GRADUATION PROJECT PROGRESS REPORT

A Real-Time, Dual-Function Sign Language Translation and Education System

Prepared by:
Salah Mohamed Salah Ibrahim Ahmed Adel Sayed Goda Ahmed
Alaa Osama Mohammed Shehab Hadeel Gamal Eldin Mohamed Ibrahim

Supervised by:
Dr. Nada Mostafa Abdelaziem Mostafa
Dr. Reham Taher Abdelmajeed Salem Al Magraby

December 16, 2025

1 Executive Summary

This report documents the developmental progress achieved during Term 9, confirming the successful conclusion of the research and initial proposal phase and the definitive transition to the system implementation stage. The foundational technical milestone—the establishment of a functional, real-time recognition engine for Arabic Sign Language (ArSL)—has been validated. This was achieved through the successful implementation and training of two distinct Artificial Intelligence paradigms: the image-based MobileNetV2 architecture and the geometry-centric MediaPipe model. The system currently demonstrates real-time translation capabilities, thereby fulfilling the core technical deliverable prerequisite for the subsequent phases involving application enclosure and user interface development.

2 1. Introduction and Background

The project aims to address significant communication barriers by developing a dual-function system that acts as both a real-time ArSL translator and an interactive educational platform. The initial stage of the project focused on selecting and validating the optimal computer vision and deep learning architectures capable of high-accuracy and low-latency performance on consumer-grade hardware. The successful implementation of the dual-model approach provides the requisite technical redundancy to ensure deployment flexibility, regardless of the target machine's computational constraints.

3 2. Project Goals and Objectives: Review of Progress

The primary objectives outlined in the original proposal remain the guiding mandates of the project. A review of the status against the established phases is provided below:

Goal 1: **Develop a High-Accuracy Recognition Model:** Achieved via the successful training of both MobileNetV2 and MediaPipe MLP models, yielding validation accuracies exceeding 90%.

Goal 2: **Implement Real-Time Inference:** Achieved through the development of the `Combined_Architectu` script, which processes live video and provides continuous predicted text output.

Goal 3: **Integrate Bidirectional Communication:** Pending. Requires the integration of external Speech-to-Text and Text-to-Speech components (Stage II).

Goal 4: **Develop an Educational Module:** In progress. The foundational geometry comparison logic (Stage I) is being planned for implementation.

Goal 5: **Deploy in a Dedicated Application:** Pending. Requires the transition from the prototype environment (Jupyter/OpenCV) to a formal GUI wrapper (Stage III & IV).

4 3. Implementation Progress

The work completed in Term 9 constitutes the entirety of Phase 1 (Model Development) and the initiation of Phase 2 (Real-Time Integration).

4.1 3.1. Environment and Hardware Configuration

Formal documentation confirms the successful institution of the foundational technical ecosystem. The NVIDIA GeForce MX150 Graphics Processing Unit (GPU) was configured for optimal operation within the TensorFlow framework, necessitating specific kernel adjustments for stable performance under load. Critical to this stage was the execution of dynamic memory growth strategies, a necessary measure to rigorously preclude the manifestation of Out of Memory (OOM) exceptions during the intensive model training regimen, given the hardware's VRAM limitations. The core computational libraries (OpenCV, MediaPipe, Keras) have been integrated into a unified software architecture, guaranteeing system integrity for real-time operations.

4.2 3.2. Architectural Model Development (Phase 1)

The strategic decision to pursue two disparate deep learning methodologies was executed to ensure technical robustness.

3.2.1. Approach A: MobileNetV2 (Transfer Learning Implementation)

The pre-trained MobileNetV2 architecture was employed as the foundational core, exploiting its deep feature-extraction capabilities. The stability of the network was augmented through the incorporation of BatchNormalization layers within the classification head, and the mandatory execution of the fine-tuning procedure on the proprietary, image-based ArSL dataset was completed. The resultant model weights are archived as `mobilenet_arabic_best_final`

3.2.2. Approach B: MediaPipe and Multi-Layer Perceptron (MLP) Integration

The geometry-centric feature extraction pipeline utilizes the MediaPipe framework for the precise extraction of twenty-one distinct 3D hand landmark coordinates, successfully reducing the input dimensionality. This feature vector extraction precedes the training of a specialized lightweight MLP classifier, detailing the resultant low-latency performance profile that renders this model exceptionally appropriate for deployment within CPU-constrained environments.

4.3 3.3. Data Curation and Real-Time Integration (Phase 2)

Quality assurance protocols were established for the ArSL dataset to guarantee semantic diversity across sign classes, encompassing variations in camera angle and background clutter. Furthermore, the unified `Combined_Architecture` control script was developed. This script facilitates concurrent video stream acquisition, implements the necessary graphic UI overlay for user feedback, and incorporates crucial temporal smoothing algorithms (based on a confidence-weighted queue) designed to actively mitigate the output text instability, specifically the detrimental phenomenon of character flickering in the predictive display.

5 4. Technical Results and Performance Metrics

A quantitative, empirical assessment of the performance efficacy exhibited by the two constructed models is provided.

Table 1: Model Performance Metrics from Training Logs

Metric	**MobileNetV2 (Image-Based)**	**MediaPipe + MLP (Landmark-Based)**	**Unit**
Training Accuracy	$\sim 92\% - 95\%$	$\sim 94\%$	Percenta
Validation Accuracy	$\sim 85\% - 90\%$	$\sim 91\%$	Percenta
Inference Speed	Moderate	Very High	Qualitat
Lighting Sensitivity	High	Low	Qualitat

Detailed Performance Observation. A critical comparative analysis meticulously contrasts the superior proficiency of the MobileNetV2 architecture in the recognition of complex static imagery (signs involving subtle shading or internal hand texture) against the MediaPipe methodology's empirically verified computational efficiency (achieving frame rates above 30 FPS). This dual implementation provides the fundamental rationale supporting the strategic necessity of adopting a dual-model deployment paradigm, allowing the system to leverage the respective strengths of each model dynamically based on runtime conditions.

6 5. Challenges Encountered and Remedial Strategies

Formal documentation of significant technical obstacles encountered during the development cycle, alongside the precise, implemented strategies utilized for mitigation.

5.1. Intra-Class Sign Confusion

Issue: A difficulty was evidenced in distinguishing between visually homologous Arabic signs (e.g., the exemplars 'K' and 'X'), which share nearly identical high-level kinematic features.
Mitigation: Remedial adoption of Test-Time Augmentation (TTA), utilizing simulated rotations and slight scale variations, was executed. This was concurrent with the systematic execution of targeted dataset expansion to increase the marginal separation between these classes in the feature space, thereby resolving the ambiguity.

5.2. Hardware Resource Constraints

Issue: The system instability and operational restrictions were directly attributable to the inherent limitations of the NVIDIA MX150 VRAM capacity, which restricted the model complexity and training parameters.
Mitigation: This was addressed through the implementation of memory growth protocols and the rigorous optimization of batch size selection parameters (e.g., maximum size 32). These measures strictly ensured reliable resource utilization without precipitating system failure.

7 6. Forthcoming Stages of Development (Timeline Review)

The following objectives are provisioned for execution during the subsequent reporting period (Term 10), providing a clear trajectory for the project's final application development and system deployment phase.

- **Stage I: Refinement of the Educational Module (Term 10 Kickoff)** Implementation of the Euclidean distance metric within the 3D landmark space, which is required for the quantitative comparison of a learner's instantaneous gesture against the established "Golden Reference" signs.
- **Stage II: Bidirectional Communication Integration (Term 10)** Integration of external Speech-to-Text (STT) and Text-to-Speech (TTS) components to finalize the system's core objective: the establishment of a complete, two-way, multimodal communication loop.
- **Stage III & IV: Application Enclosure and Deployment Strategy (Term 10)** Development of the final Graphical User Interface (GUI) and the integration of the "Combined Architecture" within this enclosure, facilitating the dynamic, run-time selection of the optimal classification model to guarantee maximal system performance.

8 7. Conclusion

A definitive concluding statement formally affirming the project's absolute adherence to the estimated schedule and established technical specifications for the completion of Phase 1 and the initial phase of Phase 2. The core recognition engine functionality is demonstrably robust, and the team is positioned for the final stages of application enclosure, comprehensive system integration, and the development of the educational features, thereby setting the stage for the final project submission.