules-based and neural approaches have their apparent benefits, none of them yields accurate or natural outputs throughout without significant manual fine-tuning, domain-specific tweaks.

In order to proceed, future systems must incorporate hybrid approaches which would fuse the linguistic strength of rule-based paradigm with the learning ability of neural networks (Forceville, 2022). Further, investment in the development of large scale, publicly available ArSL corpora including text-gloss-sign datasets is urgent to train and test more effective models. Incorporation of user feedback and collaboration with native ArSL signers will also be more than necessary to make the resultant systems culturally relevant and linguistically authentic. Once these technologies mature, they can transform the way that Arabic speaking deaf communities communicate across varied social and educational contexts.

5. Speech-to-Sign Translation for ArSL

Speech to sign translation is a recent area of work in the processing of Arabic Sign Language (ArSL) that is of high potential to be applied for real-time communication between the hearing and the deaf community. The goal of such systems is to convert the spoken Arabic language into ArSL output represented either as glosses or animated signs sequences. The viability of this translation process is especially high in public service, healthcare, and education since the real time interpretation may support inclusive communication (Yu et al., 2024). However, creating successful speech to sign applications for Arabic presents a particular set of linguistic, technical, and practical challenges because of the rich morphology of the language, diversity among the dialects, and lack of resources.

5.1 The ASR Pipeline for Arabic.

The first stage of a speech to sign system therefore follows with this being the transcription of spoken language to text by automatic speech recognition (ASR). In Arabic, this process is absolutely much more complicated as compared with such languages as the English language as the diglossia is present. Arabic includes Modern Standard Arabic (MSA) which is used formally and a collection of spoken dialects that varies much in pronunciation, vocabulary and syntax from region to region. Acoustic modeling that converts audio signals into phonetic forms is the first step for an ASR pipeline in Arabic. After that is language modeling whereby the likelihood of a sequence of words from the phonemes is estimated.

Moderately, ASR Systems have been successful in MSA because structured speech corpora are available. However, in the dialectal Arabic case, the lack of annotated data sets and standardized phonetic transcription is a serious obstacle (Chen et al., 2022). On this basis, many systems find it difficult to attain satisfactory accuracy in real world scenarios especially when it comes to handling fast or informal speech. Such recognition errors accumulate further down the pipeline in the translation process, with negative consequences for overall output quality of ArSL.

5.2 Use of Deep Learning Models: RNNs and Transformers

After having transcribed the spoken input into text, the system goes on delivering the text from Arabic to ArSL representation. More often than not deep learning models are used in this part of translation because they are able to capture complicated temporal dependencies as well as semantic nuances of language. Recurrent Neural Networks (RNNs), such as Long Short-Term Memory (LSTM) and the Gated Recurrent Unit (GRU), are conventional models used for working through sequential data, hence their use for mapping Arabic sentence structures into ArSL gloss sequences.

However, long sequences or contextual aware over long utterances pose constraints to RNNs. In order to overcome this challenge, new developments have provided Transformer-based architectures based on the attention mechanism (Zakariah et al., 2022). These models are capturing long range dependency and processing of parallel input sequences thus giving enhanced performance of mapping between spoken language and sign representations. Transformers have also eased the creation of end-to-end models that combine ASR, text-to-gloss translation, and gloss to sign rendering as a single cohesive structure.

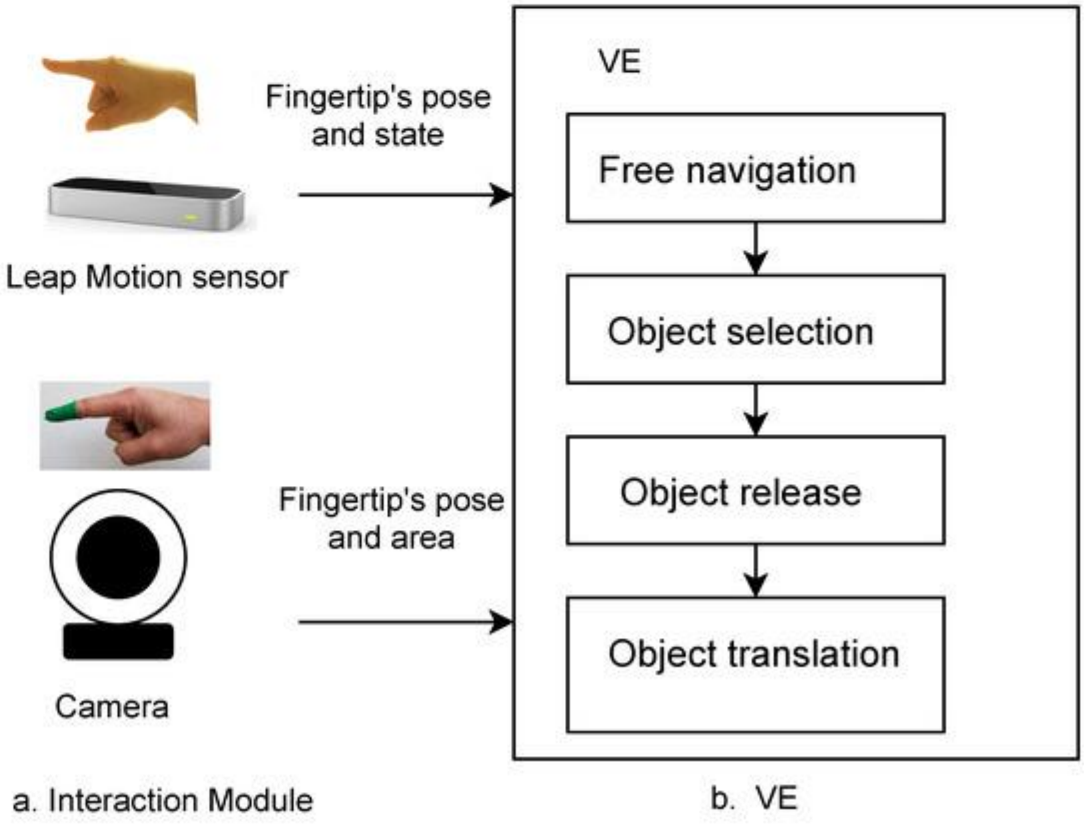

Figure 5: Workflow of Text-to-Arabic Sign Language Translation Using Gloss and Avatar Animation (e.g., Sign3D)

(Source: mdpi.com, 2025)

Even with their successes, both RNNs and Transformers are quite reliant on big, good quality datasets for training purposes. For data such as this when it comes to ArSL, they are so few, especially for dialectal Arabic, leading to overfitting and generalization issues, respectively. In addition, such models serve as black boxes, so it is not easy to reveal errors or optimize certain parts of a translation pipeline.

5.3 Dialectal Challenges and Variability

Highly dialectical application is one of the most critical issues in Arabic speech-to-sign translation due to the vast regional dialectal variation. Every Arabic-speaking country, and even entire cities, often have their own dialects, phonetics, idioms, and grammatical structures. This diversity has a major effect on the performance of ASR systems that are usually trained on standard or small dialectal corpora (Bila, Gargouri, Mahmood, & Mnif, 2024). Even if the ASR system correctly “simulates” the input speech, any dialect-specific expressions may not have point-by-point ArSL additive equivalents or may be culturally defined.

Moreover, regional dialects are generally poorly documented, and corresponding glosses in ArSL are either not standardized or simply do not exist in formalized datasets. This mismatch complicates the transformations from transcriptions of speech to sign language, which requires adaptive or regionalized models that will learn from divergent speech and signing profiles.

5.4 Real-Time Limitations and Practical Constraints

Real-time environments for speech-to-sign systems have various practical constraints. Processing latency is of great concern because any delay in transcription or translation can dictate the flow of communication (Alzubaidi, Otoom, & Abu Rwaq, 2023). Deep learning models, especially those operating on Transformers, are computationally demanding, and hardware acceleration is essential for delivering acceptable performance. Mobile and embedded deployment introduces additional restrictions in terms of power consumption and storage, and the requirement to process information on the device to increase implemented measures against infringement of user privacy.

6. Deep Learning Approaches in ArSL Translation

Improvements in deep learning have transformed ArSL translation with complex algorithms that are able to capture delicate gestures, manage sequential gestures with time and result in lifelike sign outputs. Researchers can now address such problems as differences in signing styles, time properties of gestures, and complexity of the Arabic language using architecture’s adaptability and learning from data in deep learning. The combination of DL models, including CNNs, RNNs, and Transformer architectures has become a necessity for advancing ArSL translation technologies.

6.1 Convolutional Neural Networks for Spatial Feature Extraction

ArSL researchers repeatedly refer to Convolutional Neural Networks (CNNs) for their exceptional ability to extract relevant spatial information from images and video frames. Facial expressions, hand shapes and body postures, which are fundamentals in sign language, are well-comprehended by CNNs. During ArSL recognition, CNNs are exploited to examine individual image frames to read still-hand gesture and facial emotion indication that carries grammatical or emotional meaning (Tharwat, Ahmed, & Bouallegue, 2021). Using the learning of the hierarchical feature representations, these networks can understand the nuances between similar-looking gestures that are identifiable from visual cues.

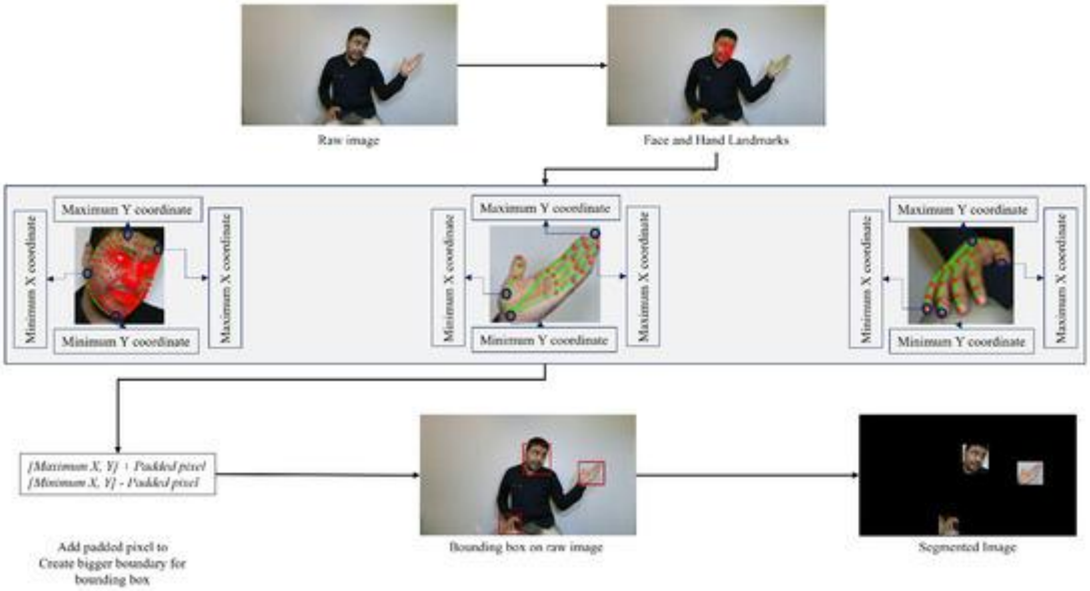

Figure 6: CNN-Based Feature Extraction in ArSL Gesture Recognition

(Source: mdpi.com, 2025)

In practice, CNNs are typically combined with pose estimation technologies such as OpenPose or MediaPipe for identifying joint locations and designating meaningful objects on the video frame. Examples of high-performance CNN models (ResNet, Inception, and VGGNet) have been further enhanced to allow for efficient ArSL data analysis and ultimately improve the gesture classification. As expected, CNNs struggle with estimating temporal context, thus it is difficult to understand how signs unfold across sequential frames.

6.2 RNNs and Temporal Modeling with LSTM and GRU

In order to remedy the limitations of CNNs in sequential data processing, Recurrent Neural Networks (RNNs) were nested into ArSL translation systems. RNNs are specifically designed to work with sequence data hence making them suitable for sequence-based analyses of temporal relationships in sign language videos. As a noteworthy point, both LSTM and GRU networks are very good at modeling extended temporal sequences and overcoming the problems related to the gradient vanishing that are common in traditional RNN methods.

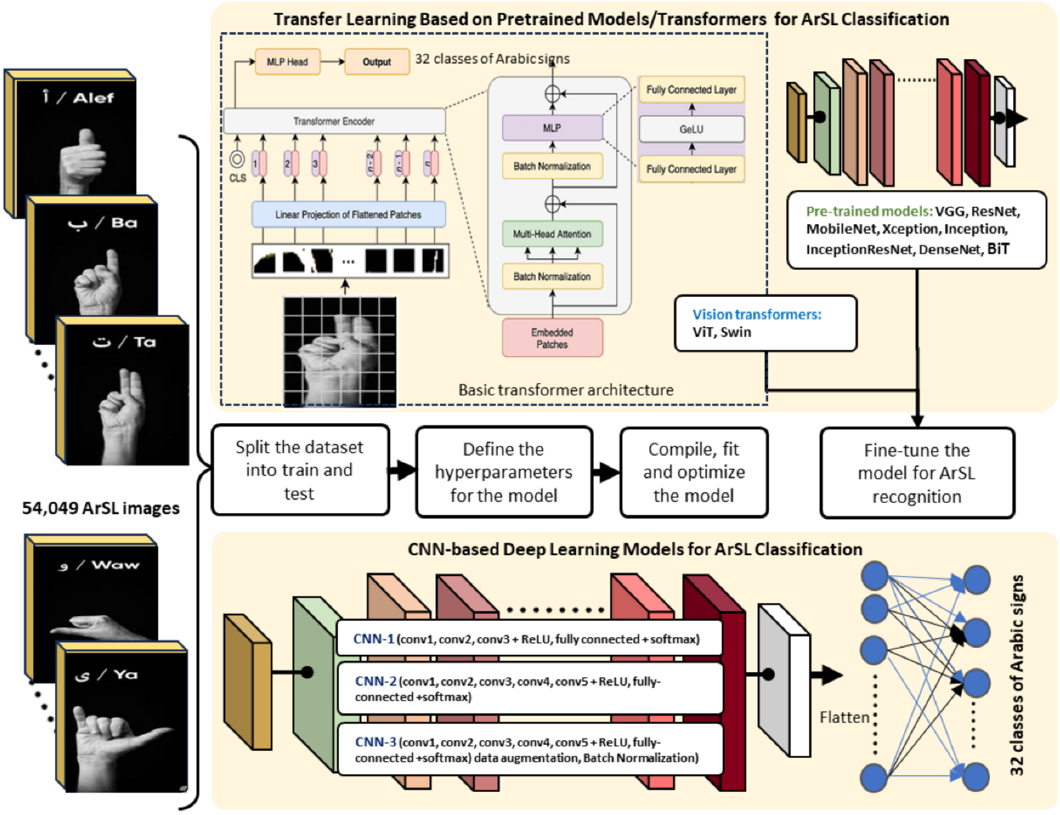
Transfer Learning Based on Pretrained Models/Transformers for ArSL Classification
MLP Head
Output
32 classes of Arabic signs
Transformer Encoder
CLS
1
2
3
n-2
n-1
n
Linear Projection of Flattened Patches
...
MLP
Batch Normalization
Multi-Head Attention
Batch Normalization
Embedded Patches
Fully Connected Layer
GeLU
Fully Connected Layer
Basic transformer architecture
Pre-trained models: VGG, ResNet, MobileNet, Xception, Inception, InceptionResNet, DenseNet, BiT
Vision transformers: ViT, Swin
أ / Alef
ب / Ba
ت / Ta
54,049 ArSL images
و / Waw
ى / Ya
Split the dataset into train and test
Define the hyperparameters for the model
Compile, fit and optimize the model
Fine-tune the model for ArSL recognition
CNN-based Deep Learning Models for ArSL Classification
CNN-1 (conv1, conv2, conv3 + ReLU, fully connected + softmax)
CNN-2 (conv1, conv2, conv3, conv4, conv5 + ReLU, fully-connected +softmax)
CNN-3 (conv1, conv2, conv3, conv4, conv5 + ReLU, fully-connected +softmax) data augmentation, Batch Normalization)
Flatten
32 classes of Arabic signs

Figure 7: Overview of Deep Learning Pipeline for Arabic Sign Language Translation

(Source: mdpi.com, 2025)

RNNs applied in ArSL translation take in sequence of visual features from either CNNs or pose estimation and learn how to convert those to glosses or written texts (Alawwad, Bchir, & Ismail, 2021). By using this method, the system can use both individual gestures and the context in which they occur. For instance, the same sign may convey an entirely different meaning once it is at different positions within the sequence, and RNNs can be able to pick differences like that.

Although RNNs prove to be excellent on many fronts, they still struggle with dealing with long-made sequences and maintaining attention on important frames. Consequently, developers have moved to Transformer models that are characterized by superior scalability and performance capabilities.

6.3 Transformer Architectures and Attention Mechanisms

Transformers models have recently become the top choice in sequence modelling in applications ranging from natural language processing and computer vision. With Transformers, unlike RNNs, that process sequences in elements one at a time, the property of self-attention is used in order to process the input as a unit which, consequently, makes it possible for Transformers to consider the relevant information regardless of its sequential order. The benefits of self-attention are especially prominent because of the requirement to use both local and global context for the interpretation of the sign language, which is a feature greatly benefiting the usage of self-attention.

respond Translating glosses into an Arabic text, generating sign movements from gloss representations, or changing visual input into gloss sequences (AlKhuraym, Ismail, & Bchir, 2022). The model can reallocate its focus with the help of attention onto things like hand posture, facial cues, or gesture path, delivering more reliable and contextually-informed translations.

Pretrained Transformer models BERT and ViT have also been fine-tuned for a particular task in the area of ArSL. These models both have wide training on large datasets followed by refinement on individual ArSL data to attain increased accuracy. Representing robustness, various forms of input, visual, textual, and auditory may be integrated by the Transformer-based models to enhance the system performance.

6.4 Transfer Learning and Cross-Lingual Adaptation

Limitations on large and labeled datasets is one of the big challenges in ArSL translation. Aware of the problem, scientists have started to use transfer learning more frequently, a technique that enables pretrained models from big datasets of sign language (such as ASL or BSL) to be fine-tuned for use on ArSL tasks. Utilized here are the common structural properties of sign languages that enable the model to be adapted for the Arabic context of a language.

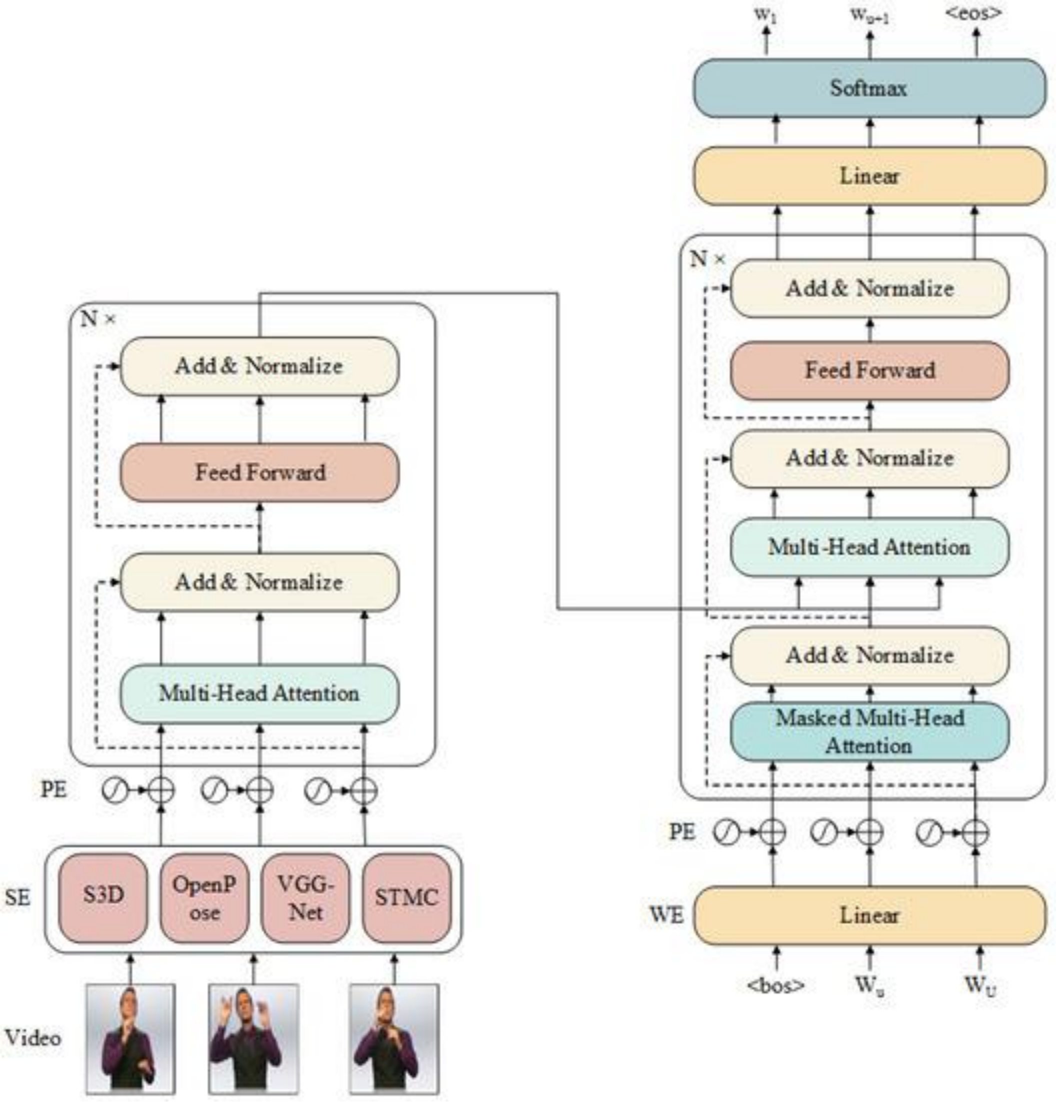

Figure 8: Transformer Attention Mechanism Applied to Sign Language Translation

(Source: mdpi.com, 2025)

It is possible to reuse a pretrained CNN-RNN architecture on American sign language, after gloves trainer had been trained on a reduced Arabic sign language corpus. As a result, the training phase now becomes faster and the performance of the system improves even if only a limited amount of training data is required (Rwelli, Shahin, & Taloba, 2022). Using methods of transfer learning, the systems for translating speech to signs for Arabic have been improved with the existing acoustic and language models optimized by using Arabic audio characteristics.

The results are improved by the combination of cross-lingual transfer and domain adaptation strategies that align the distributions of features between source and target languages. Recognition of the cultural and linguistic discrepancies that prevail in sign languages is important because the untargeted transfer could lead to misinterpretation of culturally unique signs.

6.5 Data Augmentation Using Generative Models

ArSL research is still plagued by insufficient training data. As a response, researchers have used the data augmentation approaches in order to augment the training data set. GANs are particularly effective at generating plausible synthetic sign language data including hand movements, signer figures and motion clips. GANs improve both generalization and robustness of deep learning models since it generates extra realistic training instances from a small dataset.

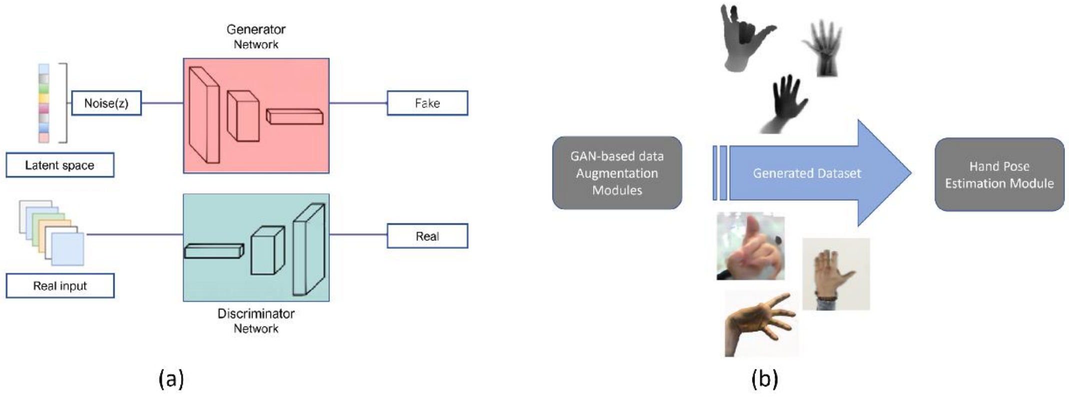

Figure 9: GAN-Based Synthetic Data Generation for Sign Language Training

(Source: mdpi.com, 2025)

The use of GANs in video-based sign language recognition enables researchers to change the signing, lighting and perspective, in the synthetic data, giving the models a broader set of examples for learning (Duwairi & Halloush, 2022). Synthetic gloss-labeled video clips can be an essential resource when original annotated data cannot be made available. However, thorough validation of data generated by GANs is needed to establish the authenticity and interpretability of this data, with a view of reducing the possibility of introducing confusion or bias into the learning process.

6.6 Multimodal Fusion Strategies

The potential of merging data from multiple modalities- visual, textual, and auditory sources- into a single deep learning paradigm is quite significant for ArSL translation. Fusion of modalities from visual, textual, and auditory sources can be reached at the input, intermediate feature, or final output level (early, mid, and late fusion, respectively). Based on the draw upon different modalities, the model can involve audio in the resolution of ambiguities in visual cues or synthesize gloss and visual elements to maximize the avatar's performance.

The latest architectures incorporate CNN, which is used for visual data processing; Transformers for handling sequence, and improved audio modules to facilitate smooth end-to-end speech to sign language conversion. These multimodal approaches are better equipped to take into account real-world variations, fit different signers and tolerate contextual ambiguity.

7. Datasets for Arabic Sign Language

Availability and quality of datasets are critical for moving machine learning or deep learning models, especially in complicated areas where sign languages are used such that the intricacies of gesture, facial expression, and spatial dynamics are a necessity. The absence of large, annotated, and standardized datasets constitutes a major barrier for research development with respect to Arabic Sign Language (ArSL). The resources that are available to ArSL relative to ASL and BSL are much fewer and of poorer quality, making this a significant issue for researchers who would like to develop high-performance, generalizable models.

7.1 Publicly Available ArSL Datasets

Some publicly available ArSL datasets have become available in the previous few years for the benefit of the research community. The ArabSign dataset is outstanding as an example as it is a multimodal resource used for continuous Arabic sign language recognition. It has video recordings from both color (RGB) and depth sensors of signers performing ArSL signs and phrases; the video can be viewed here: ArabSign. In order to apply more realism to the dataset, it concentrated on gathering different signing behavior, i.e., changes in speed, hand placement and other features, such as facial gesture. ArabSign, though trailblazing, still struggles because of its limited vocabularies sizes and narrow regional dialects coverage.

ArabicSL mark signs by their isolation and intentional use in calibrated circumstance for primarily accentuating obvious hand movement. ArabicSL is primarily composed of videos that share individual signs which are effective at classification but not forwardable to entire sentences or consecutive translations (Boukdir et al., 2021). Although it has an organized way of identifying gestures, the limited nature of the dataset about signer and background diversity restricts the applicability of these models in general.

7.2 Challenges in Annotation and Gloss Representation

Precisely annotating ArSL datasets is one of the greatest problems for researchers. Annotation of the sign language materials demands considerations that extend to various dimensions such as gloss labels, hand shapes, movement trajectories and non-manual markers. While languages that are depicted through speech or through text have characteristics and nuances, the sign languages have gestures much more characteristic – many of which are simultaneous and therefore it is difficult to discriminate and accurately annotate such movements and nuances.

In addition to this, absence of a generalized glossing system for ArSL further complicates the research in ArSL. Glosses are shortened non-inflected forms of sign language corresponding to written form, typically created from meanings of spoken language (Alyami, Luqman, & Hammoudeh, 2024). For Arabic, the diglossia and presence of different regional dialects make the problem even worse, and different sign conventions are used for the same concept in different Arab speaking countries. Lack of centralized gloss corpus degrades the achievement of uniform annotations across datasets. As a result of such contradiction both the development and prediction of models are difficult, which hinders effective comparison amongst them.

7.3 Limitations in Dataset Coverage and Diversity

This narrow parameter of available Arabic Sign Language datasets narrows the scope of the models used for development and evaluation. Sign language datasets tend to contain less than 1000 signs, and they often have a small group of signers, limiting them in realistic contexts. Recognizing systems have difficulties with robustness when they come across different styles of signing, variations in body characteristics, or shifts in lighting as a result of sparse variation of the dataset.

Thus, because the available datasets more or less only address single signs rather than continuous signing, their use for translation systems that depend on analysis of syntactic and semantic relationship in longer utterances is debilitated. Absence of multimodal data such as synchronized sound, skeletal tracking and 3D pose recordings hinders the development of multimodal or hybrids.

7.4 Need for Standardization and Future Efforts

Dealing with these issues will call for efforts to make large, accessible ArSL datasets that hold a range of signs, dialects, and signers (Alani & Cosma, 2021). They must include complete gloss annotations, Multiview footage, and non-manual marker tracking which are essential to develop robust models. Such synergies between academia and the sign language congregations and the research bodies are necessary to generate linguistically and ethically correct datasets.

Consistent gloss corpora as well as community-based annotation rules are essential for increasing interoperability and benchmarking in ArSL study. Increasing demand for barrier-free technology (BFT) necessitates the development of robust, extensive ArSL datasets as the first step toward creating effective, reliable, and culturally appropriate translation systems for sign languages.

8. Challenges in ArSL Translation

Many benefits have been achieved in terms of advancement of ArSL translation technologies, through deep learning and computer vision, but linguistic, technical, societal, and ethical issues cannot be ignored and remain the impediments to the creation of effective and inclusive systems. These challenges run along linguistic, technical, societal, and ethical dimensions, indicating that sign language processing in the world of Arabic is multi-layered and complex per se. It is of paramount importance to overcome these barriers because reliable and culturally shaped ArSL translation systems covering a broad spectrum of users in MENA are needed.

8.1 Linguistic Complexity and Dialectal Diversity

One of the key linguistic challenges when translating into Arabic Sign Language is caused by a lack of a standard language that is widely accepted. ASL, which has its codification and documentation thoroughly described, opposes ArSL, which does not constitute a single unified language, but consists of various sign languages which are being used in separate Arabic speaking countries. Dialectal diversity leads to great variation in the use of signs with the same meaning and also brings in variation in the patterns of grammar, hand shapes, orientations, and facial expressions.

This linguistic problem becomes even more complex because of diglossia, and the coexistence of MSA with various regional dialects. Even though written form of Modern Standard Arabic is observed, spoken Arabic varies significantly from one region to the other (Aloysius, Geetha, & Nedungadi, 2021). Such language diversities result in the characteristic signing tendencies that are unique to every region, which makes it difficult to develop an adequate translation system that can link diverse local dialects. The unusual syntactic specialty of Arabic—its peculiarities in verb morphology, and the subject-verb-object order—adds another layer of complexity when translating from or to spoke/written Arabic and sign glosses in that it does not quite match many Western languages.

8.2 Technical Limitations and Model Constraints

ArSL translation systems face a variety of technical challenges arising from poor and varying training data, ranging from problems with quality and quantity of data. Highly effective deep learning systems thrive on the availability of huge and well annotated datasets for them to perform with accuracy. Further, the size of the available ArSL datasets is also limited and these are not uniform in the glossing systems and annotation criteria used. Absence of consistency in annotation standards reduces trustworthiness of a model and makes cross-study comparison arduous.

In addition, the dependence on particular signers is a great challenge for both vision-based and sensor-based models. Several systems show good performance with the signers that are present in training but often lack the ability to turn gestures that are performed by the users with various physical attributes, signing methods or motion styles. As a result of this lack of independence among the singers, the ability to be adaptable to the real scenarios, on the part of the models, is significantly limited. Environmental condition noise such as poor lighting and busy backgrounds, and camera performance degrade recognition accuracy in vision-based models.

Real-time translating systems have great obstacles especially in handling latency, high computational requirements, and rapid response time. In many cases, designing real time, low-latency systems in mobile or embedded hardware are plagued by processing challenges and continuity of gesture identification.

8.3 Societal Barriers and Community Engagement

Apart from overcoming technical barriers, other wider socio-cultural issues play a major role in the development and acceptance of ArSL translation technologies. In many Arab-speaking countries deaf communities are often ignored in the creation and management of technology meaning that tools can fail to be sensitive to their ways of communicating, their cultural norms, and values (De Coster et al., 2021). Reduced availability of inclusive educational materials and devices to deaf and hearing communities causes the communication gaps to remain.

Social stigma against disability in some areas may discourage the adopters of assistive technologies. Besides, sensor-based devices can be found alarming and socially uncomfortable by some users, thereby suppressing their adoption rate. The community needs to be involved in the whole design, development and evaluation process of ArSL systems to develop a sense of trust, improve usability and remain relevant.

8.4 Ethical Considerations and Data Consent

Since ArSL systems include biometric data, video recordings, and Machine Learning, practical questions prevail. When visual and gestural information of users is captured, privacy becomes a greater issue if consent has not been clearly and consciously given. In some of the existing datasets it is ambiguous if the participants were informed how their data will be used, stored or distributed. Such practices raise questions about who owns the data, how it is being presented and who is accountable.

9. Applications for ArSL Translation Technologies

The developments in the Arabic Sign Language (ArSL) translation technologies bring the hope of making an enormous improvement in the ability of the deaf and hard-of-hearing to communicate and access information in the MENA region. The ripening of these technologies requires them to be introduced into practical settings, including education, healthcare, public services, and digital platforms in order to promote social uptake and inclusion.

It’s a significant facilitator of deaf students learning experiences in educational systems through the use of ArSL translation technologies. By means of real time translation with animated avatars or gesture recognition, the deaf students are able to go to lectures, discuss issues and access digital educational materials without the assistance of human interpreter (Xie, Zhao, & Hu, 2021). In some Gulf countries, the production of intelligent tutor systems is characterized by ArSL avatars, which have a major impact on the teaching of sign language and communication between learners and instructors.

Healthcare requires great advantages from the ArSL translation systems since communication effectiveness is the critical factor in resulting in accurate diagnoses and appropriate treatments. Technologies that combine the ArSL recognition and translation capabilities allow deaf people to interact with healthcare representatives without excessive assistance from intermediaries and maintain confidentiality. Some hospitals in the United Arab Emirates have tried out AI-powered translation kiosks that facilitate the translation of spoken Arabic into ArSL through video avatars, cutting patient wait times at check-in.

Public application of ArSL is frequent nowadays, particularly where the field is public transportation, law enforcement, and government programs (Jin, Zhao, Zhang, & Zeng, 2022). For instance, the public transport systems can offer services in digital kiosks or mobile applications which automatically translate spoken announcements or written information into ArSL such that the deaf are able to manage their travels, access legal information or accomplish administrative functions more independently. In Saudi Arabia and in Egypt, the government undertakings have looked to integrate sign language avatars onto established government websites presenting information on citizen services in accessible formats.

More and more mobile and wearable devices are making it possible to distribute compact ArSL translation tools, which promotes accessibility. Technology companies are creating apps that can translate signs captured by phones' cameras or text into animated sign language to make it simpler for day-to-day interaction (Amin, Hefny, & Ammar, 2021). These tool users are more appropriate in dealing with activities that include shopping, moving around public space and confidently communicating out of their homes.

Eventually, the advancement of the ArSL translation technologies makes digital space more inclusive. The widespread adoption of sign language in digital channels, for instance online, education or media, contributes to ensure that deaf people can participate fully in digital economy. As the uptake grows, there will be a growing need for culturally sensitive, linguistically accurate and user-friendly ArSL solutions that will advance innovation and inclusion in many spheres of activities.

10. Benchmarking and Evaluation

System evaluation in translation for Arabic Sign Language (ArSL) is essential to measure model's success, push the field forward, and ensure that the applications do not remain laboratory-relevant.

However, the fact that in practice there are no standardized benchmarks, datasets and evaluation procedures makes progress challenging. In the absence of continuous evaluation, it is hard to compare how various approaches perform and produce valuable conclusions, which diabolically hinders the progress in this important area of research.

A set of standard metrics initially designed for use in speech recognition, machine translation and classification have been borrowed to assess the performance of ArSL translation components. In order to determine the correctness of the input speech, with the aim of speech-to-text or gloss generation, Word Error Rate (WER) will frequently be used. WER measures speech recognition in terms of numbers of errors that are included in the model's output compared to the reference, including insertion of words, omission of words, and swapping words (Alethary, Aliwy, & Ali, 2022). Even though WER measures the accuracy of speech recognition, it pays little attention to semantic similarities and contextual variations, particularly with visual languages such as ArSL.

In terms of text to sign or speech to sign translation BLEU (Bilingual Evaluation Understudy) scores are commonly used to compare the automatically generated glosses or signs with reference translations. When n-gram segments in machine and reference texts are compared for their matching, BLEU makes it possible to assess fluency and linguistic accuracy of translations. Despite this, the lack of consistent gloss references and the lack of textual notation for complex non-manual components mean that BLEU does not truly evaluate sign language translations.

Consideration of both precision and recall indicates that F1 score would normally be applied in gesture classification tasks in vision-based or sensor-based ArSl recognition systems. The use of these metrics enables us to determine how well the system detects individual signs and gestures, particularly in situations that focus on one or several gestural classifications. Although a good F1 score indicates a good balance between signs that are correctly identified and those that are not, it does not guarantee system usability or the effective translation of sign language fluently.

At present, one of the main issues facing ArSL benchmarking is the lack of community-wide, open, and standardized datasets with well-defined evaluation benchmarks. As compared to English-based sign language systems, at present there is no common benchmark or evaluated task for ArSL translation systems in terms. To address these issues, researchers proposed ideas for benchmarking systems: ones with established gloss corpora, video datasets with annotations, test sets detached from signers, and real-time measures such as interpretability and interaction speed.

11. Future Research Directions

It is important to do more research on advances in Arabic Sign Language (ArSL) translation in addressing room for improvements like lack of data, real-time implementation challenge, inconsistent linguistics while building both scalable and inclusive systems. One of the noteworthy opportunities for progress is the use of GANs to produce synthetic sign language data. GANs produce genuine sign language videos and gesture sequences that can significantly expand the presence and richness of existing datasets as the style, light, and background demographics can be diversified (Alanazi, 2024). Such augmentation is critical to improving the capacity of deep learning models to succeed with a wide range of users and dialects.

Another promising domain of development addresses edge-AI applications directed at the implementation of real-time ArSL translation capabilities on mobile and embedded devices. Edge-AI technologies are a must in terms of being designed for low energy consumption and low processing times in order for them to work reliably off the grid, an imperative especially in areas experiencing scarcity in digital connectivity. Such technologies support user privacy by managing data processing on device, rather than transmission to remote servers.

Development of a common glossing system is the secret to ensuring consistency in translating spoken Arabic into sign language to be used in various dialects and places. Standardization in a gloss corpus would act as a linguistic interface between spoken Arabic and signed languages, resulting in enhanced reliability in translation and the potential for the replication of research findings.

12. Conclusion

This review highlights the state-of-the-art and hurdles in behavior of Arabic Sign language in terms of translation technologies that include evaluation of recognition methods, translation frameworks, deep learning methods, data availability and applicability. Despite impressive performance in vision-based, sensor-based, and speech-to-sign technology, the area is hampered by continued challenges such as lack of appropriate data, extreme linguistic differences, and evaluative disparities. Deep learning come improvements, due to CNNs, RNNs, Transforms, and GANs, provide greater ability to capture complex gestures and sequences. However, the lack of reliable gloss corpora and annotated datasets vastly hampers systems' ability to be accurate and scalable. Nevertheless, pragmatic obstacles such as dependence on the singer, latency, and environmental variance are important problems that remain to be solved for effective real-world use. It is crucial that future systems come into play with strong robustness and contextual insight to address these challenges through the use of multimodal approaches that combine visual, auditory and textual information. Inclusivity will also be critical, as ArSL technologies will be developed taking respect and integration of the specific needs and cultural accounts of Arab speaking hearing impaired individuals as a core premise. Achieving optimal ArSL translation systems relies upon cooperative work, significant community input, and ethically-oriented development of tools to make the area accessible, equitable, and more inclusive in digital terms for users in MENA.

13. References

Abbas, S., Al-Barhamtoshy, H. and Alotaibi, F., 2021. Towards an Arabic Sign Language (ArSL) corpus for deaf drivers. *PeerJ Computer Science*, *7*, p.e741.

Alanazi, M.S., 2024. The use of Modern Standard Arabic and colloquial Arabic in translation tasks: a new perspective. *Cogent Arts & Humanities*, *11*(1), p.2366572.

Alani, A.A. and Cosma, G., 2021. ArSL-CNN: a convolutional neural network for Arabic sign language gesture recognition. *Indonesian journal of electrical engineering and computer science*, *22*(2), pp.1096-1107.

Alani, A.A. and Cosma, G., 2021. ArSL-CNN: a convolutional neural network for Arabic sign language gesture recognition. *Indonesian journal of electrical engineering and computer science*, *22*(2), pp.1096-1107.

Alawwad, R.A., Bchir, O. and Ismail, M.M.B., 2021. Arabic sign language recognition using Faster R-CNN. *International Journal of Advanced Computer Science and Applications*, *12*(3).

Alethary, A.A., Aliwy, A.H. and Ali, N.S., 2022. Automated Arabic-Arabic sign language translation system based on 3D avatar technology. *Int J Adv Appl Sci*, *11*(4), pp.383-396.

AlKhuraym, B.Y., Ismail, M.M.B. and Bchir, O., 2022. Arabic sign language recognition using lightweight cnn-based architecture. *Int. J. Adv. Comput. Sci. Appl*, *13*(4).

Aloysius, N., Geetha, M. and Nedungadi, P., 2021. Incorporating relative position information in transformer-based sign language recognition and translation. *IEEE Access*, *9*, pp.145929-145942.

Alyami, S., Luqman, H. and Hammoudeh, M., 2024. Isolated arabic sign language recognition using a transformer-based model and landmark keypoints. *ACM Transactions on Asian and Low-Resource Language Information Processing*, *23*(1), pp.1-19.

Alzubaidi, M.A., Otoom, M. and Abu Rwaq, A.M., 2023. A novel assistive glove to convert arabic sign language into speech. *ACM Transactions on Asian and Low-Resource Language Information Processing*, *22*(2), pp.1-16.

Amin, M., Hefny, H. and Ammar, M., 2021. Sign language gloss translation using deep learning models. *International Journal of Advanced Computer Science and Applications*, *12*(11).

Architecture of Vision-Based Arabic Sign Language Recognition Using Pose Estimation and Deep Learning, mdpi.com, 2025, [Online], Retrieved From: < https://www.mdpi.com/2073-431X/13/6/153> Retrieved On: 09.05.2025

Bila, Z.S., Gargouri, A., Mahmood, H.F. and Mnif, H., 2024. Advancements in Arabic Sign Language Recognition: A Method based on Deep Learning to Improve Communication Access. *Journal of Internet Services and Information Security*, *14*(4), pp.278-291.

Boukdir, A., Benaddy, M., Ellahyani, A., Meslouhi, O.E. and Kardouchi, M., 2021. Isolated video-based Arabic sign language recognition using convolutional and recursive neural networks. *Arabian Journal for Science and Engineering*, pp.1-13.

Chen, Y., Wei, F., Sun, X., Wu, Z. and Lin, S., 2022. A simple multi-modality transfer learning baseline for sign language translation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 5120-5130).

Chen, Y., Zuo, R., Wei, F., Wu, Y., Liu, S. and Mak, B., 2022. Two-stream network for sign language recognition and translation. *Advances in Neural Information Processing Systems*, *35*, pp.17043-17056.

CNN-Based Feature Extraction in ArSL Gesture Recognition, mdpi.com, 2025, [Online], Retrieved From: <https://www.mdpi.com/1424-8220/23/16/7156> Retrieved On: 09.05.2025

Comparative Overview of Sensor Devices Used in ArSL Recognition (Myo Armband, Leap Motion, Kinect), mdpi.com, 2025, [Online], Retrieved From: < https://www.mdpi.com/1424-8220/24/11/3683> Retrieved On: 09.05.2025

Dabwan, B.A., Jadhav, M.E., Ali, Y.A. and Olayah, F.A., 2023, March. Arabic sign language recognition using efficientnetb1 and transfer learning technique. In *2023 International conference on IT innovation and knowledge discovery (ITIKD)* (pp. 1-5). IEEE.

De Coster, M., D'Oosterlinck, K., Pizurica, M., Rabaey, P., Verlinden, S., Van Herreweghe, M. and Dambre, J., 2021. Frozen pretrained transformers for neural sign language translation. In *18th Biennial Machine Translation Summit (MT Summit 2021)* (pp. 88-97). Association for Machine Translation in the Americas.

Duwairi, R.M. and Halloush, Z.A., 2022. Automatic recognition of Arabic alphabets sign language using deep learning. *International Journal of Electrical & Computer Engineering (2088-8708)*, *12*(3).

Duwairi, R.M. and Halloush, Z.A., 2022. Automatic recognition of Arabic alphabets sign language using deep learning. *International Journal of Electrical & Computer Engineering (2088-8708)*, *12*(3).

Farooq, U., Rahim, M.S.M., Sabir, N., Hussain, A. and Abid, A., 2021. Advances in machine translation for sign language: approaches, limitations, and challenges. *Neural Computing and Applications*, *33*(21), pp.14357-14399.

Forceville, C., 2022. Visual and multimodal communication across cultures. *The Cambridge Handbook of Intercultural Pragmatics/CHIP*, pp.527-551.

GAN-Based Synthetic Data Generation for Sign Language Training, mdpi.com, 2025, [Online], Retrieved From: <https://www.mdpi.com/2227-7080/10/2/43> Retrieved On: 09.05.2025

Guo, L., Lu, Z. and Yao, L., 2021. Human-machine interaction sensing technology based on hand gesture recognition: A review. *IEEE Transactions on Human-Machine Systems*, *51*(4), pp.300-309.

Jin, T., Zhao, Z., Zhang, M. and Zeng, X., 2022, May. Prior knowledge and memory enriched transformer for sign language translation. In *Findings of the Association for Computational Linguistics: ACL 2022* (pp. 3766-3775).

Luqman, H., 2023, January. ArabSign: A multi-modality dataset and benchmark for continuous arabic sign language recognition. In *2023 IEEE 17th International Conference on Automatic Face and Gesture Recognition (FG)* (pp. 1-8). IEEE.

Mohamed, N., Mustafa, M.B. and Jomhari, N., 2021. A review of the hand gesture recognition system: Current progress and future directions. *IEEE access*, *9*, pp.157422-157436.

Mohammdi, H.M. and Elbourhamy, D.M., 2023. An intelligent system to help deaf students learn Arabic Sign Language. *Interactive Learning Environments*, *31*(5), pp.3195-3210.

Nahar, K.M., Almomani, A., Shatnawi, N. and Alauthman, M., 2023. A robust model for translating arabic sign language into spoken arabic using deep learning. *Intell Autom Soft Comput*, *37*(2), pp.2037-2057.

Overview of Deep Learning Pipeline for Arabic Sign Language Translation, mdpi.com, 2025, [Online], Retrieved From: <https://www.mdpi.com/2076-3417/13/21/11625> Retrieved On: 09.05.2025

Qi, J., Ma, L., Cui, Z. and Yu, Y., 2024. Computer vision-based hand gesture recognition for human-robot interaction: a review. *Complex & Intelligent Systems*, *10*(1), pp.1581-1606.

Rwelli, R.E., Shahin, O.R. and Taloba, A.I., 2022. Gesture based Arabic sign language recognition for impaired people based on convolution neural network. *arXiv preprint arXiv:2203.05602*.

Tharwat, G., Ahmed, A.M. and Bouallegue, B., 2021. Arabic sign language recognition system for alphabets using machine learning techniques. *Journal of Electrical and Computer Engineering*, *2021*(1), p.2995851.

Transformer Attention Mechanism Applied to Sign Language Translation, mdpi.com, 2025, [Online], Retrieved From: <https://www.mdpi.com/2079-9292/12/12/2678> Retrieved On: 09.05.2025

Vision-based Arabic Sign Language recognition using camera input and deep learning techniques, researchgate.net, 2025, [Online], Retrieved From: < https://www.researchgate.net/figure/Representation-of-the-Arabic-sign-language-for-Arabic-alphabets_fig1_359510485 > Retrieved On: 09.05.2025

Workflow of Text-to-Arabic Sign Language Translation Using Gloss and Avatar Animation (e.g., Sign3D), mdpi.com, 2025, [Online], Retrieved From: < https://www.mdpi.com/2079-9292/9/12/1986> Retrieved On: 09.05.2025

Workflow of Text-to-Arabic Sign Language Translation Using Gloss and Avatar Animation (e.g., Sign3D), mdpi.com, 2025, [Online], Retrieved From: <https://www.mdpi.com/2076-3417/11/8/3439> Retrieved On: 09.05.2025

Xie, H., Qin, Z. and Li, G.Y., 2021. Task-oriented multi-user semantic communications for VQA. *IEEE Wireless Communications Letters*, *11*(3), pp.553-557.

Xie, P., Zhao, M. and Hu, X., 2021. PiSLTRc: Position-informed sign language transformer with content-aware convolution. *IEEE Transactions on Multimedia*, *24*, pp.3908-3919.

Yu, F., Xiang, Z., Che, N., Zhang, Z., Li, Y., Xue, J. and Wan, Z., 2024. Pilot-guided Multimodal Semantic Communication for Audio-Visual Event Localization. *arXiv preprint arXiv:2412.06208*.

Zakariah, M., Alotaibi, Y.A., Koundal, D., Guo, Y. and Mamun Elahi, M., 2022. Sign language recognition for Arabic alphabets using transfer learning technique. Computational Intelligence and Neuroscience, 2022(1), p.4567989.

Zhang, G., Hu, Q., Qin, Z., Cai, Y., Yu, G. and Tao, X., 2024. A unified multi-task semantic communication system for multimodal data. *IEEE Transactions on Communications*, *72*(7), pp.4101-4116.

Zhou, H., Zhou, W., Qi, W., Pu, J. and Li, H., 2021. Improving sign language translation with monolingual data by sign back-translation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 1316-1325).