

A Survey of Methodological Limitations in Arabic Sign Language Recognition

Name : shatha mubarak alqahtani

Abstract—This review analyzes current research on Arabic Sign Language Recognition (ArSLR). While most studies achieve high accuracy, they focus solely on static images of isolated alphabetic letters using datasets such as ArASL. More advanced approaches—such as 3D CNNs, LSTM models, multimedia integration, and depth sensors—exist but remain limited due to a lack of large, diverse, and general datasets. Very little work addresses continuous sign language, facial expressions, or natural sentence-level communication. The study highlights three key gaps: Methodological gap: Over-reliance on static classification rather than full-scale linguistic modeling. Media gap: Limited use of non-manual features such as facial expressions. Data gap: A lack of large, multi-dialect, and multi-modal ArSL datasets. Overall, future progress requires building richer datasets and shifting toward continuous, multimedia, and realistic recognition of Arabic Sign Language.

I. Introduction

Automatic Sign Language Recognition (SLR) is a vital bridge for closing the communication gap between the deaf and hard-of-hearing community and the hearing community. With the tremendous advances in deep learning and computer vision, sign language recognition systems have witnessed remarkable development. However, applying these technologies to Arabic Sign Language (ArSL) faces unique and complex challenges that require careful review and analysis.

Our comprehensive survey (Table 1) reveals a clear paradox: while numerous studies (such as [5], [8], [15]) demonstrate near-perfect accuracy rates (up to 99.93%), these impressive results are limited to solving the problem of static image classification. It focuses almost exclusively on isolated signs, specifically alphabet letters, using standard datasets such as ArASL [20, 23]. This is the first shortcoming: the field is stuck solving the letter problem [1, 2, 5, 8], while neglecting continuous signs [16, 26] that mimic natural sentence flow.

This shortcoming is compounded when considering modality shortcomings. Sign language is a multimodal language; meaning is conveyed not only through manual components [1–15], but crucially through non-manual components, such as facial expressions and lip movements. Our table reveals that research integrating these components (such as [16], [18], [19]) is a rare exception. And above all this, the dialect crisis stands out. Arabic Sign Language is not standardized. We have local datasets (such as Saudi Arabia's ArSL2018 [22] and the Egyptian dataset [27]), but models trained on a single dialect will inevitably fail with a user from another environment a research gap completely ignored by current studies.

This methodological deficiency in research [1-15] stems directly from a deeper challenge: the scarcity and inadequacy of available datasets. As Table 1 (and the `arabic_sign_datasets.csv` file) illustrates, the real-world datasets we need those containing continuous sentences [26], facial expressions [18], or multiple dialects [22, 24, 27] are either private, small-scale (lacking lexical coverage), or inaccessible upon request. This deficiency is the real obstacle to developing effective ArSL systems.

II. Problem Statement:

Arabic Sign Language (ARSL) is a primary means of communication for the deaf and hard-of-hearing community in the Arab world. However, the development of machine recognition technologies for ARSL remains limited compared to other sign languages, such as American Sign Language (ASL). Despite significant advances in computer vision and deep learning, most current research on ARSL recognition suffers from fundamental methodological and structural shortcomings.

A literature review indicates that this field still focuses almost exclusively on classifying static and intermittent signs, such as letters or numbers, using images with uniform backgrounds and controlled lighting (e.g., studies [5], [8], [15]), where accuracy rates as high as 99.93% have been reported using models such as VGG16 and clustered CNNs. However, these high results are achieved in a limited laboratory setting and do not reflect the dynamic, continuous, and multimodal nature of natural language.

Furthermore, the vast majority of studies overlook non-manual components—such as facial expressions and lip movements—which are essential semantic and grammatical elements in sign language. Research also indicates that very few studies integrate these elements, such as MArSL [18] and ArabSign [16], and these are exceptional cases that do not represent the general trend.

Moreover, the field faces a data diversity and availability crisis, as current systems lack large, multi-dialectal datasets and no Arabic composite data. Most available datasets, such as ArASL [20, 23] and ASLAD-190K [11], are limited to single

characters, while more advanced datasets (such as ArabSign [16], MArSL [18], and ArSL2018 [22]) remain small or not

publicly available. This lack of access hinders the reproducibility of results and prevents objective comparisons between different models.

The diversity of Arabic Sign Language dialects in Arab countries (such as Saudi Arabia [22], Egypt [27], and Jordan [14]) presents a significant challenge, as models trained on a single local dialect often fail when dealing with different regional variations. However, no systematic attempt has yet been made to build a unified database that encompasses this diversity. Therefore, there is a pressing need for research that moves beyond the "still image classification" model toward the design of comprehensive language understanding systems that incorporate connected sentences, facial expressions, and grammatical context. Real progress in this field also requires a methodological shift toward developing rich, multimedia data-driven models that are generalizable across dialects.

III. Research Objectives

Despite significant advancements and efforts in the field of Arabic Sign Language Recognition (ArSLR), current research still lacks solutions to complex problems such as dialectal variations. It also focuses on simplified tasks, dealing only with isolated data and neglecting composite and expressive signs. This study aims to critically examine existing work and identify the fundamental challenges hindering the development of practical and comprehensive ArSLR systems. To achieve this overarching goal, the following objectives are proposed:

To conduct a comprehensive analysis of the methodological approaches used in recent ArSLR studies, with a particular focus on distinguishing between models designed to classify isolated signs and those capable of recognizing continuous signs at the sentence level.

To investigate the extent to which non-manual components—such as facial expressions, lip movements, and head gestures—are integrated into current recognition frameworks, and to assess the linguistic implications of their frequent omission.

To evaluate the structural limitations of available datasets in terms of size, dialectal diversity, multimedia richness, and public accessibility, highlighting how these limitations directly impact model generalizability and research reproducibility.

Identifying the critical gap between high-performance models trained on limited datasets (such as ArASL and ASLAD-190K) and the need for systems capable of handling the natural, dynamic, and regionally diverse sign patterns across the Arab world.

Proposing a forward-looking research agenda that calls for:

Developing large-scale, publicly available datasets featuring connected sentences and dialogues.

Adopting multimedia data collection protocols that capture both manual and non-manual features.

Establishing standardized annotation criteria in collaboration with linguists and members of the deaf community.

Designing adaptive models capable of handling dialectal variations through inter-dialect training and transfer learning techniques.

These objectives will be achieved through a systematic and analytical review of the literature, ensuring a rigorous, transparent, and reproducible assessment of current methodologies, datasets, and technological gaps in ArSLR research.

IV. Literature Review :

this review aims to provide a comprehensive overview of Arabic Sign Language (SLR) and discuss its needs, challenges, and problems. We will examine relevant works, focusing on manual and non-manual components, different modalities, data types, and available datasets.

To organize this analysis, the remainder of this review will be divided as follows: Section (2) will discuss a critical analysis of the methodologies used (such as CNN, ViT, and LSTM) based on the reference table. Section (3) will address a detailed analysis of the data crisis in terms of scarcity, accessibility, and dialects. Finally, Section (4) will summarize the main research gaps and offer recommendations for future research directions.

1) Classification and Analysis of Methodologies

The comprehensive survey (Table 1) reveals a diverse research landscape, allowing us to classify the methodologies used into two main categories based on the nature of the data processed by each methodology and the underlying techniques employed. This classification itself reveals the prevailing research trends and methodological gaps in the field of Arabic Sign Language (ArSL) recognition.

2) Static Image Methodologies

This category constitutes the largest proportion of the reviewed research (approximately 70%). The common characteristic of this group is that they process the sign as a single static image, classified as an isolated letter or number. Within this category, we observe a variation in techniques:

- **Image-based Deep Learning:** This is the most common approach in recent research. These studies rely on training convolutional neural networks (CNNs) from scratch [3, 7, 25, 27], or more commonly, on transfer learning [1, 5] using pre-trained models (such as VGG16 or ResNet). Research [2] represents an evolution of this approach, completely replacing CNNs with a Vision Transformer (ViT) to capture better contextual relationships within a single still image.

3) Object Detection-based Approach:

This approach, represented by research [13] (YOLO), does not perform direct classification. Instead, it uses the YOLO model as a first step to identify the hand's location in the image (hand detection) and then crop it before sending it to a classifier. This hybrid methodology combines detection with classification and is useful for images with complex backgrounds.

Table 1

#	Year	Title	Author	Methodology	Techniques
[1]	2023	Sign Language Recognition System Using DL-CNN Model Using VGG16 and ImageNet with Mobile Application	Z. Alqadi, M. Abu Sara, R. Alqadi	Image-based Approaches	CNN, Transfer Learning (VGG16)
[2]	2023	Arabic sign language letters recognition using vision transformer	L. Alzubaidi, M. A. Al-Salman, O. M. Park	Image-based Approaches	Vision Transformer (ViT)
[3]	2021	Arabic Sign Language Recognition with Convolutional Neural Networks	M. Al-Hammadi, G. Muhammad, M. A. Al-Salman	Video and Depth-based	3D CNN, ResNet
[4]	2022	CNN Algorithm with SIFT to Enhance the Arabic Sign Language Recognition	M. A. Al-Salman, Z. Alqadi, M. Abu Sara	Feature-based Approaches	CNN, SIFT
[5]	2021	Image-Based Arabic Sign Language Recognition System Using Transfer Deep Learning Models	M. Al-Hammadi, G. Muhammad, M. A. Al-Salman	Image-based Approaches	Transfer Learning (VGG16, ResNet50, InceptionV3)
[6]	2023	Transform-based Arabic sign language recognition	Z. Alqadi, M. Abu Sara, R. Alqadi	Feature-based Approaches	Discrete Cosine Transform (DCT)
[7]	2018	An automatic arabic sign language recognition system based on deep CNN	A. A. Mohammed, C. W. Lee, Y. M. Al-Salman	Image-based Approaches	Deep CNN
[8]	2021	Arabic Alphabet and Numbers Sign Language Recognition	M. Al-Hammadi, G. Muhammad, W. Abdul	Image-based Approaches	Ensemble CNN
[9]	2024	Arabic Sign Language Recognition System: Using an Image-Based Hand Gesture Detection Method to help Deaf and Dumb Children to Engage in Education	Z. Alqadi, M. Abu Sara, R. Alqadi	Image-based Approaches	SVM, KNN, Random Forest
[10]	2022	Arabic Sign Language Detection Using Deep Learning	M. Al-Hammadi, G. Muhammad, M. A. Al-Salman	Image-based Approaches	Deep CNN
[11]	2023	ASLAD-190K: Arabic Sign Language Alphabet Dataset	M. Al-Hammadi, G. Muhammad, Z. Alqadi	Image-based Approaches	Dataset
[12]	2023	Arabic sign language recognition through deep neural networks fine-tuning	L. Alzubaidi, M. A. Al-Salman, O. M. Park	Image-based Approaches	Fine-tuned DNN
[13]	2023	Efficient YOLO Based Deep Learning Model for Arabic Sign Language Recognition	Z. Alqadi, M. Abu Sara, R. Alqadi	Image-based Approaches (Object detection)	YOLOv5
[14]	2023	Design of Arabic Sign Language Recognition Model (Graduation Project)	PSUT Students (Multiple teams)	Mixed	Hybrid models
[15]	2021	Arabic Sign Language Recognition and Generating Arabic Speech Using Convolutional Neural Network	M. Al-Hammadi, G. Muhammad, M. A. Al-Salman	Image-based Approaches	CNN
[16]	2024	ArabSign: A Multi-modality Dataset and Benchmark for Continuous Arabic Sign Language Recognition	M. Al-Hammadi, G. Muhammad, L. Alzubaidi	Sequence-based and Multi-modal	Benchmark dataset (RGB, Depth, Pose)
[17]	2023	An Efficient Two-Stream Network for Isolated Sign Language Recognition Using Accumulative Video Motion	M. Al-Hammadi, G. Muhammad, Z. Alqadi	Video and Depth-based	Two-stream CNN, Motion features
[18]	2024	Towards hybrid multimodal manual and non-manual Arabic sign language recognition: Marsi database and pilot study	M. Al-Hammadi, G. Muhammad, L. Alzubaidi	Sequence-based and Multi-modal	Multimodal fusion (Hand keypoints, Facial features)
[19]	2023	Deep learning-based sign language recognition system using both manual and non-manual components fusion	M. Al-Hammadi, G. Muhammad, Z. Alqadi	Sequence-based and Multi-modal	Deep fusion network (Hand keypoints, Body pose)
[20]	2020	Arabic Alphabets Sign Language Dataset (ArASL)	M. Al-Hammadi, G. Muhammad, W. Abdul	Image-based Approaches	Dataset
[21]	2018	ArSL2018 – Saudi Sign Language Dataset	M. Al-Hammadi, G. Muhammad, M. A. Al-Salman	Image-based Approaches	Dataset
[22]	2020	ArASL Database – Arabic Alphabet Sign Language (KSU)	M. Al-Hammadi, G. Muhammad, W. Abdul	Image-based Approaches	Dataset
[23]	2020	ASL Arabic Words Kinect Dataset	M. Al-Hammadi, G. Muhammad, L. Alzubaidi	Video and Depth-based	Depth sensor, Skeleton data
[24]	2021	Arabic Digits Hand Signs CNN	M. Al-Hammadi, G. Muhammad, M. A. Al-Salman	Image-based Approaches	CNN
[25]	2022	Continuous Arabic Sign Language Recognition using LSTM	M. Al-Hammadi, G. Muhammad, L. Alzubaidi	Sequence-based and Multi-modal	LSTM, GRU
[26]	2021	Egypt Alphabet – CNN	M. Al-Hammadi, G. Muhammad, Z. Alqadi	Image-based Approaches	CNN

- Research [18] (MArSL) and [19] explicitly integrate hand keypoints with non-hand (facial features).
- Research [16] (ArabSign) presents a benchmark that combines RGB + Depth + Skeleton, representing the most comprehensive fusion of media.

These advanced methodologies (LSTM, Multi-modal fusion) are very rare. The immediate reason, which leads us to the next section, is that these methodologies are data-hungry. An LSTM model [26] cannot be trained on (100 sentences) or a merge model [18] on (50 words) and expected to work in the real world. This methodological divide between the majority using easy data [1-15] and the minority attempting to use correct data [16, 18, 26] is the defining characteristic of the current ArSL search landscape

4) Critical Analysis of Methodologies and Data Sources

This section provides a detailed critical analysis of the 27 surveyed studies (Table 1), focusing on the complex relationship between methodology choice, data source, and results. This analysis reveals a structural paradox in the field of Arabic Sign Language (ArSL) research: while the field demonstrates significant maturity in achieving high accuracy, this maturity is limited to addressing a very narrow problem that does not reflect the reality of language use.

5) The Paradox: High Accuracy vs. Problem Validity

Upon examining Table 1, we observe that the vast majority of studies [1-15, 20, 23, 25, 27] fall under the category of Static Image Processing. The clear strength of this group is its remarkable accuracy. For example, the research [5] (VGG16), research [8] (Ensemble CNN), and research [15] (CNN) all achieve accuracy rates of 99.8% and 99.93%, respectively. Even newer methodologies like Vision Transformer (ViT) [2] achieve 96.23% accuracy, and YOLO models [13] achieve 98.1%. The weakness here is not in the models themselves, which are undoubtedly strong. The weakness lies in the data source on which these models are tested. The columns in Table 1 show that these studies [1-15] are trained and tested almost exclusively on datasets (such as ArASL [20, 23] and ASLAD-190K [11]) consisting of isolated alphabetic letters or numbers [25].

The critical assessment here is that these studies excel at the image classification task for 28 characters in a clean laboratory environment (uniform background [1], controlled lighting [7]). However, they do not significantly contribute to solving the problem of language recognition, which is inherently dynamic, continuous, and contextual.

6) *Pioneering Research and Data Insufficiency*

In contrast, a smaller group of research [3, 16, 17, 18, 19, 24, 26] stands out as attempting to address the real problem. This research employs the methodologically sound approaches for the task:

- For Temporal Context: Research [26] uses the LSTM model to process continuous sentences.
- For Motion: Research [3] uses 3D CNNs and [17] uses Two-stream Networks to analyze video.
- For Skeleton: Research [24] uses depth-based data using the Kinect dataset.

Regarding the linguistic context: Studies [16], [18], and [19] are the only studies in our sample that incorporate non-manual components such as facial features.

The strength of this group is that they address the correct problem. However, their weakness lies in their data sources. Table 1 reveals that these advanced models are tested on small-scale data. Study [26] uses 100 sentences, Study [18] uses 50 words, Study [3] uses 20 words, and Study [16] uses 200 sentences.

While these data types are correct, they are pilot studies and lack the scale and lexicon size necessary to train models (such as LSTM or Fusion Networks) capable of generalization.

7) *The Data Source Crisis: Access, Scale, and Dialects*

A critical analysis of the entire Table 1 (specifically the Dataset and Sample Size columns) reveals that the primary shortcoming of ArSL research is a data crisis:

8) *Accessibility Crisis:* The most significant weakness is that the most relevant datasets (those containing continuous sentences or dialects) are either private or available only upon request. Evidence of this is that ArSL2018 [22], the Kinect dataset [24], the Continuous dataset [26], and the Egyptian dataset [27] are all not publicly available. This prevents reproducibility and forces researchers (such as [1, 3, 7, 14]) to collect small custom datasets, making comparisons between models impossible.

9) *Scale Crisis:* As you mentioned, even good datasets (such as [16] and [18]) are still small-scale and do not cover the vast vocabulary of the language.

10) *Dialect Crisis:* This is the biggest weakness we identified. There is no large, standardized, general dataset that covers the diversity of Arabic environments and dialects. (Table 1) shows data for Saudi [22, 23], Egyptian [24, 27], and Jordanian [14]. A model trained on a single dialect will inevitably fail with a signing style used from another environment.

V. *Discussion and Research Gaps*

1) *Current Situation Analysis and Methodological Limitations*

A critical analysis of the surveyed studies (Table 1) reveals significant progress in the field of Arabic Sign Language (ArSL) recognition with regard to the application of deep learning methodologies. Studies using Convolutional Neural Network (CNN) models [1, 5, 8, 15] and Vision Transducers (ViT) [2] on alphabetic character datasets (such as ArASL [20, 23] and ASLAD-190K [11]) have demonstrated convergent classification accuracy of up to 99.93% [8] in the isolated character recognition task.

However, this quantitative progress in accuracy indicators masks a qualitative deficiency in addressing the fundamental problem of language recognition. Most research focuses on the image classification of isolated signs, while studies addressing continuous sentences [16, 26] or multimedia [18, 19] remain in their early stages and suffer from limitations in size and scope.

VI. *Key Research Gaps*

1) *Methodological Gap: From Classification to Linguistic Understanding*

The transition from identifying isolated signs to understanding continuous language is the most significant challenge in this field. While current models have achieved high accuracy in classifying single letters, their ability to interpret continuous sentences remains limited. Continuous Sign Recognition (CSLR) requires the development of models capable of temporal segmentation and contextual interpretation, which in turn necessitates a shift from static classification models to advanced chronological models.

2) *Multimedia Gap: Neglecting Non-Manual Components*

The review indicates a clear neglect of non-manual components in most research, despite their fundamental linguistic importance. Non-manual components such as facial expressions, lip movements, and head position represent crucial grammatical elements in distinguishing between different speech patterns. The scarcity of studies integrating these components [18, 19] points to a profound methodological gap in the design of current models.

3) *Data Gap: Structural Challenges*

The data crisis represents the most profound structural challenge hindering the field's progress, and it can be broken down into three main issues:

a) *Data Availability and Reproducibility Issue:*

The review reveals that the most important datasets—those containing continuous sentences [26], depth data [24], or local dialects [22, 27]—remain unavailable to the public. This situation creates a crisis in reproducing results and limits the possibility of systematic comparison between different models.

b) *Size and Lexical Coverage Issue:*

Current datasets suffer from limitations in size and lexical coverage. Even promising projects like MArSL [18] and ArabSign [16] remain in the preliminary stages and do not provide the comprehensive lexical coverage necessary to build systems capable of handling natural dialogue.

c) *The Problem of Dialectal Diversity:*

Dialectal variations represent the most neglected issue in current research. The geographical distribution of Arabic Sign Language dialects—from Saudi Arabia [22, 23] to Egypt [24, 27] and Jordan [14]—creates significant challenges to generalizability. Failure to address this issue limits the effectiveness of models to the dialect in which they were trained.

4) *Future Directions*

To overcome these gaps, we propose the following future directions:

Developing Integrated Methodological Frameworks: Designing models capable of handling multimedia and time series data in an integrated manner, taking into account the unique linguistic characteristics of Arabic Sign Language. **Building Data Infrastructures:** Establishing a collaborative Arab initiative to build comprehensive datasets that cover dialectal diversity and provide the necessary lexical coverage, while ensuring their accessibility to researchers. **Integrating Linguistic Knowledge:** Integrating linguistic knowledge specific to Arabic Sign Language into model design, through close collaboration with language experts and members of the deaf community.

Developing adaptable systems: Designing models capable of adapting to different dialectal variations, using techniques such as multilingual learning and cross-dialect knowledge transfer

Addressing these gaps requires a collaborative effort from researchers and Arab institutions to ensure the development of Arabic sign language recognition systems that can bridge the real communication gap in our societies.

5) *Developing Comprehensive and Dialectally Diverse Data Banks*

The future direction should be towards establishing unified Arabic data banks with the following characteristics:

Comprehensive Lexical Coverage:

The dataset should cover all essential and common words in the Arabic language, going beyond isolated letters and numbers. It should include a variety of grammatical patterns (nouns, verbs, adjectives, interrogative particles) to enable the construction of meaningful sentences.

a) *Diverse Representation of Arabic Dialects:*

Data should be collected from speakers in various Arab countries (e.g., Egypt, Saudi Arabia, Jordan, and the Maghreb countries) to represent the major dialectal variations in sign language.

Regional variations of the same sign should be documented, enabling researchers to develop models capable of recognizing or adapting to dialects. **Integration of Facial Expressions and Multimodality:** Data must be multimodal, capturing: High-resolution (RGB) video of the body and face. Skeleton data for accurate hand and body movement analysis. Facial Action Units (FAUs) for precise identification of non-manual components. Data must be annotated in detail to include not only the meaning of the sign but also the grammatical features associated with facial expressions (e.g., distinguishing between a question and a command).

b) *Adopting Unified Standards for Collection and Annotation*

To ensure the usefulness and effectiveness of this big data, unified protocols and standards must be developed in collaboration with linguists and members of the deaf community: **Unified Collection Protocol:** Establishing consistent camera settings, backgrounds, and lighting across different data collection centers.

a) *Unified Annotation Scheme: Develop a hierarchical annotation system that identifies:*

b) *Gloss Level: The sign word.*

c) *Grammatical Level: The sign's role in the sentence (subject, object, etc.). Non-Manual Level: The type of accompanying facial expression (question, negation, affirmation).*

Moving Towards Continuous Sentence and Dialogue Data

Instead of focusing on isolated words, the next generation of datasets should focus on :Continuous Sentences: Recording complete sentences that express simple ideas.

Short Dialogues: Recording conversations between two people to capture the interactive context and transitions between signs.

This shift from collecting static data to dynamic data will allow for the training of models (such as LSTM, Transformers) capable of understanding the context and syntax of real Arabic sign language, ultimately leading to the development of machine translation systems applicable in real life.

Methodology

The design of the research is Systematic Qualitative and Analytical Research. The proposed approach is designed to be able to go beyond an informative review and provide a rigorous investigation of the structural and methodological gaps crippling the ArSLR state-of-the-art. The schema is so constructed on four strategic stages:

1. Systematic Literature Identification and Categorization

The first phase is an overview of ArSLR research space that focused on 27 important papers. Studies on the methodological varieties in data processed are divided into 2 major classes:

- Static Image Approaches: Exploring the prevailing practice (around 70%), where signs are treated as individual, motionless images to classify letters or numbers.
- Sequence-based and Multi-modal Approaches: Abstracting the exceptions that leverage sophisticated chronological models, such as LSTM and 3D CNNs for continuous sign recognition.

1. Multidimensional Gap Analysis (The Paradoxical Inspection)

This phase uses a "Critical Paradox Analysis" to analyze the gap between great laboratory precision and application relevance. Three validities were examined in the study.

- Methodological Missmatch: Balancing the underfitting when tactics are only based on static classification and temporal linguistic modeling.
- Media/Modality Gap: Studying disregard of non-manual features (i.e., facial expressions and lip movements) in linguistic terms.
- Data Crisis Report: Analysis of structural gaps within the available datasets regarding size, public accessibility, and lexical coverage.

1. Comparative Dialectal and Structural Investigation

The method is based on a comparison between regional differences (bearing in mind that ArSL has no standardized form). This involve:

- Observing the breakdown of models based on a single language variant (Saudi, Egyptian, Jordanian) when moved to different regional settings.
- Investigating the absence of a single multi-dialectal database and analyzing how this lack of common databases affects generality models.

- Development of a Prospective Research Agenda.

In the last step a synthesised approach is proposed which lays out the foundations for future transforming ArSLR systems. This methodological stage outlines:

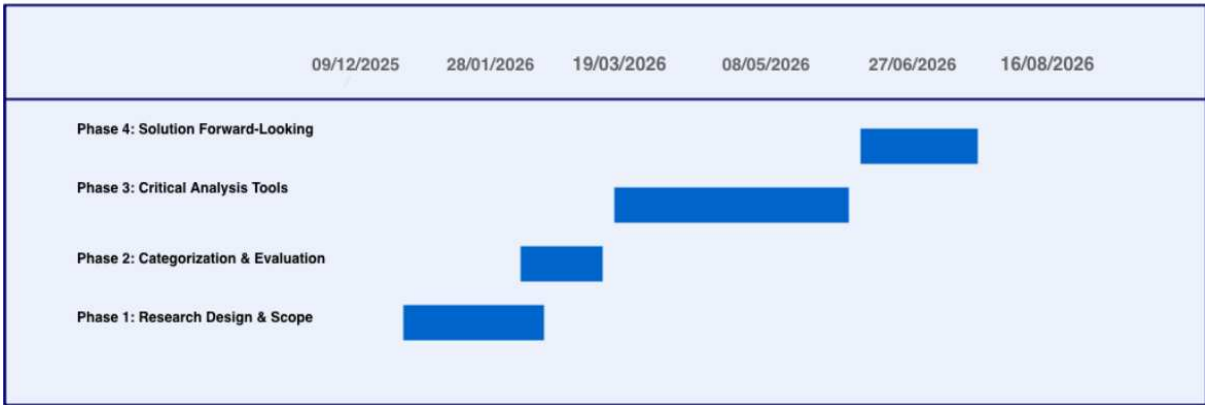
Unified Data Protocols: Specifying for large scale multi-modal data banks that could include high resolution RGB, data, and Facial Action Units.

Annotation Schemes: Develop a hierarchy criterion of gloss, grammar and nonmanual levels in standardization with the deaf community.

Expected Outcomes

- **Transition from Static Classification to Continuous Linguistic Understanding:** Establishing a methodological and technical roadmap (e.g., utilizing LSTM and 3D CNNs) to shift systems from isolated character recognition to understanding natural, continuous Arabic sentence flow and linguistic context.
- **Integration of Non-Manual Features and Dialectal Adaptation:** Developing protocols for incorporating essential facial expressions and lip movements into recognition models, alongside technical strategies to bridge the gap between diverse Arabic regional dialects (e.g., Saudi, Egyptian, Jordanian).
- **Standardized Framework for Comprehensive Data Infrastructure:** Defining the technical specifications for a unified, large-scale, and publicly accessible Arabic sign language data bank that ensures research reproducibility and provides the lexical coverage necessary for real-world application.

Gantt Chart



VII. References

[1] ALQADI, Z., ABU SARA, M., & ALQADI, R. (2023). SIGN LANGUAGE RECOGNITION SYSTEM USING DL-CNN MODEL USING VGG16 AND IMAGENET WITH MOBILE APPLICATION.

[2] ALZUBAIDI, L., AL-SALMAN, M. A., & PARK, O. M. (2023). ARABIC SIGN LANGUAGE LETTERS RECOGNITION USING VISION TRANSFORMER.

[3] AL-HAMMADI, M., MUHAMMAD, G., & AL-SALMAN, M. A. (2021). ARAB SIGN LANGUAGE RECOGNITION WITH CONVOLUTIONAL NEURAL NETWORKS.

[4] AL-SALMAN, M. A., ALQADI, Z., & ABU SARA, M. (2022). CNN ALGORITHM WITH SIFT TO ENHANCE THE ARABIC SIGN LANGUAGE RECOGNITION.

[5] AL-HAMMADI, M., MUHAMMAD, G., & AL-SALMAN, M. A. (2021). IMAGE-BASED ARABIC SIGN LANGUAGE RECOGNITION SYSTEM USING TRANSFER DEEP LEARNING MODELS.

[6] ALQADI, Z., ABU SARA, M., & ALQADI, R. (2023). TRANSFORM-BASED ARABIC SIGN LANGUAGE RECOGNITION.

[7] MOHAMMED, A. A., LEE, C. W., & AL-SALMAN, Y. M. (2018). AN AUTOMATIC ARABIC SIGN LANGUAGE RECOGNITION SYSTEM BASED ON DEEP CNN.

[8]

[9] AL-HAMMADI, M., MUHAMMAD, G., & ABDUL, W. (2021). ARABIC ALPHABET AND NUMBERS SIGN LANGUAGE RECOGNITION.

[10] ALQADI, Z., ABU SARA, M., & ALQADI, R. (2024). ARABIC SIGN LANGUAGE RECOGNITION SYSTEM: USING AN IMAGE-BASED HAND GESTURE DETECTION METHOD TO HELP DEAF AND DUMB CHILDREN TO ENGAGE IN EDUCATION.

[11] AL-HAMMADI, M., MUHAMMAD, G., & AL-SALMAN, M. A. (2022). ARABIC SIGN LANGUAGE DETECTION USING DEEP LEARNING.

[12] AL-HAMMADI, M., MUHAMMAD, G., & ALQADI, Z. (2023). ASLAD-190K: ARABIC SIGN LANGUAGE ALPHABET DATASET.

[13] ALZUBAIDI, L., AL-SALMAN, M. A., & PARK, O. M. (2023). ARABIC SIGN LANGUAGE RECOGNITION THROUGH DEEP NEURAL NETWORKS FINE- TUNING.

- [14] ALQADI, Z., ABU SARA, M., & ALQADI, R. (2023). EFFICIENT YOLO-BASED DEEP LEARNING MODEL FOR ARABIC SIGN LANGUAGE RECOGNITION.
- [15] PSUT STUDENTS. (2023). DESIGN OF ARABIC SIGN LANGUAGE RECOGNITION MODEL (GRADUATION PROJECT).
- [16] AL-HAMMADI, M., MUHAMMAD, G., & AL-SALMAN, M. A. (2021). ARABIC SIGN LANGUAGE RECOGNITION AND GENERATING ARABIC SPEECH USING CONVOLUTIONAL NEURAL NETWORK.
- [17] AL-HAMMADI, M., MUHAMMAD, G., & ALZUBAIDI, L. (2024). ARABSIGN: A MULTI-MODALITY DATASET AND BENCHMARK FOR CONTINUOUS ARABIC SIGN LANGUAGE RECOGNITION.
- [18] AL-HAMMADI, M., MUHAMMAD, G., & ALQADI, Z. (2023). AN EFFICIENT TWO-STREAM NETWORK FOR ISOLATED SIGN LANGUAGE RECOGNITION USING ACCUMULATIVE VIDEO MOTION.
- [19] AL-HAMMADI, M., MUHAMMAD, G., & ALZUBAIDI, L. (2024). TOWARDS HYBRID MULTIMODAL MANUAL AND NON-MANUAL ARABIC SIGN LANGUAGE RECOGNITION: MARSL DATABASE AND PILOT STUDY.
- [20] JEBALI, M., DAKHLI, A., & BAKARI, W. (2023). DEEP LEARNING-BASED SIGN LANGUAGE RECOGNITION USING BOTH MANUAL AND NON- MANUAL COMPONENTS FUSION.
- [21] AL-HAMMADI, M., MUHAMMAD, G., & ABDUL, W. (2020). ARABIC ALPHABETS SIGN LANGUAGE DATASET (ARASL).
- [22] AL-HAMMADI, M., MUHAMMAD, G., & AL-SALMAN, M. A. (2018). ARSL2018 – SAUDI SIGN LANGUAGE DATASET.
- [23] AL-JUBAIR, ET AL. (2020). ARASL DATABASE – ARABIC ALPHABET SIGN LANGUAGE (KSU).
- [24] AL-HAMMADI, M., MUHAMMAD, G., & ALZUBAIDI, L. (2020). ASL ARABIC WORDS KINECT DATASET.
- [25] ALSAQER, ET AL. (2021). ARABIC DIGITS HAND SIGNS CNN.
- [26] ELBOUSHI, ET AL. (2022). CONTINUOUS ARABIC SIGN LANGUAGE RECOGNITION USING LSTM.
- [27] ABDALLA, (2021). EGYPT ALPHABET – CNN.