ISSN: 1672 - 6553
JOURNAL DYNAMICS AND CONTROL
VOLUME 9 ISSUE 11: P1-9

A DEEP NEURAL NETWORK FRAMEWORK FOR ACCURATE AND NOISE-ROBUST RECOGNITION OF ARABIC SIGN LANGUAGE

Mohammed Kamel Benkaddour

Department of Computer Science and Information Technology,
Laboratory of Artificial Intelligence and Information Technologies (LINATI),
Kasdi Merbah University, Ouargla, Algeria

JOURNAL OF DYNAMICS AND CONTROL

https://doi.org/10.71058/jodac.v9i11001

A DEEP NEURAL NETWORK FRAMEWORK FOR ACCURATE AND NOISE-ROBUST RECOGNITION OF ARABIC SIGN LANGUAGE

Mohammed Kamel Benkaddour

Department of Computer Science and Information Technology,
Laboratory of Artificial Intelligence and Information Technologies (LINATI),
Kasdi Merbah University, Ouargla, Algeria.

Received on 22-05 -2025; Accepted on 25-10-2025; Published on 07-11-2025

Abstract: *This paper proposes a deep neural network framework for robust and accurate recognition of Arabic Sign Language (ArSL) gestures. The system employs a multi-layer convolutional neural network (CNN) architecture optimized for spatial feature extraction from complex hand shapes. A novel preprocessing pipeline is introduced, utilizing a hand-tracking algorithm to isolate hand regions and generate skeletal representations, thereby reducing background noise and improving recognition consistency under variable lighting conditions. The model is trained and evaluated on the ArASL2018 dataset, comprising over 54,000 labeled gesture images spanning 39 alphabetic and word classes. To enhance generalization, the training data were augmented through rotation, zoom, and shifting transformations. The proposed framework achieved an overall accuracy of 99.8% and demonstrated remarkable tolerance to Gaussian noise, confirming its ability to maintain performance in degraded visual environments.*

Keywords: *Arabic Sign Language, Convolutional neural network, ArSL, Hand gesture.*

1. Introduction

Deaf-mute individuals, who are both unable to hear and speak, form a significant part of the global population with disabilities. According to the World Health Organization (WHO) [1], around 70 million people worldwide fall into this category. More broadly, about 360 million people live with some form of hearing loss, including 32 million children. Alarmingly, by 2050, one in every four people is projected to experience some level of hearing loss. These numbers are not just statistics, but they represent families and communities who are deeply affected by these challenges. In Algeria alone, the African Sign Language Resource Center [4] reports 220,000 deaf individuals. Across the middle east's population of 350 million, over 11 million people suffer from disabling hearing loss, as noted by the Center for Strategic and International Studies (CSIS) in 2014 [12]. This issue also significantly impacts the arabic-speaking world. As arabic nations work toward greater development and strive to compete on the global stage, it becomes increasingly critical to include and empower every segment of society. Deaf-mute individuals make up a considerable percentage of the Arabic population, and marginalizing their talents and contributions would not only harm these individuals but also lead to missed opportunities for societal progress. By valuing their ideas and ensuring their inclusion, these individuals can contribute meaningfully to both their communities and the broader nation [2].

Sign language is the cornerstone of communication for the deaf-mute community. It is a powerful, expressive medium that uses visual gestures, hand movements, facial expressions, and body language to convey meaning. With its unique grammar and vocabulary, sign language enables effective communication, fosters social engagement, provides access to education, and allows deaf-mute individuals to participate fully in all areas of life. It is also a valuable alternative for communicating in foreign environments when spoken language skills are lacking. The rich cultural diversity of the arabic-speaking world, spread across regions like the middle east and north africa, has given rise to various forms of arabic sign language. These regional

https://doi.org/10.71058/jodac.v9i11001

and historical differences have resulted in distinct dialects, such as egyptian sign language and saudi sign language. Unfortunately, there is no standardized arabic sign language not even within a single continent [16]. This lack of uniformity poses a major challenge for creating tools like a unified arabic sign language recognition platform [3].

Arabic Sign Language (ArSL) is a unique form of communication within the Arabic-speaking world, relying on hand shapes, facial expressions, and body movements to convey meaning. While sharing similarities with other sign languages globally, ArSL is shaped by the cultural and regional diversity of the Arab world, resulting in variations like egyptian sign language and gulf sign language. These differences are influenced by local dialects, traditions, and historical factors. ArSL evolved naturally within deaf communities as a visual means of bridging communication gaps, reflecting both the shared experiences and cultural richness of the arab world [2][10].

2. Related Work

Researchers have made significant strides in the field of sign language recognition by leveraging deep learning techniques. Albelwi et.al [7] in 2017 developed a deep learning-based system using Convolutional Neural Networks (CNNs) to recognize arabic sign language (ArSL) gestures and hand signs. Their approach utilized wearable sensors for data collection and employed CNNs for feature extraction. The system successfully recognized 30 hand signs of ArSL with moderate accuracy, producing vocalized speech based on the input gestures. Impressively, the system achieved a recognition success rate for 90% of participants involved in the study. Similarly, Vazquez Lopez [21] created a system for recognizing American Sign Language (ASL) gestures using CNNs and depth images captured by a Kinect camera. As part of his research, he developed a new dataset comprising depth images of ASL letters and numbers. The study also included a comparative analysis with a similar dataset for Vietnamese Sign Language.

Saleem et al. [19] explored the application of neural networks to a transformed ASL dataset, emphasizing the efficiency of pre-trained models combined with detection and segmentation algorithms. His research also highlighted the benefits of machine learning techniques like transfer Learning for gesture recognition, with a focus on augmenting datasets and comparing the accuracy of various models. Balas et.al [10] proposed a functional prototype for sign language recognition using architectures inspired by speech recognition systems. His model employed a hidden markov classifier, achieving recognition accuracy between 75% and 90%. The study evaluated the system's performance in terms of response time, resource usage, and scalability for vocabularies of varying sizes.

Kamruzzaman [13] in 2020 introduced a vision-based ArSL recognition system utilizing CNNs to translate hand signs into spoken Arabic. His system achieved 90% accuracy and demonstrated the potential for further improvements using advanced devices like Leap Motion or Kinect. Meena [15] focused on recognizing 25 static gestures from the ASL alphabet using image segmentation techniques, including otsu thresholding and canny edge detection, with features extracted through contour tracking, classifiers such as Multiclass Support Vector Machine (MCSVM) and Least Square Support Vector Machine (LSSVM) achieved high recognition accuracies of 98.6% and 99.2%, respectively. Almahri [9] reviewed the challenges of data collection for ArSL recognition, citing a lack of diverse and realistic datasets. To address this, he employed media pipe for data collection via webcams, building a recognition system for emirati sign language with an LSTM model that achieved 100% accuracy on the testing dataset. Sultan et al. [20] discussed two critical tasks in sign language processing, Sign Language Recognition (SLR) and Sign Language Identification (SLID). Their research highlighted various datasets, both static and

https://doi.org/10.71058/jodac.v9i11001

dynamic, used for tasks like recognizing numerical, alphabetical, and sentence-level representations of sign languages.

3. Materials and Methods

3.1. Dataset

For this research, the Arabic Alphabets Sign Language Dataset (ArASL) 2018 was employed [9]. This extensive dataset includes 54,049 images of arabic sign language alphabets performed by over 40 individuals, covering all 32 Arabic letters. Additionally, a sample image showcasing all Arabic language signs is included. The accompanying CSV file provides the corresponding labels for each arabic sign language image based on the image file names. In total, we compiled 39 classes, with each class containing an average of 1300 to 1700 images. This robust dataset provided a solid foundation for recognizing both isolated and continuous signs in ArSL.

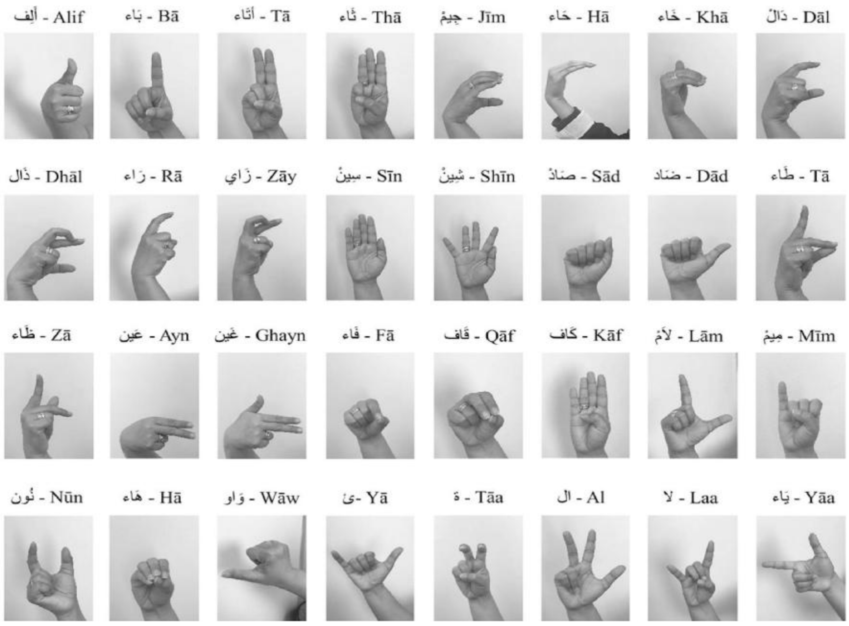

Figure 1. Representation of the Arabic Sign Language for Arabic Alphabets.

3.2. Pre-Processing

To ensure consistency and remove potential biases, we performed some pre-processing techniques on the dataset. We converted all the images to grayscale to avoid relying on skin color as a distinguishing feature. Additionally, we used the media-pipe library for hand detection and tracking. This library allowed us to accurately pinpoint the location of the hands in the captured images and crop them accordingly to eliminate any external information that could mislead the model's gesture recognition.

Furthermore, we utilized the same library to draw a shape that follows the path of the fingers, it almost represents the skeleton of the hand. This approach proved beneficial as it leverages the hand shape as a primary feature. Even when dealing with lower-quality images, the model could rely on the drawn hand shape to accurately recognize and classify the gestures since drawing the hand skeleton is pretty consistent even in bad lightning conditions.

https://doi.org/10.71058/jodac.v9i11001

We resized the images to the biggest size possible without losing necessary details for the hand-skeleton function to execute effectively. Even after the efforts of reversing the pre-processing, the hand-skeleton method failed to draw on almost 60-70% of the images, so we chose the ones with the successful operation and ignored the rest.

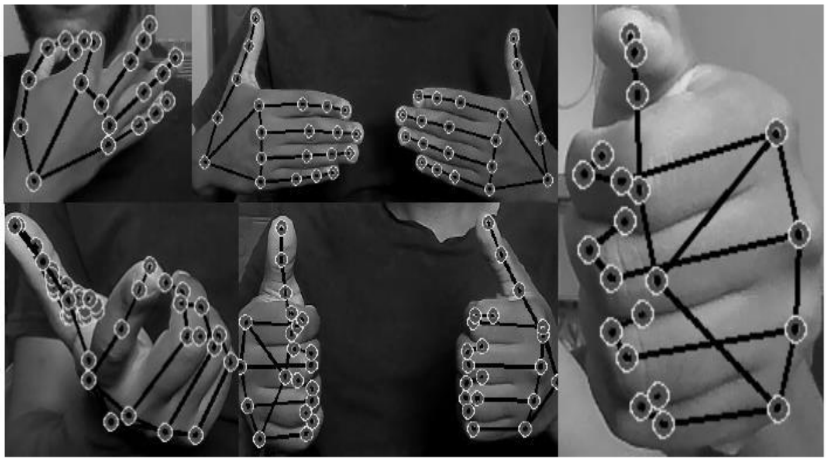

Figure 2. Some Images of alphabet and word signs from dataset after Pre-Processing.

3.3 Data-Augmentation

Since we lost a considerable amount of the alphabets dataset, and the insufficiency the numbers of the images of the words, we ended up with a small dataset, at least some classes didn't contain enough images to achieve good results. so we had to use data augmentation. In our augmentation process, we employed several key parameters. First, we used the rotation range parameter, which allowed us to rotate each image by a range of specified angles, ranging from negative to positive degrees.

- **Zoom range:** This parameter enabled us to zoom the images by a specified percentage. This zooming effect added further diversity to the dataset, as it simulated images captured from different distances or perspectives.
- **Height range** and width_range parameters: These parameters allowed us to shift and move the content of the images horizontally or vertically, creating the impression of slight displacements. This was useful in replicating different hand positions or gestures, further enriching the dataset.
- **fill mode:** This mode ensured that any empty or newly created pixels resulting from the augmentation process were filled in using the nearest neighboring pixel values. This preserved the overall appearance and integrity of the augmented images.

3.4. Proposed Architecture

The model used in this experiment is a Convolutional Neural Network [14]. CNNs are well-suited for image classification tasks due to their ability to capture spatial hierarchies and local patterns in images, in our architecture the input shape is determined by the shape of the training data (64, 64, 1), the input shape consists of the image dimensions, such as height, width, and channels. The convolutional layers begin with a Conv2D layer with 64 filters and a kernel size of (3, 3). Each Conv2D layer applies a set of learn-able filters to the input, extracting relevant features from the images, then the ReLU activation is applied after each Conv2D layer, introducing non-linearity to the network, and after each Conv2D layer, batch normalization normalizes the activations of the previous layer, helping to stabilize and speed up the training process.

In pooling layers, the MaxPooling2D is used and follow each Conv2D layer, reducing the spatial dimensions of the output feature maps while retaining the most important information, each MaxPooling2D layer uses a (2, 2) pooling window to downsample the

https://doi.org/10.71058/jodac.v9i11001

input. The Flatten layer converts the 2D feature maps from the previous layer into a 1D feature vector, preparing it for input into the dense layers. Two dense layers follow the Flatten layer, each with different units (512 and 256) and ReLU activation, the dense layers are fully connected layers, responsible for learning high-level representations based on the extracted features from the convolutional layers. Dropout layers with a rate of 0.5 are applied after each dense layer. Dropout randomly sets a fraction of input units to 0 during training, which helps prevent overfitting by introducing regularization. The final dense layer has units equal to the number of classes in the dataset, and it uses softmax activation. Softmax activation produces a probability distribution over the classes, indicating the model's predicted probabilities for each class. The model is compiled with the Adam optimizer, which is an adaptive learning rate optimization algorithm. The loss function used is sparse categorical cross-entropy, suitable for multi-class classification tasks. The accuracy metric is used to monitor the model's performance during training. The model is trained using the fit function, with a batch size of 128, 30 epochs, and validation split of 20% is used to evaluate the model's performance on a portion of the training data during training.

This architecture aims to capture and learn hierarchical features from the input images through multiple convolutional and pooling layers. The dense layers provide a high-level representation of the extracted features, leading to improved classification performance. Dropout is employed to prevent overfitting, and batch normalization enhances the stability and speed of training.

Table 1. Model Summary.

Layers	Output Shape	Parameter
Conv2D	(62, 62, 32)	320
MaxPooling2D	(31, 31, 32)	0
Conv2D	(29, 29, 64)	18496
MaxPooling2D	(14,14,64)	0
Conv2D	(12, 12, 128)	73856
MaxPooling2D	(6, 6, 128)	0
Flatten	(4608)	0
Dense	(128)	589952
Dense	(40)	5160

4. Result and Discussion

As we described earlier, our architecture uses a CNN model with three sets of convolutional layers followed by batch normalization and max pooling. It includes two fully connected layers with ReLU activation and dropout regularization. The final output layer uses softmax activation for multi-class classification. The model is trained using the Adam optimizer with categorical cross-entropy loss. We trained this architecture for 30 epochs, using a batch size of 128. After obtaining astonishing results, we experimented with the model to test it in difficult situations, so we introduced noise into the testing images with variant levels. We picked about four images form four different classes and we added noise to them using gaussian filter with different levels of intensity up to 100%. This experiment was quite interesting because of the variety of results from class to class, which enhances our understanding of the model.

4.1. Model Evaluation

In this section, we share the results of our experiments on recognizing Arabic alphabet sign language using the proposed fine-tuned CNNs. We assess the models' performance on test

https://doi.org/10.71058/jodac.v9i11001

datasets and compare their accuracy to highlight their effectiveness. We evaluated our models primarily using accuracy, a simple yet effective way to measure how well the models classify Arabic alphabet signs. To give a complete picture of the model's performance and identify any signs of overfitting or underfitting, we report accuracy, loss, recall, support and F1-scores for the training, validation, and test sets [15] [17]. The performance evaluation of our model in the classification report demonstrates its high accuracy and effectiveness in recognizing sign language gestures. The model achieved an overall accuracy of 99.8%, indicating its amazing ability to classify gestures correctly, on the first step of testing. After the experiments performed on the model, we come to a conclusion that our model is variant with noise even in extreme conditions, we saw some images that reached maximum intensity of noise which is 100%, but the model succeeded predicted it correctly which confirms the model ability to tolerate significant levels of noise and perform effectively.

The training and validation accuracy shown in figure 3 were extremely close, this means that the model is neither underfitting nor overfitting which signifies that the model is performing robustly across both the training and the unseen validation data, which indicates a well generalized model that achieved a good balance between capturing the complexities of the training data and being able to generalize well and apply those conclusions on new unseen data. Consequently, it’s very likely that the model will perform classification task effectively on new data beyond training and validation data.

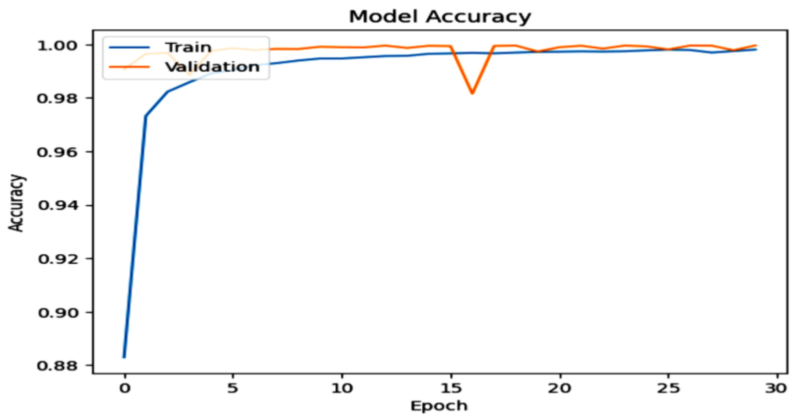

Figure 3. Training and validation accuracy on the ArASL2018 dataset.

In figure 4 the loss graph shows the same pattern with the training and the validation loss are quite close to each other, which means the model is effectively learning from the training data without excessively overfitting or underfitting.

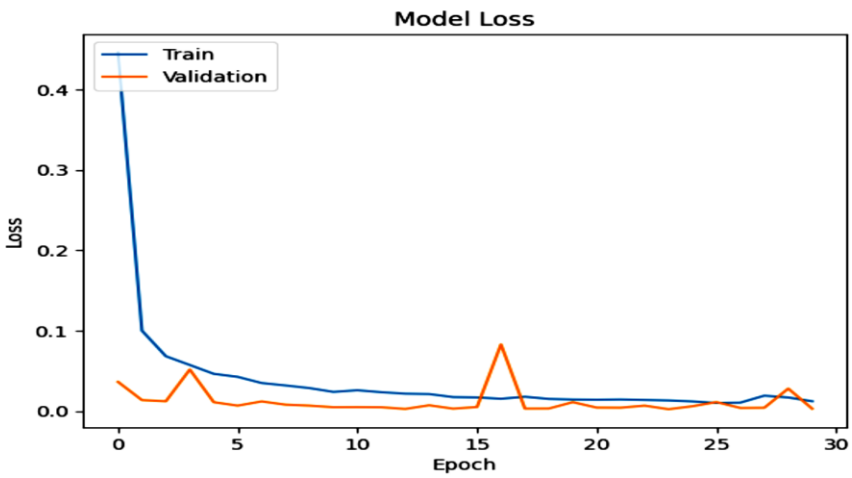

Figure 4. CNN model training and validation loss.

https://doi.org/10.71058/jodac.v9i11001

Table 2. Classification report on the ArASL2018 dataset

Class	Precision	Recall	F1-Score	Support
dal	1.0	1.0	1.0	2130
yaa	0.99	0.98	0.99	142
la	1.0	1.0	1.0	3679
ha	1.0	1.0	1.0	1367
ain	1.0	1.0	1.0	1130
aleff	0.99	1.0	1.0	165
khaa	1.0	1.0	1.0	1183
ta	1.0	1.0	1.0	2934
dha	1.0	1.0	1.0	3361
noun	1.0	0.96	0.98	132
fa	1.0	1.0	1.0	2345
dhad	1.0	0.97	0.98	191
taa	1.0	1.0	1.0	2969
meem	1.0	1.0	1.0	2781
thaa	1.0	1.0	1.0	2842
ba	0.98	0.99	0.98	127
sheen	1.0	1.0	1.0	194
gaaf	1.0	1.0	1.0	1653
ism	1.0	1.0	1.0	1858
jeem	1.0	1.0	1.0	4000
ya	0.98	0.99	0.98	146
waw	1.0	1.0	1.0	1385
saad	1.0	1.0	1.0	4177
haa	0.97	0.97	0.97	1301
seen	1.0	1.0	1.0	185
thad	0.99	0.99	0.99	1664
ra	1.0	1.0	1.0	149
ghain	1.0	1.0	1.0	2829
zay	1.0	1.0	1.0	5266
ana	1.0	1.0	1.0	1669
al	1.0	1.0	1.0	172
anta	1.0	1.0	1.0	1487
mahoua	1.0	1.0	1.0	1299
toot	0.98	0.98	0.98	122
ism	1.0	1.0	1.0	1036
asalmalykom	1.0	1.0	1.0	2289
alhamdoulilah	0.98	0.98	0.99	2863

Overall, the model achieves high precision, recall, and F1-scores for most classes, with many of them reaching a perfect score of 1.00. This indicates that the model is able to accurately predict the majority of instances for each class. The classes "dal," "la," "ha," "ain," "asalmalykom," "al," "khaa," "ta," "dha," "fa," "taa," "meem," "mahoua," "thaa," "gaaf," "ism," "jeem," "anta," "waw," "kayfa al hal," "saad," "haa," "thal," "ana," "ghain," "zay," "ism," "alhamdoulilah," and "kaaf" all have perfect precision, recall, and F1-scores of 1.00, indicating exceptional performance for these classes.

The classes "yaa," "nun," "dhad," "seen," and "ra" have slightly lower scores, but still achieve high performance overall with precision, recall, and F1-scores above 0.97. The overall accuracy of the model reached 1.00 - 100%, indicating that the model accurately

https://doi.org/10.71058/jodac.v9i11001

predicts the sign language gestures with a high degree of precision, which is not strange to classification tasks if the data is sufficient and pretty easy to differentiate between its classes.

4.2. Discussion

Our model's comprehensive evaluation highlights its outstanding accuracy and effectiveness in classifying a diverse range of Arabic sign language gestures. These findings underscore the model's potential to support effective communication for individuals relying on sign language across various applications and contexts. However, several challenges and limitations arose during the development process. The first major hurdle was the absence of a dataset specifically for Arabic words, which compelled us to create one from scratch. Additionally, we were unable to fully implement the desired preprocessing techniques on the Arabic alphabet dataset. Instead, we manually selected successfully preprocessed images, which significantly reduced the dataset size by more than half.

Another prominent challenge was the effect of lighting conditions on recognition accuracy. Despite implementing measures to enhance the model's robustness in low-light scenarios, poor lighting occasionally resulted in degraded image quality, leading to misclassifications. Furthermore, some signs that share visual similarities posed additional difficulties, particularly under suboptimal lighting conditions. These factors occasionally caused the model to struggle in distinguishing between certain gestures, thereby affecting its overall accuracy.

5. CONCLUSION

In this study, we investigated Arabic sign language recognition using deep learning techniques. Our primary goal was to develop a high-performing model capable of translating Arabic sign language gestures into coherent Arabic words and phrases. Our fine-tuned convolutional neural network (CNN) achieved remarkable results, attaining an accuracy of 99.98%. This success can be attributed to advanced preprocessing techniques that focus on the hands, removing extraneous and misleading features. Specifically, we utilized a hand-tracking approach to isolate the hands and eliminate irrelevant background information. Additionally, we incorporated a method to simulate the hand's skeletal structure, emphasizing gesture shapes over physical hand variations, such as size or proportion. The skeletal structure library we employed also demonstrated consistent performance in low-light conditions, further enhancing the model's robustness.
For future work, we plan to extend our study in two key directions. First, we aim to integrate a natural language processing (NLP) component that combines classified gestures into complete, understandable phrases, enabling the translation of full conversations conducted in sign language. Second, we intend to expand the model's vocabulary to cover a broader range of words and topics, enhancing its usability in diverse contexts and applications.

6. Conflict of Interest

The author declare that they have no conflict of interest.

7. References

[1]. World Health Organization (2024). www.who.int/newsroom/fact

[2]. Abdul Ameer, R. S., Ahmed, M. A., Al-Qaysi, Z. T., Salih, M. M., & Shuwandy, M. L. (2024). Empowering Communication: A Deep Learning Framework for Arabic Sign Language Recognition with an Attention Mechanism. *Computers*, *13*(6), 153. https://doi.org/10.3390/computers13060153

[3]. Abolmaesumi, P. (2020). Handbook of Medical Image Computing and Computer Assisted Intervention (pp. xvii–xxvi). Academic Press. https://doi.org/10.1016/B978-0-12-816176-0.00005-3

https://doi.org/10.71058/jodac.v9i11001

[4]. Adeyanju, I. A., Bello, O. O., & Adegboye, M. A. (2021). Machine learning methods for sign language recognition: A critical review and analysis. Intelligent Systems with Applications, 12, 200056. https://doi.org/10.1016/j.iswa.2021.200056

[5]. African sign languages resource center (2024) . africansignlanguagesresourcecenter.com/algeria/

[6]. Akdag, A., and Baykan, O. (2024). Multi-Stream Isolated Sign Language Recognition Based on Finger Features Derived from Pose https://doi.org/10.3390/electronics13081591

[7]. Albelwi, S., & Mahmood, A. (2017). A framework for designing the architectures of deep convolutional neural networks. Entropy, 19(6), Article 6. https://doi.org/10.3390/e19060242

[8]. Aldhahri, E., Aljuhani, R., Alfaidi, A. *et al,* (2023). Arabic Sign Language Recognition Using Convolutional Neural Network and MobileNet. *Arab J Sci Eng* 48, https://doi.org/10.1007/s13369-022-07144-2

[9]. Almahri, H. G. A. (2022). Arabic sign language recognition: A deep learning approach. https://bspace.buid.ac.ae/handle/1234/2043

[10]. Balas, V. E., Shaw, R. N., Ghosh, A. (2021). List of contributors. In Artificial Intelligence for Future Generation Robotics (pp. xi–xii). Elsevier. https://doi.org/10.1016/B978-0-323-85498-6.00018-6

[11]. Boulesnane, A., Bellil, L. & Ghiri, M.G, 2024). A hybrid CNN-random forest model with landmark angles for real-time Arabic sign language recognition. *Neural Comput & Applic*. https://doi.org/10.1007/s00521-024-10729-7

[12]. Center for Strategic and International Studies (2024). https://www.csis.org/

[13]. Kamruzzaman, M. M. (2020). Arabic sign language recognition and generating Arabic speech using convolutional neural network. Wireless Communications and Mobile Computing. https://doi.org/10.1155/2020/3685614.

[14]. Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11).https://doi.org/10.1109/5.726791

[15]. Meena, S. (2011). A study on hand gesture recognition technique. http://ethesis.nitrkl.ac.in/2887

[16]. Noor, T. H., Noor, A., Alharbi, A. F., Faisal, A., Alrashidi, R., Alsaedi, A. S., Alharbi, G., Alsanoosy, T., & Alsaeedi, A. (2024). Real-Time Arabic Sign Language Recognition Using a Hybrid Deep Learning Model. *Sensors*, *24*(11), 3683. https://doi.org/10.3390/s24113683

[17]. Podder, K. K., Ezeddin, M., Chowdhury, M. E. H., Sumon, M. S. I., Tahir, A. M., Ayari, M. A., Dutta, P., Khandakar, A., Mahbub, Z. B., & Kadir, M. A. (2023). Signer-Independent Arabic Sign Language Recognition System Using Deep Learning Model. Sensors. https://doi.org/10.3390/s23167156

[18]. Rwelli, R. E., Shahin, O. R., & Taloba, A. I. (2022). Gesture-based Arabic sign language recognition for impaired people based on convolutional neural network. arXiv. https://doi.org/10.48550/arXiv.2203.05602

[19]. Saleem, M. M. (2020). Deep learning application on American sign language database for video-based gesture recognition.

[20]. Sultan, A., Makram, W., Kayed, M., & Ali, A. A. (2022). Sign language identification and recognition: A comparative study. Open Computer Science, 12(1), 191–210. https://doi.org/10.1515/comp-2022-0240

[21]. Vazquez Lopez, I. (2017). Hand gesture recognition for sign language transcription.. https://doi.org/10.18122/B2B136