# Mobile-Based Arabic Sign Language Recognition System Using Deep Learning and 3D Avatar Rendering

## Abstract

Arabic Sign Language (ArSL) remains under-resourced compared to other sign languages, limiting effective communication tools for the Deaf community in the Arab world. This study presents a mobile-based recognition system that translates ArSL hand gestures into Arabic text and vice versa using 3D avatar animation. We explore and evaluate three deep learning architectures: a basic Convolutional Neural Network (CNN), an enhanced CNN with advanced preprocessing techniques, and a fine-tuned MobileNetV2 model using transfer learning. The ArASL dataset, consisting of 7,856 labeled images across 33 classes, was used to train and test the models. Data augmentation, normalization, and noise reduction techniques were applied to improve model robustness. Our results show that MobileNetV2 achieved the highest test accuracy of 87.77%, outperforming both CNN models. The best-performing model was integrated into a mobile application that enables two-way communication: sign-to-text and text-to-sign rendering through a 3D avatar. This system demonstrates the potential of combining deep learning and mobile technology to enhance accessibility for the Deaf in Arabic-speaking communities.

## Keywords

Arabic Sign Language, Deep Learning, MobileNetV2, Convolutional Neural Networks, Gesture Recognition, Transfer Learning, Mobile Application, 3D Avatar Animation

## 1. Introduction

Sign language is a complete and natural form of communication used by the Deaf and hard-of-hearing communities. It relies on visual gestures, hand shapes, facial expressions, and body movements to convey meaning. Among the various sign languages in use worldwide, Arabic Sign Language (ArSL) is the primary language used by the Deaf population across Arabic-speaking countries. However, compared to well-established sign languages like American Sign Language (ASL) and British Sign Language (BSL), ArSL lacks standardized datasets, recognition tools, and digital resources, making it less accessible in educational, social, and professional contexts.

With the growing integration of AI in assistive technologies, there is an urgent need to develop intelligent systems capable of translating sign language into written or spoken language and vice versa. This is particularly important for ArSL, where limited research and resources have hindered the development of effective recognition systems. Bridging this communication gap can significantly improve the quality of life, independence, and inclusion for Deaf individuals in Arabic-speaking societies.

This paper presents an end-to-end Arabic Sign Language recognition system based on deep learning techniques, integrated into a mobile application for real-time usage. The system translates hand gestures representing Arabic letters into text and converts inputted text back into sign language through a 3D avatar. We investigate the performance of three deep learning models: a basic Convolutional Neural Network (CNN), an enhanced CNN with advanced image preprocessing, and a fine-tuned MobileNetV2 model using transfer learning.

Our contributions include a comparative analysis of the three models, a unified preprocessing pipeline to handle dataset challenges, and the deployment of the best-performing model in a mobile application that supports two-way communication. The proposed system aims to provide an accessible, lightweight, and accurate tool for ArSL recognition, contributing to broader efforts in AI-based assistive technology.

## 2. Related Work

Sign Language Recognition (SLR) is a multidisciplinary research area combining computer vision, machine learning, and human-computer interaction. Traditional approaches to SLR often relied on sensor-based devices such as data gloves, which measured joint angles and hand positions. Although these methods achieved reasonable accuracy, they suffered from high costs and limited usability in real-world scenarios. With the rise of deep learning and affordable vision systems, image-based and video-based recognition methods have become dominant.

Arabic Sign Language (ArSL) poses specific challenges due to its limited resources. Unlike American Sign Language (ASL), ArSL lacks large-scale, standardized datasets and consistent notation systems. Several studies have attempted to recognize ArSL using handcrafted features such as Histograms of Oriented Gradients (HOG) combined with Support Vector Machines (SVM), or Hidden Markov Models (HMM) for sequential recognition. However, these approaches often suffer from low generalization and sensitivity to background noise.

More recent works have focused on applying Convolutional Neural Networks (CNNs) to static hand gesture images. For example, Elons et al. (2018) trained a 2D-CNN from scratch on the ArASL dataset and achieved approximately 79% accuracy. Later enhancements introduced image preprocessing techniques such as contrast normalization and data augmentation to improve performance. Transfer learning has also shown great promise. Omar and El-Sayed (2021) used pre-trained InceptionV3 models to achieve up to 88% validation accuracy on ArSL images.

Despite these advances, many studies still focus on isolated recognition without mobile deployment or user interaction. Moreover, bidirectional systems that support both recognition and sign rendering via avatars remain rare. In addition, most works have not addressed real-time implementation on mobile platforms, which is essential for practical use.

In this paper, we build upon previous work by comparing CNN architectures trained from scratch with a transfer-learning approach using MobileNetV2. Furthermore, we integrate the best-performing model into a real-time mobile application with two-way communication functionality, including a 3D avatar that animates sign language from Arabic text—an area largely unexplored in ArSL research.

## 3. Methodology

This section describes the dataset used, the preprocessing techniques applied to improve data quality and model performance, the architectures of the evaluated deep learning models, and the integration of the final model into a mobile application supporting two-way communication with 3D avatar rendering.

We used the Arabic Alphabets Sign Language Dataset (ArASL), consisting of 7,856 RGB images across 33 classes. To improve performance, we applied resizing, normalization, noise reduction, contrast enhancement, and data augmentation. Data were split into 80% training, 10% validation, and 10% testing.

Three deep learning models were developed: (1) a basic CNN with 4 convolutional layers and dense output, (2) an enhanced CNN with batch normalization and dropout, and (3) MobileNetV2 fine-tuned in two training phases using ImageNet weights. All models used categorical cross-entropy loss and the Adam optimizer. Early stopping and class balancing were employed to avoid overfitting.

The best-performing model (MobileNetV2) was exported to TensorFlow Lite and integrated into an Android mobile app. The app features live sign-to-text conversion and text-to-sign rendering via a Unity-based 3D avatar. Performance optimization included quantization and caching.

## 4. Experimental Results

The basic CNN achieved 79.06% test accuracy, while the enhanced CNN reached 82.69%. The MobileNetV2 model achieved the highest accuracy of 87.77% on the test set. Training curves showed better convergence and less overfitting with MobileNetV2. Confusion matrix analysis revealed that most misclassifications involved visually similar signs like 'س' and 'ش'.

The mobile-deployed MobileNetV2 model achieved an average inference latency of 180 ms per frame, meeting the real-time threshold. Text-to-sign animations were rendered within 300 ms due to effective caching.

## 5. Discussion and Conclusion

The results demonstrate that transfer learning via MobileNetV2 significantly outperforms models trained from scratch. Preprocessing and regularization further improved CNN performance. Integration into a mobile app validates the system's real-world applicability and potential impact for the Arabic Deaf community.

## 6. Future Work

Future work includes expanding to sentence-level recognition using LSTMs or Transformers, collecting a larger and more diverse ArSL dataset, integrating non-manual signals such as facial expressions, supporting personalized model adaptation on-device, and conducting usability testing with Deaf users.

## References

[1] S. F. Alfergani et al., “Arabic Sign Language Recognition Using HOG Features and Support Vector Machines,” Int. J. Intell. Syst. Appl., vol. 5, no. 8, pp. 12–21, 2013.

[2] D. Al-Mahadin and A. Jamous, “Continuous Arabic Sign Language Recognition via Kinematic Feature Extraction and Hidden Markov Models,” Proc. IEEE Int. Conf. Pattern Recognit., 2015, pp. 345–350.

[3] A. M. Elons and H. M. Youssef, “Deep Convolutional Neural Networks for Static Arabic Sign Language Recognition,” J. Vis. Commun. Image Represent., vol. 54, pp. 101–109, 2018.

[4] M. Hussein and F. Hassan, “Enhanced Preprocessing and Data Augmentation for Arabic Sign Language Recognition with CNNs,” Proc. IEEE ICCV Workshops, 2019, pp. 55–60.

[5] N. Khalid and S. Ahmed, “Spatio-temporal 3D-CNN for Arabic Sign Language Video Recognition,” IEEE Access, vol. 8, pp. 204561–204570, 2020.

[6] S. Omar and M. El-Sayed, “Transfer Learning with InceptionV3 for Arabic Sign Language Recognition Using Synthetic Data,” Int. J. Adv. Comput. Sci. Appl., vol. 12, no. 6, pp. 120–128, 2021.

[7] B. Abdel Halim et al., “Arabic Sign Language Recognition Using Deep Learning and MobileNetV2,” Graduation Project Report, Helwan University, 2025.