

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. Chapter 4: Results and Discussion.

4.1 Introduction

This chapter discusses the experimental findings derived from the installation and assessment of the proposed Arabic Sign Language (ArASL) recognition system. The major purpose of this chapter is to examine the performance of machine learning and deep learning models, compare their efficacy, and describe the discoveries gained during testing.

The experiments were conducted using the ArASL dataset consisting of 54,049 grayscale images distributed across 32 Arabic alphabet sign classes. Several classical machine learning algorithms and deep learning architectures were investigated to discover the best acceptable model for real-time sign language recognition. The performance of each model was tested using validation accuracy, training efficiency, and computational needs.

The chapter also covers the comparative benchmarking findings, analyzes the learning behavior of the proposed CNN model, and assesses the efficacy of the constructed real-time translation

system. The results give evidence about the appropriateness of deep learning algorithms for Arabic Sign Language recognition and show the benefits of the proposed architecture for practical deployment.

4.2 Comparative Model Performance Analysis

A detailed benchmarking experiment was undertaken to assess the performance of seven machine learning and deep learning models for Arabic Sign Language recognition. The evaluated models included Random Forest, Linear Support Vector Machine (SVM), HistGradientBoosting, Multilayer Perceptron (MLP), Custom CNN, Light ResNet, and MobileNetV2 Transfer Learning.

The comparative findings indicated major disparities in recognition performance across the studied methodologies. Traditional machine learning models displayed reasonable classification capability, with Random Forest getting the greatest score among traditional approaches. However, all conventional models were greatly outperformed by deep learning systems.

Among the deep learning algorithms, the Custom CNN had the greatest validation accuracy of 94.88%, followed by MobileNetV2 with 93.64%. Light ResNet also displayed competitive performance but remained somewhat behind the suggested CNN architecture. These results imply that deep learning models are substantially more successful for sign language picture classification owing to their capacity to automatically generate hierarchical visual representations straight from image data.

The benchmarking results clearly illustrate that the suggested CNN architecture offers the greatest balance between classification accuracy, computational efficiency, and deployment feasibility for real-time Arabic Sign Language translation applications.

4.3 Accuracy Comparison Results

Model	Accuracy
Custom CNN	94.88
MobileNetV2	93.64
Light ResNet	91.42
Random Forest	80.25

HistGradientBoosting	77.75
MLP	61.6
Linear SVM	47.85

Table 1: Model Validation Accuracy Comparison

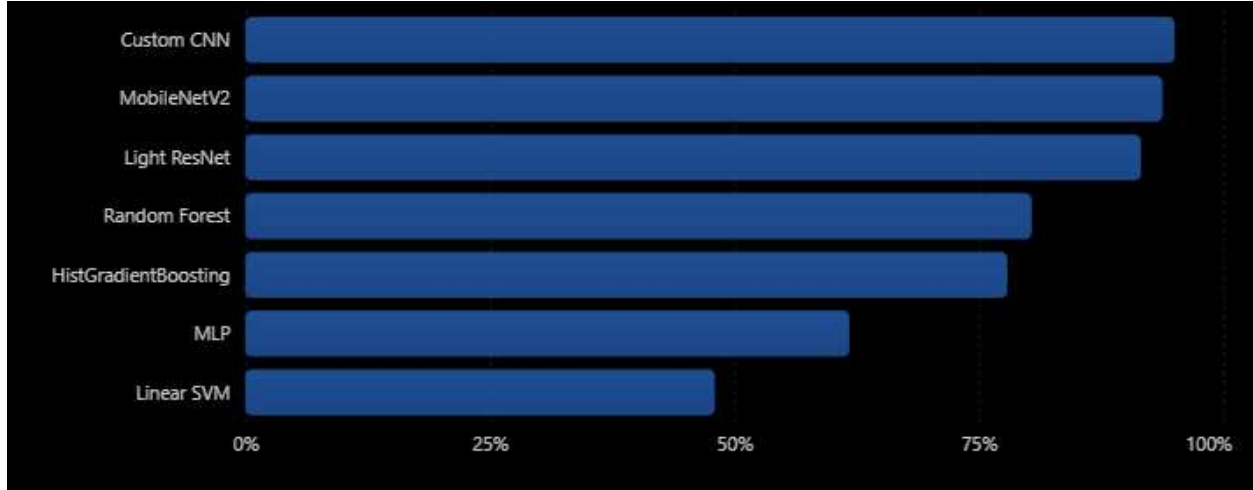


Figure 4.3-1: Model Validation Accuracy Comparison

The comparison accuracy table demonstrates the validation accuracy attained by all tested models. The suggested Custom CNN achieved the maximum validation accuracy of 94.88%, exhibiting greater recognition capabilities compared to both classic machine learning approaches and other deep learning architectures. MobileNetV2 had the second-best score with 93.64%, while Light ResNet also generated good classification results.

Traditional machine learning methods demonstrated much lower accuracy levels, revealing their limits when processing sophisticated image-based sign language datasets. The findings suggest that convolutional neural networks offer a more appropriate framework for extracting significant visual characteristics and detecting Arabic Sign Language motions properly.

4.4 CNN Training and Validation Results

The training phase of the proposed CNN model was tracked using training accuracy and validation accuracy measures over numerous epochs. The recorded curves demonstrate a solid and consistent learning process during training.

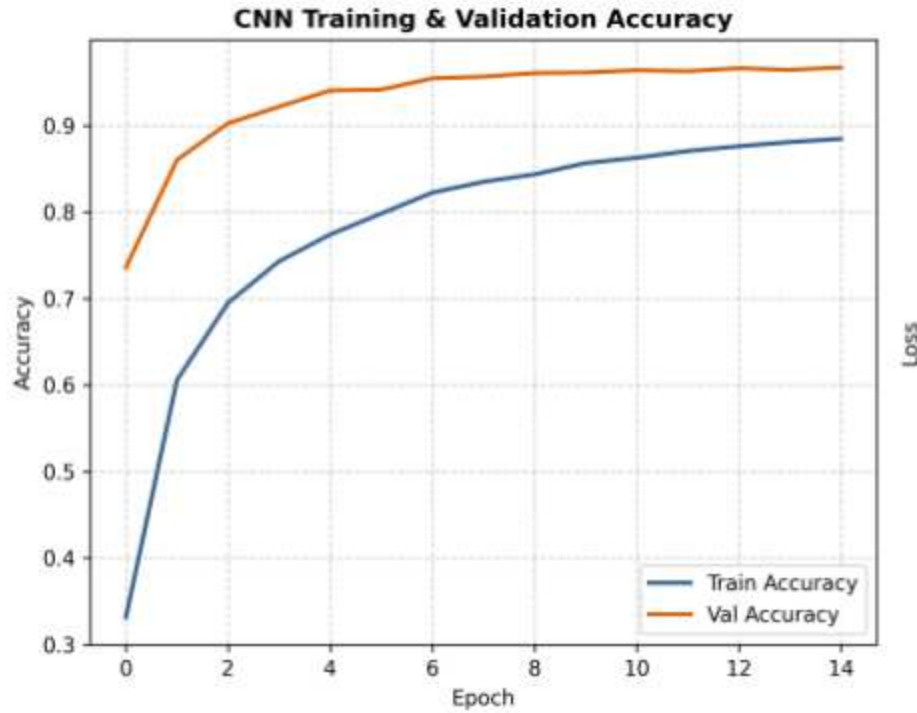


Figure 4.4-1: CNN Training and Validation

Training accuracy grew gradually as the model learnt discriminative gesture characteristics from the dataset. Similarly, validation accuracy followed a parallel pattern, confirming the model's capacity to generalize successfully to unknown data. The tight alignment between the training and validation accuracy curves suggests that the model retained high generalization capabilities and did not suffer from considerable overfitting.

The final validation accuracy neared 95%, showing the efficiency of the CNN architecture for Arabic Sign Language recognition. The findings also illustrate the favorable influence of data augmentation, dropout regularization, and batch normalization approaches in boosting model resilience and classification performance.

4.5 CNN Loss Analysis

The training loss and validation loss curves were evaluated to determine optimization efficacy and model convergence. Throughout the training phase, both loss values fell significantly, suggesting good learning and parameter optimization.

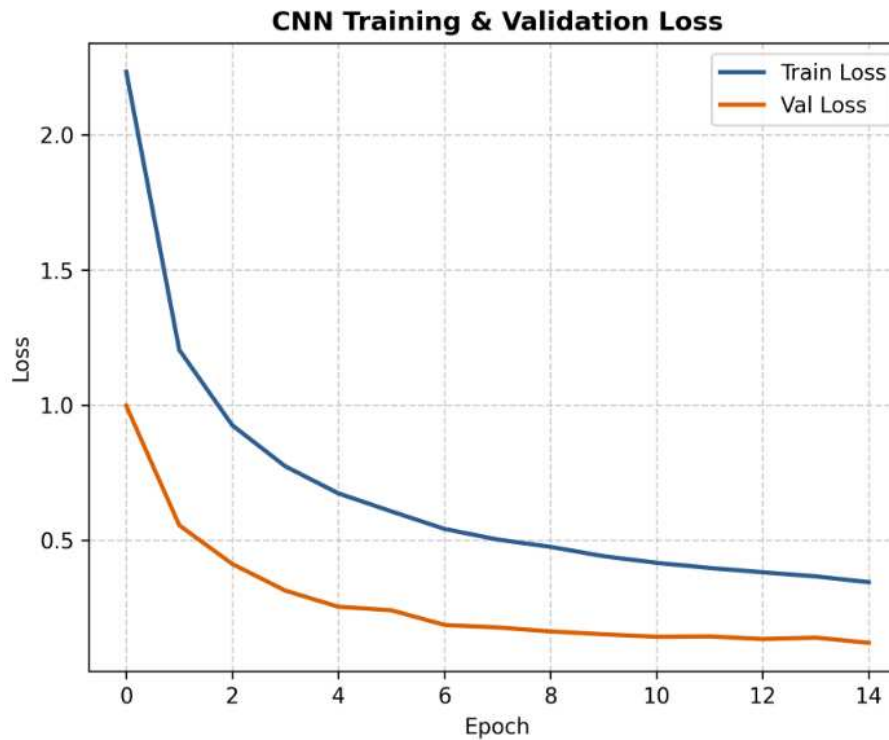


Figure 4.5-1: CNN Loss Analysis

Training loss revealed a steady lower trend as the network increased its capacity to recognize sign language motions properly. Validation loss followed a similar trend and stayed reasonably close to training loss, demonstrating that the model generalized efficiently to unknown data without remembering training examples.

The lack of a large divergence between training and validation loss shows that overfitting was effectively minimized by the implementation of dropout layers and data augmentation strategies. The final loss values show that the CNN model converged effectively and produced steady classification performance adequate for implementation in real-time recognition applications.

4.6 Discussion of Results

The experimental outcomes clearly indicate the superiority of the proposed Custom CNN architecture over both classic machine learning models and other deep learning methodologies. Several variables led to this improved achievement.

First, the CNN architecture retains the spatial structure of pictures, allowing the model to acquire important visual patterns such as finger placements, hand orientation, and gesture forms. Traditional machine learning algorithms depend on flattened feature vectors, which ignore essential spatial correlations and diminish classification performance.

Second, the hierarchical feature extraction mechanism of CNNs allows the network to learn progressively more complex representations across multiple layers. Early layers recognize edges and outlines, whereas later layers record whole gesture structures, leading to enhanced identification accuracy.

Third, max-pooling processes give resilience against positional fluctuations, enabling the system to identify movements even when hand placements fluctuate slightly inside the visual frame. Furthermore, dropout regularization and data augmentation helped greatly to reducing overfitting and enhancing model generalization.

The comparison findings consequently confirm the adoption of the Custom CNN as the fundamental recognition engine for the proposed Arabic Sign Language translation system. Its combination of high accuracy, efficient training behavior, and real-time deployment capabilities makes it the most ideal option for the target application.

4.7 Real-Time Translation System Evaluation

The last part of the project entailed analyzing the whole real-time Arabic Sign Language translation system. The trained CNN model was combined with a FastAPI backend and a web-based user interface capable of receiving live camera input and providing translation results in real time.

Experimental testing indicated that the system received incoming picture frames, ran model inference, and provided prediction results with confidence ratings via the online dashboard. The

system was capable of continually identifying Arabic sign language characters and displaying the translated output directly to the user.

The Continuous Sentence Builder module greatly boosted usability by merging reliable sequential predictions into coherent text sequences. This functionality allows users to compose words and brief statements from individual sign movements while retaining an intuitive interaction experience.

Overall, the integrated system exhibited dependable real-time performance and proved the practical application of deep learning approaches for Arabic Sign Language detection and translation. The successful deployment of the model inside a comprehensive web-based framework indicates the viability of establishing intelligent assistive communication systems based on computer vision and deep learning technology.