ISSN: 2182-2069 / E-ISSN: 2182-2077

Advancements in Arabic Sign Language Recognition: A Method based on Deep Learning to Improve Communication Access

Zaid Saad Bilal[1*], Amir Gargouri[2], Hanaa F Mahmood[3], and Hassene Mnif[4]

[1*]National School of Electronics and Telecommunications, University of Sfax, Tunisia LETI Laboratory-ENIS, Tunisia. zaidsaadd1988@gmail.com, https://orcid.org/0009-0002-0602-6220

[2]Laboratory of Signals systems Artificial Intelligence and networks" SM@RTS" Sfax University Sfax, Tunisia. amir.gargouri@enetcom.usf.tn, https://orcid.org/0009-0003-0377-4511

[3]College of Education for Pure Science, Department of Computer Science, University of Mosul, Iraq. dr.hanah@uomosul.edu.iq, https://orcid.org/0000-0001-5322-441X

[4]National School of Electronics and Telecommunications, University of Sfax, Tunisia LETI Laboratory-ENIS, Tunisia. hassene.mnif@enetcom.usf.tn, https://orcid.org/0000-0002-5912-752X

Received: July 12, 2024; Revised: August 19, 2024; Accepted: September 24, 2024; Published: November 30, 2024

Abstract

This research anticipates the issue of real-time sign language recognition as a means of facilitating the daily communication of the hearing-impaired persons. Improvement in the rate of real time sign language recognition will help in boosting interaction in the society and access to services for the hearing-impaired persons. This paper outlines a step-by-step guide on fine tuning an object detection model for the AASL using the SSD architecture. The AASL dataset which includes total of 7,857 samples of more than 200 patients are first preprocessed where the images are scaled, normalized, and augmented. In SSD model, the base of the network is the VGG16 network and few extra layers for feature extraction and for auxiliary and prediction of objects are added. Particularly in object detection, Intersection over Union (IoU), and mean Average Precision (mAP) are used in addition to the confusion matrix metrics comprising accuracy, precision, recall, and F1-score. The proposed model provides good recognition accuracy of 98 percent. This feature checks its capability of real-time identification of sign languages and shows 25% efficiency. The efficiency of the proposed method is explained when comparing the results of the modified VGG16-based SSD with other methodologies. As for the improvement of communication for the hearing-impaired individuals, the present work demonstrates that the deep learning methods can be significantly effective; however, suggesting that more research efforts should be directed to real-time solutions and that the datasets should be expanded.

Keywords: Arabic Sign Language, Deep Learning, CNN, Transfer Learning, Gesture Recognition, Deaf Communication, Machine Learning, Sign Language Technology.

Journal of Internet Services and Information Security (JISIS), volume: 14, number: 4 (November), pp. 278-291.
DOI: *10.58346/JISIS.2024.I4.017*

*Corresponding author: National School of Electronics and Telecommunications, University of Sfax, Tunisia LETI Laboratory-ENIS, Tunisia.

1 Introduction

More than 70 million people in the world are deaf and thus need sign language interpreted by and translated to the oral language they understand; interpreting it automatically would improve the communication between the sign language users and the non-sign language users significantly. In incorporate facial expressions, eye movements, hand gestures and lip movements to convey information this is what is referred to as sign language. Deaf or hard of hearing people need sign language as an important channel of communication in their everyday lives (Bin Munir et al., 2021).

The unified estimate from the WHO is that over 360 million of the world's population have moderate or profound hearing loss. This, while a small percentage, amounts to over 460 million people, 34 million of whom are children. Estimations show that over the next three decades, depending on the population, more than 900 million people will experience hearing loss, every year on the average, 1. 1 billion young people are at risk as they get exposed to noise and several other related factors. About 750 billion USD could be spent annually on untreated hearing loss on the international level (Shield & Atherton, 2019).

Severity of hearing loss is classified into mild, moderate, severe or profound. People with severe or profound IMPAIRMENT normally experience difficult in communication which if not well dealt with may lead to poor mental health; the individual feels lonely and dissatisfied. The deaf people mainly use a form of communication coined as sign language to portray information. These gestures are used by the deaf persons in communication while the rest of the society does not understand them hence a gap (Firszt et al., 2017; Sánchez-Ancajima et al., 2022).

The sign languages of the deaf comprise close to 200 languages in use now and every language is uniquely structured just as the spoken languages. There are several signal movements of sign language they include hand and finger actions, head movements, shoulder movements and facial movements. These symbols mean letters, words, or even feelings, while the entire combination is as the words in the phrases of the spoken languages. Hence, sign language is an independent natural language with the rules of syntax and grammar (Marshall et al., 2010).

Introducing the interpreting automatic sign language may help in enhancing communication between the Deaf people themselves as well as with the members of society in general, thus eradicating the crucial divide between the two groups of people (Papatsimouli et al., 2023).

For interaction, humans often use the ability to speak in their day-to-day life. However, the individuals whose voice is restricted or completely absent are featured as having difficulties in speaking, namely, people who are either deaf or mute. The making of sign language fulfils the need for a communication mode in those who use it. There is a combination of manual communication and non-manual ones which are used to pass meaning in sign language. Used globally, each sign language has got its own etiquette and features that are both, manual and non-manual (Abou-Abdallah & Lamyman, 2021).

Some of the sign languages recognized globally include ASL, BSL, LIBRAS, JSL, ArSL, ISL, and BdSL. Nonetheless, sign languages are unknown to people who can hear and speak sometimes though people with impaired hearing use them frequently. The written language relays information but only to an extent of achieving a form of interaction between the D&M and other people since majority of the D&M patients do not speak. This issue is aggravating in non-naive face-to-face operations where written communication is inconceivable (Fischer, 2015).

Among the population of Bangladesh, 16 million plus are either deaf or deafened or have some form of challenged hearing. They need to use sign language to portray something which most of the people with cannot comprehend. This is because effective communication between D&M persons and members of the public entails supping sign language into one language that everybody can understand (Tarafder et al., 2015).

When it comes to structures, there is a learning algorithm known as deep learning that comprises several nonlinear changes. Neural networks that are the basis of deep learning, continue to expand the capabilities of sound and image processing, including FaceID, computer vision, speech recognition, natural language processing, text classification (spam filtering), and a number of others, including medicine and genomics (Kumar et al., 2023). Deep learning's advantage is its ability to work with big data and the method of backpropagation that alters internal parameters for precise representation at each layer (Taye, 2023).

The main objective is to design sign language recognition that enables converting the most frequent hand gestures made by the D&M individuals into the textual data for effective communication between them and members of the hearing world.

2 Related Works

Technological changes have gone a long way in solving some of the communication barriers that affect the deaf-mute persons. This paper aims at designing an RSS system to recognize the Arabic letters to help the hearing-impaired person transfer messages. By applying a new technique of image pre-processing, the system effectively segments the hand positions from the background, while using a specifically developed deep convolutional neural network (CNN) for the depth data to be recognized. The model, as we seen, has an accuracy of 97.07% again indicates improved performance in the identification of static signs than other studies that employed the same database hence making it easy for the hearing-impaired and hearing people to easily interact (Hdioud & Tirari, 2023).

In (Hayani et al., 2019) the proposed system of automatic recognition of Arabic Sign Language has great social and humanitarian importance in the rehabilitation of the deaf-mute people. As we adopted to the complexity and variation of ArSL, this paper proposes a new system with CNNs trained on a real database. I have set the system in such a way that it is capable of detecting numbers and letters in the ArSL language (Gümüş et al., 2022). Finally, for the confirmation of its performance, the comparative study was done with the similar systems where the effectiveness and robustness of the present system was proved to be better. According to proposed CNN-based system, the average recognition rate of the numbers was 90. 02%, while the classification based on Traditional K-nearest neighbors (KNN) gave 66%, all these various Support Vector Machine (SVM) kernels gave between 84% and 88%.

Human-Computer Interaction or (HCI) is the study of people's relationship with computers and their key interaction methodologies, and sign language recognition is one of them. This research aims at creating an automatic dynamic Arabic sign language recognition using Microsoft Kinect for the disabled persons in the society particularly the hearing-impaired. The system employs two machine learning algorithms: The two commonly used techniques are Decision Tree and Bayesian Network, the former improved by the Ada-Boosting technique. Its performance was benchmarked against two direct matching methods: In the DTW case, it refers to Dynamic Time Wrapping while in the HM case it is referred to as Hidden Markov Model. Applied to 42 Arabic gestures related to the medical field, the experimental results demonstrated high recognition rates: Hence, the percent accuracy achieved in this case is 91% for Decision Tree classifier, 92% for the Bayesian classifier and 93% after the application of Ada-Boosting (Hisham & Hamouda, 2019).

In research (Al-Dawoodi, 2015; Saleh & Issa, 2020; Vujović, 2021), transfer learning and fine-tuning of pre-trained deep convolutional neural networks (CNNs) are applied to increase the effectiveness of the recognition of 32 Arabic sign language hand gestures (Nirmala, 2023). This entails repurposing models with reference to the VGG16 and ResNet152 structures by copying the pre-trained model weights into the models' layers and appending a newly created soft-max classification layer in the final place. Even with normal 2D images of Arabic sign language data, using the refined networks, it was possible to get an accuracy of up to 99%.

In (Hasasneh, 2020) a new Arabic Sign Language Recognition (ArSLR) system is proposed that uses unsupervised deep learning technique. It uses Deep Belief Network (DBN) and works with images for identification and classification of Arabic letters. Thus, through the use of deep learning, only important features are learned for easy identification process. The authors applied a total of roughly 6,000 samples of the 28 Arabic alphabetic signs, after resizing and normalization. Regarding the accuracy of the classifier a total of 83% was achieved and a sensitivity of 70 in the softmax regression classifier. 5% which shows accuracy of the predicted model and the specificity of the model was 96%.

Research (Elsayed & Fathy, 2020) presents an Arabic sign language translation system that utilizes the cross-referenced ontology and deep learning to translate the users' signs into meaningful language. Starting with static signs such as Arabic alphabets and words, the system, to address particular issues related to sign language translation, uses ontology. A Deep CNN was proposed for training and testing on the Arabic sign language dataset and a new collected dataset for increasing the recognition rate. The particular experiment showed the classification accuracy of 98 percent and the recognition accuracy of 88 percent on the pre-made dataset. The evaluation for the collected dataset included the classification accuracy of 98%, and the semantic recognition accuracy of 94%.

Deep learning has been used in research (Rwelli et al., 2021) related to ArSL and considerable progress has been attained in this field. To the extent that this study focuses on incorporating various types of communication commonly employed by the hearing-impaired, coupled with the regional linguistic comprehensiveness of ArSL, it tackles a particular research problem. The used technique for sign recognition involves CNNs along with wearable sensors and the overall system recognizes all the 30 Arabic hand sign letters.

After wearing DG5-V hand gloves with sensors, the hand motions are recorded from which features are obtained using the deep CNN. The actualized signs are then vocalised where the patient can recognize the gesture and utter the related word. They have proved the approach to be accurate to a reasonable degree; 90% of the audience was able to identify the output; thus, the endeavor is helpful for the Arabic Hearing Impaired in the local area.

It is a form of communication that is employed mainly by the hearing impaired through use of their arms and hands. An automated sign recognition system involves identification of certain aspects and subsequently classifying the input data. New trends in computer vision have led to more development of hand sign recognition using deep neural networks. This paper offers a vision-based system that includes CNN to identify the Arabic hand sign letters and convert them into Arabic language speech.

The proposed system identifies hand signs and even speaks the results, having a 90% efficiency of recognition. These types of high accuracy make the system dependable, although they can be made even more accurate Leap Motion and Xbox Kinect. The recognized letters are read out loud thus speaking out the Arabic language being a target language (Kamruzzaman, 2020; Hasan, 2023).

3 Arabic Alphabet Sign Language (AASL)

Besides oral communication the Hearing-Impaired community in the Arab world also utilize Arabic Alphabet Sign Language (AASL) which is a sign language that uses the Arabic script. It uses a set of signs for all the twenty-eight alphabets of the Arabic thus providing a way through which hearing impaired can communicate effectively and be understood. The origin and optimization of automatic recognition systems for AASL are a vital social and humanitarian contribution due to their positive impact on society's ability to communicate. New approaches in the field of computer vision and deep learning, especially CNN have improved the quality and effectiveness of these systems. With the help of image pre-processing, understanding of depth data, and transfer learning these technologies are capable of perceiving the static position of the hand and transcribing it into text or speech. If such innovations enhance the everyday communication of the hearing-impaired people, they also help to facilitate their role and acceptance in society at large. Figure (1) shows sample of signs related to Arabic alphabet language.

Figure 1: Sample of Signs Related to Arabic Alphabet Language

4 Methodology

To create, practice and assess Arabic Alphabets Sign object detection model, Arabic Alphabet Sign Language (AASL) will be first imported as the database then pre-processing which includes resizing, normalization, and augmentation of the images. The resulting set is further divided into training and testing, with the former being at 80 percent and the latter at 20 percent. Bounding boxes of the objects along with their corresponding classes are also drawn on images for confirmation. To facilitate ease when it comes to model training and evaluation we create dictionaries that map the class labels to indices and the other way round. The trained model's efficacy is assessed by applying real-time evaluation measurements such as Intersection over Union or IoU and mean Average Precision or mAP. It involves the writing of PyTorch code for MultiBox Loss; the problems that relate to localization and confidence loss; positive matches as well as the hard negative mining. Regularization hyper parameters, number of batches, numbers of epochs, and learning rate are set initially and a fast Data Loader is designed in order to manage the data during the training process. Choosing the optimization technique, Stochastic Gradient Descent (SGD) is used, and point the hyper parameters, including learning rate, momentum, and weight decay, are adjusted. Through the use of the above-defined parameters and the optimization method, the model is constantly trained and the results periodically checked on the validation set so as to note overfitting and fine tune other hyper parameters if necessary. After the training, the model draws bounding boxes and class labels on the test set, and the accuracy of the model is measured by mAP that is obtained from the detection results. Lastly, the bounding boxes and class labels are displayed on test images to verify the model's efficacy qualitatively. Figure 2 shows the Methodology Block Diagram.

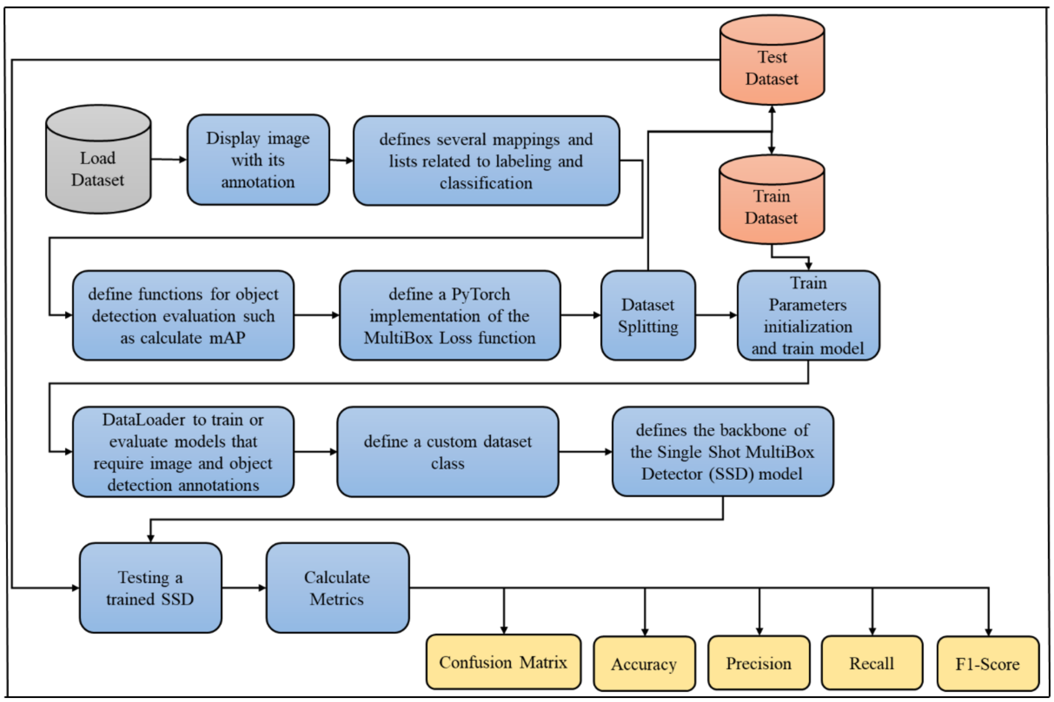

Figure 2: Methodology Block Diagram

4.1. Environment Description

The model is set up in a Google Colab environment which is an environment for logical deep learning experiments. The environment has 51 GB of system RAM to guarantee there will be enough memory to store and process large sets of data and perform intensive calculations. Furthermore, there is 15 GB of GPU memory, and this is vital especially when training the SSD model when working with large images and deep neural networks. There is enough space for datasets and other files that might be needed sometimes, as data occupies a disk space of 201.2 GB. The choice of the Google Colab environment is optimal for deep learning tasks, on the one hand, it provides a sufficient amount of computational power, on the other hand, it is very convenient in terms of developing iterations and tests.

4.2. Dataset Description

The RGB-A Issued Arabic Alphabet Sign Language (AASL) dataset is large and was collected with the help of more than two hundred participants, each of whom created one or several alphabet images.

Recorded through the webcams, digital, and phone cameras, the dataset consists of 7,857 Images with proper labeling of the Arabic sign language. These images were thoroughly supervised, validated and filtered based on the quality by a group of Arabic sign language professionals. In Table (1), each of the folders is set out with the number of images that it contains. while Figure (3) provides representative samples of images for different alphabets (Al-Barham et al., 2023).

Table 1: Dataset Distribution

No	Letter Name in English Script	Letter Name in Arabic Script	Number of Images	No	Letter Name in English Script	Letter Name in Arabic Script	Number of Images
1	ALEF	أ	287	17	ZAH	ظ	232
2	BEH	ب	307	18	AIN	ع	244
3	TEH	ت	226	19	GHAIN	غ	231
4	THEH	ث	305	20	FEH	ف	255
5	JEEM	ج	210	21	QAF	ق	219
6	HAH	ح	246	22	KAF	ك	264
7	KHAH	خ	250	23	LAM	ل	260
8	DAL	د	235	24	MEEM	م	253
9	THAL	ذ	202	25	NOON	ن	237
10	REH	ر	227	26	HEH	ه	253
11	ZAIN	ز	201	27	WAW	و	249
12	SEEN	س	266	28	YEH	ي	272
13	SHEEN	ش	278	29	TEH MARBUTA	ة	257
14	SAD	ص	270	30	AL	ال	276
15	DAD	ض	266	31	LAA	لا	268
16	TAH	ط	227				

4.3. Dataset Preprocessing

Preprocessing of data is one of the most essential steps for being prepare in a suitable format for training an object detection model in the RGB Arabic Alphabet Sign Language (AASL) Dataset.

First, the images are scaled to the certain size to avoid entering the network with images of different size, as it synthesizes the process during training. Normalization is used to scale the pixel intensities where it is benefits the gradient-based optimization algorithms. Further, the dataset is divided into training and testing data in the proportion of training data to testing data, for 8:2.

Random rotation, flip and color jittering are used to enrich the training dataset in terms of variance, thus enhancing the model's resilience. Bounding boxes and labels are suited into format of SSD, thus meaning each image's annotations will correspond to the format the SSD model expects. In addition to pre-processing the dataset for efficient training, this pre-processing pipeline improves the model's capability in handling various forms of sign language gestures. Table (2) shows Image Augmentation Parameters.

Table 2: Image Augmentation Parameters

Parameter	Value
Brightness	0.2
Contrast	0.2
Saturation	0.2
Hue	0.2

4.4. Model Description

This code snippet focuses on the construction of a Single Shot Multibox Detector (SSD) with PyTorch for real-time object detection. SSD model requires a base network from VGG16, there are extra auxiliary convolutional layers to abstract higher level features, and prediction convolutional layers to predict the location and class of the object. The model is structured as follows:

1. VGGBase Class: Describes the VGG16 network architecture from conv layers and pooling layers to FC6 and FC7 layers which are substituted by the conv6 and conv7 layers. The forward method of this class feeds the input image through these layers and outputs lower level features maps which include 'conv4_3_feats' and 'conv7_feats'. a) In this network, weights belonging to VGG16 model pre-trained on ImageNet are initialized.
2. Auxiliary Convolutions Class: Describes more convolutional layers (conv8_1 … convolution11_2) for extracting features of a higher level from the base network. The forward method takes in the output of the base network and produces features at a higher level of abstraction.
3. Prediction Convolutions Class: They have convolutional layers with the aim of predicting the bounding box coordinates (loc), and the class scores (cl). The forward method uses the feature maps coming from the base network and auxiliary convolutions to generate the locations and class scores of each prior box.
4. SSD Class: Summarizes the design pattern of the base network, auxiliary convolutions and the prediction convolutions in one model. It also contains an element that is learnable and which specifies scaling factor of the `conv4_3` feature map besides providing the default prior boxes in the SSD model. The forward method takes into account all components to perform the required task of processing the input image and coming up with the predicted locations and class scores.

Every part of the model is activated and set in a way that every path of data through the network is correct, right from the image input to the last boxes bounding and the class scores. Such a structure makes it possible for the SSD model to conduct real time object detection with efficiency and precision.

4.5. Training Parameters

The learning rate (LR), chosen as 0 0001, they regulate the step size in the optimization, which affects the balance between accuracy and training time. Implementation of BS 32 is a specification that influences memory and training characteristics whereby decreasing the batch size will provide more update frequency but might bring a slow processing speed. Momentum (0. 9) enhances gradient descent, by flattening convergence lines, and ensuring that it is not caught at local minima. Adam offers the possibilities to apply weight decay (5e-4) which punish large weights in order to minimize them and avoid overfitting. The print frequency (every epoch) contains learning information to facilitate improvements in performance on various datasets and model characteristics. These parameters are mutually adjusted to achieve the goal of model convergence as well as the rationality of training time and comparison with new data, and often combination changes depend on the particularities of the dataset and reflections on the performance of the model during training.

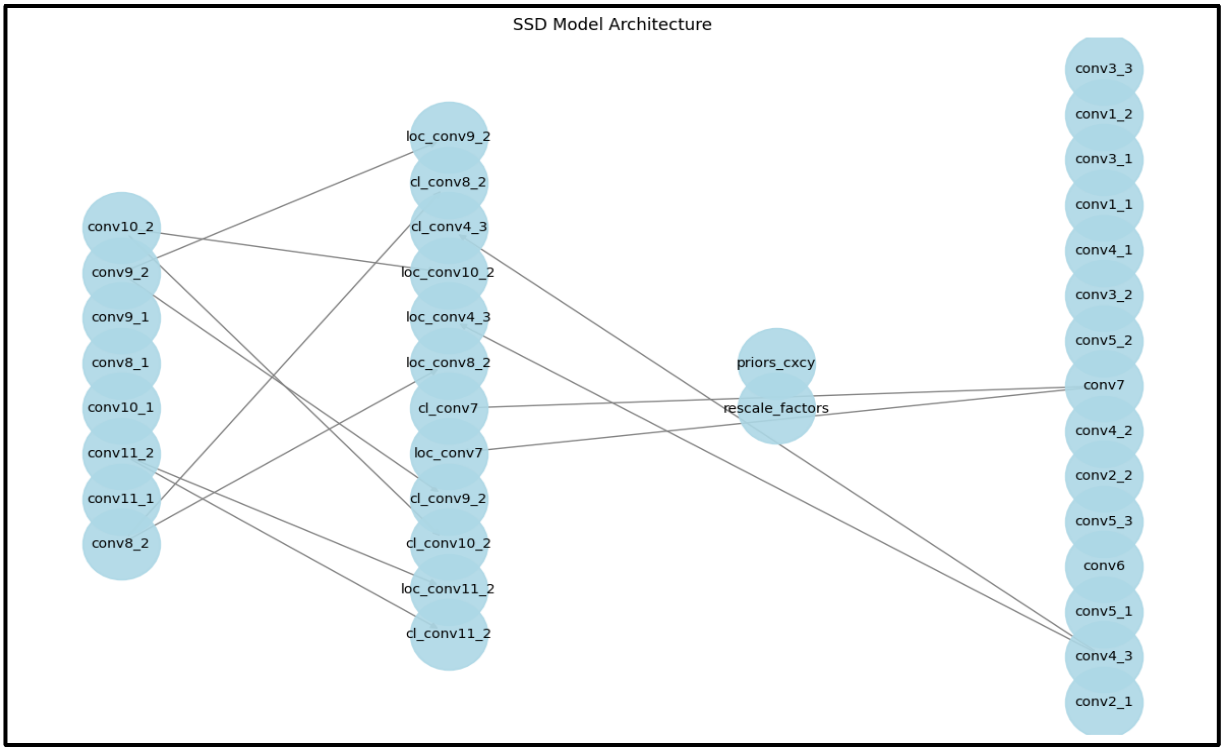

Figure 3: Proposed Model Structure

4.6. Evaluation Metrics

It is critical when training a classifier to ensure that the correct assessment scale is chosen in order to get the best emulator or classifier. The rating scale must be exercise carefully, so that the correct one is chosen and the classifier will be optimized. Additional to improving the generative classifier, this section looks into the discriminant assessment metrics for this purpose, we discuss the assessment metrics as for the generative markers of propinquity classifier, the following assessment metrics that work as discriminators are discussed in the subsequent sections. We, for instance, see that precision is applicable in many generative classifiers whereby it helps to determine the most precise solution during the training process (Hasan, 2023).

A confusion matrix (Figure 4), which has a high frequency of use, is a table that is used to assess the effectiveness of a binary classification model. It encompasses the real values of the target variable which are named as Truth (True = 1, False = 0) and the values generated by the model, named as Predictions (Positive = 1, Negative = 0). Moreover, based on the confusion matrix, it is possible to compute a number of performance indicators including accuracy, precision or recall and F1 measure.

It is typically divided into four quadrants corresponding to four possible outcomes: The four outcomes include True Positive, where the model has rightfully predicted the outcome to be a positive one when actually it is; True Negative, where the model has correctly predicted the outcome to be a negative one as is; False Positive, where the model has wrongly predicted the outcome as a positive one when in actuality it is not; and lastly False Negative where the model predicted the outcome is not positive when in real sense it is.

Class designation		Actual class	
		True (1)	False (0)
Predicted class	Positive (1)	TP	FP
	Negative (0)	FN	TN

Figure 4: Binary Classification Confusion Matrix

The values of TP, TN, FP, and FN can be used to calculate various performance metrics shown in Table (3):

Table 3: The Elements of the Evaluation Process (Variables, Definitions, and Equations)

Variable	Definition	Equation
Accuracy	it is straightforward to find the percentage of apt data from tests by dividing the total count of all accurate forecasts by total count of all predictions.	$Accuracy = \frac{Tp + TN}{TP + TN + FP + FN}$
Precision	the percentage of similar cases that remain unsolved out of all expected situations from a particular class	$Precision = \frac{TP}{TP + FP}$
Recall	the relation between the total number of occurrence and the proportion of a set that was meant to be an instance of a particular class	$Recall = \frac{TP}{TP + FN}$
F1-Score	The It is a term used in relation to the veracity of a test This is the case since every phrase is related to language and conveys information. F1-score can never be equal to 0, while the value 1 means high recall and high precision.	$F1 - Score = 2 \times \frac{percision \times recall}{Percison + recall}$

5 Results and Discussion

A confusion matrix is considered to be a key instrument that helps to define the performance of a classification model. It finds out the accuracy of the class prediction results of the model and categorizes the actual and the predicted classes. In the matrix every row defines the actual classes and each column defines the predicted class of instances. This makes it simple to understand the realism of the model in categorizing instances as belonging to the actual category or not within the various classes, and hence identify areas that require enhancement. Figure (5) shows the confusion matrix resulted in testing process.

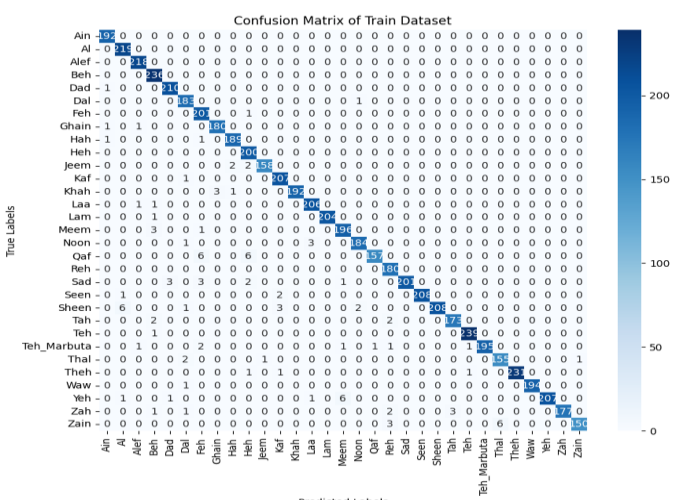

Figure 5: Confusion Matrix for Testing Model

These give the overall assessment of the model's performance on unseen data, with accuracy, precision, recall, and F1-score values of each above 98%. It is recommended that output of the results provided above indicates that the model perfectly categorize instances in the dataset. Table (4) shows the resulted metrics.

Table 4: Resulted Metrics

Metric	Value
Accuracy	0.9825
Precision	0.9828
Recall	0.9818
F1-score	0.982

In comparison, the present study modifies the VGG16 model, achieving a recognition accuracy of 98.8%. This comprehensive overview highlights the varied approaches and technologies employed across different studies, showcasing advancements and challenges in the field of sign language recognition. Table (5) shows related works comparison.

Table 5: Related Works Comparison

Reference	Techniques Used	Features	Recognition Accuracy
(Hdioud & Tirari, 2023)	Deep CNN	Image Pre-processing, Depth Data	97.07%
(Hayani et al., 2019)	CNN, KNN, SVM	Real Database	CNN: 90.02%, KNN: 66%, SVM: 84-88%
(Hisham & Hamouda, 2019)	Decision Tree, Bayesian Network, Ada-Boosting, DTW, HMM	Microsoft Kinect for Dynamic ArSL	DT: 91%, BN: 92%, Ada-Boosted: 93%
(Nirmala, 2023)	Transfer Learning, Fine-Tuning, VGG16, ResNet152	Pre-trained CNNs with Softmax Layer	Up to 99%
(Hasasneh, 2020)	Deep Belief Network (DBN)	Unsupervised Learning	83%
(Elsayed & Fathy, 2020)	Deep CNN, Ontology	Static Signs, New Dataset	Classification: 98%, Recognition: 88%
(Rwelli et al., 2021)	Deep CNN, Wearable Sensors (DG5-V hand gloves)	Hand Motion Recording, Vocalization	90%
(Gümüş et al., 2022)	CNN, Computer Vision	Hand Sign Identification, Speech Conversion	90%
Present Study	VGG16 Modified		98.8%

6 Conclusions

The comparative study included in this research highlights the efficiency of different methods in the Arabic Sign Language (ArSL) recognition. It has incorporated certain methods like image preprocessing and depth data and has even touched the efficient architectures like VGG 16 to establish its credentials and see the improved accuracy and efficiency. Particularly, the improved VGG16 model acquires a very high recognition rate of 98 percent. 8% which is higher than several other methods available in the literature of the field. This goes a long way to prove how deep learning benefits the hearing-impaired group by improving the accessibility and hence efficient communication.

For the future research perspective, it is essential to focus on the possibilities of integrating real time applications so that the communication support can happen instantaneously. Moreover, making data larger and more diverse are still going to be the crucial prerequisites for the further enhancement of the system's resistance to various adverse factors and its applicability in different contexts. If such technologies are enhanced and developed further, the impaired who rely on sign language can easily be provided with means of communication regardless of the region they belong to. Thus, the given work is not only significant for the development of new technological solutions for assistive technologies but may also encourage further improvements in social relationships for equal communication.

References

[1] Abou-Abdallah, M., & Lamyman, A. (2021). Exploring communication difficulties with deaf patients. *Clinical medicine*, *21*(4), e380-e383. https://doi.org/10.7861/clinmed.2021-0111

[2] Al-Barham, M., Alsharkawi, A., Al-Yaman, M., Al-Fetyani, M., Elnagar, A., SaAleek, A. A., & Al-Odat, M. (2023). RGB Arabic alphabets sign language dataset. https://arxiv.org/abs/2301.11932v1.

[3] Al-Dawoodi, A. G. M. (2015). An improved Bees algorithm local search mechanism for numerical dataset. *Universiti Utara Malaysia*.

[4] Bin Munir, M., Alam, F. R., Ishrak, S., Hussain, S., Shalahuddin, M., & Islam, M. N. (2021, September). A machine learning based sign language interpretation system for communication with deaf-mute people. In *Proceedings of the XXI international conference on human computer interaction* (pp. 1-9). https://doi.org/10.1145/3471391.3471422

[5] Elsayed, E. K., & Fathy, D. R. (2020). Sign language semantic translation system using ontology and deep learning. *International Journal of Advanced Computer Science and Applications*, *11*(1), 141–147. https://doi.org/10.14569/ijacsa.2020.0110118

[6] Firszt, J. B., Reeder, R. M., & Holden, L. K. (2017). Unilateral hearing loss: Understanding speech recognition and localization variability—implications for cochlear implant candidacy. *Ear and hearing*, *38*(2), 159-173. https://doi.org/10.1097/AUD.0000000000000380

[7] Fischer, S. D. (2015). Sign languages in their historical context. In *The Routledge handbook of historical linguistics* (pp. 442-465). Routledge. https://doi.org/10.4324/9781315794013

[8] Gümüş, A. E., Uyulan, Ç., & Guleken, Z. (2022). Detection of EEG Patterns for Induced Fear Emotion State via EMOTIV EEG Testbench. *Natural and Engineering Sciences, 7*(2), 148-168. https://doi.org/10.28978/nesciences.1159248

[9] Hasan, Z. (2023). Deep Learning for Super Resolution and Applications. *Journal of Galoitica: Journal of Mathematical Structures and Applications*, *8*(2), 34-42. https://doi.org/10.54216/GJMSA.080204

[10] Hasasneh, A. (2020). Arabic sign language characters recognition based on a deep learning approach and a simple linear classifier. *Jordanian Journal of Computers and Information Technology*, *6*(3), 281–290. https://doi.org/10.5455/jjcit.71-1587943974

[11] Hayani, S., Benaddy, M., El Meslouhi, O., & Kardouchi, M. (2019, July). Arab sign language recognition with convolutional neural networks. In *2019 International conference of computer science and renewable energies (ICCSRE)* (pp. 1-4). IEEE. https://doi.org/10.1109/ICCSRE.2019.8807586

[12] Hdioud, B., & Tirari, M. E. H. (2023). A deep learning based approach for recognition of Arabic sign language letters. *International Journal of Advanced Computer Science and Applications*, *14*(4), 424-429. https://doi.org/10.14569/IJACSA.2023.0140447

[13] Hisham, B., & Hamouda, A. (2019). Supervised learning classifiers for Arabic gestures

recognition using Kinect V2. *SN Applied Sciences*, *1*(7), 768. https://doi.org/10.1007/s42452-019-0771-2

[14] Kamruzzaman, M. M. (2020). Arabic sign language recognition and generating Arabic speech using convolutional neural network. *Wireless Communications and Mobile Computing*, *2020*(1), 3685614. https://doi.org/10.1155/2020/3685614

[15] Kumar, A., Joshi, P., Bala, A., Sudhakar Patil, P., Jang Bahadur Saini, D. K., & Joshi, K. (2023). Smart Transaction through an ATM Machine using Face Recognition. *Indian Journal of Information Sources and Services*, *13*(2), 7-13. https://doi.org/10.51983/ijiss-2023.13.2.3752

[16] Marshall, C. R., Kula, N. C., Botma, B., & Nasukawa, K. (2010). Sign language phonology. *The Bloomsbury Companion to Phonology*, 254-277. https://doi.org/10.4324/9781315754499-1

[17] Nirmala, M. S. (2023). Behavioural Analysis of Deaf and Mute People Using Gesture Detection. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications*, *14*(3), 125-142. https://doi.org/10.58346/JOWUA.2023.I3.010

[18] Papatsimouli, M., Sarigiannidis, P., & Fragulis, G. F. (2023). A survey of advancements in real-time sign language translators: integration with IoT technology. *Technologies*, *11*(4), 83. https://doi.org/10.3390/technologies11040083

[19] Rwelli, R. E., Shahin, O. R., & Taloba, A. I. (2021). Gesture based Arabic Sign Language Recognition for Impaired People based on Convolution Neural Network. *International Journal of Advanced Computer Science and Applications, 12*(12). https://doi.org/10.14569/ijacsa.2021.0121273

[20] Saleh, Y., & Issa, G. (2020). Arabic sign language recognition through deep neural networks fine-tuning. *International Journal of Online and Biomedical Engineering,* 16(5), 71–83. https://doi.org/10.3991/IJOE.V16I05.13087

[21] Sánchez-Ancajima, R. A., Peres, S. M., López-Céspedes, J. A., Saly-Rosas-solano, J. L., Hernández, R. M., & Saavedra-López, M. A. (2022). Gesture Phase Segmentation Dataset: An Extension for Development of Gesture Analysis Models. *Journal of Internet Services and Information Security*, *12*(4), 139-155. https://doi.org/10.58346/JISIS.2022.I4.010

[22] Shield, B., & ATHERTON, M. (2019). Hearing loss–numbers and costs. *Evaluation of the social and economic costs of hearing impairment. London: Brunel University*.

[23] Tarafder, K. H., Akhtar, N., Zaman, M. M., Rasel, M. A., Bhuiyan, M. R., & Datta, P. G. (2015). Disabling hearing impairment in the Bangladeshi population. *The Journal of Laryngology & Otology*, *129*(2), 126-135. https://doi.org/10.1017/S002221511400348X

[24] Taye, M. M. (2023). Understanding of machine learning with deep learning: architectures, workflow, applications and future directions. *Computers*, *12*(5), 91. https://doi.org/10.3390/computers12050091

[25] Vujović, Ž. (2021). Classification model evaluation metrics. *International Journal of Advanced Computer Science and Applications*, *12*(6), 599-606. https://doi.org/10.14569/IJACSA.2021.0120670

Authors Biography

Zaid Saad Bilal, is a seasoned researcher and educator with a Master's degree from Al-Hadithah University for Business and Science in Lebanon. His professional journey includes a tenure at the National School of Electronics and Telecommunications, University of Sfax in Tunisia, where he was associated with the LETI Laboratory at ENIS. With a strong focus on advanced topics in electronics and telecommunications, his interests extend to artificial intelligence. Currently, he imparts knowledge as a secondary school teacher.

Amir Gargouri, received his MSc in Electrical Engineering from the National Engineering School of Sfax, Tunisia, in 2009. Following this, he joined the Control and Energy Management Laboratory (CEM Lab) to pursue his thesis and earned his PhD in 2013. His research interests include neural network implementations, deep learning, machine learning, and Internet of Things (IoT) applications. Currently, he is a member of the Laboratory of Signals, Systems, Artificial Intelligence, and Networks (SM@RTS) and serves as an Associate Professor at the National School of Electronics and Communication, University of Sfax.

Hanaa F. Mahmood, received the B.Sc. degree in computer science from the University of Mosul, Iraq, The M.Sc. degree in Artificial Intelligent form the university of Mosul, and the Ph.D. degree in Artificial Intelligent from Loughborough university, U.K. She is a lecturer at department of computer science, college of Education for Pure Sciences, University of Mosul, Iraq. Her research interests include machine learning, deep Learning, computer vision and intelligent systems.

Hassene Mnif, was born in Sfax, Tunisia, in 1975. He received the Engineer and Master Diplomas in electrical engineering from the University of Sfax (ENIS) in 1999 and 2000, respectively, the PhD. degree in electronics from the University of Bordeaux I, France, in 2004 and the HDR degree from the University of Sfax in 2011. He is currently full Professor and Vice-President of the University of Sfax responsible for scientific research. He was the Director of this school between 2014 and 2020, where he has multiple innovative engineering education initiatives. He is a member of the Electronic and Information Technology Laboratory. His research interests include Energy Harvesting, Design of Radio-Frequency Integrated Circuits, Characterization, and compact modeling of both high frequency devices and future emerging technologies like Carbon Nanotube Field Effect Transistor (CNTFET). He also participates in research for real time image and video text extraction and micro mobility systems. He has authored and co-authored about 100 journal publications and conference papers and has gathered significant scientific coordination experience within national and international collaborative research projects. He participated in the organization of several IEEE conferences and workshops. He served as the Tunisia Section treasurer between 2011 and 2013, he is currently the IEEE Tunisia Section Chair-Elect.