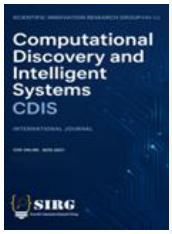

Computational Discovery and Intelligent Systems
CDIS
ISSN: 3070-5037/© 2026 CDIS. All Rights Reserved.
Journal Homepage
https://pub.scientificirg.com/index.php/CDIS/index

Automatic Arabic Sign Language Recognition: A Comprehensive and Critical Review of Methods, Datasets, Challenges, and Future Directions

Hanaa Atwa Sayed[a,1], Ahmed A. Elngar[b], Mohammed Kayed[c]
[a] *Faculty of Computer Science, Nahda University, Beni-Suef City, 62511 Egypt.*
Email: hanaa.atwasayed@gmail.com
[b] *Faculty of Computers and Artificial Intelligence, Beni-Suef University, Beni-Suef City, 62511, Egypt.*
Email: elngar_7@yahoo.co.uk
[c] *Faculty of Computers and Artificial Intelligence, Beni-Suef University, Beni-Suef City, 62511, Egypt.*
Email: mskayed@gmail.com

ABSTRACT

Sign language is the primary means of communication for people who are deaf or hard of hearing, which makes accessibility technologies increasingly important. In this context, Automatic Arabic Sign Language Recognition (ArSLR) has gained significant attention to enable effective communication between sign and spoken languages. This paper presents a comprehensive review of existing ArSLR approaches, covering the recognition of Arabic alphabets, isolated words, and continuous signs. It discusses both vision-based and sensor-based methods, explaining how they work and highlighting their strengths and limitations. The paper also reviews available datasets and commonly used evaluation methods in the literature. Several key challenges are explored, including variations between signers, complex backgrounds, and the limited availability of large and diverse datasets. In addition, less-studied areas such as backhand gestures and continuous sign recognition are highlighted as important directions for further research. Finally, the paper looks at recent trends, including the use of advanced sensors and the development of more robust, signer-independent models. By identifying current limitations and research gaps, this review aims to guide future work toward building more accurate, reliable, and practical ArSLR systems.

PAPER INFORMATION

HISTORY
Received: 2 January 2026
Revised: 5 March 2026
Accepted: 15 April 2026
Online: 25 April 2026

MSC
68T07; 68R10; 94A60; 68M15

KEYWORDS
Arabic Sign Language Recognition (ArSLR); Vision-Based Recognition (VBR); Sensor-Based Recognition (SBR); Deaf Communication; Data Glove Technology.

[1]Corresponding author: Faculty of Computer Science, Nahda University, Beni Suef, 62511 Egypt, Email:hanaa.atwasayed@gmail.com

1 INTRODUCTION

Since sign language is not familiar to people outside the deaf community, communication between the deaf and hearing people of all ages becomes particularly difficult. The development of sign language comes about naturally, just like the languages of the deaf community, and each sign language has its own rules [1]. At times, it occurs that parents

DOI: https://doi.org/10.66279/hk0z2k19

have deaf children, thereby producing a linguistic divide within the family. Additionally, it is difficult to learn sign languages for the deaf because there is no uniform sign language [1-3].
Sign Language (SL) is reputed to be an extremely expressive and natural way for people with poor hearing to engage in socialization and opportunities. Automatic Sign Language Recognition (ASLR) has improved the ability of people with speech and hearing disabilities to communicate and socialize[2, 3]. It is a multidisciplinary research area that involves natural language processing (NLP), computer vision (CV), pattern recognition, and image segmentation. Due to the complexity of the control shapes, ASLR is a general issue. It requires information on the shape, mobility, and direction of the hands [4-6]. By generating voice and text, usable ASLR systems can maximize the autonomy of deaf and hard-of-hearing individuals. The most difficult part of any ASLR system is determining, evaluating, and recognizing.

We discovered that few linguistics research has been done in ArSL and regional Arab SLs after talking to several Arabic deaf people and professionals who interpret sign language. We also discovered that the absence of a transcription or annotation system is the primary issue. Some features of ArSL, including direction, hand shape, movement, location, and facial expressions as non-manual gestures, can be expressed in text or symbol by the annotation writing system[2]. The development of an intelligent machine system that can translate Arabic into ArSL and vice versa, and is able to detect ArSL, can help bridge the gap in communication between deaf and hearing people. It is hoped that such developments would allow the deaf to acquire scientific knowledge in their own language and help them integrate into different educational environments. [7]. Considerable efforts have been made in recent years to standardize a unified sign language in Arab nations, with the documentation and registration of ArSL in the Arabic language. Similar ones have been established in the Gulf States, Egypt, Jordan, Tunisia, and other countries. They seek to propagate this amalgamated language to the deaf and interested members. The problem is precipitated by these attempts, whereby they have contributed to the generation of independent sign languages in every country with scarce concurrence alphabets as well as descriptive and non-descriptive movements. [8].

Hand gesture recognition research began with the application of gloves with multiple sensors and trackers put on [9].While such gloves provided precise data on finger movement and hand position, they necessitate fitting obstructive devices onto the individual's hand. In comparison with vision-based systems, which are provided with a natural environment. But it also carried with it several challenges, including occlusion detection or the detection and segmentation of hands and fingers. Indication-based systems employ indications such as colored gloves for both hands and colored indicators on fingers to address these problems. Despite various findings in the literature, the problem of indicator-free detection and tracking in free environments is still interesting[10].

Research work on ArSLR can be divided into two approaches: Vision-Based Recognition (VBR) and Sensor-Based Recognition (SBR), which form our primary research taxonomy. These two approaches are generally used for the Arabic alphabet, isolated words, and continuous sign language recognition tasks. This forms our secondary research taxonomy. It has been observed that the VBR approach is more widely used in ArSLR than the SBR method. Additionally, there have been many research attempts made on the Arabic alphabet and isolated words sign language recognition versus Arabic continuous sign language recognition [11].

This survey offers an organized, multi-level overview of Arabic Sign Language Recognition (ArSLR), encompassing machine learning and deep learning techniques for recognition at the alphabet, word, and sentence levels. In contrast to previous surveys, this work provides a comparative analysis of several approaches, highlighting their advantages, disadvantages, and performance variations under various input modalities, such as vision-based and sensor-based systems. Additionally, it provides future research routes toward creating reliable and practical ArSLR applications and highlights important research gaps, especially in continuous sign recognition and signer-independent systems.

Major contributions of this paper include:

1. Comprehensive Review: A comprehensive survey of Arabic Sign Language Recognition studies that cover a broad range of techniques.
2. Structured Taxonomy: A well-structured taxonomy of existing research based on input modality, which can be classified into Vision-Based and Sensor-Based, and recognition level, which can be classified into alphabet, isolated words, and continuous sentences.
3. Gap Identification: Identification of significant research gaps in the area of continuous sign recognition, data sets, and multi-modal approaches.
4. Future Research Directions: Highlighting potential areas for future investigation in the field.

The rest of the paper is structured as follows: The methodology is described in Section 2. In Section 3, machine learning and deep learning techniques for sentence, single word, and letter recognition are reviewed, along with an analysis and comparison of the findings. An overview of Arabic Sign Language, including its difficulties, principles, databases, and uses, is given in Section 4. The suggested methodology, including data collection, preprocessing, feature extraction, classification, and learning strategies, is explained in Section 5. The primary difficulties in Arabic Sign Language recognition are highlighted in Section 6. The main findings are covered in Section 7, and future study directions and the conclusions are presented in Sections 8 and 9, respectively.

2 METHODOLOGY

This survey reviews studies on Arabic Sign Language Recognition (ArSLR) using a structured approach. Major databases, such as IEEE Xplore, ScienceDirect, Springer, and Google Scholar, were searched for relevant papers using keywords like "Arabic Sign Language," "ArSLR," "sign language recognition," "machine learning," and "deep learning." Research from 2000 to 2026 was considered. Studies offering novel research on the recognition of Arabic alphabets, isolated words, or continuous sentences utilizing vision-based or sensor-based techniques were the main emphasis of the inclusion criteria. Excluded were studies that did not address ArSLR or lacked experimental validation.

A total of 150 papers were first evaluated throughout the selection process. Excluded were studies that were deemed irrelevant, lacked experimental validation, or did not directly address Arabic Sign Language Recognition. 111 studies were kept and examined after a thorough evaluation, guaranteeing that the survey includes significant advancements in the field while upholding methodological rigor. **Figure 1.** Study selection process for the Arabic Sign Language Recognition (ArSLR) survey.

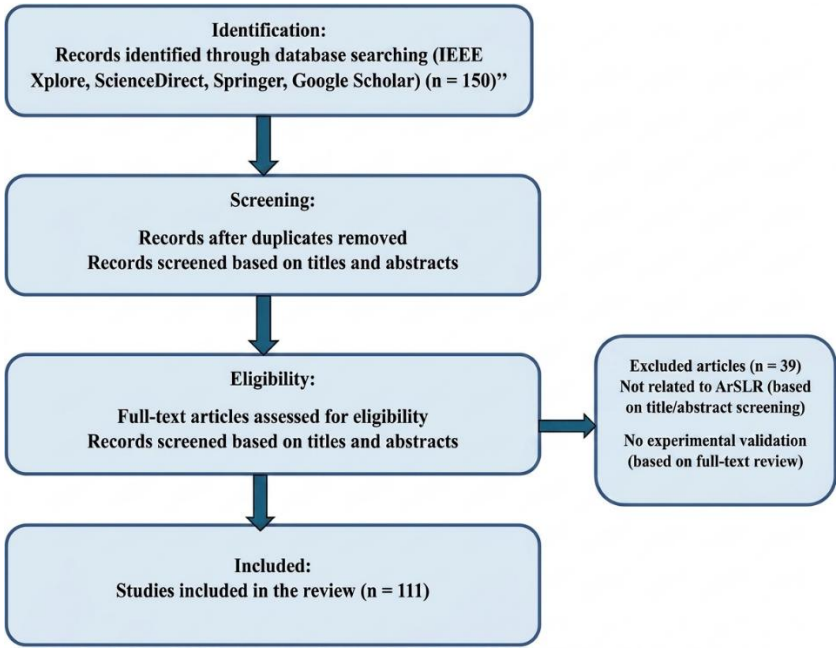

Figure 1. PRISMA flow diagram illustrating the study selection process for the ArSLR survey

3 LITERATURE REVIEW

According to the recognition unit, the relevant work in Arabic Sign Language recognition can be generally divided into three levels: sentence level, word level, and alphabet level. In alphabet-level studies, researchers focus on sign recognition, while in word-level studies, researchers try to recognize a whole word, and in sentence-level studies, researchers try to decipher a sentence. In addition, the research examined is discussed in terms of learning methods used in each study, i.e., transfer learning, machine learning, and deep learning.

3.1 Machine Learning Techniques for Alphabet Recognition

Detection of the ArSL alphabet implies that a sequence of gestures corresponding to the Arabic alphabet letters must be recognized. It is possible to use a vision-based and/or sensor-based approach. One of the methods of detection of the ArSL alphabet is to recognize each letter as a separate gesture and use machine learning algorithms for recognizing the gestures from visual or sensor features. In this section, several approaches to the image-based recognition of Arabic alphabet signs are presented[10].

Authors in [12]. Presented a method for identifying signals in the Arabic alphabet and translating them into spoken language. Though the system is not capable of real-time recognition, it is far more appropriate for the real world. It focuses on easy moveable movement and static gestures. Color images of gestures provide the system with an input signal. The skin patches are segmented based on the YCbCr color model. The shape of the hand is also described by using the Prewitt edge detector once the data has been extracted. The principal component analysis is used to transform images into feature vectors (PCA). In the classification stage, the k-nearest neighbor (KNN) method is used with a success rate of 97Authors in [13] have suggested a real-time ArSL recognition from high-resolution video. Preprocessing operations like skin detection and normalization of size are carried out by the system. Fourier descriptors are used during the feature extraction stage, while KNN is used during the classification stage. The above process has 90.55 percent recognition efficiency.

Authors in [14] develops an automated system that enables the detection of the ArSL letters. SVMs are used for classification, and moment invariants are used for feature extraction. An 87 percent recognition rate is achievable. Authors in [15] Translate hand pictures using a neuro-fuzzy algorithm. The sign image was taken by a PC camera without gloves. With variations in the gesture position, size, and direction in the picture, the model remained uniform. Their work achieves a 93.55% identification rate. To identify letters, Authors in [16] Design a flexible neuro-fuzzy interference system. For simplifying the segmentation process, the hand region can be segmented simply by using a multi-colored glove. A recognition rate of 9.55% is obtained. Authors in [17] Use polynomial classifiers to carry out the alphabet acknowledgement. They learn a new 2nd-order polynomial classifier in each class at training time and use it to generate feature vectors. Training and test error rates are identical: 1.6% on the training data and 6.59% on the test data.

Authors in [1] Suggest a machine learning-based ArSL alphabet automated recognition system. They account for 28 alphabets and 2800 images in total, with ten students per class. They have 100 letters' photos relating to each letter, making 2800. They utilize MLP and KNN for handling classification techniques, whereas feature extraction relies on a hand shape description, where every hand image is represented by a vector with 15 values describing important point locations. The test's accuracy is 97.548%.
As shown in **Table 1.** illustrates some related work for ArSL alphabet recognition.

Table 1. Machine Learning Techniques for Alphabetic Recognition

Reference - Year	Recognition System	Techniques	Input source	Recognition Rate (Accuracy%)	Limitation	Contribution
[14] - 2001		SVM (Manual features)	Images	87%	-Limited dataset size -Evaluation of a controlled environment	Early ArSL work showed that automated recognition was feasible by employing image processing and pattern recognition for Arabic alphabets.

[9] - 2010	Image-based	MLP and MDC (edge-detection and feature-vector creation)	Images of bare hands	91.3%	Dependence on controlled input conditions	developed the ARSLAT system, which uses computer vision techniques to automatically recognize Arabic sign language alphabets.
[12] - 2010		PCA (Prewitt edge detector)	Images	97%	- Reliance on edge-based feature extraction - Limited to static alphabet recognition	proposed an edge-based system that can recognize the Arabic Sign Language alphabet and convert it to voice.
[18] - 2012		KNN (sign histogram for surface behavior detection)	Images	Bare fists have a hit rate of 50%, red gloves up to 75%, black gloves up to 65%, and white gloves up to 80%	Limited scalability and generalization	used the K-Nearest Neighbor (KNN) algorithm to identify hand motions in an Arabic Sign Language recognition system.
[19] - 2020		DBN followed by SoftMax/SVM (DBN-based feature extraction)	image-signs	83.32%	- Sensitivity to environmental conditions - Focus on static hand gestures only	concentrated on employing computer vision techniques to recognize static hand gestures as a foundation for interpreting sign language.

3.2 Machine Learning Techniques for Isolated-Word

In contrast to letter sign identification, single-word recognition is concerned with extracting the important frames from the input video sign [4]. Authors introduce feature extraction techniques for ArSL recognition in user-independent mode [20]. In user hands separation by color segmentation, schemes employ colored gloves. The prediction errors encountered when predicting each successively segmented image from its predecessor are combined to produce two constant images. To maintain motion in the correct direction, the built-up prediction errors are given a directed and weighted bias. User-independent motion information can be eliminated through the computation of the bounding box of the overall prediction errors. Zone-coded discrete cosine transform (DCT) coefficients of the bounded images are employed to generate the feature vectors. KNN and polynomial networks are employed to establish that the suggested approach is reliable. 87% of the users fall under the category of user-independent mode, as per reports. Experiments show that, for feature vector sizes greater than 30, KNN performs superior to 2nd order polynomial networks on numerical grounds.

A new benchmark dataset and technique for sign language recognition are published by the authors in [21]. Along with hand segmentation, the ArSL technique also specifies body motions, classifies signs, and specifies the hand shape sequence. Canonical correlation and RF classifiers are used in sign classification. The technique uses 150 signs, which

were gathered from 21 signers using a Kinect v2 sensor. There are 7,500 samples altogether. Finally, the algorithm's solution on a public data set is state-of-the-art. Identification accuracy, based on 150 ArSL signs, is 55.57%.

Authors in [22] emphasize finding ArSL via the combination of hardware and machine learning methods. For capturing and processing hand motions, the system uses hardware gadgets such as a jump motion controller and a latte panda. KNN and SVM are the two machine learning algorithms used by the system when it is in recognition mode. The algorithms are used by the jump motion controller for processing and identifying the hand motions that it records. Ada-boosting is used to build a more robust and accurate classifier, which enhances the accuracy of these methods. Dynamic time wrapping (DTW) is a direct matching method that the system uses in conjunction with machine learning algorithms. The hand movements recorded are compared and matched to the known patterns of the dataset by DTW. For better accuracy of the system in identifying ArSL gestures, DDW is used along with Ada-boosting. The test dataset consists of 30 hand movements, 10 double-hand movements, and 20 single-hand movements. Experiments show that the DTW approach attains 88% accuracy on single-hand movement detection and 86% on double-hand movement detection. Using Ada-Boosting, the proposed method has 92.3% accuracy on single-hand movement detection and 93% accuracy on double-hand movement detection. The study[23]The authors proposed a system that makes use of Arabic Sign Language postures that correspond to words in the Quran for teaching the Quran to deaf and hard-of-hearing students. The study used 2,054 labeled images for evaluating pose keypoint classifiers such as MLP, SVM, RF, and an image classification ResNet50 model. The study was affected by a single sura, a small dataset, static frame images rather than signatures, but it achieved good results.
As shown in **Table 2.** some works are for isolated word recognition using machine learning.

Table 2. Some works for isolated word recognition using machine learning

Reference - Year	Recognition System	Techniques	Input source	Recognition Rate (Accuracy%)	Limitation	Contribution
[23] - 2026	Image-based	MLP, SVM, RF (Pose keypoint–based features)	2,054 labeled static images (ArSL words – Surah Al-Ikhlas)	High performance (slightly lower than ResNet50)	Dependence on body pose estimation accuracy. - Limited generalization across users and environments .	suggested an automated system that combines body posture categorization with Arabic Sign Language to assist hearing-impaired individuals in learning the Qur'an.
[22] - 2020	Sensor-based	KNN, SVM (filter feature selection)	30 hand gestures	Single-handed: 92.3%; Double-handed: 93%	- Dependence on Leap Motion Controller hardware - Sensitivity to hand positioning and sensor range	Suggested a method for recognizing Arabic Sign Language that uses an AdaBoost classifier and a Leap Motion Controller to record and categorize gestures.
[20] - 2007	Sensor-based	KNN, Polynomial Networks (Zonal-coded DCT coefficients)	Signer wears colored gloves	87%	-Lower accuracy in user-independent mode -Limited feature representation	enhanced generality across various signers by proposing a user-independent Arabic Sign Language recognition system.

[21] - 2018	Video-based	RF (HOG–PCA, CCA, Cov3DJ+)	150 videos	55.57%	-Increased system complexity due to multi-modality. -Dependence on multiple sensors or data sources.	To increase recognition accuracy and robustness, a multi-modality Arabic Sign Language recognition system incorporating various input sources was proposed.
[24] - 2020	Sensor-based	SVM, RF, KNN (sensor-based features)	gesture words	83 % for SVM	Focus on a limited set of dynamic gestures.	carried out a comparative analysis of various models for dynamic gesture recognition in Arabic Sign Language, emphasizing variations in method performance.

3.3 Machine Learning Techniques for Sentence Recognition

The identification of continuous signs is far more complicated than the above two methods. This method is plagued by motion detection, hand tracking, feature extraction, and a large vocabulary. Classification is simple once the respective feature vector has been extracted because of the number of techniques that may be applied, i.e., KNN and HMM[10]. In comparison to the two earlier modes (alphabet and individual words), continuous sign language recognition is harder since it relies on performing and interpreting entire phrases. Continuous sign language recognition is more adequate and efficient in real life. Continuous ASLR systems are more appropriate for the Deaf community and for all those with hearing disabilities. With a high recognition rate, a real-time, available, and reliable ongoing ArSLR should be in place. In continuous sign language recognition, the identification of the extra movements from the transition from a set of signs, recognition, and modeling are regarded as the main challenges[3].

Authors in [25] 23 ArSL two-handed phrases encompass differences in attire, hand size, and camera distance. Log-Gabor, Fourier, and Hartley convert an accumulated image. KNN, MLP, and SVM models are contrasted with Fourier, Hartley, and Log Gabor transforms. KNN, SVM, and MLP classifiers are used in uncontrolled environments. The recognition success rate of SVM is 98.8%. The study reveals that SVM and Hartley transform work well. Authors in [26]. This article is prepared to develop a communication aid through a wearable smart glove that was integrated with bend sensors to predict hand orientation and finger movements for data processing that will be communicated wirelessly while performing machine learning prediction. Models such as SVM-FE models and LSTM from a dynamic set of hand signs were developed for this purpose. An accuracy of 99.6% was found for the LSTM model, and the development of a real-time recognition application showed its effectiveness in practice.

3.4 Deep Learning Techniques for Alphabet Recognition

Authors in [27], proposed a CNN-based vision system for the recognition of Arabic hand sign-based letters to read out the recognized letter in Arabic. A deep learning technique is pursued for automatic recognition of the hand sign letters and reading them out in Arabic. As far as the segmentation method is concerned, this proposal recognizes the Arabic hand sign-based letters with the highest recognition rate of 90%. Arabic language, upon the rendering based on the accepted hand signs of the letters, is an input with outcomes to the speech engine. The output from the speech engine will be the Arabic language in sound format as its output.

Authors in [28] present a new ArSL recognition system with a faster region-based CNN (R-CNN) for detecting and recognizing the alphabet of ArSL. Faster R-CNN is used to perform the extraction and mapping of image features for

detecting the area where the hand is positioned in an image. Normal phone cameras are used to capture 15,360 images of hand gestures on different backgrounds. From the captured images, the proposed model is approximated. The recognition rate of 93% has been achieved from the obtained ArSL image dataset by the combination of the proposed model with the ResNet and VGG-16 approaches.

Authors in [29] have created a mobile app that translates signs into the Arabic alphabet according to the input hand provided to it. Taking RGB-colored images as input, the authors have tried a CNN-based supervised machine learning that recognizes edges and shapes for them. This includes an open camera image from a hand gesture and some of the learned images from internal storage. They have tested the model performance using the TensorFlow package and TensorFlow Lite APIs, which support mobile applications. The accuracy is 72.5% using a new image taken from the camera and 91.1% using a trained image from internal storage.

Authors in [30] employed CNN architecture for ArSL classification. They provide examples of how the SMOTE Oversampling technique improved the data's accuracy. In the case of the dataset for ArSL2018, the highest without oversampling, the new ArSLCNN model achieved a classification accuracy of 97.29% and 96.59%. Every convolution has employed a pooling layer to reduce the size of the feature mapping space. The difficulties of Arabic Sign Language (ArSL) recognition in real-world settings, where elements like background noise, illumination changes, and hand occlusions may impact recognition performance, were examined by the authors in [31]. They presented ASLDetect, a deep learning-based model that combines a U-Net architecture for gesture segmentation with ResNet for feature extraction. The ArASL2018 and ArASL2021 datasets were used to assess the method, which also used several preprocessing approaches, such as image scaling, normalization, and data augmentation. According to experimental results, the suggested model beat several current methods, including ResNet34, T-SignSys, and UrSL-CNN, and achieved excellent recognition accuracy.
Table 3. shows some related work for alphabet recognition using deep learning techniques.

Table 3. Some related work for alphabet using deep learning techniques

Reference - Year	Recognition System	Techniques	Input source	Recognition Rate (Accuracy)	Limitation	Contribution
[32] - 2023	Image-based	CNN, Transfer Learning (VGG, ResNet, MobileNet, Xception, Inception, DenseNet), Vision Transformer (Data Augmentation)	54,000 images	99%	- High computational and memory requirements. - Dependence on large, labeled datasets.	improved feature extraction and classification performance by looking at Vision Transformers and transfer learning techniques for Arabic Sign Language recognition.
[28] - 2021		Faster R-CNN (CNN)	Images	93%	- High computational cost and slower inference. - Limited handling of complex backgrounds or occlusions.	suggested an object detection-based gesture recognition method for Arabic Sign Language utilizing Faster R-CNN.

[33] - 2021		Faster R-CNN (CNN feature maps and filters)	Images of (ArSL2018)	96.59% to 97.29% after using SMOTE	- Focus on static gestures. - Limited generalization to unseen users.	Introduced ArSL-CNN, a model for Arabic Sign Language gesture recognition based on convolutional neural networks that enhances automatic feature learning and classification precision.
[29] - 2022		CNN (TensorFlow & TensorFlow Lite APIs)	Images	91.1% (trained images); 72.5% (new camera images)	- Dependence on mobile device hardware. - Limited gesture recognition scope.	Using Android Studio, a mobile application for Arabic Sign Language detection was created with the goal of enabling real-time communication for people with hearing impairments.
[31] - 2025		ResNet, U-Net-like components	Images	ASLDetect: 99.35%; ResNet34: 99.08%; T-SignSys: 97.92%; UrSL-CNN: 98%	- High computational and memory requirements. - Potential sensitivity to variations in hand position and background	evaluated various models for dynamic gesture recognition in Arabic Sign Language and assessed their efficacy and performance.

3.5 Deep Learning Techniques for Isolated Word Recognition

Authors in [34] provided different frames as input to the developed system, which utilized a 3D CNN in extracting both temporal and spatial features with which to identify 25 ArSL signals in an intelligent edge detector-based approach. Video is fed into the system and divided into frames for scoring and down-sampling purposes. A normalized depth video stream is taken as input to the system, where feature extraction of spatial-temporal inputs is carried out. The 3D deep architecture has been fine in its testing. The new data accuracy is 85%, while the seen data is 98% accurate.

Authors in [35] propose an SLR system using deep learning architectures for a range of tasks, irrespective of the signer. The system utilizes deep recurrent neural networks, hand form feature encoding, and hand semantic segmentation. Hand segmentation is effectively handled by using a state-of-the-art semantic segmentation model, DeepLabv3+, based on spatial pyramid pooling and Resnet-50 as a backbone encoder network. One-layer convolutional SOM learns and extracts hand shape features. The average accuracy of hand semantic segmentation of the proposed system on the Arabic benchmark set is 89.5% using DeepLabv3+, and 69.0% without DeepLabv3+.
Authors in [36] propose a CNN augmented with an attention mechanism to regain the spatial information. Furthermore, they recommend using bio-inspired deep learning methods to extract temporal information, i.e., BI-LSTM architecture. The different conditions used in testing this model include the variations in lighting conditions, clothes, and distances from the camera. This model learns faster because of the smaller number of deep learning layers and parameters. This model achieves an accuracy of 85.6%, 86.6%, and 95.8%, respectively, on the 79-sign ArSL dataset, the 86-sign NVIDIA gesture database, and the 25-sign Jester dataset. Bi-directional model performance is

superior to that of LSTMs. in the study[23]The authors presented a system that employs Arabic Sign Language postures that represent words from the Quran for teaching the Quran to deaf and hard-of-hearing students. The study employed 2,054 labeled images for evaluating pose keypoint classifiers such as MLP, SVM, RF, and image classification ResNet50 models. The study was affected by a single sura, a small dataset, static frame images rather than signatures, but good results were obtained. **Table 4.** shows some related work for isolated words recognition using deep learning.

Table 4. Some related work for isolated word recognition using deep learning techniques

Reference – Year	Recognition System	Techniques	Input source	Recognition Rate (Accuracy%)	Limitation	Contribution
[23] - 2026	Image-based	ResNet50 (ResNet50 image-based features)	2,054 labeled static images (ArSL words – Surah Al-Ikhlas)	≈100% (near perfect word classification)	- Reliance on accurate body pose estimation. - Limited generalization across users and environments.	suggested an automated framework that combines Arabic Sign Language and body position classification to assist hearing-impaired people in learning the Qur'an.
[37] - 2024	Video-based	Fully connected + Softmax (MediaPipe pose estimator)	100 videos	99.74% (signer-dependent); 68.2% (signer-independent)	Dependence on accurate landmark keypoint detection	enhanced feature representation and classification performance by proposing a transformer-based model for isolated Arabic Sign Language identification using landmark keypoints.
[38] - 2023	Video-based	CNN + RNN (Double CNNs)	8,467 videos	98% (testing); 92% (validation)	- High dependency on computational resources. - Potential variability in performance across users.	improved generalization across various signers by proposing a vision-based deep learning method for interpreting Arabic Sign Language with user-independent recognition capacity.
[36] - 2021	Video-based	CNN + BiLSTM (CNN-based features)	NVIDIA Gesture, Jester, and ArSL datasets	ArSL: 85.6%; Jester: 95.8%; NVIDIA: 86.6%	Sensitivity to gesture variations and background conditions.	proposed a sophisticated real-time Arabic Sign Language classification system that successfully captures spatial and temporal variables by combining BiLSTM with an attention-

						based Inception network.
[39] - 2021	Video-based	LSTM + Softmax (CNN features)	6748 videos	72.4% (signer-independent); 99.7% (signer-dependent)	- Complexity of multimodal integration. - Limited dataset size and diversity.	presented the MArSL database and a hybrid multimodal method that combines non-manual and manual elements to recognize Arabic Sign Language.

3.6 Deep Learning Techniques for Sentence Recognition

Authors in [40] utilize motion and spatial information in cascaded design of CNN and LSTM models to provide novelty for sign language gesture recognition. To this extent, the key challenge is how to model the temporal dynamics and spatial information of the gestures efficiently to capture and integrate such dynamics for enhancing SLR performance. As will be shown with experiments on benchmark datasets in the present paper, the work is better than state-of-the-art and therefore can potentially find its way into real-world use in true instances of SLR. The proposed approach attained over 99% identification accuracy and outperformed competing approaches.

Authors in [41]. Recent research in Arabic sign language recognition has focused on the complexity of recognizing continuous sentences, which was sidestepped by earlier research using isolated signs. A new two-stage deep learning approach is introduced that creates motion images and uses motion compensation algorithms to predict the words in a phrase. Continuous sign language recognition issue was resolved using the system provided, which outperformed the word recognition rate as 97.3% and sentence recognition rate as 92.6%, using biLSTM layers for classification. Results are radical improvements over the compared methods in comparison to their past versions.

Authors in [42] The related work in sign language recognition introduces the ASODCAE-SLR model that employs Atom Search Optimization to perform hyperparameter tuning, Capsule Networks, and Deep Convolutional Autoencoders. The model achieves accuracy above 99% on the Arabic Sign Language dataset through weighted average filtering on preprocessed input frames. The model is better compared to the existing techniques, that is, GRU-LSTM and RNN, thus making it efficient for real-time use that would better facilitate the communication of speech as well as hearing-impaired patients.

Authors in [43]Introduces a new multi-modality dataset and benchmark for continuous ArSL recognition. The dataset was created to help researchers in the Arabic-speaking world overcome the challenges of developing improved algorithms in sign language recognition. There are three types of data for each syllable: color, depth, and skeleton joint points captured by a Kinect v2 camera. He has also published several models, including encoder-decoder and attention models, for sign language recognition. The sentences are taken as input to the suggested techniques after separating the spatial details from them through two pre-trained approaches. The top performing model also achieved a WER of 0.50, which is extremely low and lays broad paths for future improvement.

M. Saied Abdel-Wahab et al. [44] Use an ANN model for the recognition stage after breaking down continuous motions into more static postures. They display the gesture sequence as a graph. The graph matching approach oversees gesture recognition. 90.5% is the final recognition rate for the test set. To recognize continuous sentences in ArSL, Tolba and Abul-Ela [45] introduce a novel graph matching technique. They develop an approach that makes use of connected sequence gesture recognition. For thirty consecutive sentences composed of one hundred gestures, the recognition accuracy is more than 70%. The assessment shows that using this method produces some encouraging outcomes. **Table 5.** presents some related work for sentences.

Table 5. Some related work for sentences using deep learning techniques

Reference - Year	Recognition System	Techniques	Input source	Recognition Rate (Accuracy %)	Limitation	Contribution
[40] - 2022	Video-based	CNN + LSTM (spatio-temporal and motion features)	Videos	99%	- Lack of Non-Manual Features. - Limited Dataset	proposed a cascaded CNN–LSTM framework that improves temporal modeling performance by using motion and spatial information for sign language gesture identification.
[43] - 2023	Video-based and sensor-based	Encoder–decoder model (Pretrained CNN)	sentence by Kinect	WER of 0.50	- High Computational Cost. - Lack of Real-Time Implementation	ArabSign, a multi-modality dataset and benchmark for continuous Arabic Sign Language recognition, was introduced to help studies on continuous sign understanding in the real world.
[46] - 2024	Image-based	CNN, LSTM (CNN, LSTM features)	20 different words, resulting in 4000 images	CNN: 94.40%; LSTM: 82.70%	- Real-Time Constraints Vs Accuracy Trade-off. - Limited Generalization.	suggested a mixed deep learning model to recognize Arabic Sign Language in real time, increasing the precision and effectiveness of gesture classification.
[47] - 2024	Image-based	Transfer Learning (CNN)	The ASL-DS-I, ASL-DS-II, ASL-DS-III	ASL-DS-I: 96.25%; ASL-DS-II: 95.85%; ASL-DS-III: 97.02%	- Lack of Temporal Modeling. - Sensitivity to Background and Lighting.	suggested a CNN-based method for Arabic Sign Language identification that is improved by transfer learning to increase feature extraction and classification accuracy.

3.7 Results and Analysis

The literature on Arabic Sign Language Recognition (ArSLR) is critically analyzed in this part. Accuracy outcomes are compared with several methods, such as vision-based and sensor-based systems, machine learning, and deep learning. Important elements influencing performance are covered for each recognition level: alphabet, isolated words, and continuous signs, including input conditions, feature extraction techniques, and model selection. The study provides a cohesive picture of the field rather than just descriptive reporting by highlighting the reasons why some methods perform better than others, pointing out the shortcomings of each category, and highlighting unexplored regions.

3.7.1 Alphabet Recognition – Machine Learning Approaches

The recognition of the Arabic Sign Language (ArSL) alphabet has been extensively studied using vision-based methods, in which each letter is identified using machine learning techniques and treated as separate gestures.

According to [10] Most early systems used pattern recognition algorithms for categorization after collecting visual information from images. Conventional methods mostly rely on classifiers like KNN, SVM, and MLP in conjunction with manually created feature extraction techniques, including edge detection, shape descriptors, and statistical transformations. These approaches attained comparatively good accuracy, as noted in [12-14], but their effectiveness is heavily reliant on controlled conditions and predetermined feature representations.

Additionally, to increase resilience against changes in hand position and orientation, several studies, including [15, 16], used clever strategies such as neuro-fuzzy systems. However, these methods frequently need extra limitations, like the usage of colored gloves or simplified backgrounds, which restrict their usefulness in practical situations. Furthermore, a lot of research ignores the dynamic aspect of sign language in favor of static gesture recognition, as noted in [19]. Although increasingly sophisticated classifiers and feature representations have brought about considerable gains, these systems' scalability is still constrained, especially when working with bigger vocabularies or heterogeneous datasets [18].

Deep learning-based methods, which seek to automatically extract discriminative characteristics from data, have been the focus of recent advancements. Deep architectures like Deep Belief Networks (DBN) have demonstrated encouraging outcomes. Nevertheless, these approaches continue to face difficulties like sensitivity to environmental factors, a lack of diversity in datasets, and inadequate generalization to practical applications.

3.7.2 Isolated Word Recognition – Machine Learning Approaches

Arabic Sign Language (ArSL) isolated word recognition, as opposed to alphabet-level recognition, concentrates on deriving significant temporal information from video sequences, where precise interpretation depends on the identification of crucial frames and motion patterns [4]. Because both spatial and temporal aspects are involved, the task is more complicated, as documented in the literature. To lessen reliance on particular signers, user-independent recognition has been the subject of numerous studies. Motion-based feature extraction methods, such as zonal-coded DCT coefficients and prediction error accumulation, have been used to extract dynamic information, as stated in [20]. Nevertheless, these methods frequently depend on regulated configurations, like colored gloves, which restrict their usefulness in practical settings.

Additionally, multimodal methods have been developed to improve recognition performance. Combining hand shape, motion, and body movement data using various feature extraction and classification algorithms enhances representation capabilities, as stated in [21]. However, when applied to bigger and more varied vocabularies, these systems often suffer from increased complexity and relatively low accuracy. Furthermore, there have been proposals for hybrid systems that combine machine learning algorithms with hardware. According to [22], Sensor-based methods that integrate classifiers like KNN and SVM with devices like Leap Motion controllers can reach comparatively high accuracy. Nevertheless, their reliance on specialized hardware limits practical deployment and decreases scalability.

Additionally, some recent research focuses on word-level recognition using pose-based and image-based representations. Pose keypoint-based techniques work well when paired with classifiers like MLP, SVM, and RF, as was covered in [23]. Nevertheless, these methods continue to be limited by small datasets, a dependence on static frames, and a lack of generalizability across various environments and users. In general, isolated word recognition systems have progressed from manually designed motion-based methods to multimodal and hybrid approaches, as noted in [20, 24]. Notwithstanding these developments, real-world deployment is still severely hampered by issues including reliance on controlled environments or hardware, a lack of diversity in datasets, excessive system complexity, and inadequate generalization.

3.7.3 Sentences Recognition – Machine Learning Approaches

Since continuous Arabic Sign Language (ArSL) recognition requires reading entire sentences rather than individual gestures, it is far more difficult than alphabet and isolated word identification. According to [10], A variety of machine learning methods, including KNN and Hidden Markov Models (HMM), can be used for categorization after feature vectors are retrieved. However, effectively capturing temporal connections and segmenting continuous sign streams are more difficult than categorization itself. Continuous recognition is more appropriate for real-world communication settings than previous recognition levels since it requires handling motion detection, hand tracking, and a large

vocabulary [3]. One of the main issues is the existence of transitional movements between signs, which makes segmentation more difficult and reduces identification accuracy, as the literature has pointed out.

Numerous strategies have been tried to tackle these issues by utilizing various extraction and categorization methods. According to [25], even in somewhat unrestricted contexts, high recognition rates can be attained by combining classifiers like KNN, SVM, and MLP with frequency-based transformations like Log-Gabor, Fourier, and Hartley. Similarly, as noted in [26]The modeling of temporal sequences is greatly improved, and recognition accuracy is increased when sensor-based data from wearable devices is integrated with machine learning models, especially LSTM networks. Continuous ArSL recognition systems still have a number of drawbacks despite these developments. High accuracy frequently requires controlled settings or specialized gear, including wearable gloves. Furthermore, maintaining reliable segmentation of continuous motions and managing extensive vocabularies continue to be difficult problems.

3.7.4 Alphabet Recognition – Deep Learning Approaches

Arabic Sign Language (ArSL) letter recognition has greatly evolved as a result of deep learning, which provides automatic feature extraction and better classification performance than conventional techniques. Convolutional neural networks (CNNs) and their variations, such as region-based CNN (R-CNN), ResNet, VGG, and U-Net architectures, are frequently used to recognize hand movements that correlate to Arabic characters, as documented in [27, 31].

To reduce class imbalance, recent research uses large-scale picture datasets to train deep models, which are occasionally improved with data augmentation or oversampling techniques, such as SMOTE [30, 32]. According to the research, these methods attain high recognition rates, with claimed accuracy typically surpassing 90% and reaching up to 99% when paired with transfer learning or vision transformers [31, 32]. Furthermore, several techniques have been modified for mobile apps utilizing frameworks like TensorFlow Lite, allowing for useful deployment on devices with limited resources [29].

Despite these developments, there are still several issues with deep learning-based ArSL alphabet recognition
These include:

- high memory and processing demand, especially for huge datasets and complicated systems.
- dependence on sizable, labeled datasets for efficient training, which restricts flexibility in situations with limited resources.
- Sensitivity to changes in illumination, backdrops, and hand occlusions can diminish resilience in real-world settings.
- Particularly when models are trained on datasets, there is little generalization to users who have not been observed.
- Concentrate on static motions and use fewer methods for dynamic or continuous sequences.

3.7.5 Isolated Word Recognition – Deep Learning Approaches

The goal of deep learning techniques for isolated Arabic Sign Language (ArSL) word recognition is to extract temporal and spatial information from pictures or video sequences. To extract spatiotemporal information and increase classification accuracy, 3D CNNs, recurrent networks, and attention-augmented architectures are frequently employed, as documented in [34, 36]. To improve feature extraction, a number of methods use semantic hand segmentation. For instance, as noted in [36], DeepLabv3+ and ResNet-50 successfully segment hand regions, allowing for increased identification accuracy even in signer-independent settings. In a similar vein, dynamic movements across frames are captured by modeling temporal relationships using attention processes and Bi-LSTM networks [36]. These methods lessen the impact of changes in lighting, attire, and camera distance. Furthermore, other techniques rely on pose-based attributes that are taken from the body's and hands' keypoints. Pose-based deep learning models, such as ResNet50 or MediaPipe estimators, can achieve near-perfect recognition on tiny, controlled datasets, as demonstrated in [23, 37]. The difficulty of generalization is shown by the fact that their performance declines in significantly independent conditions or when applied to more varied datasets. The majority of studies claim great accuracy, although there are still several limitations, such as:

- reliance on precise assessment of hand or body position.
- restricted environmental and user generalization, especially for signer-independent recognition.
- high memory and processing demand, particularly for CNN-RNN or Bi-LSTM models.
- sensitivity to differences in dataset size, background conditions, and gestures.

3.7.6 Sentences Recognition – Deep Learning Approaches

Because it requires collecting temporal dynamics, motion transitions, and contextual information across sequences of gestures, continuous or sentence-level Arabic Sign Language (ArSL) recognition is more difficult than alphabet or isolated word recognition [40]. To successfully model spatiotemporal dependencies, deep learning architectures that combine CNNs with recurrent models (LSTM or Bi-LSTM) have been widely used, according to the literature[40, 41]. Two-stage and multi-modal methods to improve performance are the subject of recent research. For example, sentence-level recognition rates have been significantly increased above word-level recognition rates with the use of motion-compensated pictures and biLSTM layers [41]. Similarly, hybrid models that combine deep convolutional autoencoders and capsule networks with optimization methods (such as Atom Search Optimization) have demonstrated great accuracy (over 99%) for real-time gesture detection [42].

The use of multi-modal datasets offers more detailed information for sentence recognition. Encoder-decoder and attention-based models can achieve exceptionally low word error rates (WER = 0.50) using datasets that combine color images, depth maps, and skeleton joint points recorded by Kinect sensors, as shown in [43], indicating the possibility for further research. Despite fewer datasets, graph-based techniques have also been put forth to describe gesture sequences as structured data, with encouraging recognition rates [44, 45].

Continuous ArSL recognition still has several drawbacks despite these developments:

- Model generalization is limited by the scarcity of large and varied datasets.
- high computational expense, especially for encoder-decoder or multi-modal models.
- difficulties with real-time deployment, necessitating compromises between inference speed and accuracy.
- restricted control over non-manual characteristics, such as body posture and facial expressions.
- Robustness may be impacted by sensitivity to background, illumination, and temporal fluctuations.

3.8 Comparative Discussion and Insights

The advantages and disadvantages of machine learning and deep learning methods for Arabic Sign Language Recognition (ArSLR) at various recognition levels are discussed in the preceding subsections. For the recognition of alphabets and isolated words, machine learning techniques typically work effectively, especially when features are thoughtfully created and environmental factors are managed. However, their inadequate ability to describe temporal dependencies, gesture transitions, and coarticulation effects makes it difficult for them to recognize continuous sentences.

With the use of automatic hierarchical features learning and temporal modeling with architectures like CNNs, Bi-LSTMs, encoder-decoder networks, and attention mechanisms, deep learning techniques greatly increase recognition accuracy. Robustness to environmental changes, signer differences, and occlusions is further improved by multi-modal inputs, which combine color, depth, and skeletal joint data. Despite these advancements, several issues still need to be resolved, including real-time deployment in real-world scenarios, scalability to huge vocabularies, and signer-independent generalization. In summary, the comparative analysis reveals that:

1. In machine learning, automatic feature extraction using deep learning is superior to manual feature creation.
2. Temporal modeling is essential because sequential dependencies that classical machine learning is unable to effectively capture are necessary for continuous recognition.
3. Robustness is provided via multimodal and hybrid models: When vision-based, sensor-based, and attention-enhanced architectures are combined, real-world performance is improved.
4. Evaluation frameworks are inconsistent: To accurately evaluate approaches, a single benchmark and standardized measurements are required.

Overall, the report shows how ArSLR research has advanced, offering a comprehensive grasp of the area, pointing out important obstacles, and placing the chances for developing fully functional and useful solutions for the Deaf population in context.

4 SIGN LANGUAGE RECOGNITION: CONCEPTS, CHALLENGES, TECHNIQUES, AND DATASETS

The most common form of communication for the deaf and hearing-impaired is Sign Language (SL) [48]. They employ it as their mother tongue, and their number is roughly 18 million in the universe of Arabic signers. There are no norms and grammatical structure necessary to generate an ArSL sentence, even though there exists a significant number of deaf persons[2].

4.1 Challenges in the Arabic Language

Due to its problems and sociopolitical significance, the Arabic language is greatly sought after by the NLP community. Its problems include its rich morphology, dialect variation, opaque orthography, and diglossia [49]. Arabic is therefore one of the most difficult natural languages, particularly when machine learning is applied. The following factors contribute to the Arabic language's intricacy, as shown in **Figure 2**.

- The three diverse varieties of Arabic language, modern standard, classical Arabic, and dialectal Arabic, each possess unique qualities [3, 50-52].
- The potential for variations in letter order based on letter placement due to the absence of some letters for representing short vowels [3, 53].
- Depending on where a letter appears in a word at the start, middle, end, or standalone, there are multiple ways to write an Arabic letter, which can produce two or more forms[54].
- Many NLP applications, such as POS tagging, NER (named entity recognition), parsing, tokenization, and many more, are significantly impacted by the lack of capitalization and the irregular and inconsistent use of punctuation [49, 53].
- The creation of Arabic words entails suffix and prefix concatenation onto the word stem based on various linguistic rules. This creates structure limitation in analysis and root recovery, especially in computerization and linguistic theory [3] [49].

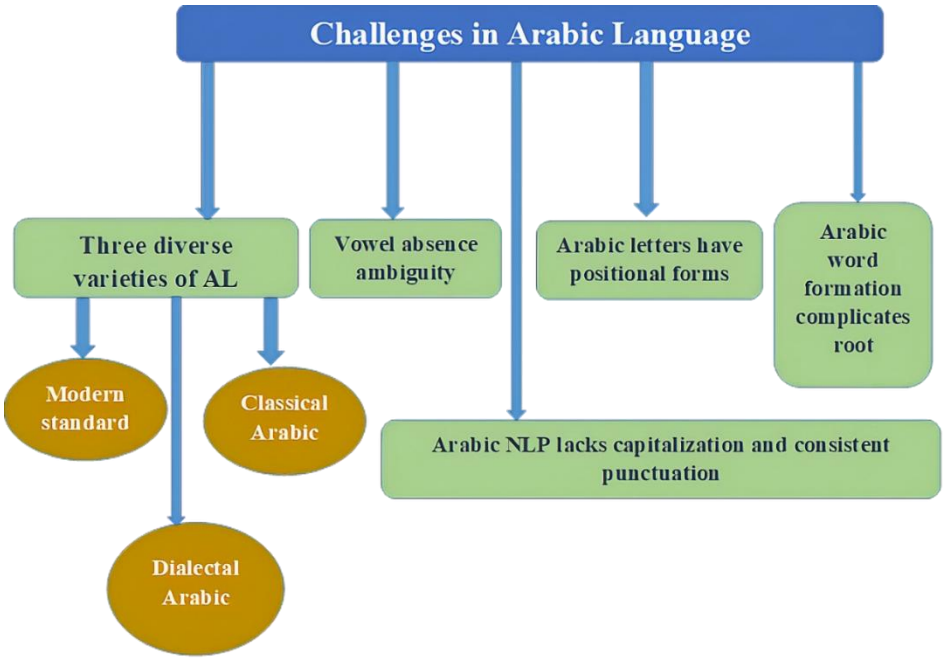

Figure 2. Key challenges in Arabic language processing include dialects, vowel ambiguity, positional letter forms, complex word formation, and punctuation issues.

4.2 Challenges in Arabic Sign Language Recognition

As sign language is not an international language, it may differ depending on the country or even the region. There have been many attempts to develop and standardize sign language in the Gulf States' Arabic nations, Jordan, Tunisia, Algeria, Iraq, Palestine, Syria, Sudan, and Djibouti to make it more widely used by the deaf population living in those

countries. Therefore, in all the Arabic-speaking nations, various sign languages have been developed with similar sign letters [9, 55].

The primary objective of opening hard-of-hearing schools in Arab nations is the availability of appropriate environments for deaf children. A single sign (gesture) is formed by combining motions, palm orientations, and specific, recognizable hand shapes. Sign language is also regarded as a native language among the hearing handicapped and the deaf. ArSL is regarded as an official language among deaf individuals in Arab nations [3]. Examples of the challenge issues that ArSL scholars typically run into when translating into Arabic Sign Language are given below [8, 56] and as shown in **Figure 3**.

- Absence of research on Arabic sign language morphology, syntax, and language.
- The enormous volume of translations is used in the process of developing an Arabic sign language translation method.
- The difficulty in displaying the results of an Arabic sign language translation.
- The challenge of locating a technique for assessing Arabic sign language translation results.

Arabic sign language is a spatial and gestural one, much like other sign languages. However, the Arabic sign language morphology, syntax, and linguistic structure do not align with those of the spoken Arabic language. As spoken languages lack a proper representation for notions in these languages, their corresponding sign counterparts turn out to be difficult to relate to, as with other languages.

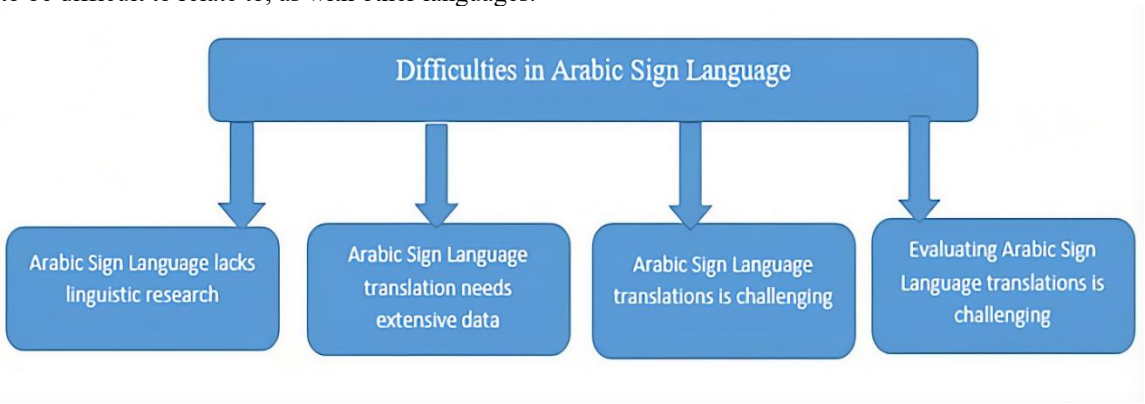

Figure 3 Key challenges in Arabic Sign language processing

4.3 Fundamentals of Sign Language Recognition

Static movements, which incorporate hand postures and positions, and dynamic movements, which consist of hand motions of a defined nature, such as waving, in a precise space and moment without causing any movement, tend to be compensated for by hand position and alignment. One also uses one orientation of the hand with no motion for static motion [57] [3].

A dynamic gesture uses ongoing sequences of signs from a movie as input, whereas a static gesture uses isolated frames of signals. Besides, the hand posture does not change during the gesture, whereas the static gesture typically depends on the fingers' form and angular position. With a dynamic movement. Conversely, the hand position in a dynamic hand gesture is constantly changing with respect to time, and the order of stroke phases comprises the message. The three phases of movement in the dynamic gesture are preparation, retraction, and the stroke phase[58] ,**Figure 4.** illustrates the static and dynamic gestures.

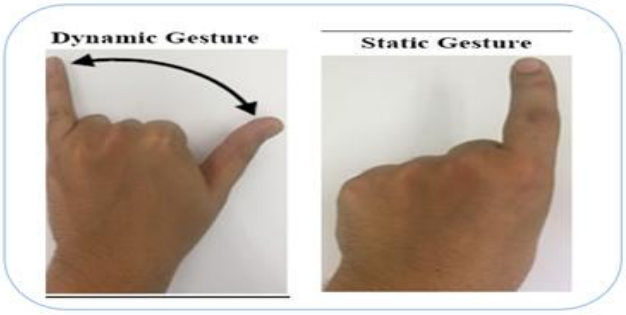

Figure 4. Example for static and dynamic gesture

4.4 Hand gesture recognition approaches

The ability of hand gestures to enable communication and offer a natural form of interaction that can be readily employed in a wide range of applications makes them an exciting area of research. Wearable sensors mounted on the hand via gloves were previously used to recognize hand gestures. These sensors used hand gestures or finger bending to detect a physiological response. The information that was gathered was then processed through a computer that was interfaced with the glove. This glove sensor system had the potential to be made portable by interfacing a sensor to a microcontroller [59].

- **Data glove-based approaches:**

In the data glove-based approach, a glove-like device that can detect hand position, hand movement, and finger flexion is used. In this method, the user puts on a device resembling a glove that employs sensors to monitor the movement of his/her hand or fingers and send the data to the computer[60].As shown in **Figure 5**.

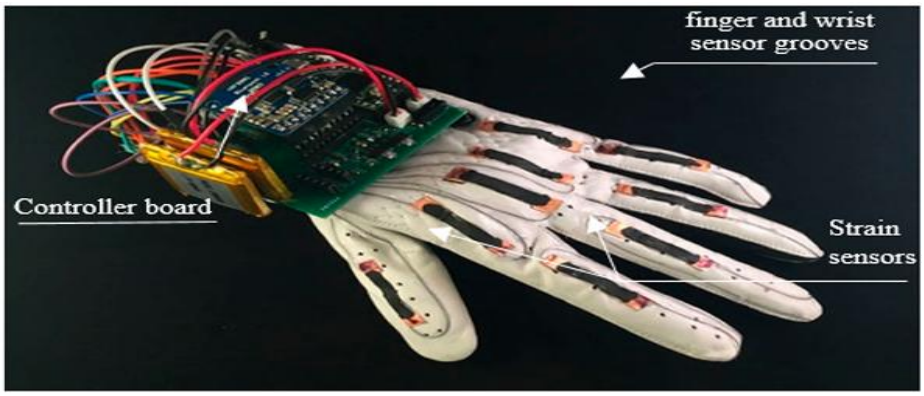

Figure 5. Sensor-based data glove from Site: Smart glove translates sign language into digital text – Physics World

- **Color glove-based approaches**

As seen in **Figure 6**. [61], this technique employs a camera to observe the hand motion while wearing a glove with colored markings. The technique has been employed in interaction with 3D models with some processing such as zoom, movement, sketching, and typing using a virtual keyboard offering good flexibility [62]. The camera sensor can see and identify the location of the palm and fingers because of the colors on the glove and thus can recover a geometric model of the shape of the hand.

Figure 6. Color-based recognition using a glove marker

- **Vision-based approaches**

By employing this method, the user does not need to wear anything. To facilitate human-computer interaction, the system requires one or more cameras to record hand images. By employing a vision-based method, it is convenient, intuitive, and easy [63]. Lamination variations, background noise, partial or complete occlusion, and other issues still must be resolved.

- **Appearance-based methods**

Sometimes referred to as view-based methods, define the movements as a series of views and model the hand using the intensity of 2D images. Using a template database, these models obtain the parameters straight from the photos or videos, eliminating the need for a spatial representation of the body.
Some of them are based on the hands' malleable 2D templates. Because it is simpler to extract information from 2D images, appearance-based approaches are thought to be simpler than 3D model approaches[60].As shown in **Figure 7.** hand gesture recognition approaches.

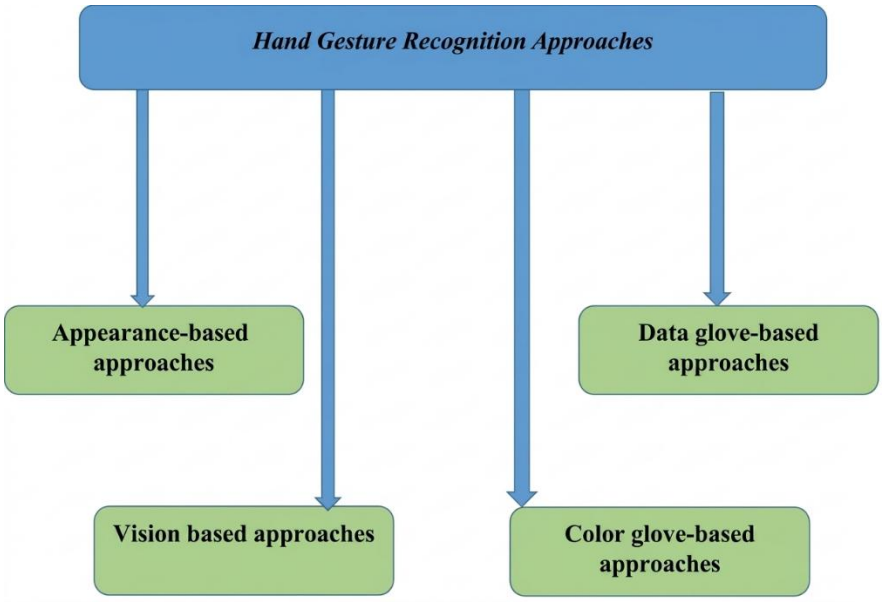

Figure 7. Hand Gesture Recognition Approaches

4.5 Other sign languages

It is necessary to have an extensive corpus of signs for a particular standard language before embarking on the design and development of any ASLR system. Any ASLR system cannot be developed without adequate amounts of prepared and usable training and test data. Superior databases from all over the world are listed alphabetically in this section. As many standard SL corpora as available from all around the globe are included. These corpora were created to support machine-aided ASLR system development. **Table 6** lists the primary standard SL corpora in each nation[3].

Table 6. corpus of some Sign Languages

Reference /year	Sign language	SL Code	Dataset Name	Description
[3, 64]/ 2020-2002	American Sign Language	ASL	The Purdue RVL-SLLL Database	RVL-SLLL Database is a vast database of ASL motion primitives, gestures, words and sentences. It was recorded by fourteen singers. The database contains 2576 clips of 39 motion primitives, 62 hand shapes, and sentences.
[65]/2013	British Sign Language	BSL	British Sign Language Corpus	The BSL Corpus consists of videos of dialogues of two hundred forty-nine participants. The corpus includes annotations of 6330 signers' gestures from the conversational dataset.
[66]/2012	Italian Sign Language.	LIS	The A3LIS-147	From LIS. DB is split into 6 groups, depending on various daily life circumstances. This DB has been conducted by ten signers.
[67]/2008	Pakistani Sign	PSL	Pakistani Sign Language Database	A small database. It has thirty-seven signs. These gestures are developed from Urdu. Signs in the corpus are the fingerspelling alphabet.
[68]/ 2007	Turkish Sign Language	TSL/TSM	The Buhmap	The BUHMAP consists of one hundred thirty-two video segments of eight dynamic gestures. The DB is performed by eleven signers.

4.6 ARSL Datasets

High-variability and large-sized datasets that represent sign language gesture diversity and richness are required to design and test ArSL recognition systems. Those datasets are normally in the form of sensor or video recordings of sign language gestures, along with the annotations explaining the meaning or label of a particular gesture. The unannotated character and the unsegmented symbols make the video data, although online available, unsuitable for training ArSL recognition models. ArSL recognition systems are faced with the fact that there is no large-scale benchmarking dataset. It is difficult to find a full dataset that satisfies the requirements for ArSL recognition. Two of the factors contributing to this are the lack of certified ArSL professionals and the cost and time it takes to gather sign language data. [69] In **Table 7,** some Arabic Sign language datasets.

Future work should concentrate on developing extensive, annotated, and standardized Arabic Sign Language corpora that represent the variety of regional dialects and signer variants to overcome the shortcomings of current datasets. Research institutions can reduce individual costs and effort while collecting high-quality data through collaborative activities. To record hand gestures, face expressions, and body movements, these datasets may include recordings from RGB cameras, depth sensors, motion capture systems, or wearable technology. The creation of such extensive resources will lay the groundwork for the advancement of Arabic Sign Language recognition research and is necessary for training reliable models capable of real-time, multimodal recognition.

Table 7. Some Datasets of Arabic Sign Languages

Reference /Year	Dataset Description
[70] - 2023	It consists of 7,856 RGB images of the ArSL alphabet. Samples are taken from over 200 people under a wide range of shooting conditions (such as, but not limited to, lighting, background, orientation, size, and resolution of images).
[71]- 2021	There are 220000 images in the data set, spread across 44 unique classes (32 letters, 11 digits ranging from 0 to 10, and 1). There are 5000 images in total, all of which have been captured by different people, of each of the fixed signs.
[69] - 2021	There are eleven chapters with 502 signs making up the words in the ArSL lexicon. Three signers work for each sign. There are 75300 samples in total, which are a result of 50 repetitions of each sign by each signer.
[1] - 2021	It consists of 9240 images of the Arabic alphabet from 10 locations and age groups. The images are categorized into four different datasets.
[72] - 2020	44 signs (29 one-handed and 15 two-handed) are signed by a 5-signer group, with 80% being assigned for training and 20% being assigned for testing.

4.6.1 Limitations of the Existing Datasets:

Although these datasets offer useful tools for Arabic Sign Language identification, there are a few drawbacks. The small number of signers in most datasets may limit the model's ability to generalize to new users. Some datasets lack continuous sentence-level gestures and solely concentrate on alphabets or individual signs. Variations in background, lighting, and recording settings are frequently insufficient to capture the richness of the real world. Furthermore, a lot of datasets are small-scale in comparison to the demands of contemporary deep learning, and multimodal data like body position or facial expressions is rarely included. To develop reliable and thorough ArSL recognition systems, it is imperative to address these constraints.

4.7 Application based on Sign Language Recognition

1. Virtual Reality Perhaps the most prevalent computer gestures are those for virtual and augmented reality. Virtual reality interaction employs gestures to facilitate realistic hand manipulations of virtual objects, either for 2D presentations that mimic 3D interaction or for full 3D virtual worlds. [73] or 3D display interactions [74] [75].
2. Desktop and Tablet PC apps: Gestures can be used as a substitute for the keyboard and mouse in desktop computing software. A lot of gestures for desktop computing tasks include editing and annotating documents with pen-based gestures or modifying visuals [76].
3. A good example of a communicative gesture is sign language. Sign languages are ideal for vision algorithm testing since they have very high structural complexity[77]. Nevertheless, they can also be an effective means of assisting those with disabilities with disabilities by using computers.

5 FRAMEWORK OF ARABIC SIGN LANGUAGE RECOGNITION ARSL

A system that uses computer vision and machine learning techniques to recognize and understand hand gestures executed in ArSL is called an ArSL recognition system [78]. As seen in **Figure 8**, the framework usually comprises several stages, including data collection, preprocessing, feature extraction, and classification. As seen in Figure 8 [10]ASLR consists of four main phases to identify the appropriate gestures: 1) data acquisition, 2) pre-processing and segmentation, 3) feature extraction, and 4) classification. Another representation of the SLR and ArSLR process model or cycle model phases is presented in [3, 48]. The image is divided to recognize the hand location in each of the body parts, starting from the background, after the hand image has been taken with a suitable input device. The location image is then pre-processed to remove noise, recognize features, and provide a proper model. Once more, after pre-processing of the images and movements, the feature extraction process begins.

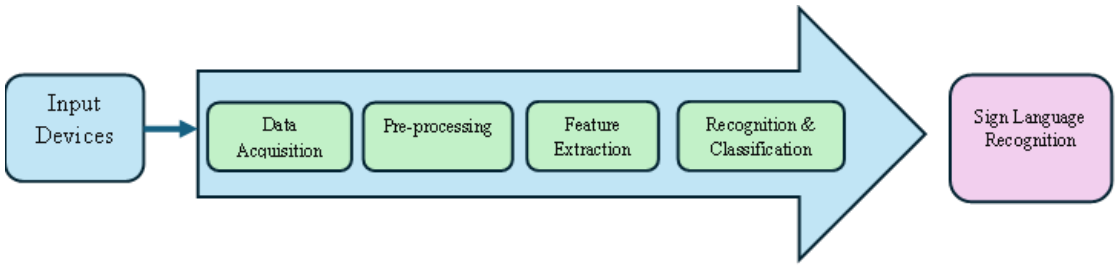

Figure 8. Framework of Arabic Sign Language Recognition

5.1 Data Acquisition Phase

The camera is a crucial component of the sign language recognition method (SLR), which is employed as the input method. A moving picture that is simple for a camera to record serves as the SLR's input data. However, some researchers take pictures with regular cameras [79, 80]. To minimize the hassle of using sensor-based gloves, some researchers say that they are using cameras instead of gloves. Since most cameras offer a variety of video formats, we must use a Digitizer Configuration Format (DCF) file to specify the default format as well as the preferred format. Since the image captured by the webcam is fuzzy, some researchers have used cameras of higher quality. Real-time 30 frames-per-second video was recorded using a camera, and every frame was analyzed one by one to look for dynamic gestures. The area of skin is isolated by the system with the assistance of a skin filter, and each frame of the image is then transformed to the HSV color space[81].

Additionally, there is a different gadget called Microsoft Kinect [82, 83] that are being used in the process of image capturing. Kinect is used nowadays by researchers because of its features. Kinect can record both color and depth video streams at the same time. From the depth data of depth, segmentation of the background is a simple process that can be achieved using Kinect with signal language recognition.

5.1.1 Sensor-based approach

Sensor-based recognition method processes data obtained from Smart Gloves, which is sensor-based. Power-Glove, Data-Glove, and Cyber-Glove have been used primarily for ArSLR. According to data obtained from smart gloves, vast numbers of features can be extracted, such as finger flexing, hand orientation, movement, rotation, and position. Additionally, the classification algorithm uses those features in the detection and identification of the optimal sign[3]. Here are some key points focused specifically on **sensor-based** sign language recognition systems:

1. **Sensor-Based Systems**: These systems rely on sensors embedded in devices like smart gloves to detect hand movements and gestures.
2. **Flex Sensors**: Smart gloves are equipped with flex sensors to capture finger bends and hand positions, which are crucial for recognizing sign language gestures.

3. **Advantages of Sensor-Based Systems**:
    - **Mobility and Flexibility**: Sensor-based gloves are lightweight, portable, and flexible, making them convenient to wear and use.
    - **No Environmental Limitations**: Unlike vision-based systems that can be affected by background or lighting conditions, sensor-based systems work reliably in any environment.
4. **Recognition of Gestures**: These systems can precisely measure finger configuration, hand orientation, and articulation points to convert gestures into meaningful information (e.g., text or voice).
5. Inertial measurement unit: this measurement is accomplished by an accelerometer and gyroscope to calculate the position, the recognition strength, and the acceleration of the fingers. The orientation and motion data of the user's hands can be precisely obtained by these sensors at a high frame rate (e.g., Xsens MTw IMU has a 50Hz frame rate)[84].

Sensor-based SLR is becoming increasingly popular owing to wearable, low-cost sensor devices such as the SEMG (Surface Electromyogram), gyroscope, and accelerometer (ACC). Data Gloves (sensor-based): It is an ADC converter that is used to convert an analog signal into a digital representation. It employs a group of sensors for the identification of hand signs and gestures. It is a fusion of bend signal detection with a flex sensor and an accelerometer. Gyroscope utilizes the assistance of an accelerometer to obtain orientation, angular, and acceleration data. Flex sensor data on bending fingers. IMUs, which are the abbreviation of inertial measurement units, are implemented to estimate hand movement. Electromyography, abbreviated as EMG, is utilized by attaching or implanting electrodes on human muscle. The electrodes attached allowed it to capture muscle action in electrical form. The Surface Electromyogram or SEMG is applied in finger movement discrimination and capturing [85].

5.1.2 Vision-based approach:

To employ a vision-based recognition approach, you need to provide a set of static and dynamic images, i.e., video. To create high-quality signs and to help the segmentation procedure, signers are usually asked to move rapidly between signs. The main benefit of the vision-based approach is one benefit of ArSL is that users need not wear the cumbersome Data Glove. However, there are several challenges in using the vision-based approach in ArSL, including lighting, face and hand segmentation, image background, and hand segmentation. Moreover, the computational cost of segmenting lips, facial expressions, and hand gestures is high. Real-time segmentation can now be used and implemented using methods and algorithms. The vision-based recognition method, however, is still limited and needs more to be established [3]. The following types of cameras are employed in the vision-based approach: a device that is invasive (body marker method): LED lights are one example.

1. Gloves with color and script band.
2. Kinect and Leap motion sensors are examples of active devices.
3. Record depth data with a stereo camera, also known as a depth camera.
4. One-camera devices include smartphones, thermal cameras, webcams, and video cameras.
5. LMC (Leap Motion Controller): The LMC is equipped with two cameras and three infrared LEDs, capable of tracking light at a wavelength of 850 nm within a range of up to 60 cm (2 feet). It detects hand movements and transforms them into appropriate computer command formats. The device uses Leap Motion service software to collect raw grayscale images. Merit: Low accuracy.
6. It creates motion and skeleton images from three-dimensional picture data. RGB camera, multi-array microphone, and depth sensor are the parts making up parts of the sensor Kinect. Cons: There has to be additional space (6–10 feet) between the sensor and the signer[85]. **Figure 9**. illustrates Sign Language Recognition Modalities[85].

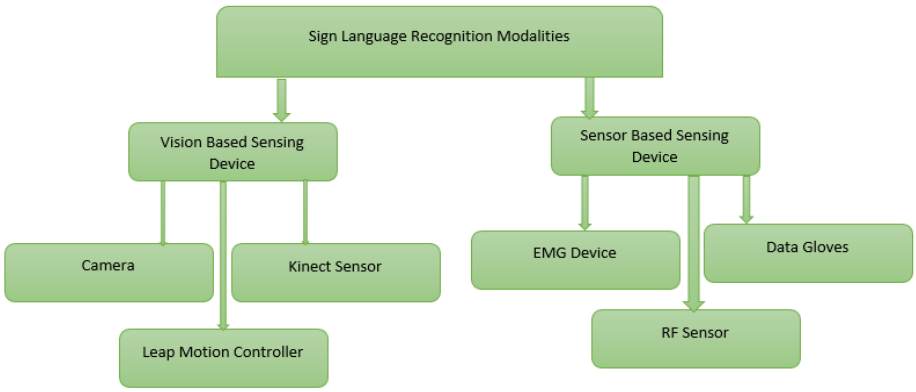

Figure 9. Classification of Sign Language Recognition Modalities Based on Vision and Sensor Technologies

5.2 Pre-processing and Segmentation phase

5.2.1 Noise Reduction Phase

The image pre-processing process enhances the ability of the system to alter images and videos. The most popular techniques used for noise removal in input videos or images are the use of media and Gaussian filters. According to research [86-88], Only the picture pre-processing stage uses median filtering and morphological techniques [89] They are widely utilized to eliminate undesirable information from the input. As an example, in the pre-processing stage, authors in [79] and the authors in [90] Threshold the input image to binary and then use K-means clustering with morphological processing to eliminate noise. An adaptive histogram [91] is employed to enhance the contrast of input photos obtained from various settings.

5.2.2 Phase of segmentation

The goal of this stage is to divide the image into numerous shapes in such a manner that the ROI (region of interest) will be detected from neighboring images. Non-contextual segmentation and contextual segmentation are two such stages of segmentation. Contextual segmentation deals with geometric relationships between features, such as the detection of edge techniques. The non-contextual form in the order hand arranges pixels according to global properties [92]. When it comes to Vision-based ASLR, hand segmentation is the most crucial and difficult stage compared to gesture recognition. **Figure 10**. illustrates some segmentation techniques [3].

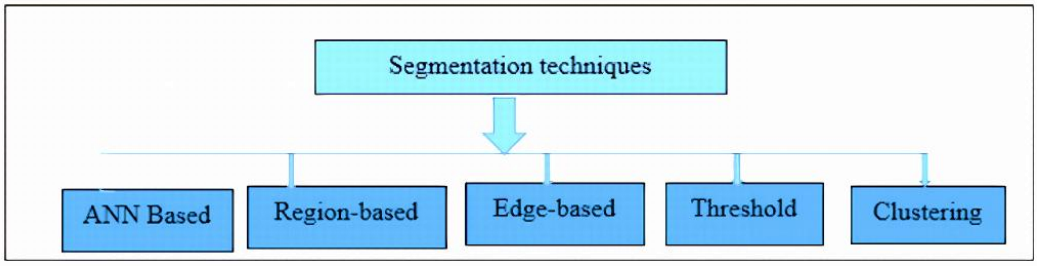

Figure 10. General image segmentation techniques commonly used in image processing

5.3 Feature Extraction Phase

The process of extracting various elements from an image is known as feature extraction. Image scaling, translation, shape, rotation, angle, coordinates, and picture background are some of the features. The ideal feature extraction process is the result of an ideal segmentation process [93, 94]. The properties are taken to be the components to be used in constructing hand gesture identification. For gesture recognition, the feature extraction needs to include pertinent data derived from the input of hand movements. But this characteristic is better defined as the gesture identity preserved for purposes of classification, distinct from the gestures obtained from the other limbs.

Shift-invariant feature transform (SIFT), principal component analysis (PCA), accelerated robust features (SURF), linear discriminant analysis (LDA), convexity defects, and K-curvature are significant feature extraction techniques used in sign language[92, 95, 96].

1. SIFT: Lowe [97] presented a feature extraction technique referred to as rotation invariant and scale, which applies several scale techniques for detection. Furthermore, the image is defined by its interest point, and at each pyramidal level, a Gaussian function is employed to rescale and blur the image [92]. The scale space is created by convolving the image with a Gaussian function using **Equation 1**.

$$L(x, Y, \sigma) = I(X, Y) * G(X, Y, \sigma) \quad (1)$$

Where $I(X, Y)$ represents the input image intensity at coordinates (X, Y), $G(X, Y, \sigma)$ is a Gaussian kernel with standard deviation σ\sigmaσ, used to blur the image at each pyramidal level.

Integration with Recognition of ArSL:

- A crucial stage in SIFT feature extraction, this Gaussian scale-space allows the algorithm to recognize important locations in Arabic Sign Language gestures.
- The model becomes resilient to changes in hand size, distance from the camera, and rotation by evaluating features at several scales.

2. PCA is a numerical process that utilizes orthogonal revolution to transform the value of a balanced variable into a sequence of values for an imbalanced variable (also known as a basic component)[98].
3. LDA: By augmenting the adaptive class, this technique finds the precise combination of attributes that optimally divides the object classes [92, 99]. Furthermore, the LDA approach is frequently applied in dimension reduction and as definite classifiers[92]. It should be highlighted that PCA focuses on figuring out the order of the biggest variation between the characteristics rather than on class differences [100]. Nonetheless, in order to identify the precise feature combination that provides a detailed explanation of the data, both PCA and LDA algorithms might be used.
4. SURF: This method is based on shift-invariant feature transformation, which computes numerous scale pyramids and then searches the space of scales for the local extremum by rotating the lower and upper scales of the image using a difference-of-Gaussian operator. Furthermore, instead of following an iterative procedure, SURF uses a filtering approach to reduce the size of the image to a minimum. By approximating the LoG (Laplacian of Gaussian) by difference-of-Gaussian (DoG), scale-space can be achieved in SIFT [92]. It was calculated using **Equation 2** [95].

$$H(X, Y, \sigma) = \begin{bmatrix} L_{xx}(X, Y, \sigma) & L_{xy}(X, Y, \sigma) \\ L_{xy}(X, Y, \sigma) & L_{yy}(X, Y, \sigma) \end{bmatrix} \quad (2)$$

Where X, Y are Spatial coordinates, σ is the scale parameter

Integration with Recognition of ArSL:

- To find reliable keypoints in hand movements, SURF employs the Hessian matrix. These keypoints are then utilized to compute descriptors.
- This approach is appropriate for real-time Arabic Sign Language identification systems since it is faster than SIFT because it approximates the Laplacian of Gaussian (LoG) using a filter.
- Using $H(X, Y, \sigma)$, Keypoints were retrieved. The invariance of H (X, Y,σ) to rotation, scaling, and tiny affine transformations is essential for precise gesture identification in a variety of camera settings.

As shown in **Figure 11**. illustrates some Feature Extraction Techniques.

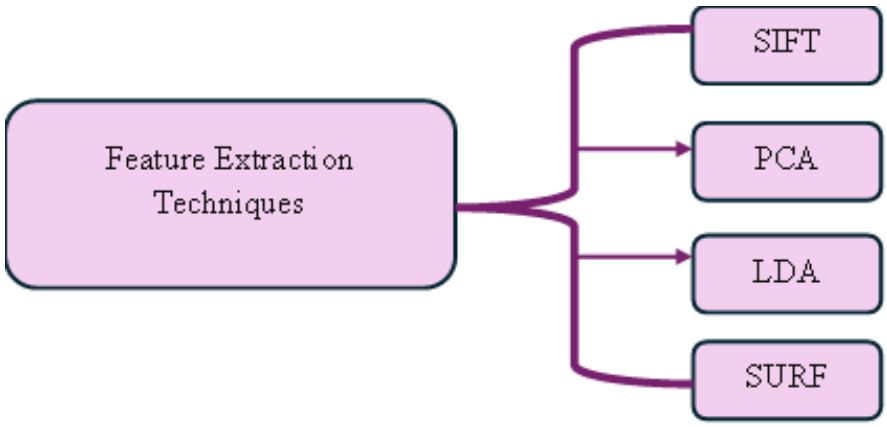

Figure 11. Some Feature Extraction Techniques Used in Arabic Sign Language Recognition Systems

5.4 Classification Phase

The classification of hand gestures is the final stage of the recognition system. This stage needs to be considered to have an effective classification method and recognition technique, both of which are beneficial in various gesture recognition research studies. This stage proceeds in parallel with the pattern recognition domain and artificial intelligence. Rule-based and machine learning-based methods of hand gesture classification have been proposed [94] [75], and they will be covered in more detail in the sections that follow.

Rule-based Approach: The methodology establishes some of the manually coded relations referenced as rules between the feature inputs. As a result, characteristics of the input gesture are derived and compared with coded rules. Eventually, the ensuing rule is transformed into a gesture. The method has the weakness that the achievement of the recognition process is bound by human capacity to code rules [75].

Machine Learning Approach: As described in the previous section, rule-based approaches suffer from a limitation when it comes to gesture identification. As a result of this limitation, many researchers have sought to use machine learning to discover gesture-set mappings over collections of high-dimensional features. Machine learning-based gesture recognition interprets gestures as outputs of random processes[3].

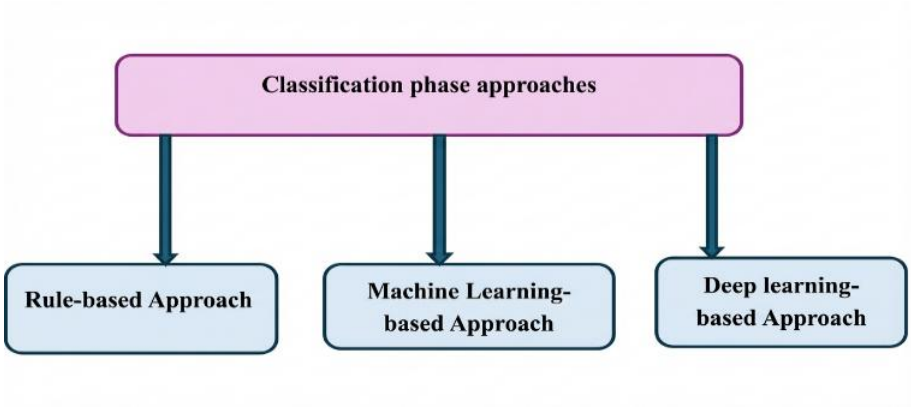

Figure 12. Classification approaches

Deep learning Approach: Later years have witnessed more significant designs with multiple layers and transferring information in vector format across layers, which have essentially become substitutes for plain machine learning techniques. The architecture increasingly refines the estimation towards positive recognition. While much more complicated, these algorithms, which are most frequently referred to as deep learning systems or deep neural networks, function according to principles analogous to those of machine learning techniques already discussed. Two architectures are usually employed for other tasks based on the network's structure: convolutional neural networks (CNNs), which include at least one convolutional layer, and recurrent neural networks (RNNs), which include at least

one recurrent layer. The networks can have different properties and are generally applicable to a variety of tasks based on the number and nature of layers. The algorithm's performance also depends a great deal on the training procedure. More intensive network training is enabled by larger and more specific datasets; hence, the training set quality plays a very critical role. As a rule, adjusting some suitable hyperparameters that define the training procedure makes it feasible to optimize a model further [101, 102]. **Figure 12** illustrates Classification phase approaches.

5.5 Arabic Sign Language Recognition Techniques

Machine learning (ML) and deep learning (DL) are playing a vital role in developing Arabic Sign Language (ArSL) recognition, closing the communication gap between deaf and hearing communities. With extensive datasets of Arabic gestures, body language, and facial expressions, ML and DL can teach models to effectively recognize and translate ArSL in real time. Machine learning techniques, including support vector machines (SVMs) and random forests, are utilized for recognizing single signs, while deep learning methods, specifically convolutional neural networks (CNNs) and recurrent neural networks (RNNs), are more appropriate for modeling complex patterns and sequences. These DL models are particularly useful since they can handle hand shape and position variations, making them more appropriate for real-world usage. With these advancements, ArSL recognition systems can now translate gestures into text or speech, increasing accessibility to education, services, and everyday communication for Arabic-speaking deaf communities. However, one of the primary challenges is the collection and annotation of high-quality ArSL datasets, which is essential to further develop the accuracy and performance of the systems.

5.5.1 Machine learning techniques

Understanding the Arabic phonetic alphabet entails understanding a series of movements that represent the letters in the ArSL alphabet. Both vision-based and/or sensor-based methods can be used to accomplish this. Using machine learning techniques and treating each letter as a unique gesture is one method of recognizing the ArSL alphabet, to categorize the motions according to sensor or visual characteristics. A variety of methods for image-based Arabic alphabet sign recognition are covered in this section.

5.5.1.1 Support Vector Machine (SVM)

Each of the several classifiers predicts and selects the appropriate class using a distinct set of parameters. By identifying the hyperplane that optimizes the margin between the two classes, the SVM carries out classification. The support vectors are the vectors (cases) that define the hyperplane [103]. To categorize data, the SVM algorithm clearly distinguishes between observation data, which each of the several classifiers predicts and selects the appropriate class using a distinct set of parameters. By identifying the hyperplane that optimizes the margin between the two classes, the SVM carries out classification. The support vectors are the vectors (cases) that define the hyperplane [103].
We find the weight vector w and Bias b by solving the following objective function using Quadratic Programming using **Equations 3 and 4**.

$$\min \frac{1}{2} \left|\left|w\right|\right|^2 \quad (3)$$

$$y_{i(w.x_i+b)\geq \forall x_i} \quad (4)$$

Where y_i is label, x_i is the input feature vector (sample i), $|w|$ is the **weight vector** defining the orientation of the hyperplane.is the **weight vector** defining the orientation of the hyperplane.

Integration with Recognition of ArSL:

- In Arabic Sign Language identification, retrieved gesture features (from SIFT, SURF, or CNN descriptors) are categorized into distinct sign classes using SVM classifiers.
- Several gesture classes by optimizing the Even in cases where features slightly overlap, SVM guarantees robust separation between hyperplanes.
- To effectively solve this optimization, quadratic programming is utilized.

5.5.1.2 K-Nearest Neighbor (KNN)

Identification of the most similar feature vector for the new vector from the reference feature set is the primary objective of classification. KNN is one of the most utilized techniques in sign language recognition systems. For the accomplishment of KNN in a dimensional space, it uses feature vectors that are created during training. The feature vector is classified by a majority vote of its neighbors. A collection of objects whose correct class is known is used to select neighbors [19]. The number of approximate nearest neighbors is returned by calculating the difference between the query and the target shape feature vectors using Euclidean distance metrics [13]. Using **Equation 5** [104].

$$ED(x, y) = \sqrt{\sum_{i=}^{n} | x_i - y_i|^2} \qquad (5)$$

Where x, y are vectors, n is the number of dimensions, and i is the index of dimension.
Integration with ArSL Identification:

- To find the closest movements in the reference feature set for classification in Arabic Sign Language recognition, KNN employs ED.
- To ensure robust classification, the query feature vector is allocated to the class of the majority of its nearest neighbors once distances have been calculated.
- In gesture recognition systems, ED is a key statistic for feature-based methods (such as SIFT, SURF, or CNN descriptors).

5.5.2 Deep Learning Techniques

5.5.2.1 Convolutional Neural Networks (CNNs)

Convolutional Neural Networks (CNNs) are a type of deep learning model with remarkable power that are commonly utilized for image classification. CNNs use convolutional layers to hierarchically and automatically learn spatial features from low-level edges and textures in early layers to high-level object parts and patterns in later layers. CNNs recognize unique features in images by using learnable filters and, therefore, are highly capable of classification tasks [105] .using **Equation 6.** [106].

$$S(i, j) = \sum_m \sum_n I\ (i - m, j - n\ )K(m, n) \qquad (6)$$

Where $I(i, j)$ represents the input image pixel values at coordinates (i, j) , $K(m, n)$ is the learnable filter (kernel) applied over the image. $S(i, j)$ is the output feature map after applying the convolution.

Integration with Recognition of ArSL:

- This convolution process is used by CNNs in Arabic Sign Language recognition to extract spatial characteristics from hand motions.
- Higher-level features like finger combinations and hand forms are captured by deeper layers, whilst lower-level features like edges and textures are detected by early layers.
- These feature maps are then utilized to classify motions, either as input to sequential models like LSTM for dynamic gestures or through fully connected layers like SoftMax.

5.5.2.2 Recurrent Neural Networks (RNNs)

Recurrent Neural Networks (RNNs) are a kind of artificial neural network for processing sequential data, including speech, text, and time-series data. RNNs possess a memory element enabling them to capture information from the past and apply it in the present processing, which makes them handy in language modeling, machine translation, and speech recognition tasks. However, conventional RNNs fail to learn long-term dependencies due to the vanishing gradient problem, which weakens their ability to recall information in lengthy sequences. To remedy this, more complex models like Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU) were developed, which improved performance by regulating the flow of information. RNNs find extensive applications in natural language processing (NLP), chatbots, music generation, and predictive analytics across various industries [107] using **Equations 7and 8** [108].

$$h_t = f(W_h h_{t-1} + W_x x_t + b_h) \quad (7)$$
$$L = \sum_t l(y_t, \hat{y}_t) \quad (8)$$

Where h_t is the hidden state at time step t, which stores information about previous inputs, h_{t-1} is the hidden state from the previous time step. x_t is the current input vector at time step t. The W_h and W_x These are weight matrices for the hidden state and input, respectively. b_h is the bias term. f is **a** non-linear activation function (e.g., tanh or ReLU). L represents the total loss over the sequence, where y_t is the true output and $\hat{y}_t$ is the predicted.

ArSL Recognition Integration:

- Dynamic gesture recognition uses RNNs (and their derivatives, LSTM/GRU), where gestures are frame sequences.
- The concealed state h_t recognizes movements that change over time by capturing temporal connections between frames.
- The network's prediction of the proper gesture label sequence from input video frames is guided by the loss function.
- In continuous sign language identification, where individual motions or phrases must be recognized within a sequence, this is crucial.

5.5.3 Transfer Learning

Transfer learning is a machine learning process in which a model trained on a task is reused and applied to another, yet related, task. Transfer learning enables models to carry learned knowledge from a source domain to a target domain, which is helpful when the target domain has limited data. Mainly, there are two types of transfer learning: inductive, where the target task and the source task are different, but the model is transferred to the new task, and transductive, where the task remains the same, but the domain of the data is different. Transfer learning has been applied across various fields, e.g., in image classification, where pre-trained models on huge datasets like ImageNet are fine-tuned to specific image tasks, and natural language processing, where pre-trained models like BERT or GPT are fine-tuned to perform sentiment analysis or question answering. The procedure can drastically enhance performance, diminish the requirement for enormous labeled data sets, and speed up training [109].using **Equations 9 and 10.**

Source domain[110]

$$D_S = \{X_S, P(X_S)\} \, wit \; task \; T_S = \{Y_S, P(Y_S | X_S)\} \quad (9)$$

Target domain [110]

$$D_T = \{X_T, P(X_T)\} \, with \; task \; T_T = \{Y_T, P(Y_T, P(Y_T | X_T)\} \quad (10)$$

Where D_S and D_T represent the source and target domains, respectively, X_S , X_T are the input feature spaces in the source and target domains. $P(X_S)$, $P(X_T)$ These are the probability distributions of the inputs. T_S , T_T denote the tasks in the source and target domains. Y_S , Y_T These are the output labels for each domain.

Integration with Recognition of ArSL:

- Pre-trained models (such as CNNs trained on massive gesture datasets or ImageNet) can be refined on smaller ArSL datasets for Arabic Sign Language recognition.
- This method decreases training time and reliance on big, labeled datasets while increasing recognition accuracy.
- Transfer learning is very beneficial for:
  - Adding sentence-level gestures to word-level models
  - Models learned on one dataset or signer can be modified for another

6 CHALLENGES IN ARABIC SIGN LANGUAGE RECOGNITION

The identification of Arabic Sign Language (ArSL) is also difficult for various reasons. First, the diversity in the styles used by singers and the dialects used in different places makes it hard to generalize the models used to identify the sign language. The complexity associated with hand movements, where fingers are used to convey different meanings, also poses a challenge to the identification process. Moreover, the limited available datasets also hinder the development of models to identify sign language. The constant movement associated with sign language also poses a significant challenge to the identification process, especially because of the co-articulation effect. The non-manual aspects associated with sign language, which include facial and body movements, are also very important in conveying meaning in sign language. The need to perform real-time recognition also poses a significant challenge to the identification process. These points outline the main challenges of Arabic Sign Language recognition:

- **Variability in Signing Styles:**
  Depending on hand size, motion speed, and personal style, various people can execute the same sign in different ways. Further differences in hand shapes, orientations, and movement patterns are introduced by regional dialects of Arabic Sign Language (ArSL). These variations make it impossible for systems to effectively understand hand gestures across a variety of users without access to huge and comprehensive datasets, which poses serious challenges for training robust recognition models.
- **Complex Hand Gestures**
  The majority of hand gestures used in Arabic Sign Language (ArSL) require rapid hand motions, many joints, and complex finger configurations. Achieving a high level of precision in feature extraction is crucial for effective recognition since even small changes in finger positions or hand orientations might result in completely distinct meanings.
- **Background and Lighting Conditions:**
  Systems that recognize hand gestures using vision are extremely sensitive to changes in the background, lighting, and shadows. Models find it difficult to reliably identify and recognize motions when lighting conditions change, especially when using edge or contour-based approaches.
- **Occlusions:**
  In situations involving two-handed sign language, hands may be obscured by other hands or body parts. The efficacy of gesture recognition systems is diminished by these occlusions, which make it difficult to precisely extract hand keypoints or image-based information.
- **Limited Datasets:**
  The majority of Arabic Sign Language (ArSL) datasets that are accessible to the general public are minimal and usually contain only a few words or solitary hand gestures. Furthermore, a lot of these datasets are limited to particular signers or controlled circumstances, which makes it difficult for models to properly generalize to a variety of users and real-world situations.
- **Continuous Signing and Co-articulation:**
  Unlike isolated signals, continuous signing requires seamless transitions between successive motions, which complicates segmentation and recognition. Co-articulation effects, in which a sign's beginning and conclusion are affected by nearby signs, add even more uncertainty and make precise interpretation extremely difficult.
- **Facial Expressions and Body Movements:**
  Such other sign languages, Arabic Sign Language (ArSL), rely heavily on non-manual signs, such as body postures, head motions, and facial expressions, to transmit semantic meaning. However, the majority of recognition algorithms ignore these non-manual indications in favor of hand movements, which might result in insufficient comprehension and challenges with precise interpretation.

- **Real-time Processing Requirements:**
  Systems must be able to immediately identify signals and function in real time for assistive technologies to use sign language recognition efficiently. To do this, efficient algorithms and optimized model designs are needed, which are frequently aided by hardware acceleration methods like GPUs to guarantee quick and dependable performance.

7 DISCUSSION

Arabic Sign Language (ArSL) alphabet and word recognition have been greatly improved through the application of various machine learning and deep learning techniques. Traditional classification techniques such as Support Vector Machine (SVM), k-Nearest Neighbors (KNN), and Multi-Layer Perceptron (MLP) have been successfully used for static hand sign language recognition. Some techniques, such as neuro-fuzzy systems, have proved to be successful in handling changes in hand positions. However, most of these techniques have used additional accessories, such as colored gloves and preprocessing techniques to improve precision in the results. The use of dynamic hand gestures is still challenging for most of these techniques, as they are not capable of accurately tracking the movement of the hand.

The introduction of deep learning technology, specifically Convolutional Neural Networks (CNN) and object detection models such as Faster R-CNN, has helped improve recognition capabilities. Transfer learning with pre-trained models such as ResNet and MobileNet has helped improve model generalization and efficiency in adapting to new datasets. This shows the possibilities for implementing ArSL recognition in various applications.

Despite all these improvements, there are still challenges to be addressed. One of them is dealing with imbalanced datasets, the performance variability of sign language signs by different individuals, and the difficulty in dealing with continuous signing. Another difficulty is the variety of Arabic hand signs and their different shapes and orientations. Another aspect is dealing with models that can recognize continuous signs instead of isolated signs.

To address these challenges, future studies need to be directed towards improving the efficiency of models, optimizing systems in terms of real-time processing, and gathering more data. The implementation of ArSL recognition in practical assistive technologies, such as smart gloves, mobile apps, and other AI-based communication devices, has the potential to improve accessibility and bridge the communication gap between deaf and hearing communities. Further developments in this area have the potential to bring about greater inclusiveness and enable individuals to engage more in communication.

8 FUTURE RESEARCH DIRECTIONS

Although Arabic Sign Language Recognition (ArSLR) has advanced significantly in recent years, there are still several obstacles that prevent it from reaching its full potential. To enhance recognition accuracy, generalization, and practical applicability, specific study is needed in some crucial areas, according to the analysis of previous studies. Resolving these issues will improve ArSLR systems' usability in assistive technology, instructional resources, and communication platforms for the deaf and hard-of-hearing community, in addition to advancing their technological advancement. We list eight important avenues for further investigation in this area below:

- **Sentence-Level and Continuous Recognition:** Create reliable models that can identify continuous signals and complete phrases while taking temporal dependencies and co-articulation effects into consideration.
- **Extensive and Various Datasets:** To improve model generalization and scalability, gather and standardize datasets encompassing various areas, dialects, and signer variants.
- **Integrating Multimodal Features:** To fully convey the linguistic and grammatical context of signals, use non-manual elements like body position, head motions, and facial expressions.
- **Optimizing in Real Time:** To enable effective real-time recognition, create lightweight architectures and make use of hardware acceleration (such as GPUs and edge computing).
- **Domain Modification for Local Dialects:** Create strategies to modify recognition models for various Arabic dialects and regional variances, guaranteeing wide applicability.
- **Synthetic Data Generation and Data Augmentation:** To reduce dataset scarcity and enhance recognition performance, use augmentation techniques to create synthetic data.
- **Combining Assistive Technologies:** For practical use, incorporate ArSLR systems into mobile apps, educational resources, and AI-powered communication devices.
- **Interpretable and Explainable Models:** Develop interpretable models that yield comprehensible results, boosting user and stakeholder trust and promoting adoption.

9 CONCLUSION

For those who are deaf or hard of hearing, Arabic Sign Language recognition (ArSLR) is essential to communication and social inclusion. ArSLR has been greatly improved over time by machine learning and deep learning techniques. While deep learning models, such as CNNs and Transformer-based architectures, provide more adaptable solutions for dynamic and sentence-level recognition, traditional classifiers, like SVM and MLP, perform well for static hand gestures. Many obstacles still exist despite these developments. Temporal dependencies and co-articulation effects continue to restrict continuous gesture recognition. Additional obstacles to model generalization include signer volatility, imbalance, and scarcity of datasets. Furthermore, most existing systems ignore non-manual elements that are essential for capturing the complete linguistic and grammatical context of signs, such as body language and facial expressions. Due to computational limitations, real-time deployment is still difficult. The analysis shows that the shift from word-level to continuous recognition is mostly constrained by temporal modeling and dataset complexity, whereas the primary performance difference between ML and DL techniques is found in feature representation.

ACKNOWLEDGMENTS

The authors sincerely thank the referees, Associate Editor, and Editor-in-Chief for their valuable comments and suggestions, which have greatly improved this paper. I would like to express my gratitude to ChatGPT by OpenAI for helping to organize and enhance the research content.

FUNDING

The authors state that no outside funding was received for this study.

DISCLOSURE STATEMENT

No potential conflict of interest was reported by the author(s).

REFERENCES

[1] G. Tharwat, A. M. Ahmed, and B. Bouallegue, "Arabic sign language recognition system for alphabets using machine learning techniques," *Journal of Electrical and Computer Engineering,* vol. 2021, no. 1, p. 2995851, 2021. https://doi.org/10.1155/2021/2995851.

[2] A. M. Ahmed *et al.*, "Towards the design of automatic translation system from Arabic Sign Language to Arabic text," in *2017 International Conference on Inventive Computing and Informatics (ICICI)*, 2017, pp. 325-330: IEEE. https://doi.org/10.1109/ICICI.2017.8365365.

[3] A. S. Al-Shamayleh, R. Ahmad, N. Jomhari, and M. A. Abushariah, "Automatic Arabic sign language recognition: A review, taxonomy, open challenges, research roadmap and future directions," *Malaysian Journal of Computer Science,* vol. 33, no. 4, pp. 306-343, 2020. https://doi.org/10.22452/mjcs.vol33no4.5.

[4] A. M. Ahmed, R. Abo Alez, G. Tharwat, M. Taha, B. Belgacem, and A. M. Al Moustafa, "Arabic sign language intelligent translator," *The Imaging Science Journal,* vol. 68, no. 1, pp. 11-23, 2020. https://doi.org/10.1080/13682199.2020.1724438.

[5] M. Mohandes, J. Liu, and M. Deriche, "A survey of image-based arabic sign language recognition," in *2014 IEEE 11th International Multi-Conference on Systems, Signals & Devices (SSD14)*, 2014, pp. 1-4: IEEE. https://doi.org/10.1109/SSD.2014.6808906.

[6] X. Zabulis, H. Baltzakis, and A. A. Argyros, "Vision-Based Hand Gesture Recognition for Human-Computer Interaction," *The universal access handbook,* vol. 34, p. 30, 2009.

[7] M. A. Ahmed, B. B. Zaidan, A. A. Zaidan, M. M. Salih, and M. M. B. Lakulu, "A review on systems-based sensory gloves for sign language recognition state of the art between 2007 and 2017," *Sensors,* vol. 18, no. 7, p. 2208, 2018. https://doi.org/10.3390/s18072208

[8] M. A. Abdel-Fattah, "Arabic sign language: a perspective," *Journal of deaf studies and deaf education,* vol. 10, no. 2, pp. 212-221, 2005. https://doi.org/10.1093/deafed/eni007.

[9] N. El-Bendary, H. M. Zawbaa, M. S. Daoud, A. E. Hassanien, and K. Nakamatsu, "Arslat: Arabic sign language alphabets translator," in *2010 international conference on computer information systems and industrial management applications (CISIM)*, 2010, pp. 590-595: IEEE. https://doi.org/10.1109/CISIM.2010.5643519.

[10] A. Moustafa *et al.*, "Arabic Sign Language Recognition Systems: A Systematic Review," *Indian Journal of Computer Science and Engineering,* vol. 15, pp. 1-18, 2024. https://doi.org/10.21817/indjcse/2024/v15i1/241501008.

[11] N. Tubaiz, T. Shanableh, and K. Assaleh, "Glove-based continuous Arabic sign language recognition in user-dependent mode," *IEEE Transactions on Human-Machine Systems,* vol. 45, no. 4, pp. 526-533, 2015. https://doi.org/10.1109/THMS.2015.2406692.

[12] E. E. Hemayed and A. S. Hassanien, "Edge-based recognizer for Arabic sign language alphabet (ArS2V-Arabic sign to voice)," in *2010 International Computer Engineering Conference (ICENCO)*, 2010, pp. 121-127: IEEE. https://doi.org/10.1109/THMS.2015.2406692.

[13] N. R. Albelwi and Y. M. Alginahi, "Real-time arabic sign language (arsl) recognition," in *International Conference on Communications and Information Technology*, 2012, pp. 497-501: International Conference on Communications and Information Technology.

[14] M. Mohandes, "Arabic sign language recognition," in *International conference of imaging science, systems, and technology, Las Vegas, Nevada, USA*, 2001, vol. 1, pp. 753-9.

[15] O. Al-Jarrah and A. Halawani, "Recognition of gestures in Arabic sign language using neuro-fuzzy systems," *Artificial Intelligence,* vol. 133, no. 1-2, pp. 117-138, 2001. https://doi.org/10.1109/THMS.2015.2406692.

[16] M. Al-Rousan and M. Hussain, "Automatic recognition of Arabic sign language finger spelling," *International Journal of Computers and Their Applications,* vol. 8, pp. 80-88, 2001.

[17] K. Assaleh and M. Al-Rousan, "Recognition of Arabic sign language alphabet using polynomial classifiers," *EURASIP Journal on Advances in Signal Processing,* vol. 2005, pp. 1-10, 2005. https://doi.org/10.1155/ASP.2005.2136.

[18] R. Naoum, H. H. Owaied, and S. Joudeh, "Development of a new Arabic sign language recognition using k-nearest neighbor algorithm," *Journal of Emerging Trends in Computing and Information Sciences,* vol. 3, no. 8, pp. 1173-1178, 2012.

[19] T. Messer, "Static hand gesture recognition," *University of Fribourg, Switzerland,* 2009.

[20] T. Shanableh and K. Assaleh, "Arabic sign language recognition in user-independent mode," in *2007 International Conference on Intelligent and Advanced Systems*, 2007, pp. 597-600: IEEE. https://doi.org/10.1109/ICIAS.2007.4658457.

[21] M. Elpeltagy, M. Abdelwahab, M. E. Hussein, A. Shoukry, A. Shoala, and M. Galal, "Multi-modality-based Arabic sign language recognition," *IET Computer Vision,* vol. 12, no. 7, pp. 1031-1039, 2018. https://doi.org/10.1049/iet-cvi.2017.0598.

[22] B. Hisham and A. Hamouda, "Arabic sign language recognition using Ada-Boosting based on a leap motion controller," *International Journal of Information Technology,* vol. 13, no. 3, pp. 1221-1234, 2021. https://doi.org/10.1007/s41870-020-00544-4.

[23] H. AbdElghfar, H. A. Youness, M. Wahba, and H. M. Abdelaal, "An automated framework for qur'anic education of the hearing-impaired using body pose classification and Arabic sign language integration," *Scientific Reports,* vol. 16, no. 1, p. 5939, 2026. https://doi.org/10.1038/s41598-026-5939-3.

[24] M. A. Almasre and H. Al-Nuaim, "A comparison of Arabic sign language dynamic gesture recognition models," *Heliyon,* vol. 6, no. 3, 2020.

[25] H. Luqman and S. A. Mahmoud, "Transform-based Arabic sign language recognition," *Procedia Computer Science,* vol. 117, pp. 2-9, 2017. https://doi.org/10.1016/j.procs.2017.10.087.

[26] M. Almaazmi, S. Elkadi, L. Elsayed, L. Salman, and T. Shanableh, "Motion Images with Positioning Information and Deep Learning for Continuous Arabic Sign Language Recognition in Signer Dependent and Independent Modes," *IEEE Access,* 2024. https://doi.org/10.1109/ACCESS.2024.3485131.

[27] M. Kamruzzaman, "Arabic sign language recognition and generating Arabic speech using convolutional neural network," *Wireless Communications and Mobile Computing,* vol. 2020, no. 1, p. 3685614, 2020. https://doi.org/10.1155/2020/3685614.

[28] R. A. Alawwad, O. Bchir, and M. M. B. Ismail, "Arabic sign language recognition using Faster R-CNN," *International Journal of Advanced Computer Science and Applications,* vol. 12, no. 3, 2021. https://doi.org/10.14569/IJACSA.2021.0120380.

[29] N. A. N. Azhar *et al.*, "Development of Mobile Application for Arabic Sign Language based on Android Studio Software," *Journal of Algebraic Statistics,* vol. 13, no. 3, pp. 3152-3160, 2022.

[30] A. Hasasneh, "Arabic sign language characters recognition based on a deep learning approach and a simple linear classifier," *Jordanian Journal of Computers and Information Technology,* vol. 6, no. 3, 2020.

[31] N. Alasmari and S. Asiri, "ASLDetect: Arabic sign language detection using ResNet and U-Net like component," *Scientific Reports,* vol. 15, no. 1, p. 18012, 2025. https://doi.org/10.1038/s41598-025-18012-0.

[32] N. M. Alharthi and S. M. Alzahrani, "Vision transformers and transfer learning approaches for arabic sign language recognition," *Applied Sciences,* vol. 13, no. 21, p. 11625, 2023. https://doi.org/10.3390/app132111625.

[33] A. A. Alani and G. Cosma, "ArSL-CNN: a convolutional neural network for Arabic sign language gesture recognition," *Indonesian journal of electrical engineering and computer science,* vol. 22, 2021. https://doi.org/10.11591/ijeecs.v22.i1.pp259-267.

[34] M. ElBadawy, A. Elons, H. A. Shedeed, and M. Tolba, "Arabic sign language recognition with 3d convolutional neural networks," in *2017 Eighth international conference on intelligent computing and information systems (ICICIS)*, 2017, pp. 66-71: IEEE. https://doi.org/10.1109/INTELCIS.2017.8260028.

[35] S. Aly and W. Aly, "DeepArSLR: A novel signer-independent deep learning framework for isolated arabic sign language gestures recognition," *IEEE Access,* vol. 8, pp. 83199-83212, 2020.

[36] W. Abdul *et al.*, "Intelligent real-time Arabic sign language classification using attention-based inception and BiLSTM," *Computers and Electrical Engineering,* vol. 95, p. 107395, 2021. https://doi.org/10.1016/j.compeleceng.2021.107395.

[37] S. Alyami, H. Luqman, and M. Hammoudeh, "Isolated arabic sign language recognition using a transformer-based model and landmark keypoints," *ACM Transactions on Asian and Low-Resource Language Information Processing,* vol. 23, no. 1, pp. 1-19, 2024. https://doi.org/10.1145/3631981.

[38] M. M. Balaha *et al.*, "A vision-based deep learning approach for independent-users Arabic sign language interpretation," *Multimedia Tools and Applications,* vol. 82, no. 5, pp. 6807-6826, 2023. https://doi.org/10.1007/s11042-022-14432-3.

[39] 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, 2021. https://doi.org/10.3390/electronics10141739.

[40] H. Luqman and E. ELALFY, "Utilizing motion and spatial features for sign language gesture recognition using cascaded CNN and LSTM models," *Turkish Journal of Electrical Engineering and Computer Sciences,* vol. 30, no. 7, pp. 2508-2525, 2022. https://doi.org/10.55730/1300-0632.3952.

[41] T. Shanableh, "Two-stage deep learning solution for continuous Arabic Sign Language recognition using word count prediction and motion images," *IEEE Access,* vol. 11, pp. 126823-126833, 2023. https://doi.org/10.1109/ACCESS.2023.3332250.

[42] R. Marzouk, F. Alrowais, F. N. Al-Wesabi, and A. M. Hilal, "Atom search optimization with deep learning enabled arabic sign language recognition for speaking and hearing disability persons," in *Healthcare*, 2022, vol. 10, no. 9, p. 1606: MDPI. https://doi.org/10.3390/healthcare10091606.

[43] H. Luqman, "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)*, 2023, pp. 1-8: IEEE. https://doi.org/10.1109/FG57933.2023.10042720.

[44] M. S. Abdel-Wahab, M. Aboul-Ela, and A. Samir, "Arabic sign language recognition using neural network and graph matching techniques," 2006: Proceedings of the 6th WSEAS International Conference on Applied Informatics ….

[45] M. Tolba, A. Samir, and M. Abul-Ela, "A proposed graph matching technique for Arabic sign language continuous sentences recognition," in *2012 8th International Conference on Informatics and Systems (INFOS)*, 2012, pp. MM-14-MM-20: IEEE.

[46] T. H. Noor *et al.*, "Real-Time Arabic Sign Language Recognition Using a Hybrid Deep Learning Model," *Sensors,* vol. 24, no. 11, p. 3683, 2024.

[47] S. Al Ahmadi, F. Muhammad, and H. Al Dawsari, "Enhancing Arabic Sign Language Interpretation: Leveraging Convolutional Neural Networks and Transfer Learning," *Mathematics,* vol. 12, no. 6, p. 823, 2024. https://doi.org/10.3390/math12060823.

[48] A. M. Ahmed, R. A. Alez, M. Taha, and G. Tharwat, "Automatic translation of Arabic sign to Arabic text (ATASAT) system," *Journal of Computer Science and Information Technology,* vol. 6, pp. 109-122, 2016.

[49] A. Soudi, G. Neumann, and A. v. d. Bosch, "Arabic computational morphology: knowledge-based and empirical methods," in *Arabic computational morphology: Knowledge-based and Empirical Methods*: Springer, 2007, pp. 3-14.

[50] M. A. Abushariah, "TAMEEM V1. 0: speakers and text independent Arabic automatic continuous speech recognizer," *International Journal of Speech Technology,* vol. 20, pp. 261-280, 2017. https://doi.org/10.1007/s10772-017-9411-5.

[51] M. A. Abushariah, R. N. Ainon, R. Zainuddin, M. Elshafei, and O. O. Khalifa, "Phonetically rich and balanced text and speech corpora for Arabic language," *Language resources and evaluation,* vol. 46, pp. 601-634, 2012. https://doi.org/10.1007/s10579-011-9171-9.

[52] M. Elmahdy, R. Gruhn, W. Minker, and S. Abdennadher, "Survey on common Arabic language forms from a speech recognition," 2009.

[53] A. Farghaly and K. Shaalan, "Arabic natural language processing: Challenges and solutions," *ACM Transactions on Asian Language Information Processing (TALIP),* vol. 8, no. 4, pp. 1-22, 2009. https://doi.org/10.1145/1644879.1644881.

[54] H. Althobaiti and C. Lu, "A survey on Arabic optical character recognition and an isolated handwritten Arabic character recognition algorithm using encoded freeman chain code," in *2017 51st Annual conference on information sciences and systems (CISS)*, 2017, pp. 1-6: IEEE. https://doi.org/10.1109/CISS.2017.7926062.

[55] S. M. Halawani, "Arabic sign language translation system on mobile devices," *IJCSNS International Journal of Computer Science and Network Security,* vol. 8, no. 1, pp. 251-256, 2008.

[56] A. Almohimeed, M. Wald, and R. I. Damper, "Arabic text to Arabic sign language translation system for the deaf and hearing-impaired community," in *Proceedings of the second workshop on speech and language processing for assistive technologies*, 2011, pp. 101-109. https://doi.org/10.3115/2017171.2017183.

[57] A. S. Al-Shamayleh, R. Ahmad, M. A. Abushariah, K. A. Alam, and N. Jomhari, "A systematic literature review on vision based gesture recognition techniques," *Multimedia Tools and Applications,* vol. 77, pp. 28121-28184, 2018. https://doi.org/10.1007/s11042-018-5971-z.

[58] P. K. Pisharady and M. Saerbeck, "Recent methods and databases in vision-based hand gesture recognition: A review," *Computer Vision and Image Understanding,* vol. 141, pp. 152-165, 2015. https://doi.org/10.1016/j.cviu.2015.08.004.

[59] M. Oudah, A. Al-Naji, and J. Chahl, "Hand gesture recognition based on computer vision: a review of techniques," *journal of Imaging,* vol. 6, no. 8, p. 73, 2020. https://doi.org/10.3390/jimaging6080073.

[60] R. B. Dan and P. Mohod, "Survey on hand gesture recognition approaches," *structure,* vol. 15, p. 17, 2014.

[61] R. Y. Wang and J. Popović, "Real-time hand-tracking with a color glove," *ACM transactions on graphics (TOG),* vol. 28, no. 3, pp. 1-8, 2009. https://doi.org/10.1145/1531369.1531326.

[62] L. Lamberti and F. Camastra, "Real-time hand gesture recognition using a color glove," in *Image Analysis and Processing–ICIAP 2011: 16th International Conference, Ravenna, Italy, September 14-16, 2011, Proceedings, Part I 16*, 2011, pp. 365-373: Springer. https://doi.org/10.1007/978-3-642-24085-0_38.

[63] P. Garg, N. Aggarwal, and S. Sofat, "Vision based hand gesture recognition," *International Journal of Computer and Information Engineering,* vol. 3, no. 1, pp. 186-191, 2009.

[64] A. M. Martínez, R. B. Wilbur, R. Shay, and A. C. Kak, "Purdue RVL-SLLL ASL database for automatic recognition of American Sign Language," in *Proceedings. Fourth IEEE International Conference on Multimodal Interfaces*, 2002, pp. 167-172: IEEE. https://doi.org/10.1109/ICMI.2002.1166987.

[65] A. Schembri, J. Fenlon, R. Rentelis, S. Reynolds, and K. Cormier, "Building the British sign language corpus," 2013.

[66] M. Fagiani, E. Principi, S. Squartini, and F. Piazza, "A new Italian sign language database," in *Advances in Brain Inspired Cognitive Systems: 5th International Conference, BICS 2012, Shenyang, China, July 11-14, 2012. Proceedings 5*, 2012, pp. 164-173: Springer. https://doi.org/10.1007/978-3-642-31561-9_18.

[67] S. Kausar, M. Y. Javed, and S. Sohail, "Recognition of gestures in Pakistani sign language using fuzzy classifier," in *Proceedings of the 8th conference on Signal processing, computational geometry and artificial vision*, 2008, pp. 101-105: World Scientific and Engineering Academy and Society (WSEAS).

[68] O. Aran *et al.*, "A database of non-manual signs in turkish sign language," in *2007 IEEE 15th Signal Processing and Communications Applications*, 2007, pp. 1-4: IEEE. https://doi.org/10.1109/SIU.2007.4298708.

[69] A. A. I. Sidig, H. Luqman, S. Mahmoud, and M. Mohandes, "KArSL: Arabic sign language database," *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP),* vol. 20, no. 1, pp. 1-19, 2021. https://doi.org/10.1145/3423420.

[70] M. Al-Barham *et al.*, "RGB Arabic alphabets sign language dataset," *arXiv preprint arXiv:2301.11932,* 2023. https://doi.org/10.48550/arXiv.2301.11932.

[71] M. H. Ismail, S. A. Dawwd, and F. H. Ali, "Static hand gesture recognition of Arabic sign language by using deep CNNs," *Indonesian Journal of Electrical Engineering and Computer Science,* vol. 24, no. 1, pp. 178-188, 2021. https://doi.org/10.11591/ijeecs.v24.i1.pp178-188.

[72] A. Alnahhas, B. Alkhatib, N. Al-Boukaee, N. Alhakim, O. Alzabibi, and N. Ajalyakeen, "Enhancing the recognition of Arabic sign language by using deep learning and leap motion controller," *Int. J. Sci. Technol. Res,* vol. 9, no. 4, pp. 1865-1870, 2020.

[73] T. Starner, J. Auxier, D. Ashbrook, and M. Gandy, "The gesture pendant: A self-illuminating, wearable, infrared computer vision system for home automation control and medical monitoring," in *Digest of Papers. Fourth International Symposium on Wearable Computers*, 2000, pp. 87-94: IEEE. https://doi.org/10.1109/ISWC.2000.888469.

[74] R. Sharma *et al.*, "Speech/gesture interface to a visual-computing environment," *IEEE Computer Graphics and Applications,* vol. 20, no. 2, pp. 29-37, 2000. https://doi.org/10.1109/38.824531.

[75] G. Murthy and R. Jadon, "A review of vision based hand gestures recognition," *International Journal of Information Technology and Knowledge Management,* vol. 2, no. 2, pp. 405-410, 2009.

[76] D. Stotts, J. M. Smith, and K. Gyllstrom, "Facespace: endo-and exo-spatial hypermedia in the transparent video facetop," in *Proceedings of the fifteenth ACM conference on Hypertext and hypermedia*, 2004, pp. 48-57. https://doi.org/10.1145/1012807.1012827.

[77] C. Valli, *Linguistics of American Sign Language: An Introduction*. Gallaudet University Press, 2000.

[78] A. M. Ahmed *et al.*, "Gestures Arabic Sign Language Conversion to Arabic Alphabets," in *2018 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*, 2018, pp. 1-6: IEEE. https://doi.org/10.1109/ICCIC.2018.8782315.

[79] P. C. Badhe and V. Kulkarni, "Indian sign language translator using gesture recognition algorithm," in *2015 IEEE international conference on computer graphics, vision and information security (CGVIS)*, 2015, pp. 195-200: IEEE. https://doi.org/10.1109/CGVIS.2015.7449921.

[80] P. A. Nanivadekar and V. Kulkarni, "Indian sign language recognition: database creation, hand tracking and segmentation," in *2014 International conference on circuits, systems, communication and information technology applications (CSCITA)*, 2014, pp. 358-363: IEEE. https://doi.org/10.1109/CSCITA.2014.6839287.

[81] S. Subburaj and S. Murugavalli, "Survey on sign language recognition in context of vision-based and deep learning," *Measurement: Sensors,* vol. 23, p. 100385, 2022. https://doi.org/10.1016/j.measen.2022.100385.

[82] S. G. M. Almeida, F. G. Guimarães, and J. A. Ramírez, "Feature extraction in Brazilian Sign Language Recognition based on phonological structure and using RGB-D sensors," *Expert Systems with Applications,* vol. 41, no. 16, pp. 7259-7271, 2014. https://doi.org/10.1016/j.eswa.2014.05.024.

[83] E. Gani and A. Kika, "Albanian Sign Language (AlbSL) Number Recognition from Both Hand's Gestures Acquired by Kinect Sensors," *arXiv preprint arXiv:1608.02991,* 2016. https://doi.org/10.48550/arXiv.1608.02991.

[84] S. Ruffieux, D. Lalanne, and E. Mugellini, "ChAirGest: a challenge for multimodal mid-air gesture recognition for close HCI," in *Proceedings of the 15th ACM on International conference on multimodal interaction*, 2013, pp. 483-488. https://doi.org/10.1145/2522848.2532590.

[85] D. M. Madhiarasan, P. Roy, and P. Pratim, "A comprehensive review of sign language recognition: Different types, modalities, and datasets," *arXiv preprint arXiv:2204.03328,* 2022. https://doi.org/10.48550/arXiv.2204.03328.

[86] P. Kishore and P. R. Kumar, "A video based Indian sign language recognition system (INSLR) using wavelet transform and fuzzy logic," *International Journal of Engineering and Technology,* vol. 4, no. 5, p. 537, 2012.

[87] S. Shivashankara and S. Srinath, "American sign language recognition system: an optimal approach," *International Journal of Image, Graphics and Signal Processing,* vol. 10, no. 8, p. 18, 2018. https://doi.org/10.5815/ijigsp.2018.08.03.

[88] K. M. Lim, A. W. Tan, and S. C. Tan, "A feature covariance matrix with serial particle filter for isolated sign language recognition," *Expert Systems with Applications,* vol. 54, pp. 208-218, 2016. https://doi.org/10.1016/j.eswa.2016.01.047.

[89] R. Akmeliawati, M. P.-L. Ooi, and Y. C. Kuang, "Real-time Malaysian sign language translation using colour segmentation and neural network," in *2007 IEEE Instrumentation & Measurement Technology Conference IMTC 2007*, 2007, pp. 1-6: IEEE. https://doi.org/10.1109/IMTC.2007.379311.

[90] M. Krishnaveni and V. Radha, "Classifier fusion based on Bayes aggregation method for Indian sign language datasets," *Procedia Engineering,* vol. 30, pp. 1110-1118, 2012. https://doi.org/10.1016/j.proeng.2012.01.970.

[91] G. A. Rao and P. Kishore, "Selfie sign language recognition with multiple features on adaboost multilabel multiclass classifier," *Journal of Engineering Science and Technology,* vol. 13, no. 8, pp. 2352-2368, 2018.

[92] M. J. Cheok, Z. Omar, and M. H. Jaward, "A review of hand gesture and sign language recognition techniques," *International Journal of Machine Learning and Cybernetics,* vol. 10, pp. 131-153, 2019. https://doi.org/10.1007/s13042-017-0705-5.

[93] A. G. Bairagi, "Y.. Kapse,"Survey on Sign language to Speech Conversion,"" *Int. J. Innov. Res. Comput. Commun. Eng,* vol. 6, no. 1, pp. 267-274, 2018.

[94] R. Z. Khan and N. A. Ibraheem, "Hand gesture recognition: a literature review," *International journal of artificial Intelligence & Applications,* vol. 3, no. 4, p. 161, 2012.

[95] H. Bay, "Surf: Speeded up robust features," *Computer Vision—ECCV,* 2006.

[96] J. Rekha, J. Bhattacharya, and S. Majumder, "Hand gesture recognition for sign language: A new hybrid approach," in *Proceedings of the International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV)*, 2011, p. 1.

[97] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision,* vol. 60, pp. 91-110, 2004. https://doi.org/10.1023/B:VISI.0000029664.99615.94.

[98] G. Kumar and P. K. Bhatia, "A detailed review of feature extraction in image processing systems," in *2014 Fourth international conference on advanced computing & communication technologies*, 2014, pp. 5-12: IEEE. https://doi.org/10.1109/ACCT.2014.74.

[99] K. Delac, M. Grgic, and S. Grgic, "Independent comparative study of PCA, ICA, and LDA on the FERET data set," *International Journal of Imaging Systems and Technology,* vol. 15, no. 5, pp. 252-260, 2005. https://doi.org/10.1002/ima.20059.

[100] M. Suriya, N. Sathyapriya, M. Srinithi, and V. Yesodha, "Survey on real time sign language recognition system: an LDA approach," in *International conference on exploration and innovations in engineering and technology, ICEIET*, 2016, pp. 219-225.

[101] M. Al-Qurishi, T. Khalid, and R. Souissi, "Deep learning for sign language recognition: Current techniques, benchmarks, and open issues," *IEEE Access,* vol. 9, pp. 126917-126951, 2021. https://doi.org/10.1109/ACCESS.2021.3110912.

[102] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature,* vol. 521, no. 7553, pp. 436-444, 2015.

[103] S. Sayad, *Real time data mining*. Self-Help Publishers Cambridge, 2011. https://doi.org/10.1038/nature14539.

[104] V. Prasath *et al.*, "Distance and similarity measures effect on the performance of K-nearest neighbor classifier--a review," *arXiv preprint arXiv:1708.04321,* 2017. https://doi.org/10.48550/arXiv.1708.04321.

[105] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems,* vol. 25, 2012. https://doi.org/10.1145/3065386.

[106] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE,* vol. 86, no. 11, pp. 2278-2324, 1998. https://doi.org/10.1109/5.726791.

[107] I. Goodfellow, "Deep learning," ed: MIT press, 2016.

[108] H. E. T. Siegelmann, *Foundations of recurrent neural networks*. Rutgers The State University of New Jersey, School of Graduate Studies, 1993.

[109] J. Devlin, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805,* 2018. https://doi.org/10.48550/arXiv.1810.04805.

[110] F. Zhuang *et al.*, "A comprehensive survey on transfer learning," *Proceedings of the IEEE,* vol. 109, no. 1, pp. 43-76, 2020. https://doi.org/10.1109/JPROC.2020.3004555.