Salud, Ciencia y Tecnología. 2025; 5:2309
doi: 10.56294/saludcyt20252309

ORIGINAL

Deep Learning Approach for Arabic Sign Language Alphabet Recognition

Enfoque de Aprendizaje Profundo para el Reconocimiento del Alfabeto de la Lengua de Señas Árabe

Abdelfatah Maarouf[1], Otman Maarouf[1,2], Rachid El Ayachi[1], Mohamed Biniz[3]

[1]Faculty of Science and Technology Sultan Moulay Slimane University, Department of Computer Science. Beni Mellal, Morocco.
[2]Faculty of Sciences Ibnou Zohr University, Center of Excellence, LabSIV laboratory, Department of Computer Science. Agadir, Morocco.
[3]Polydisciplinary Faculty Sultan Moulay Slimane University, Department of Computer Science. Beni Mellal, Morocco.

Cite as: Maarouf A, Maarouf O, El Ayachi R, Biniz M. Deep Learning Approach for Arabic Sign Language Alphabet Recognition. Salud, Ciencia y Tecnología. 2025; 5:2309. https://doi.org/10.56294/saludcyt20252309

Submitted: 28-03-2025 **Revised:** 21-06-2025 **Accepted:** 14-10-2025 **Published:** 15-10-2025

Editor: Prof. Dr. William Castillo-González

Corresponding Author: Abdelfatah Maarouf

ABSTRACT

Introduction: sign language plays a crucial role in enabling communication for individuals with hearing impairments. Among the various sign languages, Arabic Sign Language (ArSL) is one of the most widely used in the Arab world. It consists of two main forms: word-based ArSL and alphabetic ArSL (ArSLA), where each Arabic letter is represented by a specific hand sign.
Objective: this study aims to develop an effective and robust classification model for recognizing Arabic Sign Language Alphabets to enhance communication accessibility for the hearing-impaired community.
Method: a Convolutional Neural Network (CNN) architecture was designed and trained on a dataset of Arabic Sign Language Alphabet images. The model's performance was evaluated using accuracy metrics on both training and testing datasets.
Results: the proposed CNN model achieved an accuracy of 99,4 % on the training set and 96,57 % on the test set, demonstrating its strong generalization ability in recognizing Arabic sign alphabets.
Conclusions: the findings confirm the effectiveness of a simple CNN-based approach for Arabic Sign Language Alphabet recognition. This work highlights the potential of deep learning methods to promote accessibility and social inclusion for individuals with hearing disabilities.

Keywords: Arabic Sign Language Alphabets; Deep Learning; Hearing Disabilities; Hearing-Impaired Community; Convolutional Neural Network; ArSL2018.

RESUMEN

Introducción: el lenguaje de señas desempeña un papel fundamental en la comunicación de las personas con discapacidades auditivas. Entre los diversos lenguajes de señas, el Lenguaje de Señas Árabe (ArSL) es uno de los más utilizados en el mundo árabe. Consta de dos formas principales: el ArSL basado en palabras y el ArSL alfabético (ArSLA), en el cual cada letra árabe se representa mediante un signo manual específico.
Objetivo: este estudio tiene como objetivo desarrollar un modelo de clasificación eficaz y robusto para el reconocimiento del alfabeto del Lenguaje de Señas Árabe, con el fin de mejorar la accesibilidad a la comunicación para la comunidad con discapacidad auditiva.
Método: se diseñó y entrenó una arquitectura de Red Neuronal Convolucional (CNN) utilizando un conjunto de datos de imágenes del alfabeto del Lenguaje de Señas Árabe. El rendimiento del modelo se evaluó mediante métricas de precisión en los conjuntos de entrenamiento y de prueba.
Resultados: el modelo CNN propuesto alcanzó una precisión del 99,4 % en el conjunto de entrenamiento y del 96,57 % en el conjunto de prueba, demostrando una gran capacidad de generalización en el reconocimiento de los alfabetos de señas árabes.

Conclusiones: los resultados confirman la eficacia de un enfoque sencillo basado en CNN para el reconocimiento del alfabeto del Lenguaje de Señas Árabe. Este trabajo destaca el potencial de los métodos de aprendizaje profundo para promover la accesibilidad y la inclusión social de las personas con discapacidades auditivas.

Palabras clave: Alfabeto de la Lengua de Señas Árabe; Aprendizaje Profundo; Discapacidades Auditivas; Comunidad con Discapacidad Auditiva; Red Neuronal Convolucional; ArSL2018.

INTRODUCTION

Communication is a fundamental human need, and for individuals with hearing impairments, sign language serves as an essential means of expression and interaction. The absence of accessible communication tools often leads to social and educational barriers for the deaf and hard-of-hearing community. In the Arab world, Arabic Sign Language (ArSL) is widely used, yet remains underrepresented in technological and linguistic research compared to other sign languages such as American Sign Language (ASL) or British Sign Language (BSL). Developing automated recognition systems for ArSL can therefore play a vital role in improving inclusion, accessibility, and equal opportunities for individuals with hearing disabilities.

Over the past decade, advances in computer vision and deep learning have significantly enhanced gesture and image recognition tasks, opening new possibilities for sign language interpretation. Several researchers have proposed innovative solutions to tackle this challenge. For instance, [1] introduced a glove-based recognition system for ASL, while [2] applied multiple kernel learning to recognize Pakistan Sign Language gestures. Deep learning techniques, particularly Convolutional Neural Networks (CNNs), have proven especially effective in identifying spatial patterns in hand gestures.[3,4] Comprehensive reviews, such as that by [5], emphasize CNNs' potential to improve classification performance while identifying key challenges related to dataset diversity and computational efficiency.

Despite these advancements, limited research has focused on Arabic Sign Language, particularly at the alphabetic level. Many existing models are either computationally expensive or rely on large-scale datasets that are not always available for ArSL. Therefore, there is a pressing need for lightweight yet accurate models that can effectively classify Arabic Sign Language Alphabets, even in resource-constrained environments.

This research addresses this gap by developing a simple yet robust CNN-based model for recognizing Arabic Sign Language Alphabets (ArSLA). The model aims to balance high accuracy with computational efficiency, contributing to ongoing efforts to create practical and accessible sign language recognition systems tailored to Arabic users.

METHOD

Dataset

Source: https://ars.els-cdn.com/content/image/1-s2.0-S2352340919301283-gr1.jpg

Figure 1. Selected dataset

A comprehensive and fully-labeled dataset, named ArSL2018,[6] has been curated to advance research in the domain of Arabic Sign Language (ArSL) recognition. The primary objective of this dataset is to facilitate the development of automated systems that cater to the needs of the deaf and hard-of-hearing population through the application of machine learning, computer vision, and deep learning algorithms.

The ArSL2018 dataset comprises a substantial collection of 54 049 images representing 32 Arabic sign language signs and alphabets. These images were meticulously gathered from 40 participants spanning different age groups, ensuring diversity in the dataset. The dataset captures real-world complexities by intentionally incorporating variations in lighting conditions, angles, timings, and backgrounds, which reflect the intricacies of Arabic Sign Language gestures.

The images were captured at Prince Mohammad Bin Fahd University and in the Khobar Area, Kingdom of Saudi Arabia, with volunteers positioned approximately 1 meter away from a smart camera mounted on a tripod. During the data collection process, variations were deliberately introduced to enhance realism, encompassing diverse environmental conditions and participant movements.

To ensure uniformity and optimize the dataset for classification and recognition tasks, a series of pre-processing techniques were applied. All images were resized to a fixed dimension of 64 x 64 pixels and converted to grayscale format, standardizing pixel values to a range of 0 to 255 (figure 1). These pre-processing steps were crucial in preparing the dataset for effective utilization in deep learning models, ensuring consistency and suitability for classification tasks.

Data Pre-processing

Data pre-processing is crucial for preparing the dataset for training. The Arabic Sign Language (ArSL) dataset underwent several key steps:

Dataset Splitting

The dataset was split into training (80 %, 43 240 images) and testing (20 %, 10 809 images) subsets using the train test split function from the sklearn.model selection module.

Image Transfer

Images were organized into class-specific folders for both training and testing sets to maintain structured classification.

Data Augmentation

To enhance the diversity of the training set, rescaling and zoom range augmentations (0,2) were applied using ImageDataGenerator from the tensorflow.keras.preprocessing.image module.

Generator Initialization

Image generators were set up for both training and testing datasets, resizing all images to a uniform 150x150 pixels to ensure consistency and computational efficiency during model training.

Label Mapping

A mapping between class indices and labels was created to ensure accurate interpretation of the model's predictions.

Visualization

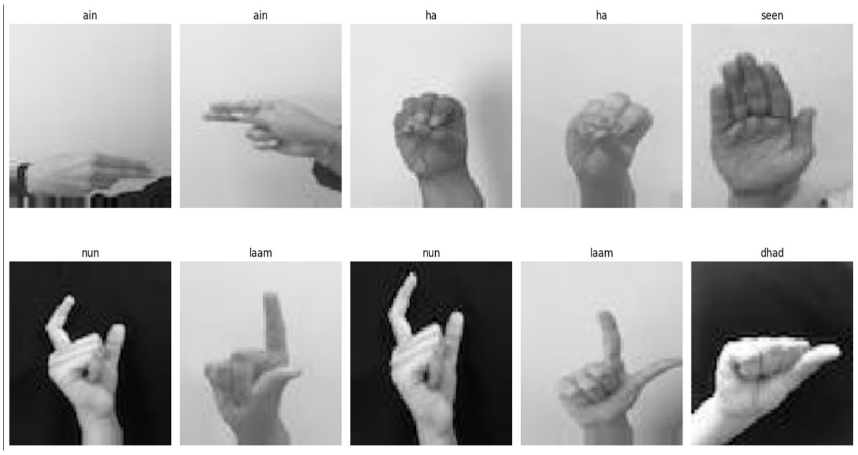

Figure 2. Training samples after pre-processed

Sample images from the training set were visualized using matplotlib to provide insights into the dataset and its diverse sign language representations (figure 2).

These pre-processing steps ensured the dataset was effectively prepared for the classification task, contributing to more reliable model training.

Model Architecture

The proposed model architecture is a simple Convolutional Neural Network (CNN) designed for the efficient classification of Arabic Sign Language alphabets (figure 3). It is composed of multiple convolutional and dense layers that progressively extract and process meaningful features from the input data. The architecture begins with a convolutional layer with 128 filters, a kernel size of (5, 5), and ReLU activation. This layer is followed by a MaxPooling layer with a pool size of (2, 2), which reduces the spatial dimensions, and Batch Normalization, which helps stabilize the training process and speed up convergence.

Layer (type)	Output Shape	Param #
conv2d_6 (Conv2D)	(None, 60, 60, 128)	9728
activation_6 (Activation)	(None, 60, 60, 128)	0
max_pooling2d_6 (MaxPooling2D)	(None, 30, 30, 128)	0
batch_normalization_6 (BatchNormalization)	(None, 30, 30, 128)	512
conv2d_7 (Conv2D)	(None, 28, 28, 64)	73792
activation_7 (Activation)	(None, 28, 28, 64)	0
max_pooling2d_7 (MaxPooling2D)	(None, 14, 14, 64)	0
batch_normalization_7 (BatchNormalization)	(None, 14, 14, 64)	256
conv2d_8 (Conv2D)	(None, 12, 12, 32)	18464
activation_8 (Activation)	(None, 12, 12, 32)	0
max_pooling2d_8 (MaxPooling2D)	(None, 6, 6, 32)	0
batch_normalization_8 (BatchNormalization)	(None, 6, 6, 32)	128
flatten_2 (Flatten)	(None, 1152)	0
dense_4 (Dense)	(None, 256)	295168
dropout_2 (Dropout)	(None, 256)	0
dense_5 (Dense)	(None, 32)	8224

```
Total params: 406272 (1.55 MB)
Trainable params: 405824 (1.55 MB)
Non-trainable params: 448 (1.75 KB)
```

Figure 3. Model architecture

The second convolutional layer consists of 64 filters with a kernel size of (3, 3). It incorporates L2 regularization with a penalty of 0,00005 to prevent overfitting. This layer is also paired with MaxPooling and Batch Normalization, ensuring consistent feature extraction and improved training stability. Similarly, the third convolutional layer uses 32 filters with a kernel size of (3, 3), along with MaxPooling, Batch Normalization, and L2 regularization to enhance the model's ability to learn intricate patterns in the data.

Following the convolutional layers, the feature maps are flattened into a one- dimensional vector, preparing the data for the fully connected layers. The first dense layer consists of 256 units with ReLU activation, which extracts higher-level features. To prevent overfitting, a Dropout layer with a rate of 0,5 is applied, randomly deactivating neurons during training. The final dense layer, with 32 units, uses a Softmax activation function to output class probabilities corresponding to the 32 Arabic Sign Language alphabet classes.

The model is compiled using the Adam optimizer with a learning rate of 0,001, chosen for its ability to dynamically adapt the learning process for faster and more stable convergence. The loss function is set to categorical cross-entropy, which is ideal for multi-class classification tasks, while accuracy is used as the primary evaluation metric. This compilation setup ensures that the model is optimized for both learning and generalization.

The training process spans 40 epochs and uses a batch size of 32, balancing computational efficiency and model performance. A ReduceLROnPlateau callback dynamically adjusts the learning rate whenever the validation loss plateaus, ensuring smoother convergence. Additionally, model checkpoints are saved at regular intervals, providing a mechanism for resuming or fine-tuning the training process as needed.

Overall, this simple CNN architecture balances performance and computational efficiency. It is well-suited for the ArSL2018 dataset, effectively leveraging its ability to learn spatial and hierarchical features while maintaining feasibility for deployment in resource-constrained environments.

RESULTS

This section presents the experimental outcomes of the proposed simple CNN model using the ArSL2018 dataset, which contains 54 049 images representing a variety of Arabic Sign Language alphabet gestures. The evaluation was conducted using key performance metrics such as accuracy, loss, and learning rate dynamics, accompanied by graphical and tabular summaries.

Model Training Performance

The model was trained over 40 epochs, demonstrating stable learning behavior throughout the process. The accuracy plot (figure 4) shows a steady upward trend for both training and validation sets, indicating effective learning and consistent performance.

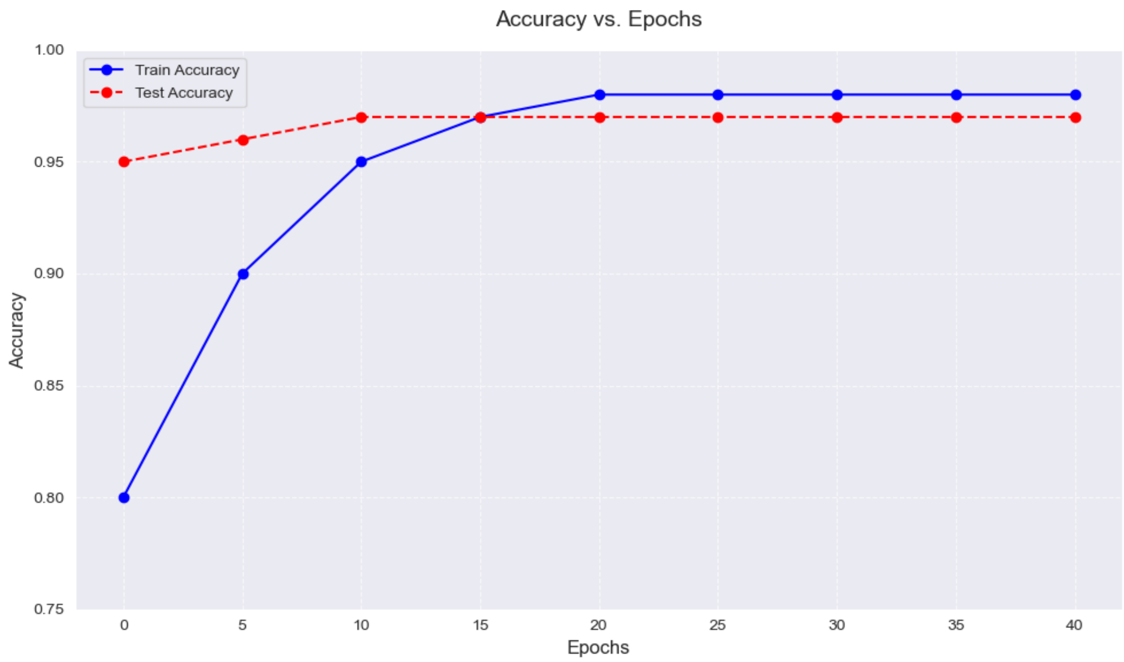

Figure 4. Accuracy vs Epochs

Loss Behavior

As shown in figure 5, both training and validation losses decreased progressively over time, confirming

https://doi.org/10.56294/saludcyt20252309 ISSN: 2796-9711

efficient convergence and the model's ability to minimize prediction errors.

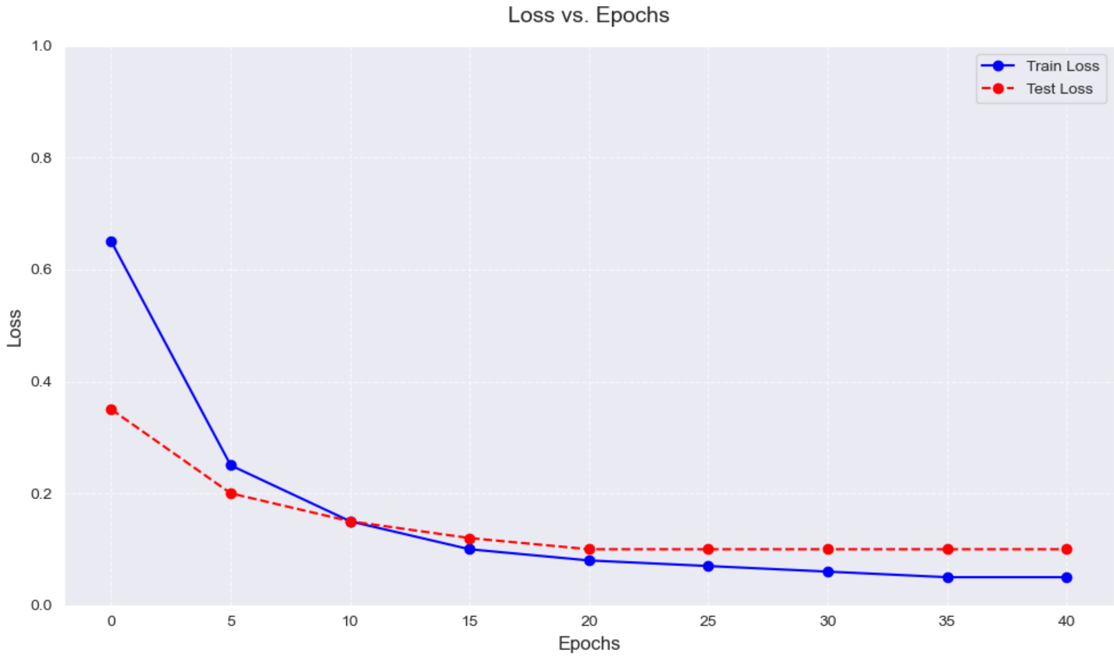

Figure 5. Loss vs Epochs

Learning Rate Dynamics

The learning rate schedule (figure 6) started at 0,001 and was adjusted dynamically based on validation loss plateaus. This adaptive strategy facilitated smooth convergence and prevented overshooting during optimization.

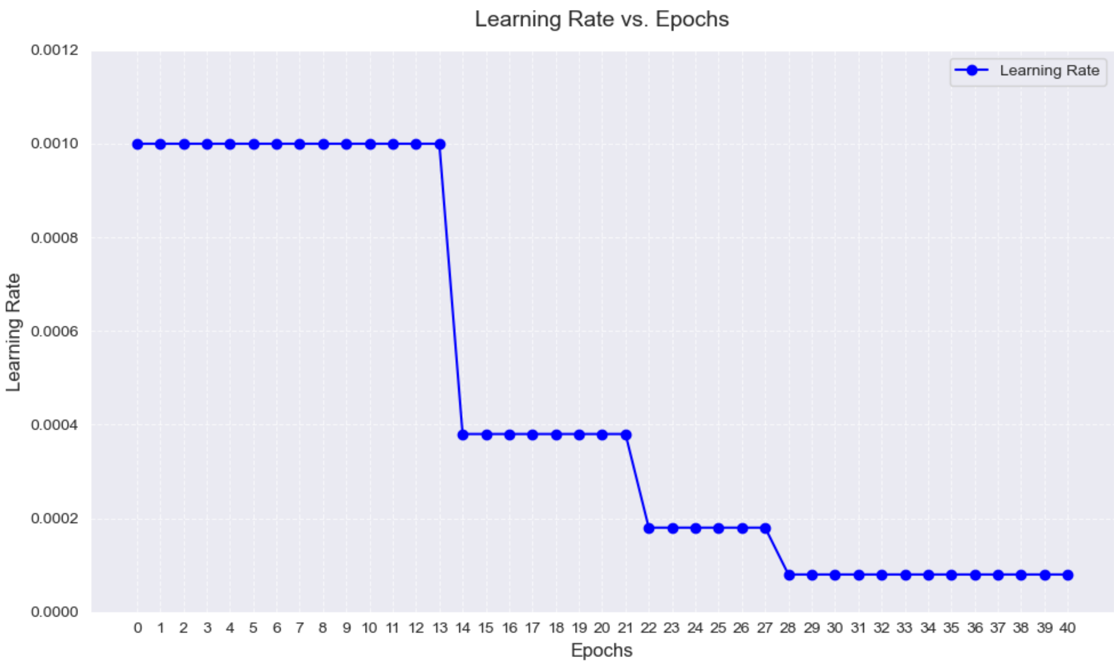

Figure 6. Learning Rate vs Epochs

Confusion Matrix and Misclassifications

The confusion matrix (figure 7) reveals a strong diagonal distribution, indicating that most predictions were correct. Minor misclassifications occurred between visually similar gestures, such as "DHA" and "TA" (figure 8), where seven instances of "TA" were incorrectly classified as "DHA" and four instances of "DHA" were misclassified as "TA".

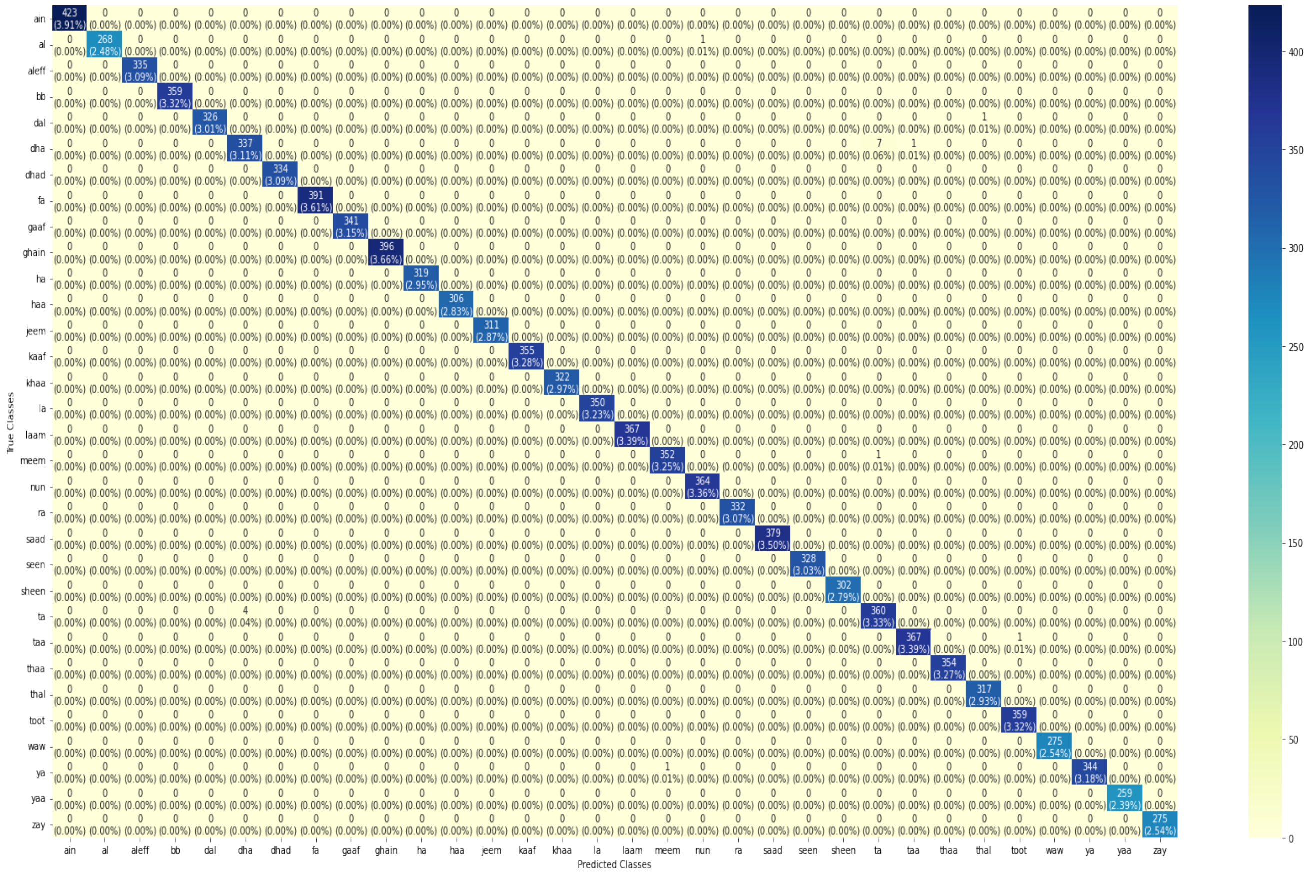

Figure 7. Confusion matrix

Figure 8. "TA" and "DHA"

Performance Metrics

Table 1. summarizes the model's quantitative performance on the training and test datasets

Metric	Training Set	Test Set
Accuracy	0,994	0,9657
Precision	0,995	0,97
Recall	0,9963	0,967

These findings collectively confirm that the proposed CNN model achieves strong recognition performance on the Arabic Sign Language alphabet dataset.

DISCUSSION

Summary of Key Findings

This research aimed to develop an efficient and accurate deep learning model for Arabic Sign Language Alphabet (ArSLA) recognition. The proposed simple CNN achieved high performance, with 99,4 % accuracy on the training set and 96,57 % on the test set. These results demonstrate that even a lightweight CNN can effectively classify complex hand gesture patterns associated with Arabic sign letters.

Interpretation of Findings

The obtained results align with existing research highlighting the effectiveness of CNN-based models for sign language recognition. Similar high accuracies were reported by [7] using ResNet[8] architectures and by [4,9] in hybrid CNN frameworks. However, compared to these studies, the present work employs a simpler architecture that achieves comparable accuracy while maintaining lower computational costs.

The slight misclassifications observed between "DHA" and "TA" reflect the inherent visual similarities among certain Arabic signs, a phenomenon also noted in prior works.[10] This suggests that feature overlap between similar gestures remains a persistent challenge for vision-based recognition systems.

Implications

The results demonstrate that robust Arabic Sign Language Alphabet recognition can be achieved using lightweight CNN architectures. This is particularly significant for real-world applications in educational tools, communication aids, and low-resource deployment environments, such as mobile or embedded systems. The findings also reinforce the importance of developing language-specific datasets to enhance inclusion within the Arabic-speaking deaf community.

Limitations

Despite the promising results, this study has some limitations. The dataset used, while extensive, may not capture the full diversity of lighting conditions, backgrounds, and signer variations encountered in real-world settings. Additionally, the model focuses on static alphabet signs, excluding dynamic gestures or continuous signing, which limits its applicability to full-sentence recognition. Future research should explore temporal models such as LSTM-CNN hybrids or Transformer-based approaches to handle sequential signing patterns.

Conclusion and Future Directions

This work validates the potential of a simple CNN for Arabic Sign Language Alphabet recognition, achieving competitive accuracy while maintaining computational efficiency. Future studies should aim to expand the dataset to include more signers, varied environments, and dynamic gestures. Integrating real-time detection and user feedback mechanisms would further enhance the system's usability and impact in promoting accessibility for the hearing-impaired community.

CONCLUSIONS

This study developed and evaluated a simple Convolutional Neural Network (CNN) model for recognizing Arabic Sign Language Alphabets (ArSLA) using the ArSL2018 dataset. The model achieved a high classification accuracy of 96,57 % on the test set, confirming its effectiveness and robustness in identifying Arabic sign gestures.

The results demonstrate that a lightweight CNN architecture can deliver strong performance while maintaining computational efficiency, making it suitable for real-world applications such as educational tools and assistive communication systems for the hearing-impaired community.

Although the findings are promising, future research should focus on optimizing the model for deployment on low-resource devices, expanding the dataset to include more signers and dynamic gestures, and integrating temporal models to handle continuous signing.

In conclusion, the proposed CNN provides a reliable foundation for advancing Arabic Sign Language recognition and contributes to ongoing efforts to enhance accessibility and social inclusion for individuals with hearing disabilities.

BIBLIOGRAPHIC REFERENCES

1. Amin MS, Amin MT, Latif MY, Jathol AA, Ahmed N, Tarar MIN. Alphabetical Gesture Recognition of American Sign Language using E-Voice Smart Glove. In: 2020 IEEE 23rd International Multitopic Conference (INMIC). Bahawalpur, Pakistan: IEEE; 2020. https://ieeexplore.ieee.org/document/9318185/

2. Shah F, Shah MS, Akram W, Manzoor A, Mahmoud RO, Abdelminaam DS. Sign Language Recognition Using Multiple Kernel Learning: A Case Study of Pakistan Sign Language. IEEE Access. 2021;9:67548–58.

3. Zhang Y, Jiang X. Recent Advances on Deep Learning for Sign Language Recognition. CMES. 2024;139(3):2399–450.

4. Buttar AM, Ahmad U, Gumaei AH, Assiri A, Akbar MA, Alkhamees BF. Deep Learning in Sign Language Recognition: A Hybrid Approach for the Recognition of Static and Dynamic Signs. Mathematics. 2023 Aug 30;11(17):3729.

5. Al-Qurishi M, Khalid T, Souissi R. Deep Learning for Sign Language Recognition: Current Techniques, Benchmarks, and Open Issues. IEEE Access. 2021;9:126917–51.

6. Latif G, Mohammad N, Alghazo J, AlKhalaf R, AlKhalaf R. ArASL: Arabic Alphabets Sign Language Dataset. Data in Brief. 2019 Apr;23:103777.

7. Alasmari N, Asiri S. ASLDetect: Arabic sign language detection using ResNet and U-Net like component. Sci Rep. 2025 May 23;15(1):18012.

8. Targ S, Almeida D, Lyman K. Resnet in Resnet: Generalizing Residual Architectures. arXiv; 2016. http://arxiv.org/abs/1603.08029

9. Kamruzzaman MM. Arabic Sign Language Recognition and Generating Arabic Speech Using Convolutional Neural Network. Wireless Communications and Mobile Computing. 2020 May 23;2020:1–9.

10. Pan W, Zhang X, Ye Z. Attention-Based Sign Language Recognition Network Utilizing Keyframe Sampling and Skeletal Features. IEEE Access. 2020;8:215592–602.

FINANCING
The authors did not receive financing for the development of this research.

CONFLICT OF INTEREST
The authors declare that there is no conflict of interest.

AUTHORSHIP CONTRIBUTION
Conceptualization: Abdelfatah Maarouf, Otman Maarouf.
Data curation: Abdelfatah Maarouf, Otman Maarouf.
Formal analysis: Abdelfatah Maarouf, Otman Maarouf.
Research: Abdelfatah Maarouf, Otman Maarouf
Methodology: Abdelfatah Maarouf, Otman Maarouf, Rachid El Ayachi, Mohamed Biniz.
Project management: Rachid El Ayachi, Mohamed Biniz.
Resources: Abdelfatah Maarouf, Otman Maarouf.
Software: Abdelfatah Maarouf.
Supervision: Rachid El Ayachi, Mohamed Biniz.
Validation: Rachid El Ayachi, Mohamed Biniz.
Display: Abdelfatah Maarouf, Otman Maarouf, Rachid El Ayachi, Mohamed Biniz.
Drafting - original draft: Abdelfatah Maarouf
Writing - proofreading and editing: Abdelfatah Maarouf, Otman Maarouf, Rachid El Ayachi, Mohamed Biniz.