

# Deep Learning Methods for Arabic Sign Language Recognition and Translation

Zaid Saad Bilal<sup>1,\*</sup>, Amir Gargouri<sup>2</sup>, Hanaa F Mahmood<sup>3</sup> and Hassene Mnif<sup>1</sup>

<sup>1</sup>National School of Electronics and Telecommunications, LETI Laboratory-ENIS, University of Sfax, Tunisia

<sup>2</sup>Laboratory of Signals systems aRtificial Intelligence and neTworkS "SM@RTS", University of Sfax, Tunisia

<sup>3</sup>College of Education for Pure Science, Department of Computer Science, University of Mosul, Iraq

## Abstract

The methods of machine learning and deep learning neural networks are in use in big data analysis to improve the training of the algorithm for the necessary appreciation of human natural language. Such algorithms are capable of analyzing the data characteristics and determining the relevant patterns, new information, and different aspects of modern society. Complex and deep learning algorithms are highly beneficial solutions in translation, voice search, and customer feedback analysis. The concentration of this research lies on realizing the utilization of deep learning algorithms in helping the deaf and hard of hearing especially in translation of what is being said in sign language. Thus, sign languages play an essential role in the exchange of information and interaction between the deaf and the hard of hearing. While translators and sign language interpreters mitigate the issue of communication barriers, they face some limitations – for example, translators are not always available, and the number of sign languages continues to expand. In this area, there is a need to enhance the current traditional practices and make the systems developed more accessible through older technologies. In line with this, this research includes a comparative analysis of deep learning algorithms for Arabic sign language recognition and translation thereby assessing the suitability of the technology in serving the needs of the deaf and hard of hearing people in real-time interaction. The evidence supports the hypothesis on the effectiveness of deep learning techniques compared to other AI methods in translating Arabic sign language.

## Keywords

Deep learning, Machine learning, Neural networks, Sign language recognition Arabic sign language Deaf and hard of hearing

## 1. Introduction

Today the world wants to implement deep learning algorithms for understanding, translation, voice search, or to serve the public and know their feelings and perceptions about the services offered by corporations. Thus, it grew necessary to use deep learning to extend humanitarian service to a section of the society that needs a hand to help take its hand so as to improve on its ability to communicate and communication with people. This community will benefit from the above highlighted capabilities and Vonore State help to bring this community out of the isolation it currently experiences.

Self organisation and communication assurance is among the most basic and vital functions in people's existence. It is a simple yet functional means of communication and expressing oneself as well as getting informed on the events occurring in individual areas and throughout the world. However, it is important to note that there is a significant portion of the world's population who are unable to do so [1], for instance, due to hearing impaired conditions. Instead, weak speech, or both Weak speech means that a party cannot effectively communicate or argue its case; or they have valid but feeble grounds to support their case. Hearing loss can be defined as being the partial or complete loss of hearing in one

---

WAISS'2024: 1st Euro-Mediterranean Workshop on Artificial Intelligence and Smart Systems, October 15, 2024, Djerba, Tunisia (Co-located with the 17th International Conference on Verification and Evaluation of Computer and Communication Systems (VECoS'2024), October 15-18, 2024, Djerba, Tunisia)

\*Corresponding author

✉ zaidsaadd1988@gmail.com (Z. S. Bilal); amir.gargouri@enetcom.usf.tn (A. Gargouri); dr.hanah@uomosul.edu.iq (H. F. Mahmood); hassene.mnif@enetcom.usf.tn (H. Mnif)

🌐 <https://orcid.org/0000-0002-5964-0382> (Z. S. Bilal)

🆔 0009-0002-0602-6220 (Z. S. Bilal); 0009-0003-0377-4511 (A. Gargouri); 0000-0001-5322-441X (H. F. Mahmood); 0000-0002-5912-752X (H. Mnif)



© 2025 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

ear, or both ears. However, mutism is a disability that affects the ability to speak and this can make some affected individuals unable to talk if for example deafness and mutism happens at childhood since it leads to language impairment which is also known as 'mutism'. These diseases are among the most prevalent disability known in the world today (Hassan and colleagues 2020). Firstly, the statistical data on children with physical disabilities for the past decade provides an indication of an emergence of defective hearing loss among the newborns, which distorts the communication process between the infants and rest of the society [2].

According to WHO report, the population of people with hearing impairment in year 2005 has approximate to 278 million[3]. Within this period, this figure has risen to 360 million, a figure that has been arrived at by an appreciation of approximately 14 % [3]. Therefore, the number has since been on the rise. The current WHO report shows that 466 million people had hearing loss in 2019, 5% of the world's population, 432 million (5%) being adults, and 34 million 5% being children. Children [4]. WHO also predicted that by 2050 it will almost double ( reaching 900 million individuals [5]. This population is on the rise and growing very fast. Thus, there exist a communication gap, which hinders the lives of the deaf and mute people and their interactions with their fellow beings. The deaf are a dynamic segment of society, and it is from here that there developed a need for a way of enabling the deaf and mute to engage with different segments. In the communities that they live, therefore sign language.

Most of the deaf and hard of hearing individuals use sign languages as their primary mode of communication [6]. It is a useful method of coping with the lack of communication and interpersonal interactions between them and others, and the sign language interpreters assist in eradicating the communication disparity with the hearing impaired through sign language to spoken languages and spoken languages to sign languages. However, these interpreters may not always be present to assist the deaf and hard of hearing people also there are many sign languages being used around the world. As stated by the World Federation of the Deaf, there is over 300 sign languages in over 7 million deaf communities in the world today [7]. Thus, there is a need to have a technological-based system to assist the work of the sign language interpreters, top of that, the availability of this technology to the deaf and hard of hearing people at the right time and place. In this study, the aspects that will be understood include machine learning and the strategies needed in the recognition of sign language with their models and the application of it to build a speaking model of the Arabic sign language.

## 2. Related Works

In [8] study they designed a method/ system for visual manual data collection utilizing Arabic Sign Language; and translating the data processed in this visual format into text. The data used that has been incorporated into the work encompasses 54,049 images of the Arabic sign language alphabet, whereby every given category consists of 1,500 images and the total of 25 categories or meanings that are represented by a hand gesture or sign. Several transformations and image enhancements were performed on the input images during the preprocessing stage. EfficientNetB4 is a heavy-weight architecture that takes totally more complex form but achieves 98% training accuracy and 95% testing accuracy.

In [9], the aim was to describe a good technique for the automatic recognition of Arabic sign language that employs convolutional neural networks (CNNs) for feature extraction, and the long-term memory (LSTM) for classification. CNN named Alex Net is incorporated to extract deep features from the input image while LSTM to maintain sequential contextual information of video frames. The method was used to analyse fifty iterations of one hundred and fifty tags typical of daily activities of the three signers securing the dataset. The accuracy of the overall recognition reached 95 per cent, while the proposed method significantly decreased the average number of iterations of the optimization algorithm. 9% for the signer-dependent case and 43 for the signer-independent case based on the modified Hidden Markov Model. And the signer-independent results are 86% for the non-word sequences and 62% for the more challenging case.

In the study [10] a novel system namely ArSLR is introduced for localizing and recognizing the

various Arabic sign language alphabets using a Faster Region-based CNN. Successfully, Faster R-CNN is intended for the identification and the correspondence of features in a given image and training of the position of the hand. For the purpose of experiment to verify the proposed R-CNN-based fast sign recognition system, VGG-16, and ResNet-18 models are employed, and openly available a real ArSLR image dataset is gathered. The proposed approach used in the modelling gave a accuracy level of 93 percent. In [11] study, the authors planned to review scientific papers on intelligent systems in sign language recognition discussed during the last twenty years. The research study analysed a total of 649 publications of decision support and Intelligent Systems for Sign Language Recognition (SLR) derived from Scopus database. Using triangulation, the extracted publications were analyzed using VOS Viewer plyometric software. It is believed that this work will also advance understanding and build the intelligence-driven SLR in various subject fields, and offer directions to readers, academics, and professionals[12]. study developed a novel ArSLR scheme utilizing an unsupervised DNN based approach for Arabic alphabet letters recognition and classification. It is important to note that after resizing and normalization 6000 samples of 28 Arabic alphabetic signs were used to extract the features. In the process of classification, SoftMax regression was considered and reflected an 83% average accuracy level. 32%.

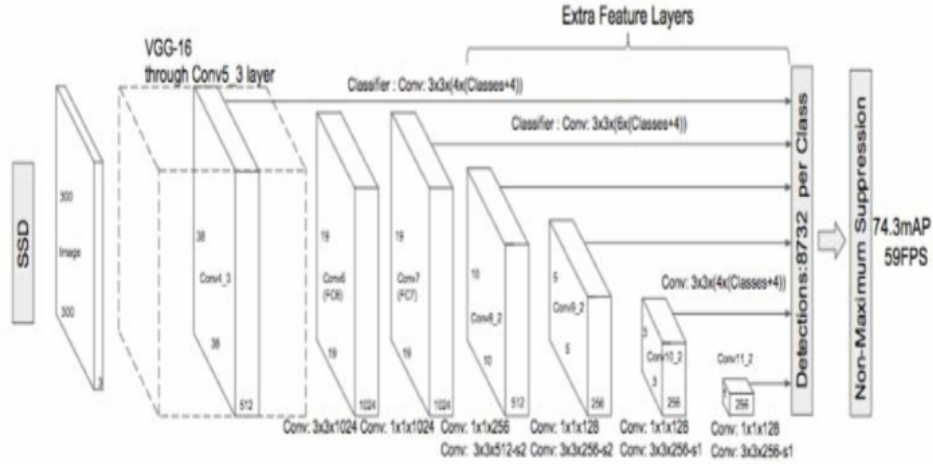
In the study, [13] described a new technique in identifying the word in the Arabic sign language using the Leap Motion device, which helps in developing three dimensional gloves models of human hand by using the infrared rays. Their approach based on identification of features that are mathematical extracted from Leap Motion controller. Such as, the results in the presented research reveal that the proposed method yielded outcomes as high as 89%, proving its efficiency. The recognition rate ranges from 65% for one-handed gestures, 96% for two-handed gestures.

When carried out the study delineated by [14], I provided a new system which does not compell the deaf individuals to put on uncomfortable devices like gloves to enable the recognition of hands. The system is founded on motion descriptors from 4D image planes, to be exact. This task is done under Scale-invariant feature transform (SIFT) technique since it can extract strong similar rotation and occlusion features. Moreover, the dimensionality of the extracted feature vectors is handled in LDA (Lineral Discriminant Analysis) technique, and by making increase in distance between the classes, the accuracy of the input system is increased. Hence, the classifiers are classified as Arabic Sign characters by using different methods namely; Support Vector Machine (SVM), k-nearest Neighbor (k-NN), and minimum distance. Several experiments were carried out to adopted in testing the efficiency of the system which indicates that the accuracy of the results obtained is in the rate of 99%. There were also experiments which showed how the proposed system is reliable even if there was any rotation and exhibited identification rate of about 99 %. Furthermore, the evaluation demonstrated that the developed system has the potential to be on the same level with other systems developed in related works.

### 3. Methodology

This section includes data collection and also some simple enhancement of the data by applying simple transformations to the image. A popular algorithm for detecting entities in images that we employed in our study is the Single Shot MultiBox Detector (hereinafter SSD). This algorithm combines several components: This algorithm combines several components:

1. VGGBase: This part generates the lower-level feature maps, and the model adopted here is the Convolutional Neural Network based on the VGG16.
2. Auxiliary convolutions: This particle carries out feature maps which VGGBase layer has performed to produce additional Feature Maps.
3. Prediction convolutions: These convolutions predict translation boxes and the classes present in the boxes for the video.
4. MultiBoxLoss: This is the loss function for training the SSD model, which will take in an input image and output the locations of the boxes that contain the objects of interest. It contains loss of location and loss of assurance.



**Figure 1:** Single Shot Multibox Detector (SSD) structure

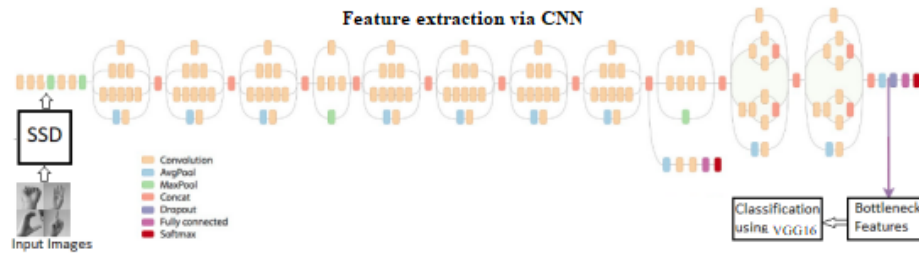
5. SSD: This is the primary MEC-SSD model that includes all the features stated earlier. The last stage takes the input image and generates the final estimate of the locations of the objects in the image and the probability of their separation.

The algorithm follows the SSD architecture, where the basic network extracts features, auxiliary convolutions generate additional feature maps, prediction convolutions predict object locations and classes, and a MultiBox loss function is used to train the model. The hand is first detected using the Single Shot Multibox Detector (SSD) [15], and then the detected hand serves as the input for the classification system. To achieve high performance and precision in solving object detection challenges, [16] uses a machine learning model called SSD. As seen in Figure I, the SSD structure in this paper is based on a VGG-16 structure. Fully connected layers are eliminated from the classification system and a series of convolutional layers are introduced for feature extraction.

In this way, the CNN progressively downsamples the input to the next layers but meanwhile extracts useful features within different sizes. In Multi Box, we explored the continuous dimension-limiting prerogatives that have already been pre-calculated in order to ascertain that they approximate the distribution of the original ground truth boxes as closely as possible. These priorities are selected so that the likelihood of ending up with a figure that is greater than 0 was higher. The data sets have been collected from the Kaggle as the dataset from different source which is available for research purpose. We have accumulated a large amount of data of the movements that are used in Arabic Signing Language. Still and motion clips of some movements are also in this collection. There is a Selected Collection in the Arabic Alphabet RGB Sign Language (AASL) comprising 7,857 raw, fully labelled RGB Sign Language images of Arabic alphabets; the only such dataset available on Kaggle as of now [17]. The proposed dataset targets to assist the people who prefer to develop models for different Arabic sign language scenarios. Surveys were completed by more than 200 participants in different settings such as Lighting, Background, orientation of the picture, size of the picture, and resolution of the picture. Potential errors and incorrect images were thus avoided after having the collected images moderated, validated, and filtered by a panel of experts in the field. Figure II shows mode we present here is based on the results collected from students in and outside class settings.

We then again applied data sanitizing preprocesses including scaling and normalization as well as data splitting into training with validation data sets. Data was preprocessed with data cleaning and augmentation to enhance the quality of the datasets as well as the size of the datasets.

There were few, yet notable changes we made to adjust the dataset obtained from the Kaggle website for the Single Shot Multibox Detector (SSD) algorithm. First, we enclosed the objects of interest with rectangles on each image of the dataset and focused on hand gestures likely to depict the Arabic numerals. Due to this, a rectangular box had to be decided on a case-by-case basis in order to wrap



**Figure 2:** The Proposal model

around the hand signs so that the SSD could localize such gestures within the images. We also worked on the truth table for all bounding boxes and annotations to the proper Arabic letter. The third point reveals that SSD has to understand the correlation between the hand sign visual patterns and the corresponding Arabic letters, and this truth table is helpful for that process. We incorporated these adjustments in a way that enhanced the suitability of the dataset to be placed under SSD for effective real-time gesture recognition and Arabic sign language interpretation.

The selected dataset presents a challenge since it lacks an annotation file, which is necessary for the creation of bounding box labels for object detection. However the location and label of the hand are the only information the SSD network outputs; additionally, training data with two outputs (labels and annotations) needs to be generated. The researcher proposed the following steps to add annotation to the dataset to fix this issue: all photos in the dataset are shrunk to 200x200 pixels; additionally, to add annotation to the dataset, bounding box information must be added. This suggests that in addition to labeling the objects or areas of interest in our pictures, we also need to note the bounding box coordinates each item corresponds to. There should be a matching bounding box for each image in the dataset. We must create code to load the bounding box data for every image in our training set. We labeled the bounding boxes of items in images using the Labeling annotation tool to add an annotation. We Check and Examine Annotations. To make sure the annotations are accurate, it is essential to review them. We verify that the class names are correct and that the bounding boxes around the objects of interest are placed correctly.

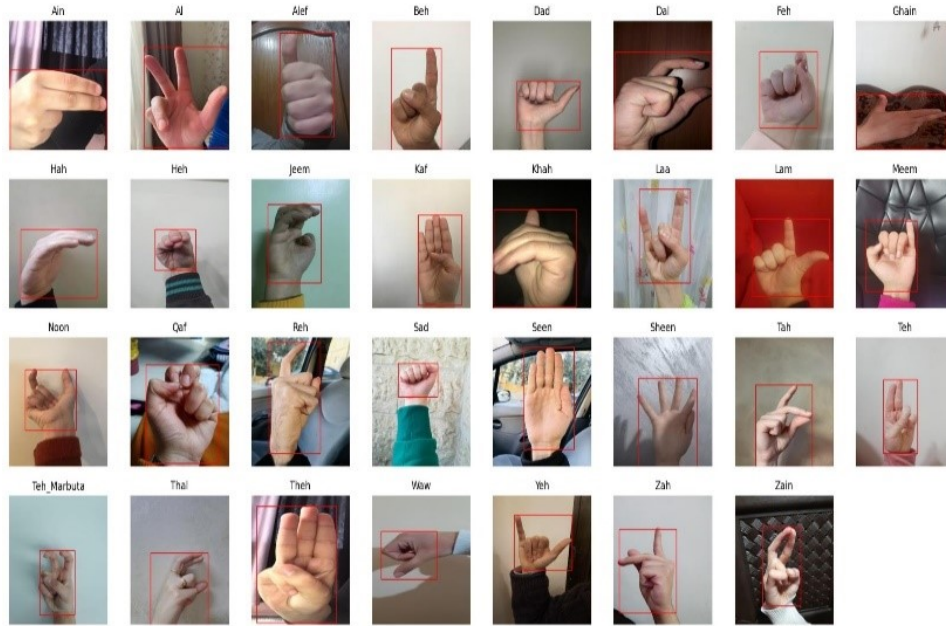
Figure 3 shows the sign language for each letter of the Arabic language. We have labeled the Arabic letters with numbers so that the algorithms can recognize them in training and testing.

## 4. Results and Discussions

The following section shows the result of the proposed models in terms of regular DL metrics so that the overall performance of our model can be compared to other models. To achieve better results, we developed other codes whose framework uses different deep learning models that other researchers have deemed accurate in the classification of images into multiple classes. The test record count of different classes which have been accurately predicted by the proposed classification model as well as the count of instances where misclassification has occurred are employed for evaluating the model. Hence, from the confusion matrix a clearer picture of a predictive model's performance especially in regard to which classes are being forecasted rather accurately or inaccurately, the kinds of errors that are being made and so on can easily be determined. The performance output gauge is evaluated by Accuracy, Recall, Precision, and F1-Score as follows: Accuracy is the ratio of True Positives to all the predictions. The recall is the ratio of True Positives to all the positives in Dataset. The Precision is the ratio of True Positives to all the positives predicted by the model. F1-Score provides a single score that balances both the concerns of precision and recall in one number. A good F1 score means that you have low false positives and low false negatives.

In Figure 4 we show the Confusion Matrix for training, and in Figure V we show the Confusion Matrix for testing, and in Figure III we show an illustration of the accuracy and loss levels during





**Figure 3:** The sign language for each letter of the Arabic language

training on the database. During training, our proposed model achieved impressive scores of 98%, 98%, 98%, and 98% in the parameters of accuracy, precision, recall, and F1 score, respectively. Interestingly, during testing, our proposed model achieved scores of 94%, 95%, 94%, and 94% for the same criteria. The confusion matrix displayed shows the performance of the model on the training dataset. It compares the true labels (actual classes) on the vertical axis with the predicted labels (model's output) on the horizontal axis. Each number within the matrix represents how many instances were classified into a particular true versus predicted label combination.

The diagonal values of the matrix indicate the number of correct predictions made by the model for each class. For instance, a value of 236 in the row for the "Alef" class and the column for the "Alef" class shows that the model correctly predicted 236 instances of the "Alef" class. These diagonal elements reflect the model's success in accurately classifying instances of each class.

The off-diagonal values represent misclassifications. For example, if the value at row "Alef" and column "Dad" is 3, it means the model mistakenly predicted 3 instances of "Alef" as "Dad". Such misclassifications can help in understanding where the model is confused and where it needs improvement.

Overall, the matrix shows that the model performs reasonably well with many high diagonal values, suggesting that the classifier is accurately predicting most of the classes. However, examining the misclassifications (off-diagonal elements) provides a deeper understanding of where the model struggles, indicating areas for improvement or potential issues in class overlap. The performance metrics derived from this matrix, such as accuracy, precision, recall, and F1-score, can help quantify the model's effectiveness and guide future optimizations.

This confusion matrix represents the performance of the classifier on the test dataset, similar to the previous one. It compares the true labels (actual classes) on the vertical axis with the predicted labels (model's output) on the horizontal axis, showing the model's ability to classify the data correctly.

Like in the training confusion matrix, the **\*\*diagonal elements\*\*** indicate the number of correct predictions made by the model for each class. For example, the value 55 in the row for "Ain" and the column for "Ain" shows that 55 instances of the "Ain" class were correctly identified. This is similar to the diagonal behavior observed in the training dataset.

The **\*\*off-diagonal values\*\*** indicate misclassifications. For example, the value 3 in the row for "Hah" and column for "Teh Marbuta" means that 3 instances of "Hah" were incorrectly predicted as "Teh Marbuta." These off-diagonal elements are crucial for diagnosing where the model is confusing classes.

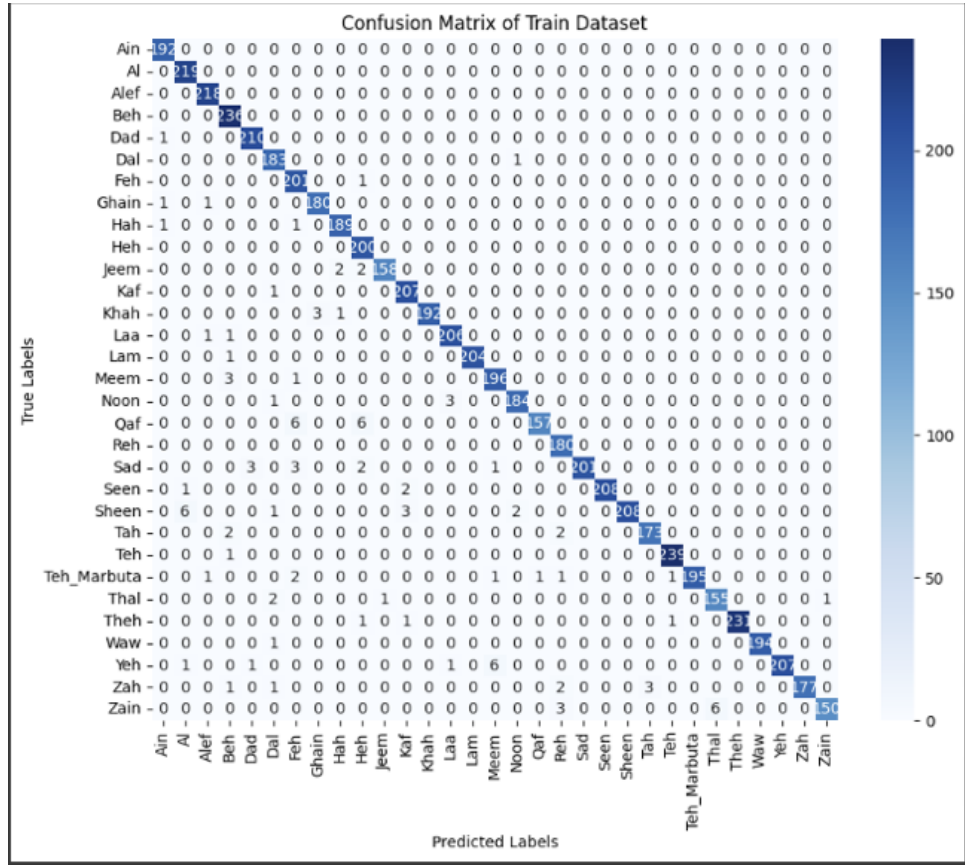


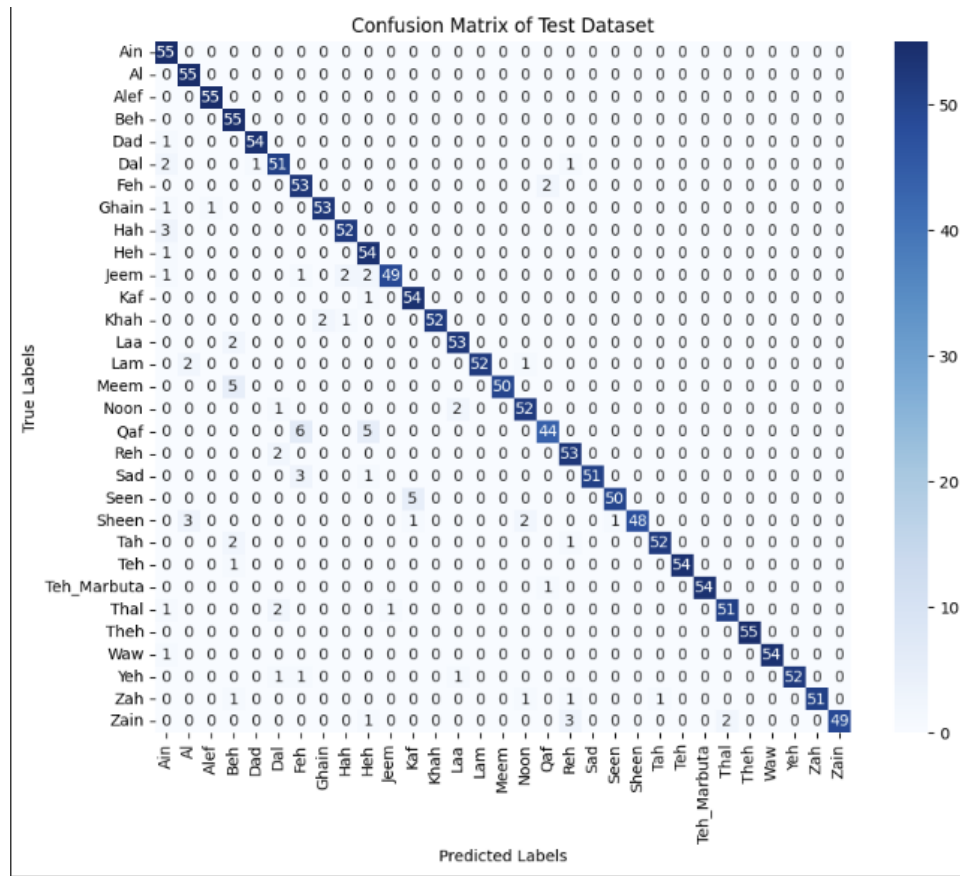
Figure 4: Confusion Matrix of Training

Overall, the matrix shows that the classifier performs similarly on the test dataset as it did on the training dataset. The diagonal values are predominantly high, indicating that the model is correctly predicting the majority of the classes. However, the off-diagonal values highlight instances where misclassifications occur. A detailed analysis of these misclassifications can provide insights into the model's weaknesses and areas for improvement. This confusion matrix can also be used to calculate evaluation metrics such as accuracy, precision, recall, and F1-score, which are vital for understanding the classifier's overall performance.

In the Figure 7, we show the test images of the trained model in one image selected from each label and show the image with its real box and the image with the box predicted by the trained model.

As a result, the study's thorough assessment of the proposal model under various training circumstances demonstrates its resilience and adaptability. These findings emphasize the significance of taking into account model performance under various training setups and offer useful insights for practitioners and researchers working on related deep-learning problems. Table 1 presents the outcomes of the competitive models' performances in translating text into sign language. The studies that provided precise results have been taken into consideration. The authors' choice of datasets was used to run the simulations. As demonstrated, certain research employ datasets in Arabic sign language, American Sign Language, others use datasets in Indian sign language, and still others use datasets in Indonesian sign language.

In Figure 4 we show the Confusion Matrix for training, and in Figure V we show the Confusion Matrix for testing, and in Figure III we show an illustration of the accuracy and loss levels during training on the database. During training, our proposed model achieved impressive scores of 98%, , and 98% in the parameters of accuracy, precision, recall, and F1 score, respectively. Interestingly, during testing, our proposed model achieved scores of 94%, 95%, 94 and 94% for the same criteria. The confusion matrix displayed shows the performance of the model on the training dataset. It compares the true labels



**Figure 5: Confusion Matrix of Testing**

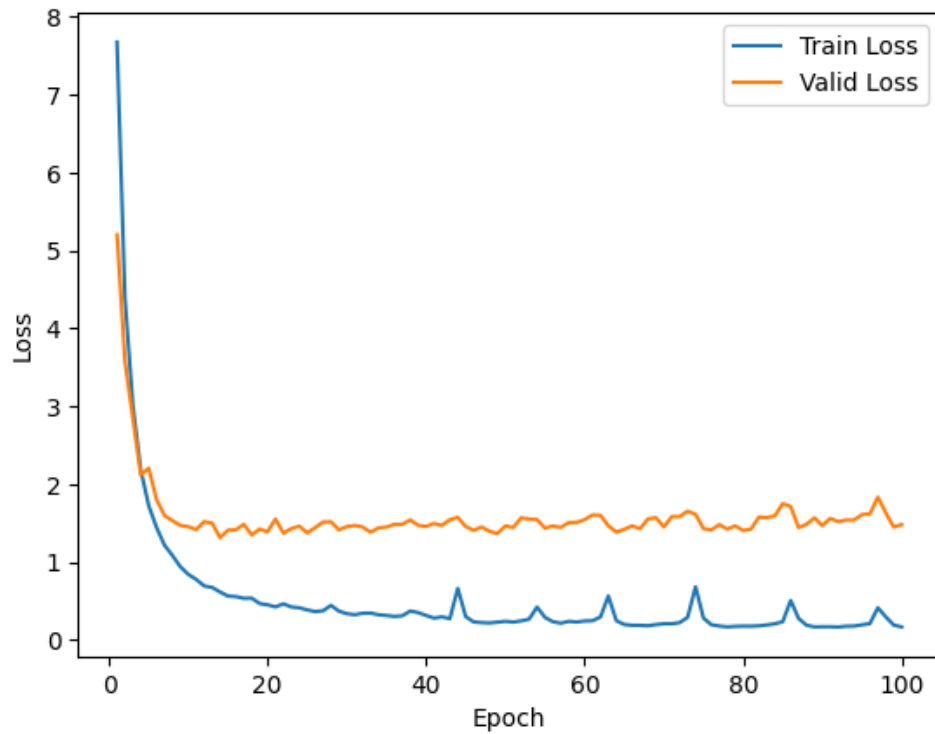
(actual classes) on the vertical axis with the predicted labels (model's output) on the horizontal axis. Each number within the matrix represents how many instances were classified into a particular true versus predicted label combination.

This confusion matrix in Figure (5) represents the performance of the classifier on the test dataset, similar to the previous one. It compares the true labels (actual classes) on the vertical axis with the predicted labels (model's output) on the horizontal axis, showing the model's ability to classify the data correctly. Like in the training confusion matrix, the diagonal elements indicate the number of correct predictions made by the model for each class. For example, the value 55 in the row for "Ain" and the column for "Ain" shows that 55 instances of the "Ain" class were correctly identified. This is similar to the diagonal behavior observed in the training dataset.

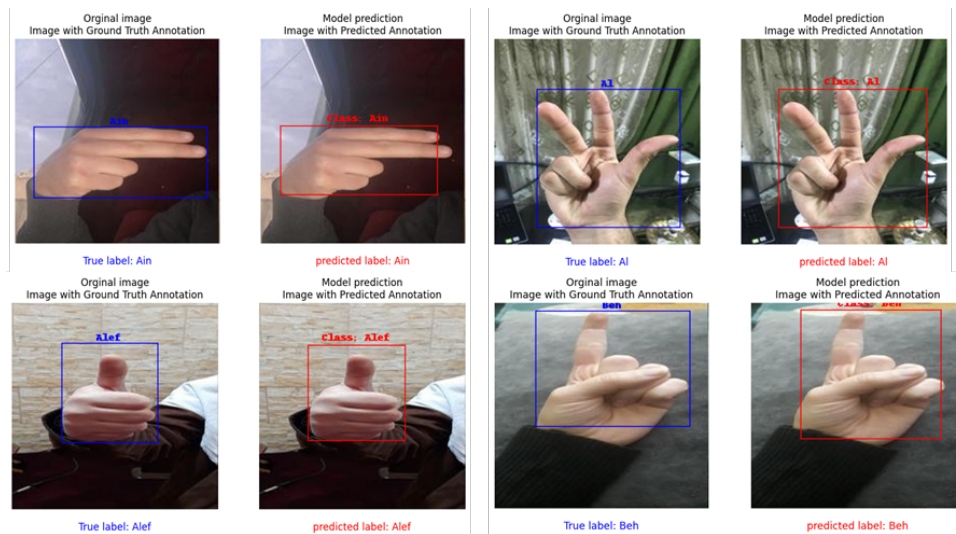
Figure 6 shows the loss curves for both the training and validation datasets over the course of 100 epochs. The blue line represents the training loss, which initially drops sharply before stabilizing at a lower value as the model improves. The orange line represents the validation loss, which starts higher than the training loss and decreases gradually, though it remains consistently higher throughout. The gap between the two lines suggests that while the model is improving on the training data, it is not generalizing as well to the validation data, a potential sign of overfitting. The oscillations observed in both loss curves after initial stabilization indicate that the model's performance on the training and validation sets reaches a plateau after a certain point, implying that further training may yield diminishing returns without significant improvements.

The study presented in the paper effectively uses deep learning methods (SSD, VGG16, and CNN) for Arabic Sign Language (ArSL) recognition and achieves high accuracy (94-98%); however, it has several significant limitations that can be further improved to increase its impact on the field. There is one major drawback in the current programs, and it is the emphasis on the static recognition of the alphabet, which does not capture the dynamic nature of sign language. It is important to understand



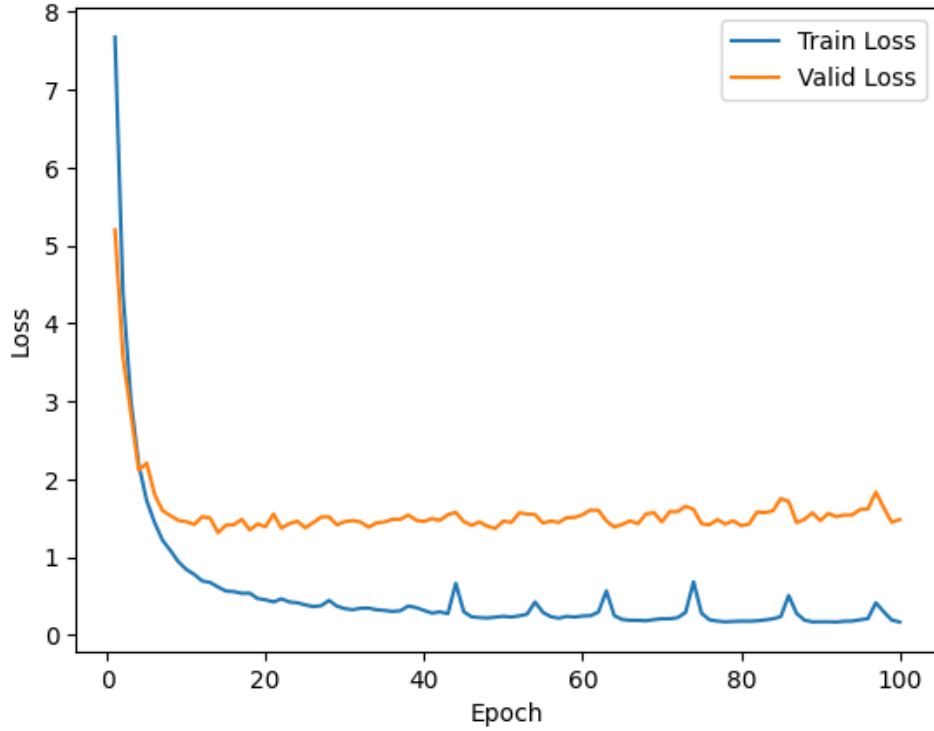


**Figure 6:** Accuracy And Loss Levels During Training



**Figure 7:** The test images of the trained model in one image selected from each label and show the image with its real box and the image with the box predicted

that sign languages are not just a set of fixed positions of the hands but rather dynamic movements, facial expressions, and other contextual features. The use of temporal models like LSTM networks or RNNs could help enhance real-time gesture translation since sign language is temporal and sequential in nature. the dataset used in the study is small and only focused on the Arabic alphabet, which makes it difficult to apply the model to more complex sign languages that include sentences, phrases, or context information. Incorporating more signs and real life sign language applications would improve the model and its versatility in different situations. The last issue that can be mentioned as a weakness of the study is the use of manual annotations, which can be subjective and contain mistakes. The training data could be improved by employing automated or semi-automated methods of annotation,



**Figure 8:** Confusion Matrix of Testing

and possibly crowd-sourced validation. the absence of performance statistics and comparisons with other models in real-time also poses a drawback to the study. Measures such as inference time, latency, and computational complexity would give information on the model's suitability for use in practical applications. Moreover, a more detailed comparison of the proposed model with other approaches would provide a better understanding of the advantages of the proposed model and its weaknesses. Describing why the specific model architectures and techniques were chosen and discussing their weaknesses would enhance the paper's originality and significance.

In the Figure 7, we show the test images of the trained model in one image selected from each label and show the image with its real box and the image with the box predicted by the trained model. As a result, the study's thorough assessment of the proposal model under various training circumstances demonstrates its resilience and adaptability. These findings emphasize the significance of taking into account model performance under various training setups and offer useful insights for practitioners and researchers working on related deep-learning problems.

Table 1 presents the outcomes of the competitive models' performances in translating text into sign language. The studies that provided precise results have been taken into consideration. The authors' choice of datasets was used to run the simulations. As demonstrated, certain research employ datasets in Arabic sign language, American Sign Language, others use datasets in Indian sign language, and still others use datasets in Indonesian sign language.

## 5. Conclusion

In conclusion, this study emphasizes the importance of deep learning neural networks in addressing the communication barriers faced by the deaf and hard-of-hearing community. Our study focuses on leveraging deep learning algorithms to facilitate sign language interpretation, to enhance accessibility and communication for individuals who rely on sign languages. In our study, we used the integration of SSD, Vgg16, and CNN, and an Arabic Sign Language Translator is implemented on The Arabic Alphabet RGB Sign Language (AASL) from Kaggle. During training, our proposed model achieved

**Table 1**

The outcomes of the competitive models' performances in translating text into sign language.

Study	Method	Dataset	Accuracy %
[18]	VGGNet	The ArSL2018	97%
[19]	CNN	NZ ASL	82%
96%		Neural Network	ASL alphabet
[20]	CNN, BiLSTM	SSL	94.79%
[21]	CNN	Kaggle ASL	92%
[22]	Neural Network	DSL10	92%

impressive scores of 98%, 98%, 98%, and 98% in the parameters of accuracy, precision, recall, and F1 score, respectively. Interestingly, during testing, our proposed model achieved scores of 94%, 95%, 94%, and 94% for the same criteria. The results of the study demonstrate the effectiveness of deep learning algorithms, especially in the context of recognizing and translating Arabic sign language. Despite the slight variation in performance between the training and testing phases, the proposed model achieved commendable scores across key evaluation metrics, demonstrating its ability to facilitate seamless communication for deaf and hard-of-hearing communities.

## Declaration on Generative AI

The authors have not employed any Generative AI tools.

## References

- [1] A. Hassan, Arabic sign language characters recognition based on a deep learning approach and a simple linear classifier, *Jordanian Journal of Computers and Information Technology (JJCIT)* 06 (2020).
- [2] A. Alnahhas, et al., Enhancing the recognition of arabic sign language by using deep learning and leap motion controller, *International Journal of Scientific & Technology Research* 9 (2020).
- [3] M. Zakariah, et al., Sign language recognition for arabic alphabets using transfer learning technique, *Hindawi Computational Intelligence and Neuroscience* (2022).
- [4] A. Tharwat, et al., Sift-based arabic sign language recognition system, in: *Afro-European Conference for Industrial Advancement. Advances in Intelligent Systems and Computing*, volume 334, 2015.
- [5] U. Michelucc, *Applied Deep Learning: A Case-Based Approach to Understanding Deep Neural Networks*, APress, Switzerland, 2018.
- [6] R. Alawwad, et al., Arabic sign language recognition using faster r-cnn (2021).
- [7] L. Deng, D. Yu, *Deep learning: Methods and applications*, *Foundations and TrendsR in Signal Processing* (2014).
- [8] W. Suliman, et al., Arabic sign language recognition using deep machine learning, in: *4th International Symposium on Advanced Electrical and Communication Technologies (ISAECT)*, 2021.
- [9] I. Adeyanjua, et al., Machine learning methods for sign language recognition: A critical review and analysis, *Intelligent Systems with Applications* 12 (2020).
- [10] B. Abul, Survey on evolving deep learning neural network architectures, *Journal of Artificial Intelligence and Capsule Networks* 01 (2019).
- [11] K. Yasaka, et al., Deep learning with convolutional neural network in radiology, *Japanese Journal of Radiology* 36 (2021).
- [12] M. Nielsen, *Neural Networks and Deep Learning*, ACADEMIA, 2018.
- [13] J. DeMuro, What is a neural network?, *TechRadar* (2019).

- [14] I. C. Education, Neural networks, <https://www.ibm.com/ae-ar/cloud/learn/neural-networks#toc-----cFaMGv3g>, 2020.
- [15] W. Liu, et al., Ssd: Single shot multibox detector, in: Proc. of European Conference on Computer Vision, 2016, pp. 21–37.
- [16] C. Szegedy, et al., Going deeper with convolutions, in: Proc. of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 1–9.
- [17] M. Al-Barham, et al., Rgb arabic alphabets sign language dataset, arXiv (2023). URL: <https://arxiv.org/abs/2301.11932>.
- [18] R. M. Duwairi, Z. A. Halloush, Automatic recognition of arabic alphabets sign language using deep learning, International Journal of Electrical and Computer Engineering (IJECE) 12 (2022) 2996–3004. doi:10.11591/ijece.v12i3.pp2996-3004.
- [19] V. Bheda, D. Radpour, Using deep convolutional networks for gesture recognition in american sign language, CoRR abs/1710.06836 (2017).
- [20] L. A. Khuzayem, et al., Efhamni: A deep learning-based saudi sign language recognition application, MDPI Journal 24 (2024). doi:10.3390/s24103112.
- [21] J. B. Idoko, Deep Learning-Based Sign Language Translation System, Ph.D. thesis, Near East University, 2020.
- [22] A. M. Buttar, et al., Deep learning in sign language recognition: A hybrid approach for the recognition of static and dynamic signs, Mathematics 11 (2023). doi:10.3390/math11173729.