

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. Chapter three: Working Plan (Methodology): Proposed Research Methodology

This document presents the proposed working plan for the methodology to be followed in developing a real-time Arabic Sign Language (ArASL) recognition and translation system based on deep learning. It outlines the planned stages of the research, from dataset preparation through model development to system deployment, prior to the execution and reporting of results. The plan is intended to guide the upcoming implementation phase and to be reviewed and approved before work proceeds.

3.1 Research Objectives

The proposed methodology aims to achieve the following objectives:

- Design and prepare a reliable Arabic Sign Language dataset suitable for training and evaluating recognition models.
- Develop a Custom Convolutional Neural Network tailored to Arabic alphabet sign recognition.
- Benchmark the proposed model against several traditional machine learning and deep learning baselines under a unified protocol.
- Integrate the selected model into a real-time web-based translation system.

3.2 Proposed Methodology Overview

The study will follow a quantitative, experimental methodology consisting of six planned stages: dataset collection and preparation, data preprocessing, data augmentation, model design, model training and benchmarking, and system integration. Each stage is outlined below.

3.3 Dataset Collection and Preparation (Planned)

The Arabic Sign Language (ArASL) dataset, comprising grayscale images across 32 Arabic alphabet sign classes, will be obtained and audited for integrity. This step will involve verifying

that all image files are valid and correctly associated with their corresponding labels, and organizing the dataset by class in preparation for preprocessing.

3.4 Data Preprocessing (Planned)

All images will be resized to a fixed resolution suitable for model input, converted to grayscale where applicable, and their pixel values normalized to improve training stability. Class labels will be encoded numerically, and the dataset will be split into training and validation subsets using a stratified approach to preserve class balance.

3.5 Data Augmentation (Planned)

To improve the robustness and generalization of the trained models, data augmentation techniques such as random rotation, translation, and zoom will be applied during training. This is intended to expose the models to realistic variations in hand position, orientation, and scale that may occur during real-world use.

3.6 Proposed Model Architecture

A Custom Convolutional Neural Network will be designed specifically for Arabic Sign Language recognition, consisting of multiple convolutional blocks with batch normalization, max-pooling, and dropout layers, followed by a fully connected classification head. The architecture will be designed to balance recognition accuracy with computational efficiency suitable for real-time inference.

3.7 Candidate Models for Benchmarking

In order to identify the most effective approach, the proposed CNN will be benchmarked against a set of traditional machine learning and deep learning baselines, planned as follows: Random Forest, Linear Support Vector Machine, HistGradientBoosting, Multilayer Perceptron, Light ResNet, and MobileNetV2 (transfer learning). All models will be trained and evaluated under the same data split and evaluation protocol to ensure a fair comparison.

3.8 Training and Evaluation Plan

Each model will be trained using appropriate optimization algorithms and loss functions suited to multi-class classification. Training and validation performance will be monitored throughout to

assess convergence and generalization. Validation accuracy will serve as the primary criterion for model comparison, supported by analysis of training and validation loss curves.

3.9 Planned System Integration

The best-performing model will be integrated into a real-time translation system composed of a FastAPI backend and a web-based user interface. The planned system will accept live camera input, perform preprocessing and inference on incoming frames, and display the predicted Arabic letter together with a confidence score, enabling continuous recognition and text construction.

3.10 Proposed Timeline

The following timeline outlines the planned phases of work and their estimated durations:

Phase	Planned Activity	Estimated Duration
Phase 1	Dataset acquisition, integrity auditing, and preprocessing (resizing, normalization, label encoding)	1–2 weeks
Phase 2	Design and implementation of the proposed Custom CNN architecture and baseline models	1–2 weeks
Phase 3	Model training, validation, and comparative benchmarking of all candidate models	2–3 weeks
Phase 4	Selection of the best-performing model and integration into a FastAPI backend	1–2 weeks
Phase 5	Development of the real-time web interface and end-to-end system testing	1–2 weeks
Phase 6	Results documentation, evaluation, and final reporting	1 week

3.11 Expected Outcome

Upon completion of the proposed methodology, it is expected that a trained and validated deep learning model will be identified as the most suitable architecture for Arabic Sign Language recognition, and that this model will be successfully deployed within a functioning real-time translation system. Detailed results, performance comparisons, and discussion will be reported separately following the execution of this plan.