# Enhancing Communication Accessibility: Real-Time Recognition and Synthesis of Arabic Sign Language Gestures Using Long Short-Term Memory

Mina Nagy Gaber Sorial[1], Rodaina Abdelsalam[2]*, Mayar Ali[3]*, Hesham Hassan[4]
Faculty of Computers and Information Technology, Egyptian E-learning University, Cairo, Egypt[1, 2, 3]
Faculty of Computing and Artificial Intelligence, Cairo University, Cairo, Egypt[4]

***Abstract*—The Arabic Sign Language Recognition Research aims to develop a real-time system that accurately recognizes Arabic Sign Language (ArSL) gestures and translates them into both text and speech. This Research leverages the KArSL-502 Dataset, which contains 502 unique Arabic signs, to train a deep learning model using Bidirectional Long Short-Term Memory (LSTM) networks. LSTMs are particularly suited for capturing the temporal patterns of sign language gestures, which often involve sequential hand movements. The system integrates advanced image processing techniques such as Mediapipe and Handtrack for detecting and extracting hand landmarks, followed by key point adjustments to ensure consistency across gestures. The model's performance was evaluated using categorical accuracy, achieving a training accuracy of 98% and a testing accuracy of 96%, demonstrating the model's ability to generalize well to unseen data. Additionally, the proposed system includes text-to-speech functionality via Google Text-to-Speech (Gtts), enabling real-time vocalization of recognized gestures, thus facilitating communication between sign language users and non-sign language speakers. The system's high accuracy and fast processing time (measured in milliseconds per gesture) make it suitable for real-time applications.***

***Keywords*—Arabic Sign Language; bidirectional LSTM; machine learning; text-to-speech; real-time processing; accessibility; sign language translation**

## I. Introduction

### A. Background

Speaking and listening represent fundamental civil liberties, enabling individuals to exchange information and fully participate in society. Nevertheless, for those persons who are deaf or have hearing loss, effective communication through speaking and listening is impossible. However, for these people, sign language is the major way of interaction, which enables them to make use of language and communicate effectively. Nevertheless, there are some known and critical restraints to the use of sign languages; these are the absence of accessible and efficient resources, especially in parts of the world where the deaf population is discriminated against or neglected [1].

Arabic Sign Language (ArSL) is the major sign language used by many deaf people in many Arabic-speaking nations.

ArSL, in the same manner as any other sign language in the world, is grounded in its syntactic and semantic properties that are dissimilar to the Arabic language. It has a significant function of enabling deaf people to communicate with people with disabilities in their community. Nevertheless, some constraints defeat the functionality and utilization of ArSL in its entirety. One of the primary questions is the presence of regional subtypes within ArSL [2]. Thus, even within the national language, different countries or even regions may use some other signs to represent the same idea, which makes it impossible to unify all the details of the language.

Further, there is a scarcity in the available resources that tutor and teach ArSL in the many Arabic-speaking parts of the globe. The present lack of educational literature, training courses, and qualified sign language interpreters also limits the opportunities of deaf people and their educational achievements, as well as social integration. Still, most deaf people can only do so with the assistance of their relatives or friends, which is very unreliable, though in business or legal circumstances. The absence of structured materials also ensures that the hearing community does not learn ArSL [3].

Moreover, technological communication aids for the sign language users and non-users are critically rare. Despite tremendous progress in technology and its fines in different areas, deaf people still have hardly any opportunities to have flawless communication systems. This gap keeps deaf people at a disadvantage when trying to access basic services, including healthcare, education, and job opportunities. For instance, in medical emergencies, the deaf person may be unable to explain their or their relatives' ailments or respond to questions due to a lack of a sign language interpreter, and this causes misunderstandings which lead to improper treatment [4].

Lack of real-time sign language recognition as well as synthesis hinders the integration of the two parties (deaf and hearing impaired). Professionally provided sign language interpretation is often not guaranteed and too frequently is not in real time or in all environments. That is why the demand for high ideas like signing recognition of Arabic Sign Language gestures and translating them into text or voice in real time is put forward [5]. Such technologies would make it easier for the deaf to share information more easily in social, career, and academic situations.

### B. Problem Statement

In many Arabic-speaking countries, the lack of accessible communication tools between the deaf and hearing communities has led to exclusion and social isolation. While sign language

*Corresponding author.

interpretation services are available in some cases, they are often not enough to ensure real-time, fluid communication in everyday interactions [6]. The absence of real-time recognition and synthesis of ArSL gestures further limits communication, creating a need for advanced technological solutions.

This research contributes significantly to enhancing accessibility for the deaf and hard-of-hearing community, providing a robust framework for Arabic Sign Language recognition. Furthermore, it serves as a foundation for future advancements, including multimodal gesture recognition and expansion to other sign languages, ultimately promoting inclusion and better communication in various domains such as healthcare, education, and customer service.

### C. Research Objectives

The primary objective of this study is to enhance communication accessibility for deaf individuals in Arabic-speaking communities through the development of a real-time recognition and synthesis system for Arabic Sign Language gestures. The specific goals of this research include:

- Developing a system that can accurately recognize ArSL gestures in real-time.
- Creating a method for synthesizing these gestures into text or spoken language for non-signing users.
- Investigating the technical challenges and solutions related to the recognition of ArSL, including the use of machine learning and computer vision technologies.
- Proposing a framework for integrating these systems into everyday life to promote inclusion and accessibility.

This research has the potential to revolutionize communication for the deaf community by providing a real-time, reliable system for translating Arabic Sign Language into accessible formats. By focusing on Arabic Sign Language, the study aims to address a critical need in Arabic-speaking regions where deaf individuals often encounter language barriers that prevent them from fully participating in society. The findings of this research could not only improve the quality of life for deaf individuals but also foster a more inclusive society that values communication diversity.

## II. Literature Review

As shown in Table I, Arabic Sign Language Recognition (ASLR) has seen interesting development lately as many studies focus on raising accuracy, speed, and accessibility. Al Abdullah et al. [1] provided a broad overview of the field, covering present technological achievements and challenges without specifically providing a model or dataset.

A deep learning framework containing an attention mechanism and a CNN-LSTM hybrid model was presented by Abdul Ameer et al. [2]. Their approach, using a special dataset, fit dynamic Arabic signs with an incredible accuracy of 96.3%. Great accuracy in spotting challenging motions was the main advantage, even if the dataset was tiny and not publicly available.

Al-Hammadi et al. [3] reported a CNN-RNN network tailored for efficient hand gesture representation. Emphasizing its real-time characteristics, their model obtained an accuracy of 94.7% trained on RGB and LeapMotion datasets. Still, the approach proved inappropriate for motions at the phrase level. Bencherif et al. [4] proposed a skeleton-based recognition model by using 2D hand and body joints enabled by Kinect sensors. Their approach showed greater performance on continuous sign language but requires expensive technology; it proved 95.2% correct.

Alsulaiman et al. [5] made a significant contribution by building the largest Saudi Sign Language dataset to date. Though they did not propose a model, their dataset increased the training possibilities for forthcoming deep learning models. Rajest [6] developed a CNN-based real-time recognition system, but the dataset details were not particularly clear, and the accuracy was somewhat low at roughly 91%.

Al Ahmadi et al. [7] reported 94.8% accuracy using CNNs for ArSL interpretation with transfer learning. This approach has tremendous strength in decreased training time and improved generalization, even if it only addresses a limited range of motions. Alzubaidi et al. [8] proposed to interpret Arabic signs into voice by means of a wearable glove fitted with ML algorithms. Although depending on outside hardware was a drawback, this method provided mobility and 93.5% accuracy.

Al-Shamayleh et al. [9] methodically collected gesture recognition methods based on eyesight. Their study presented a useful taxonomy and classification, even if they lacked original experimentation. Aiouez et al. [10] reached 92.4% accuracy in real-time video recognition of Arabic signals using YOLOv5. Though its vocabulary was limited, the approach was rapid and effective.

Natarajan et al. [11] created a complete end-to-end deep learning pipeline for sign detection, text translation, and video generation. Although the system has a significant computing load, utilizing datasets like RWTH-PHOENIX, it achieved ~94% accuracy. Similarly, Ismail et al. [12] produced a fresh dataset for Egyptian Sign Language and used a comprehensive pipeline translating signs to Arabic speech. Although regional, it lacked a large gesture set and attained about 90% accuracy.

Buttar et al. [13] looked at hybrid models combining CNN and RNN architectures in order to classify both stationary and dynamic gestures with 95.6% accuracy. Still, dataset details were scant. Although the method was difficult, Marzouk et al. [14] improved training for ArSL identification and reached 96% accuracy by using deep learning with Atom Search Optimization.

Tharwat et al. [15] focused on conventional machine learning Arabic script recognition. Even limited to alphabet signs, their simple yet efficient approach achieved 93.7% accuracy. Abdulhamied et al. [16] employed LSTM with hand detection for American Sign Language, obtaining 94.5% accuracy and proving LSTM's temporal modeling ability.

Though not immediately relevant to ArSL, Hussain et al. [17] offered insights on Indian Sign Language systems, which can be useful for comparison research. Mukhedkar et al. [18] developed a gesture translation system with limited scalability but ~92% accuracy using CNN.

TABLE I. COMPARISON OF RELATED WORKS IN ARABIC SIGN LANGUAGE RECOGNITION

| Ref | Model Type | Dataset Type | Accuracy | Strengths | Limitations |
|---|---|---|---|---|---|
| [1] | Survey / Overview | - | - | Broad overview of the field, challenges & progress | No model or dataset provided |
| [2] | CNN + LSTM + Attn | Custom ArSL | 96.3% | High accuracy, dynamic support | limited dataset |
| [3] | CNN-RNN | LeapMotion, RGB | 94.7% | Real-time, efficient features | Not for full sentences |
| [4] | Skeleton-based | Kinect skeleton | 95.2% | Includes body context | Needs a depth camera |
| [5] | Dataset Construction | : Saudi Sign Language dataset | - | Largest SSL dataset to date | No model proposed |
| [6] | CNN (Real-time) | unclear | 91% | Real-time processing | Unclear Dataset details, lower accuracy |
| [7] | Transfer Learning | Custom ArSL | 94.8% | Fast training, generalization | Few gestures |
| [8] | Glove + ML | Hardware-based | 93.5% | Portable hardware | Requires gloves |
| [9] | Survey / Taxonomy | Vision-based approaches | - | Useful Taxonomy & Classification | No original experimentation |
| [10] | YOLOv5 (Object Det) | Custom ArSL video | 92.4% | Real-time processing | Small vocabulary |
| [11] | End-to-End DL | RWTH-PHOENIX, others | ~94% | Full pipeline | High computation |
| [12] | End-to-End DL Pipeline | Egyptian SL dataset | 90% | Dataset + speech integration | Small dataset, limited gesture set |
| [13] | Hybrid CNN-RNN | Custom ArSl | 95.6% | Works on both static & dynamic gestures | Scarce Dataset details |
| [14] | DL + Optimization | Custom ArSL | 96% | Fast convergence | Complex implementation |
| [15] | Classical ML (alphabet recognition) | Arabic alphabet signs | 93.7% | Simple, efficient | Limited to alphabet signs only |
| [16] | LSTM + Detection | ASL dataset | 94.5% | Temporal modeling | Not for Arabic |
| [17] | Survey / comparative | Indian SL systems | - | Useful insights for comparison | Not relevant to Arsl |
| [18] | CNN-based gesture translation | Custom dataset | 92% | Functional translation system | Limited scalability |
| [19] | HMM-based | Multilingual system | - | Gesture-to-voice translation | Outdated, lacks naturalism |
| [20] | Gesture synthesis + prosody | Synthetic coordination | - | Focused on realism in synthesis | Experimental, lacks practicality |
| [21] | Gesture synthesis (tracking-based) | - | - | Improved alignment via tracking algorithms | Computationally expensive |
| [22] | Attention-CNN | Arabic alphabet (112 static) | 96% | High accuracy for static letters | Limited to isolated characters |
| [23] | UNet3+ + Attention (Hybrid) | Isolated ArSL words | 97.4% | Strong accuracy | Dataset details unclear |
| [24] | CNN | ~1,500 ArSL words (static) | 95% | Good accuracy on word-level recognition | Static only, no sequence modeling |

Yang et al. [19] invented gesture-to–voice translating by using HMMs in a multilingual system. Though basic, the method is today out of date and devoid of naturalism. Working on the coordination of gesture synthesis with vocal prosody to boost realism, Fernández-Baena et al. [20], Lu and Song [21] improved synthesis alignment using better gesture tracking algorithms, while their technique was computationally costly.

Almufareh et al. [22] investigated alphabetic recognition in the Arabic script with an attention-focused CNN model from 112 static signs with 96% accuracy. However, this investigation applied to isolated characters without considering the sequences of continuous signs. Mohamed et al. [23] employed a hybrid attention-based UNet3+ for isolated ArSL word recognition, with an accuracy of 97.4%. However, dataset specs were not adequately recorded, thereby limiting reproducibility. Alzu'bi et al. [24] trained a CNN model on approximately 1,500 signed Arabic words with approximately 95% accuracy. The static-only sign recognition system was not equipped with modelling of the temporal sequence.

Especially in Arabic environments, it shows the diversity of methods and the rapid advancement in ASLR. Deep learning models—especially LSTM and CNN hybrids—rule the field, supported by growing datasets and coupling with TTS for comprehensive communication systems.

## A. Research Gap

From the studies, there are still major gaps, even if several studies show how effectively deep learning can identify Arabic Sign Language. Most modern systems either focus on limited dynamic indicators or stationary gestures and lack the suppleness to manage whole-sentence communication. Second, many of the data sets used are either small, domain-specific, or not publicly available, therefore limiting the generalizability of the models. Moreover, despite various research aiming at end-to-end systems, few offer seamless integration between gesture recognition, text generation, and natural-sounding speech synthesis. These flaws draw attention to the need for a single, scalable system able to control continuous Arabic Sign Language and provide high-quality audio outputs.

This research seeks to address the critical gap in real-time sign language recognition and synthesis systems, specifically tailored for Arabic Sign Language, to bridge the communication divide between deaf and hearing individuals.

## III. Methodology

### A. Introduction

The proposed approach followed in creating and assessing a real-time Arabic Sign Language (ArSL) recognition and synthesis system. The method comprises thorough procedures addressing the employed dataset, implementation environment, system architecture, data preprocessing, model training, and evaluation methods. The main objective is to build a dependable and accurate system able to identify ArSL motions and convert them into understandable text and voice in real time.

### B. Description of the Dataset

Training and testing in this research were carried out with the KArSL-502 dataset, a multimodal Arabic Sign Language (ArSL) (Fig. 1). The dataset consists of 502 unique Arabic sign gestures, encompassing extensive collections of static and dynamic signs that are applicable to everyday communication. Each gesture is executed by three native Deaf signing individuals, with three repetitions per signing person, producing 4,518 video samples and more than 135,000 frames. The dataset was strategically partitioned into 80% for training and 20% for testing to evaluate the model's performance. The test set was strictly held out from all tuning decisions to ensure that the reported accuracy reflects genuine generalization. The dataset consists of synchronized RGB video frames (1920×1080 resolution), depth maps (512×424), and 3D skeletal joint coordinates (25 upper-body joints per frame). This multimodal composition enables models to learn from dense spatial, time-aware, and depth-conscious patterns. Since the sign gestures are time-sequential, particularly dynamic signs that consist of prolonged motions, a Long Short-Term Memory (LSTM) network was utilized to extract time-related dependencies and motion paths between frames. The LSTM topology is particularly apt for dealing with the time-series nature of the information, allowing the model to store context across multiple time steps while discovering gesture-related motion patterns. While in this research only the LSTM model was utilized, the structure of the dataset is open to future exploration with multimodal integration as well as deeper topologies. Some limitations in the dataset are that gestures are isolated, and recording is controlled indoors, which can decrease the generalizability to everyday environments. However, the high-resolution imagery, fine-grained 3D skeleton annotations, as well as diverse signers, in this dataset render the dataset a comprehensive pilot study and a good starting point to study deep temporal sequence learning in Arabic Sign Language recognition tasks.

Fig. 1. Sample sequence from the KArSL-502 dataset illustrating the dynamic sign for the word “يحب”.

### C. Implementing Context

Python 3.6+ with many supporting libraries—TensorFlow 2.15.0, Keras, OpenCV, Mediapipe, Pandas, NumPy, Pillow, GTTS, Pygame, Tkinter, and Handtrack—was used in the model. These instruments allowed deep learning, real-time video processing, hand tracking, GUI development, and text-to-speech conversion. Along with a webcam (minimum 720p), microphone, and speaker, the system was designed and tested on a machine running an Intel Core i7 processor, 16GB RAM, and an NVIDIA RTX 3060 GPU to ensure high-performance inference. While the current implementation utilizes the Google Text-to-Speech (GTTS) library, which requires an active internet connection for high-quality cloud-based synthesis, the system's architecture is designed to integrate offline TTS engines in future iterations to ensure functionality in low-connectivity environments.

### D. Data Preparation and Preprocessing

By utilizing OpenCV and Pillow, the preprocessing workflow began with loading images from the dataset, where we examined attributes such as height, width, color channels—typically RGB—format (JPEG/PNG), and file size. Data cleansing, resizing, normalizing, and transformation included standard preprocessing actions. Using Mediapipe or Handtrack, hand landmark detection extracted key points for both hands. Techniques for landmark correction were used to reduce data noise and inconsistencies, therefore guaranteeing strong gesture representation, see Fig. 2.

### E. Architectural Model Design

Long Short-Term Memory (LSTM) layers are added before the dense layers, as these are particularly optimal when working with sequential information. Unlike standard RNNs, LSTMs do not suffer from the vanishing gradient problem and can maintain long-term dependencies, which is important for dynamic sign recognition, where the meaning of a gesture is likely to depend on its motion path over more than one frame. Here, two bidirectional LSTM layers were applied, each containing 64 units, allowing the model to capture temporal dependencies in both forward and backward directions. Through this bidirectional approach, the system can observe fine differences between similar gestures as well as improve robustness in general. This architecture was chosen as it demonstrated superior performance compared to standard LSTM and CNN-based approaches reported in prior literature using the KArSL-502 dataset, providing a more refined balance between temporal awareness and classification accuracy. After the extraction of temporal features through the LSTM layers, another dense layer containing 32 units with ReLU activation refines the learned representations. A classification dense layer containing 502 units, which is the same as the number of target classes, then calculates a softmax activation function to obtain probability distributions over all outputs that are possible. Fig. 2 below shows the proposed methodology for Arabic Sign Language.

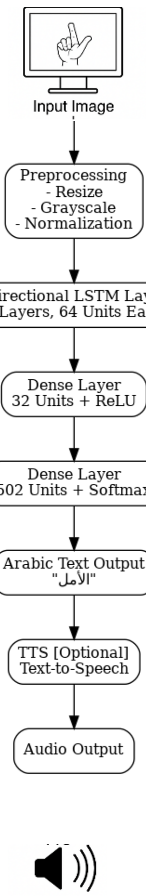


Fig. 2. Proposed methodology for Arabic Sign Language.

### F. Graphical User Interface (GUI)

A graphical user interface (GUI) was built to improve user interaction and give a useful presentation of the Arabic Sign Language Recognition system. With Text-to-Speech (TTS) capability, the interface lets users start real-time gesture detection, read the translated Arabic text, and perhaps listen to the result.

Built using Python tools including Tkinter, Pygame, and OpenCV, the GUI boasts a neat, understandable style. It comprises elements for webcam input choice, live video feed, detected sign display, recognized Arabic text, and an audio output speaking button.

The following screenshots show important phases of user involvement:

Fig. 3 shows the first GUI screen that the user chooses to start real-time Arabic Sign Language detection from the webcam.

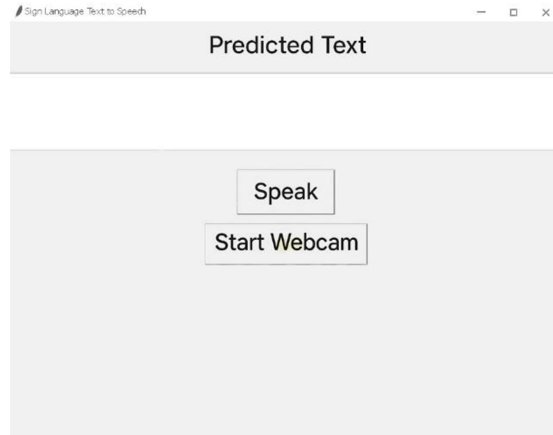


Fig. 3. First GUI Screen

Real-time gesture recognition examples include the detection of “السلام عليكم” as recorded and handled by the system (Fig. 3), identification of the sign for “شاب” with its corresponding Arabic text displayed (Fig. 4), recognition of the gesture for “متكبر” with live interactive feedback (Fig. 5), and the successful detection and translation of “أناني” into readable text (Fig. 6). Fig. 7 shows the recognized sign for “أنانى” (selfish), where the system captures the gesture and displays the corresponding word in text form for user interpretation.


Fig. 4. Sign for “السلام عليكم”


Fig. 5. Sign for “شاب”

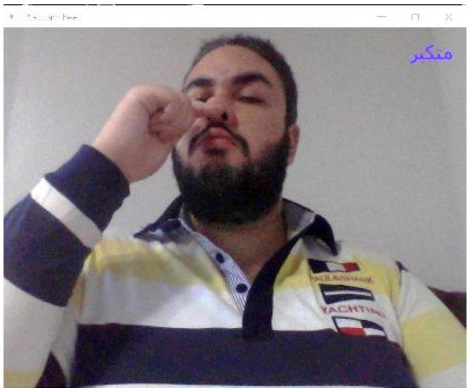


Fig. 6. Sign for "متكبر"

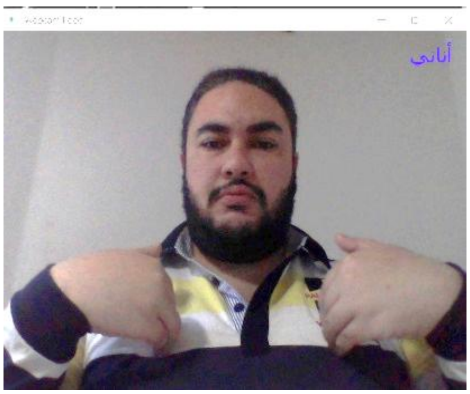


Fig. 7. Sign for "أنانى"

Fig. 8 shows the GUI displays the whole sentence together with an active Speech button so users may translate the identified text into spoken audio. This element makes the system realistic for real-world implementation among deaf and hearing people equally, since it greatly enhances its accessibility and usability.

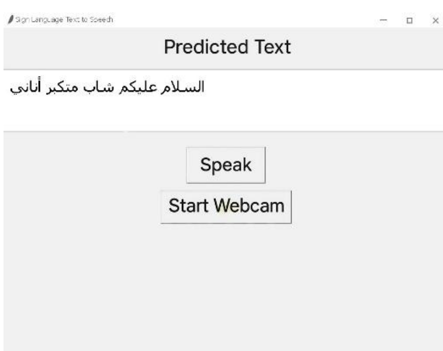


Fig. 8. Second GUI screen.

## G. Model Performance Results

502 Arabic Sign Language classes (KARSL-502) make up the dataset used in this investigation. As is common in sequence classification tasks, an 80:20 training–testing split was used for the experiments. The suggested BiLSTM-based model showed a strong ability to learn temporal dependencies across 48-frame gesture sequences during training, achieving a 98% categorical accuracy with a loss of 0.0480. The model achieved 96% categorical accuracy with a 0.1850 loss on the independent test set. Despite a slight decrease in test accuracy compared to training, the difference was still minimal, suggesting strong generalization to new data. To evaluate the trained model thoroughly, standard evaluation metrics such as accuracy, precision, recall, F1-score, and the confusion matrix were used. While the full confusion matrix for 502 classes is visually impractical to include, a detailed per-class analysis revealed that most misclassifications occurred between signs with highly similar 3D motion trajectories or minimal structural differences. These metrics were computed on the test set after convergence to provide a holistic evaluation of system performance. Furthermore, the real-time processing capability was verified through latency testing on the previously specified hardware (Intel i7, 16GB RAM, RTX 3060), yielding an average inference time of approximately 35ms per gesture (median latency), which comfortably satisfies real-time interaction requirements. Overall, the findings demonstrate that the suggested framework provides a dependable, scalable, and instantaneous Arabic Sign Language recognition solution, which can greatly improve the deaf and hard-of-hearing community's accessibility to communication. Fig. 9 shows the total accuracy verses the validation accuracy.

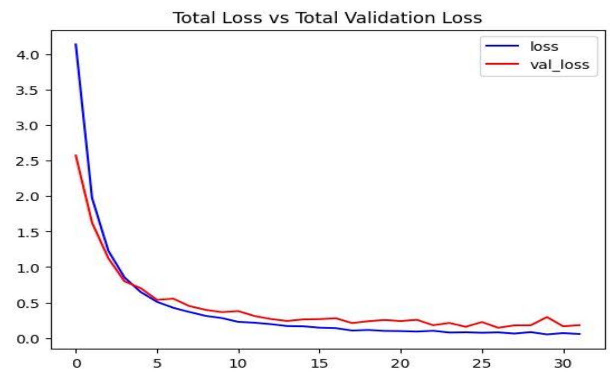


Fig. 9. Total accuracy vs. Validation accuracy.

The training accuracy (blue) increases rapidly over time, reflecting that the model is learning to recognize gestures in the training data. Similarly, the validation accuracy (red) rises, demonstrating that the model is generalizing well to unseen data. The close alignment of both curves towards 100% accuracy confirms that the model is performing well on both the training and validation datasets, without overfitting.

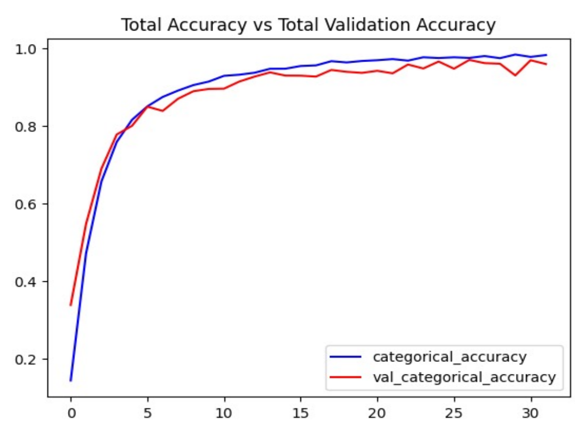


Fig. 10. Training loss and validation loss.

Both the training loss (blue) and validation loss (red) decrease steadily over the epochs, indicating that the model is improving and learning effectively. The training loss decreases faster at first, but both losses stabilize as the model reaches a point of convergence, suggesting good generalization without overfitting (Table II).

TABLE II. TRAINING AND TESTING PERFORMANCE METRICS OF THE PROPOSED LSTM MODEL

| Training Metric | Value | Testing Metric | Value |
|---|---|---|---|
| **Loss** | 0.048 | Loss | 0.185 |
| **Accuracy** | 98% | Accuracy | 96% |

## IV. DISCUSSION

The outcomes and assessment of the ASL Recognition model suggest that it reached acceptable performance and confirmed that it can detect motions and change them into text and vocalization with significant efficiency. The proposed model should be highly praised because the use of graphs and tables enables us to have a much better look at the learning process of a model, how it is performing, how well it is generalizing, among others. Loss curves training and validation proved the model to have a strong capability to acquire necessary patterns in relation to training data. In the beginning, the loss values are quite high, but as training continues, the training and validation losses are reduced greatly and stay at a lower rate. This seems to suggest that the model is fine-tuning its accuracy of predictions with the training data set, as well as enhancing the accuracy of extrapolation on unknown validation data. The predominant signs of falling values for both losses, with their difference not increasing over time, assure us that overfitting is not occurring when the model excessively tunes into the training set and does not generalize well to the test set.

On the same note, the graphs of the training and validation accuracy support the model as well as the identification of ineffective neurons. The two accuracies rise steeply in the initial epochs, as is common, because the model is learning patterns of the data set. That training accuracy and validation accuracy are in sync and touch on the 100 percent mark indicates the model's ability to learn from data and generalize to unseen data sets. The direction of both curves also reinforces this idea because generally, overfitting means that the accuracy of the training rate keeps on rising as that of the validation rate plummets or at best stagnates.

The loss and accuracy matching between the training and validation set shows model generalization and is suggestive. In machine learning, overfitting is a widespread problem in which a model learns the training data set and does not generalize to other data sets. Fig. 10 above shows that no large gap exists between the training loss and validation loss, and the training accuracy and validation accuracy are almost identical, which demonstrates that the model has achieved superior performance on the issue of overfitting. This shows that the proposed model is able to generalize to new, unseen sign language gestures, and the elaboration of variability present both in gestures and signing style makes it robust for real-world applications.

## V. CONCLUSION AND FUTURE WORK

Based on the outcomes, the proposed model appears well-suited to real-world sign language recognition and synthesis since the high accuracy of the trained and validation datasets is nearly identical. This research serves as a pilot study to demonstrate that capturing gestures of Arabic Sign Language in real time and translating them to texts and speech will presumably benefit the deaf and hard-of-hearing. The addition of Text-to-Speech (TTS), therefore, improves the existing system by allowing non-sign language users to more easily understand the gestures that have been recognized. This can then result in attaining equal and improved communication in areas involving accessibility to people with disabilities, like healthcare, education, and in customer services where communication with the hearing impaired is crucial.

This highlights the need for improvements to make even with the impressive performance of this model. While the 96% accuracy is high, future work will include signer-independent testing to ensure robustness. The weak link that can be identified is the fact that sign languages don't do well in mimicking the more complicated hand signs and complicated signs that demand things like facial expressions and body movements. These aspects of sign language are crucial for full comprehension but remain barely satisfactorily contained in popular sign language and therefore can be tough to register using current hand-tracking technologies. Using communication modes, such as visual (facial expressions, posture, etc.), the system design could enhance its ability to manage such elaborate gestures. Furthermore, expanding the training dataset to include more variations in sign language gestures, regional dialects, and environmental factors (e.g., different lighting or backgrounds) could enhance the model's robustness. Continuous user feedback and testing in real-world scenarios would also help fine-tune the system and ensure that it meets the needs of a diverse group of users.

## VI. FUTURE WORK AND CONTRIBUTION

The Arabic Superformance Recognition system has been developed and evaluated, giving satisfactory performance; however, the following areas are open to further study. Further research may involve exploring advanced architectures like Transformers and Conformer models and extending the system to incorporate additional data modalities, including face pictures, head motion, and pose, that are important for accurately interpreting sign language. Increasing the type of signers, dialects, and conditions during data collection will improve the generalization ability of the proposed model across various conditions. Also, real-time processing speed and latency time reduction will enhance the system's effectiveness. The Research could also be extended to support the identification of multiple sign languages, for instance, American Sign Language (ASL), thus increasing its applicability.

Using AR or VR in the system would still be an added advantage, building on the aspect of communication. Interacting with users will enable the improvement of the system to incorporate different signing patterns, hence making it friendly to users.

Accessibility is an area improved by this Research, especially in the Arabic context, where little has been done to improve recognition of sign languages. When gesture recognition is integrated with text-to-speech, it helps in real-time communication between a sign language user and a normal language user. Bidirectional LSTM networks for gesture recognition are novel; this creates the basis for other improvements and brings a considerable influence on inclusive interaction. With some refinements, this technology can greatly improve the quality of life for the deaf and hard-of-hearing population by decreasing the time needed for sign translation for common exchanges.

## REFERENCES

[1] Al Abdullah, B., Amoudi, G., & Alghamdi, H. (2024). Advancements in Sign Language Recognition: A Comprehensive Review and Future Prospects. *IEEE Access*.

[2] Abdul Ameer, R. S., Ahmed, M. A., Al-Qaysi, Z. T., Salih, M. M., & Shuwandy, M. L. (2024). Empowering communication: A deep learning framework for Arabic sign language recognition with an attention mechanism. *Computers*, *13*(6), 153.

[3] Al-Hammadi, M., Muhammad, G., Abdul, W., Alsulaiman, M., Bencherif, M. A., Alrayes, T. S., & Mekhtiche, M. A. (2020). Deep learning-based approach for sign language gesture recognition with efficient hand gesture representation. *IEEE Access*, *8*, 192527-192542.

[4] Bencherif, M. A., Algabri, M., Mekhtiche, M. A., Faisal, M., Alsulaiman, M., Mathkour, H., ... & Ghaleb, H. (2021). Arabic sign language recognition system using 2D hands and body skeleton data. *IEEE Access*, *9*, 59612-59627.

[5] Alsulaiman, M., Faisal, M., Mekhtiche, M., Bencherif, M., Alrayes, T., Muhammad, G., ... & Alfakih, T. (2023). Facilitating the communication with deaf people: Building the largest Saudi sign language dataset. *Journal of King Saud University-Computer and Information Sciences*, *35*(8), 101642.

[6] Rajest, S. (2024). S. Enhancing Communication: Real-Time Sign Language Detection with Deep Learning and Python. *Central Asian Journal of Medical and Natural Science*, *5*(4), 940-951.

[7] Al Ahmadi, S., Muhammad, F., & Al Dawsari, H. (2024). Enhancing Arabic Sign Language Interpretation: Leveraging Convolutional Neural Networks and Transfer Learning. *Mathematics*, *12*(6), 823.

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

[9] Al-Shamayleh, A. S., Ahmad, R., Abushariah, M. A., Alam, K. A., & Jomhari, N. (2018). A systematic literature review on vision-based gesture recognition techniques. *Multimedia Tools and Applications*, *77*, 28121-28184.

[10] Aiouez, S., Hamitouche, A., Belmadoui, M. S., Belattar, K., & Souami, F. (2022, April). Real-time Arabic Sign Language Recognition based on YOLOv5. In *IMPROVE* (pp. 17-25).

[11] Natarajan, B., Rajalakshmi, E., Elakkiya, R., Kotecha, K., Abraham, A., Gabralla, L. A., & Subramaniyaswamy, V. (2022). Development of an end-to-end deep learning framework for sign language recognition, translation, and video generation. *IEEE Access*, *10*, 104358-104374.

[12] Ismail, Y., Tarek, A., Zakaria, O., Taha, S., Masoud, S., Sobh, K., ... & Khoriba, G. (2023, July). From Gestures to Audio: A Dataset Building Approach for Egyptian Sign Language Translation to Arabic Speech. In *2023 Intelligent Methods, Systems, and Applications (IMSA)* (pp. 400-405). IEEE.

[13] Buttar, A. M., Ahmad, U., Gumaei, A. H., Assiri, A., Akbar, M. A., & Alkhamees, B. F. (2023). Deep learning in sign language recognition: a hybrid approach for the recognition of static and dynamic signs. *Mathematics*, *11*(17), 3729.

[14] Marzouk, R., Alrowais, F., Al-Wesabi, F. N., & Hilal, A. M. (2022, August). Atom search optimization with deep learning enabled Arabic sign language recognition for persons with speaking and hearing disability. In *Healthcare* (Vol. 10, No. 9, p. 1606). MDPI.

[15] Tharwat, G., Ahmed, A. M., & Bouallegue, B. (2021). An Arabic sign language recognition system for the alphabet using machine learning techniques. *Journal of Electrical and Computer Engineering*, *2021*(1), 2995851.

[16] Abdulhamied, R. M., Nasr, M. M., & Abdulkader, S. N. (2023). Real-time recognition of American Sign Language using a long short-term memory neural network and hand detection. *Indonesian Journal of Electrical Engineering and Computer Science*, *30*(1), 545556.

[17] Hussain, A., Saikia, N., & Dev, C. (2023). Advancements in Indian Sign Language Recognition Systems: Enhancing Communication and Accessibility for the Deaf and Hearing Impaired. *Asian Journal of Electrical Sciences*, *12*(2), 37-49.

[18] Mukhedkar, M., Deshmukh, D. P., Chaudhari, O., Shinde, H., Shinde, A., & Kadam, A. (2024, June). Efficient Development of Gesture Language Translation System using CNN. In *2024, the 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)* (pp. 1-6). IEEE.

[19] Yang, H., An, X., Pei, D., & Liu, Y. (2014, September). Towards realizing gesture-to-speech conversion with an HMM-based bilingual speech synthesis system. In the *2014 International Conference on Orange Technologies* (pp. 97-100). IEEE.

[20] Fernández-Baena, A., Montaño, R., Antonijoan, M., Roversi, A., Miralles, D., & Alías, F. (2014). Gesture synthesis adapted to speech emphasis. *Speech communication*, *57*, 331-350.

[21] Lu, R., & Song, Q. (2023). Research on the improved gesture tracking algorithm in sign language synthesis. *The Journal of Supercomputing*, *79*(1), 867-879.

[22] M. F. Almufareh, S. Tehsin, and M. Humayun,(2025) "Attention-Based Approach for Arabic Sign Language Recognition, Supporting Differently Abled Persons," Journal of Disability Research, 2025.

[23] A. A. Mohamed, A. Al-Saleh, S. K. Sharma, and G. Tejani, (2025) "Attention-based hybrid deep learning model with CSFOA optimization and G-TverskyUNet3+ for Arabic sign language recognition," Scientific Reports

[24] A. Alzu'bi, A. Albashayreh, and L. B. Younis, "Arabic Sign Language Recognition Using Image-Based Deep Learning and Edge User-Centered Interface," in Emerging Technologies for Computing, Communication and Smart Cities, Springer, 2025, pp. 209–220