

Collage of Informatics

Midocean University

UAE



Research Title

Speech Recognition System Using Deep Learning Techniques for Arabic
Sign Language

Prepared By

Eng. Sultan Ali Senan

ID No.: 253002563

Submitted in fulfillment of the requirements for the degree Master of
Science in Artificial Intelligence

Supervisor Name: Dr. Amjad Qatish

Academic year:

2026

1.

2.

3.

4.

5. Chapter 5: Conclusion and Future Work

5.1 Introduction

This chapter presents the conclusion of the research and summarizes the major findings obtained from the development and evaluation of the Arabic Sign Language (ArASL) recognition system. The chapter reviews the research objectives, discusses the key outcomes of the study, highlights the limitations encountered during implementation, and provides recommendations for future improvements. The findings presented in this chapter demonstrate the effectiveness of deep learning techniques for real-time Arabic Sign Language recognition and translation.

5.2 Summary of the Research

The primary objective of this research was to develop an intelligent Arabic Sign Language recognition system capable of translating sign gestures into meaningful outputs using machine learning and deep learning techniques. To achieve this objective, an extensive experimental

framework was designed involving dataset preparation, preprocessing, model training, performance evaluation, and real-time system deployment.

The study utilized the Arabic Sign Language dataset containing 54,049 images distributed across 32 Arabic alphabet classes. Multiple preprocessing operations including image resizing, normalization, label encoding, and data augmentation were applied to improve data quality and model generalization.

A comparative evaluation was conducted using seven machine learning and deep learning models. The evaluated models included Random Forest, Linear SVM, HistGradientBoosting, Multilayer Perceptron, Custom CNN, Light ResNet, and MobileNetV2. Experimental results demonstrated that the proposed Custom CNN achieved the highest validation accuracy of 94.88%, outperforming all other evaluated models.

Following model selection, the CNN architecture was integrated into a FastAPI-based real-time translation system. A web-based dashboard was developed to capture webcam input, perform sign recognition, display confidence scores, and generate translated outputs. The completed system successfully demonstrated real-time Arabic Sign Language recognition and translation capabilities.

5.3 Achievement of Research Objectives

The research objectives established at the beginning of this study were successfully achieved. The first objective, which focused on preparing and organizing a reliable Arabic Sign Language dataset, was accomplished through data validation, preprocessing, and augmentation procedures.

The second objective involved the development and evaluation of multiple recognition models. This objective was achieved through the implementation and benchmarking of seven machine learning and deep learning algorithms. Comparative analysis identified Custom CNN as the most effective architecture for sign language recognition.

The third objective is to develop a real-time recognition framework capable of processing live webcam input. This goal was successfully accomplished through the integration of the trained CNN model with a FastAPI backend and a web-based user interface.

Finally, the objective of evaluating system performance was fulfilled through extensive experimentation using validation accuracy, training behavior analysis, and comparative benchmarking. The results confirmed the effectiveness of the proposed approach and validated the suitability of deep learning techniques for Arabic Sign Language recognition.

5.4 Key Findings

The study produced several important findings regarding Arabic Sign Language recognition and translation. First, deep learning architectures significantly outperforms traditional machine learning models in image-based sign language classification tasks.

Second, Custom CNN architecture achieved the highest validation accuracy of 94.88%, demonstrating superior feature extraction and classification capabilities. The results indicate that task-specific CNN architectures can be highly effective for sign language recognition applications.

Third, data augmentation, dropout regularization, and batch normalization contributed significantly to improving model robustness and reducing overfitting. These techniques enhanced the generalization capability of the proposed model.

Finally, the integration of the trained CNN model within a FastAPI-based web application demonstrated the feasibility of implementing a practical real-time Arabic Sign Language translation system. The developed platform successfully recognized hand gestures and generated translation outputs in an interactive environment.

5.5 Limitations of the Study

Although the proposed system achieved promising results, several limitations should be acknowledged. The study focused primarily on recognizing Arabic alphabet signs rather than complete words, sentences, or continuous sign language conversations.

Additionally, the system performance may be affected by environmental conditions such as lighting variations, camera quality, background complexity, and hand positioning. These factors can influence image quality and recognition accuracy during real-time operation.

The experiments were conducted using a specific dataset, which may not fully represent all possible variations of Arabic Sign Language gestures. Therefore, additional data collected from diverse users and environments could further improve model robustness and generalization.

Furthermore, the developed system was evaluated mainly within a controlled experimental environment. Large-scale real-world testing involving diverse user groups was beyond the scope of this research.

5.6 Future Work

Several directions can be explored to enhance the capabilities of the proposed system in future research. One potential improvement involves extending the recognition framework beyond individual alphabet signs to support complete words, phrases, and continuous sign language translation.

Future studies may also investigate the use of advanced deep learning architectures such as Vision Transformers (ViT), hybrid CNN-Transformer models, and attention-based networks to further improve recognition accuracy.

Another promising direction is the development of mobile and edge-device implementations to enable portable and accessible sign language translation applications. Integration with text-to-speech technologies may also provide more comprehensive communication assistance for deaf and hard-of-hearing individuals.

In addition, future research could utilize larger and more diverse datasets containing different users, environments, and gesture variations to improve model robustness and real-world applicability.

