ALGSL89: Bir Cezayir İşaret Dili Veri Seti

***

ALGSL89: An Algerian Sign Language Dataset

Ahmed KHELDOUN[1]

Imene KOUAR[2]

El Bachir KOUAR[3]

DOI:10.33461/uybisbbd.1339892

Öz

Makale Bilgileri

Makale Türü:
Araştırma Makalesi

Geliş Tarihi:
09.08.2023

Kabul Tarihi:
07.12.2023

Otomatik İşaret Dili Tanıma (ASLR), sağır ve işiten insanlar arasında iletişimi kolaylaştırmayı amaçlayan aktif bir araştırma alanıdır. Özellikle Cezayir İşaret Dili (ALGSL) bağlamında işaret dili tanıma, henüz kapsamlı bir şekilde incelenmemiş benzersiz zorluklar sunmaktadır. Bildiğimiz kadarıyla, şimdiye kadar ALGSL Tanıma üzerine bir çalışma yapılmamıştır. Bu durum, büyük ölçüde mevcut veri setlerinin eksikliğinden kaynaklanmaktadır. Bu zorluğun üstesinden gelmek için, ALGSL araştırmalarında öncü bir çaba olarak ALGSL89 veri setini öneriyoruz. ALGSL89 veri seti, 10 konu tarafından kaydedilen 89 farklı ALGSL işaretini kapsayan 4885 video içermektedir. Bu veri seti, Cezayir işaret dili topluluğuna özgü ASLR araştırmalarını ilerletmek için temel bir kaynak olarak hizmet etmektedir. Ek olarak, el şekilleri, pozisyonlar, yörüngeler ve işaret hareketlerinin dinamik yönleri dahil olmak üzere, karakteristiklerinin kapsamlı bir analizini sunuyoruz. Bu detaylar, araştırmacıların veri setini nüanslı bir şekilde anlamalarını ve ASLR çalışmalarında etkili bir şekilde kullanmalarını sağlamak için hayati öneme sahiptir. Veri setimizin geçerliliğini test etmek amacıyla, derin öğrenme modelleri uygulayarak elde ettiğimiz sonuçları sunuyoruz. Son olarak, Otoenkoder modeline dayanan yenilikçi bir ALGSL tanıma sistemi olan SignAtlas'ı sunuyoruz.

Anahtar Kelimeler: *Cezayir İşaret Dili, İşaret Dili Tanıma, El Şekli Tanıma, Otoenkoder, Derin Öğrenme.*

Abstract

Article Info

Paper Type:
Research Paper

Received:
09.08.2023

Accepted:
07.12.2023

Automatic Sign Language Recognition (ASLR) is an area of active current research that aims to facilitate communication between deaf and hearing people. Recognizing sign language, particularly in the context of Algerian Sign Language (ALGSL), presents unique challenges that have yet to be comprehensively explored. So far, to the best of our knowledge, no study has considered the ALGSL Recognition. This is mainly due to the lack of available datasets. To overcome this challenge, we propose the ALGSL89 dataset, a pioneering effort in ALGSL research. The ALGSL89 dataset encompasses 4885 videos, capturing 89 distinct ALGSL signs, recorded by 10 subjects. This dataset serves as a foundational resource for advancing ASLR research specific to the Algerian signing community. In addition, we provide a comprehensive analysis of its characteristics, including statistical insights and detailed information on handshapes, positions, trajectories, and the dynamic aspects of sign movements. These details are crucial for researchers to gain a nuanced understanding of the dataset, ensuring its effective utilization in ASLR studies. In order to test the validity of our dataset, we provide the results obtained by applying a set of deep learning models. Finally, we present SignAtlas, an innovative ALGSL recognition system based on Autoencoder model.

Keywords: *Algerian Sign Language, Sign language recognition, handshape recognition, Autoencoder, Deep learning.*

Atıf/ to Cite (APA): Kheldoun A., Kouar I. & Kouar E.B. (2023). ALGSL89: An Algerian Sign Language Dataset. International Journal of Management Information Systems and Computer Science, 7(2), 128-141. DOI:10.33461/uybisbbd.1339892

[1] Dr., Department of Mathematics and Computer Science, University of Yahia Fares, kheldoun.ahmed@univ-medea.dz, Medea, Algeria.

[2], Department of Mathematics and Computer Science, University of Yahia Fares, kouar.imene@gmail.com, Medea, Algeria.

[3] , Department of Mathematics and Computer Science, University of Yahia Fares, kouar.elbachir@gmail.com, Medea, Algeria.

1. INTRODUCTION

Sign languages, as highlighted by (Ruben, 2005) exhibit significant differences from spoken languages regarding their lexicons and linguistic grammars. Consequently, hearing individuals face substantial difficulties communicating through sign languages without proper training, creating a communication gap between the deaf and hearing communities. To bridge this communication divide, sign language recognition has emerged as a developing area of research. To bridge this communication divide, sign language recognition has emerged as a developing area of research. The goal is to interpret the meaning conveyed by signers through their movements and gestures. The history of SLR traces back to the 18th century, initially focusing on the recognition of simple finger-spellings. Over time, research has evolved, with numerous studies exploring the potential solutions offered by technologies and applications of sign language recognition (Obi et al., 2022) (Srivastava et al., 2022) (Vargas et al., 2011). However, sign language recognition represents a challenging research field, especially in the context of Algerian Sign Language (ALGSL). So far, to the best of our knowledge, no study has considered the ALGSL recognition. This is mainly due to the lack of available datasets.

There are numerous publications dealing with the automatic recognition of sign languages (AL-Qurishi et al., 2021) (Adeyanju et al., 2021). The full task of recognizing a sign language depends on a fundamental stage which is the pre-process step. The pre-process step is the task of preparing and refining the input video before feeding it into the main machine learning model. Indeed, it is reasonable to assume that optimizing the pre-processing of the input video is critical in improving convenient real-time machine learning solutions for sign language recognition further. In the context of video-based recognition, the pre-process can be simplified as :

- Tracking and segmenting the hands in every frame of the video.
- Recognizing the shapes of the hands, the movements they made and their positions.

The main contribution of this work is to propose a new ALGSL video-based sign dataset. For that, we took a proactive approach and visited several deaf schools across Algeria, immersing ourselves in the language and culture of the deaf community. This involved actively engaging with deaf individuals, sign language interpreters, and native signers to capture a wide range of ALGSL signs, gestures, and expressions. By actively involving the deaf people and gathering a diverse dataset, we ensured that our research reflects the linguistic and regional variations within ALGSL. In addition, we collaborated closely with the deaf community, valuing their expertise and incorporating their feedback to ensure the accuracy and cultural authenticity of the collected data.

The second contribution of this work is to provide some statistics of the signs to better understand the nature and challenges of our dataset. These statistics, need to be calculated during the pre-process step, can be useful for benchmarking a variety of computer vision and machine learning methods. Indeed, these statistics related to handshapes, positions, trajectories and amount of movement of the signs, in each frame of the video.

The final contribution of this work is to test the validity of the proposed dataset by applying several deep learning models. The obtained results indicates that the Autoencoder model outperforms all the other proposed models. Based on the Autoencoder model, we have developed the first ALGSL recognition system, named SignAtlas.

The remainder of this paper is organized as follows. Section 2 gives un overview about Algerian sign language. In section 3, we present the new video-based sign dataset for ALGSL. Section 4 presents some statistics of the proposed dataset. In section 5, we describe some deep learning models and results obtained with our dataset. In section 6, we present the developed ALGSL recognition system. Finally, Section 7 concludes and gives some further research directions.

2. ALGERIAN SIGN LANGUAGE

The ALGSL is a gestural language used by deaf individuals in Algeria as a means of communication through signs. The ALGSL is officially recognized by the law of May 8, 2002, as the first language of the deaf community in Algeria, making it the only country in the Arab world and in Africa to officially recognize sign language (Lanesman, 2016).

The ALGSL can be traced back to its origin from French Sign Language (LSF). It is noteworthy that Algerian Sign Language is not uniform, and comprises various regional variations (Hicham, 2021). These variations are not exhaustive and include, but are not limited to:

– Algerian Jewish Sign Language, AJSL or Ghardaia Sign Language since it was mainly developed and used in the village of Ghardaia by the Algerian Jewish individuals at that time (Lanesman & Meir, 2012).
– Algerian Sign Language of Laghouat, which is used by many Deaf people in Laghouat province and other cities (or villages) around it (Amal, 2016).
– Algerian Sign Language of Oran, it is used by the Deaf in the North of Algeria, particularly in the city of Oran (Sghier, 2007).
– Algerian Sign Language of Adrar, which is used by the Algerian Deaf community in Adrar, in the South of Algeria (Abdelouafi, 2019).

The ALGSL is a gestural language that relies entirely on signs, which are produced through different parts of the body, including one or both hands, the face, shoulders, or even the entire body. Despite its unique mode of expression, the ALGSL is considered to be a language like any other, possessing a distinct vocabulary and organized syntax. To effectively learn and understand the ALGSL, one must simply become familiar with its alphabet. In fact, every sign within the ALGSL alphabet is generated by one or two particular hand postures, highlighting the precision and complexity of the language's gestural system (Nekkaa, 2015).

ALGSL is composed of a total of 42 signs for alphabet (Nekkaa, 2015), which include 35 static and 6 dynamic signs. It is noteworthy that these signs are produced using a single hand gesture (refer to Figure 1). Moreover, each static sign is identified by two distinct features: configuration and orientation.

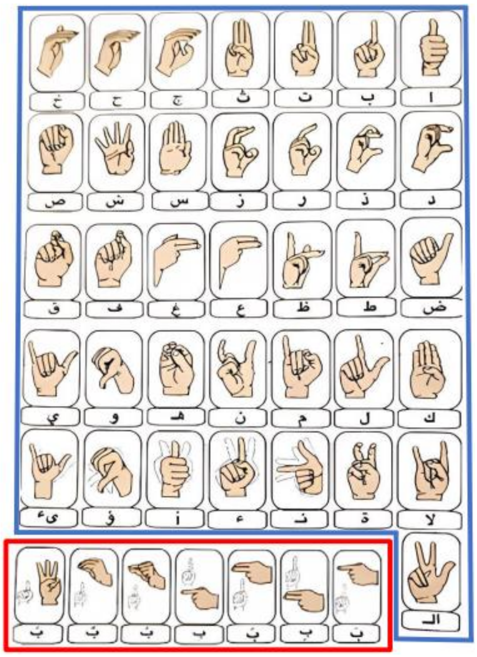


Figure 1. Algerian Sign Language Alphabet (the ones with red border are the dynamic signs)

As an example, the sign for the letter "ب" in Algerian Sign Language is defined by two parameters (Nekkaa, 2015):

- Configuration: making a fist with the index finger extended.
- Orientation: the palm is facing upwards (or with the wrist in a downwards position).

3. ALGERIAN SIGN LANGUAGE DATASET (ALGSL89)

The development of an Algerian Sign Language dataset, aimed at facilitating the creation of a dictionary and training an automatic sign recognition system. This dataset contains 4885 videos showcasing ten non-expert subjects and one deaf individual each perform five repetitions of 89 distinct sign types. The selected signs, comprising both verbs and nouns, represent the most commonly used elements of the ALGSL lexicon and are classified into nine thematic categories: time, colors, places, justice-related terms, medical terminology, months, interrogatives, family, general communication, and specific verbs. Figure 2 provides several examples of these signs.

Figure 2. Snapshots of six frames random distinct signs extracted from the ALGSL89 dataset

Table *1* provides a detailed list of the signs used in our ALGSL89 dataset. It showing the ID, name, and hand(s) used in each sign of the dataset. The "H" column specifies whether the sign was performed with the right hand ("R"), left hand ("L"), or both hands ("B").

Table 1. List of the signs in ALGSL89 dataset, along with the ID and hand(s) used for each sign

ID	Name	H	ID	Name	H	ID	Name	H	ID	Name	H
00	Semaine	B	24	Aéroport	B	48	Juillet	R	72	Prénom	B
01	Dimanche	R	25	Banque	B	49	Aout	R	73	Nom de famille	B
02	Lundi	R	26	Restaurant	B	50	Septembre	B	74	Oui	R
03	Mardi	R	27	Hôtel	R	51	Octobre	B	75	Non	R
04	Mercredi	R	28	Rue	B	52	Novembre	B	76	Bonjour	R
05	Jeudi	B	29	École	B	53	Décembre	B	77	Merci	R
06	Vendredi	R	30	Université	R	54	Mère	R	78	Derien	R
07	Samedi	R	31	Loi	B	55	Père	R	79	Question	R
08	Année	B	32	Avocat	B	56	Frère	R	80	Message	B
09	Heur	B	33	Juge	B	57	Sœur	R	81	Toilette	R
10	Minute	B	34	Liberté	B	58	Fille	R	82	Quand	B
11	Couleur	R	35	Témoin	B	59	Garçon	R	83	Comment	B
12	Blanc	R	36	Héritage	B	60	Entrer	B	84	Ou	R
13	Noir	R	37	Malade	B	61	Sortir	B	85	Combien	B
14	Rouge	R	38	Vaccination	R	62	Aimer	R	86	Hier	R
15	Bleu	R	39	Médecin	R	63	Acheter	B	87	Demain	R
16	Jaune	R	40	Médicament	R	64	Manger	R	88	Rendez-vous	B
17	Vert	R	41	Premier secours	B	65	Chercher	B			
18	Rose	R	42	Janvier	R	66	Demande	B			
19	Hôpital	B	43	Février	R	67	Écrire	B			
20	Police	B	44	Mars	B	68	Appeler	B			
21	Tribunal	B	45	Avril	B	69	Perdre	B			
22	Mosquée	B	45	Mai	R	70	Trouver	R			
23	Pharmacie	R	47	Juin	R	71	Traduire	B			

The ALGSL89 dataset was captured in eleven different sessions, primarily within the indoor setting of our university classrooms. Each session focused on a specific set of signs, with most signs being captured in a single session. However, some signs required additional sessions for refinement and correction. More repetitions than required were captured during each session to account for potential errors. We utilized indoor lighting for most of the recordings and supplemented it with an artificial white projector to ensure consistent lighting across all sessions, regardless of the different recording times. All recordings were captured using a OnePlus 8 smartphone, which records videos at 4K 30fps with a 16:9 ratio. A tripod was set 1.6 meters away from the wall and at a height of 1.4 meters to maintain consistency across all recordings (more details illustrated can be seen in Figure 3).

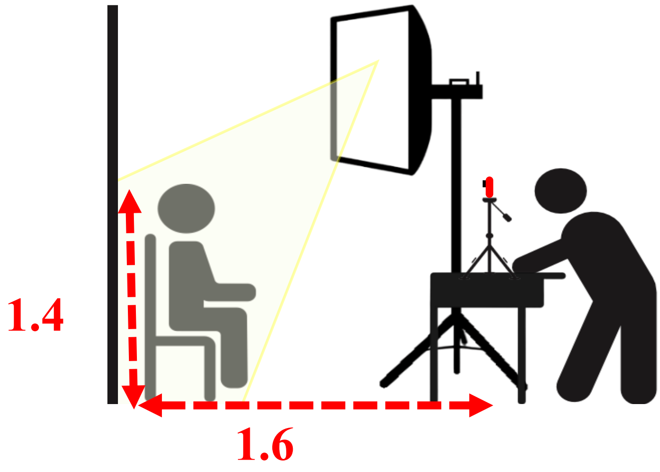


Figure 3. The configuration of the recording setup

Due to unforeseen circumstances, we had to adjust the recording location for deaf Subject, who was unable to participate in the same session as the others. We arranged a session at the deaf school in Algeria, where the individual was located. Despite this change, we strove to maintain the same recording conditions to preserve the dataset's quality and integrity.

The majority of the subjects wore dark-colored clothing and performed the signs against a white wall background to maintain consistency across all videos. To enhance the database's diversity and realism, we imposed minimal constraints on the subjects while performing the signs, ensuring a natural execution and a more realistic representation of real-world signing conditions. We believe that incorporating non-expert subjects added a greater level of diversity to the dataset.

In Table 2. *Comparison of datasets* we present a comparison of our dataset with the other video-datasets existing in the literature. We consider only the datasets for isolated word recognition. The advantage of our dataset is it provides the maximum number of samples per class of sign. This may be useful for a good training of a machine learning model.

Table 2. Comparison of datasets

Dataset	Subjects	Classes	Samples
LSA64 (Rosete et al., 2016)	10	64	3200
Boston ASL LVD (Neidle et al., 2012)	6	3300+	9800
GSL 20 (Helen et al., 2011)	6	20	840
PSL Kinect 30 (Tomasz et al., 2015)	1	30	300
PSL ToF 84 (Tomasz et al., 2015)	1	84	1680
ALGSL89 (our)	10	89	4885

4. STATISTICS OF THE ALGSL89 DATASET

We present some statistics for the ALGSL89 dataset such as : Handshapes, Positions, Trajectories and amount of movement of the signs. These details are crucial for researchers to gain a nuanced understanding of the dataset, ensuring its effective utilization in Automatic Sign Language Recognition studies. All these statistics has been computed using the Mediapipe library (Lugaresi et al., 2019), which provides a real-time human body tracking.

4.1. Handshapes

In sign languages, handshapes refer to the configuration of the hands at a specific moment, defined by the positioning of the fingers and palm. Figure 4 and Figure 5 showcase the distinct handshapes of the right and left hand, respectively, for each sign class. For both figures, we show only the handshape of the first frame of the video-sign. We note that numerous handshapes exhibit repetition, although their 2D projection may vary due to hand rotation. It means that signs in the ALGSL89 dataset possess significant overlap in terms of types of handshape.

Figure 4. Images of segmented right hand as captured in the ALGSL89 dataset

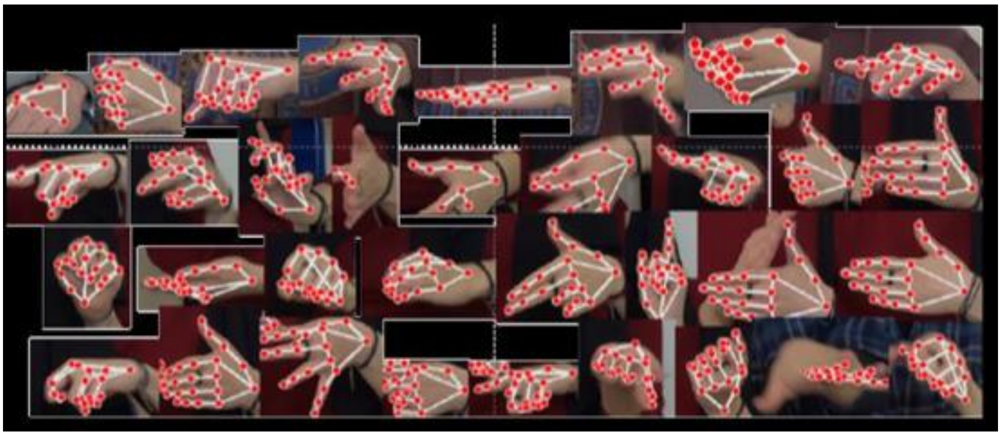


Figure 5. Images of segmented left hand as captured in the ALGSL89 dataset

4.2. Positions

In Figure 6, the mean initial and final positions for each hand, along with the covariance, are depicted. Although some signs can be distinguished based on their positions, there is substantial overlap observed in the majority of cases (i.e. positions are very close to each other).

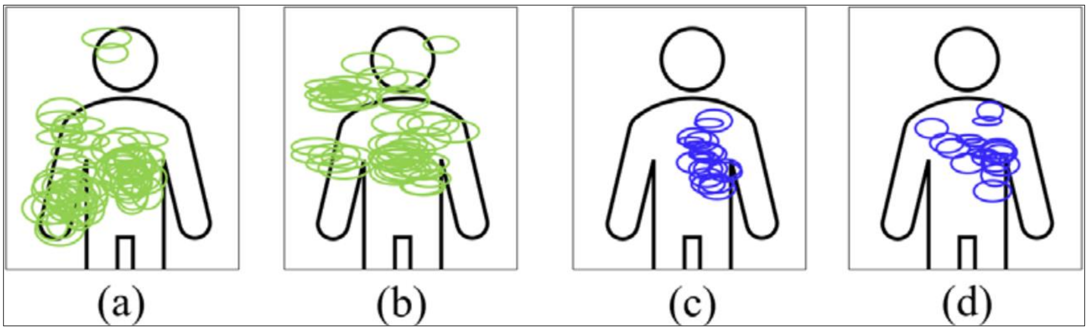


Figure 6. Means for the initial and final positions of the right hand for each sign (a and b), and also for the left hand (c and d)

4.3. Trajectories

Trajectory refers to the motion or path the hands take when producing a sign. Figure 7 shows sample trajectories of each sign, as performed by subject 6. There is a noticeable amount of overlap in the movements of both one-handed and two-handed signs. Noteworthy examples of one-handed signs with similar trajectories include signs 2, 3, 48, 40, and 15. Similarly, two-handed signs like signs 52, 53, and 54 also exhibit comparable movement patterns.

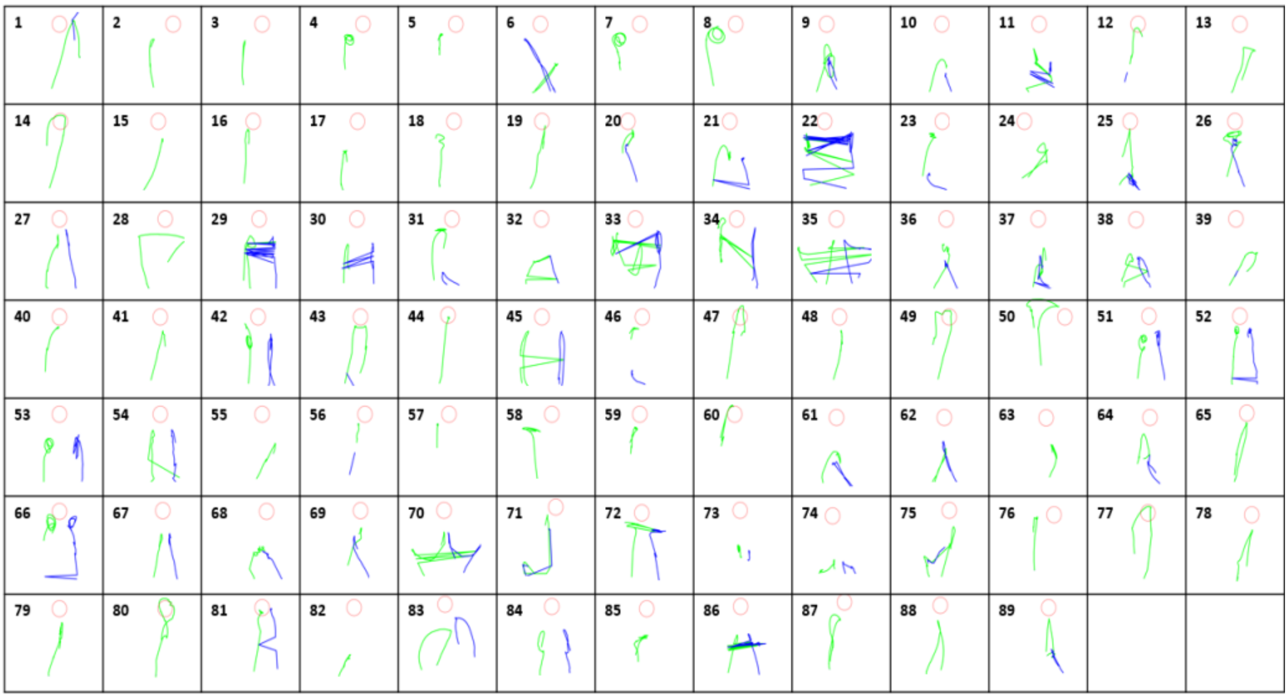


Figure 7. Sample trajectories for each sign in ALGSL89 dataset. The left-hand trajectory is shown in light green, the right-hand one in blue, and the head position as a red circle

4.4. Amount of movement

Figure 8 shows the amount of movement for each hand, measured with coordination of the central landmark in the trajectory of the hand. The movement in the left hand (blue bars) is significantly smaller than that of the right hand (green bars) in many signs, consistent with the fact that the right hand is the dominant one for all the subjects. Therefore, in ALGSL89 dataset, the right hand is more used then the left one for producing the different class of signs.

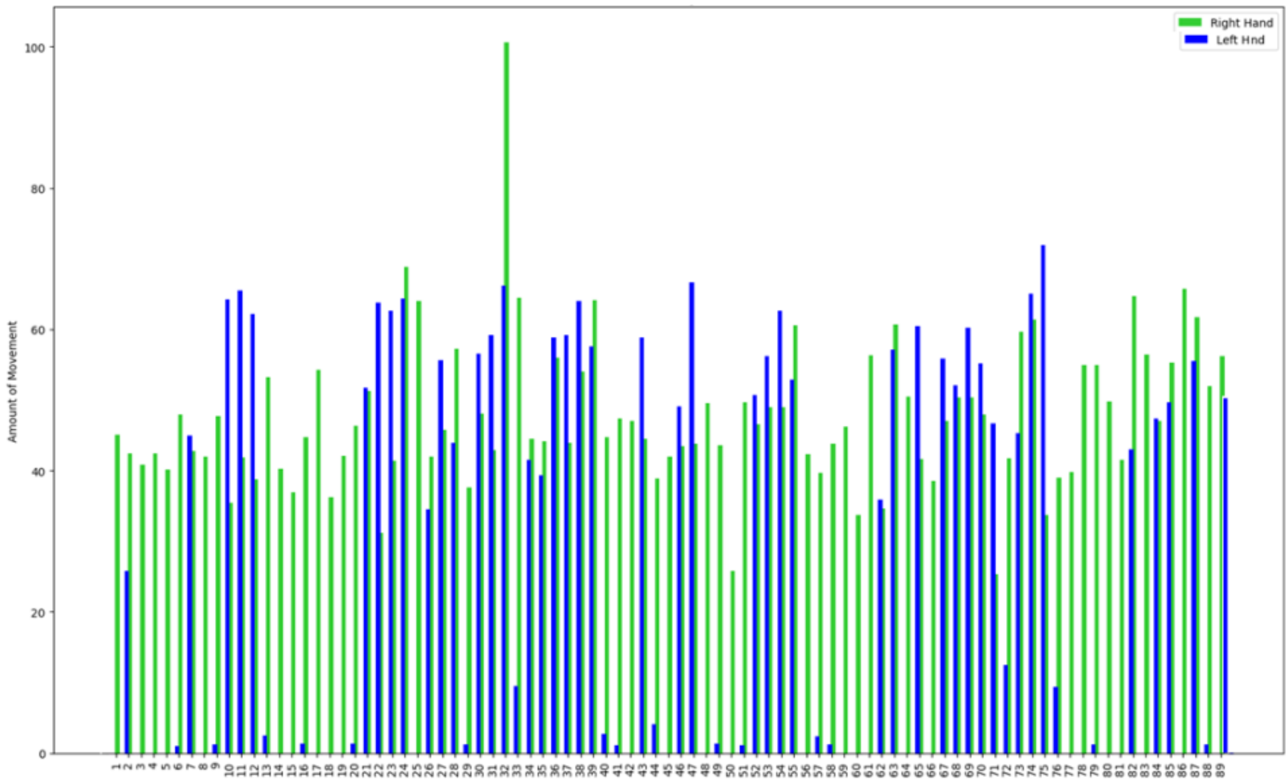


Figure 8. Amount of movement for each class of sign in ALGSL89 dataset.

5. APPLYING DEEP LEARNING TECHNIQUES

In this section, we describe some deep learning models and results obtained with the ALGSL89 dataset. First, we present the pre-process step, which consists in extracting features from input sign video. The pre-process step is a fundamental stage in any deep learning technique (LeCun et al., 2015). In Figure 9, we show the proposed pre-process step. From the input video, we extract only non blurred frames using Laplacian variance. Then, from each selected frame, we extract an image of each hand and 21 landmarks per hand. The result of this pre-processing step is a sequence of selected frames, where for each frame we extract the handshape and positions of hands.

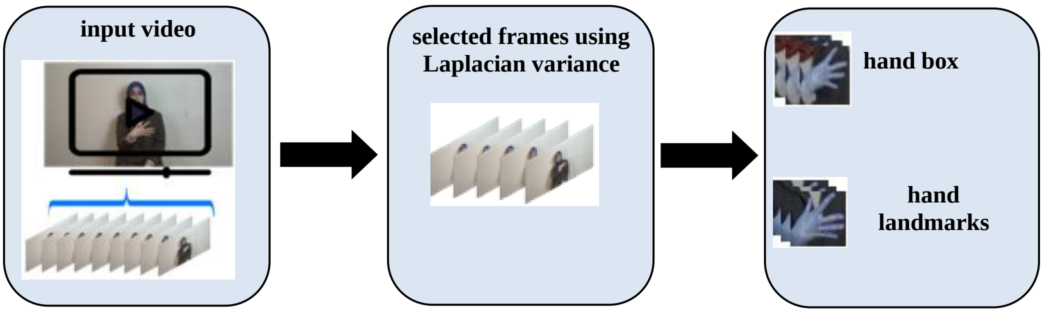


Figure 9. Pre-processing step

The architecture of the developed deep learning models are described in the Appendix. The models using the well known CNN (Alzubaidi et al., 2021) (Chang & Jin, 2017) and LSTM (Lindemann et al., 2021) (Siriak et al., 2019) or a combination of them. Table 3 presents the obtained results of different models on the ALGSL89 dataset. The reported metrics include the loss and accuracy values. It is noted that the Autoencoder model outperformes all the other models. It achieves encouraging performance. The Autoencoder combines both convolutional neural networks

(2DCNNs) and long-short-term memory (LSTM) networks along with an attention mechanism (Vaswani et al., 2017). In fact, the CNN model is used to capture spatial information in frames, the LSTM model used to capture sequential dependencies in a series of frames and the attention concept is used to focus on important parts in frames.

Table 3. Performance Comparison of different models on ALGSL89 dataset

Model	Accuracy (%)	Loss	Validation Acc (%)	Validation Loss
LSTM	92.28%	0.2259	60.56%	2.3433
2DCNN-LSTM	71.48%	0.9300	65.00%	0.9395
ConvLSTM	80.48%	0.6930	73.48%	0.9395
Autoencoder	**100%**	**0.00005**	**98.99%**	**0.0015**

6. GRAPHICAL TOOL

We have developed an automatic algerian sign language recognition system based on Autoencoder model, named SignAtlas. This system, as shown in Figure 10. *SignAtlas*, allows users to use webcam to start recording gestures, and the results are displayed in both written and sound formats. Furthermore, SignAtlas allows users to upload pre-recorded videos and predicted the corresponding gestures.

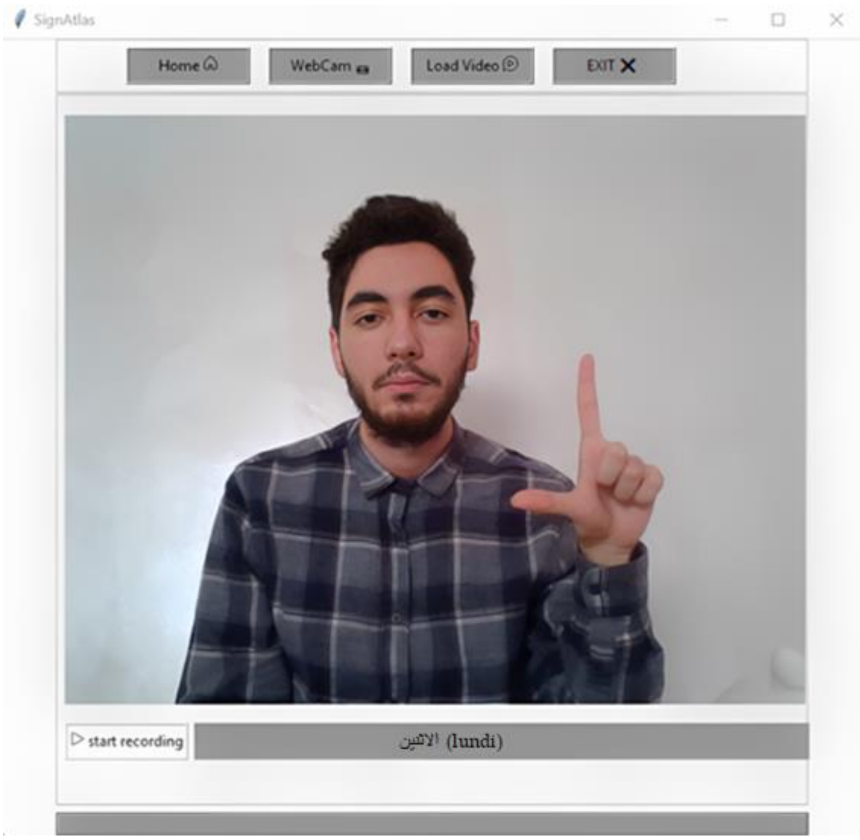


Figure 10. SignAtlas

7. CONCLUSION AND FUTURE WORKS

The major contribution of this work is the proposed of a new ALGSL video-based sign dataset which consisting of 89 distinct signs. We have presented a set of statistics and extra information to characterize the dataset and allow researchers to easily understand its nature. This work is a result of visiting several deaf schools across Algeria and collaboration with the deaf community. The development of this dataset aimed at facilitating the creation of a dictionary and training an automatic sign recognition system for ALGSL. To the best of our knowledge, no study has considered the creation of dataset for ALGSL.

In order to test the validity of our dataset, we have developed several deep learning models where the Autoencoder model outperforms all the other proposed models. Based on the Autoencoder model, we have introduced SignAtlas, an innovative ALGSL recognition system.

There are several perspectives to consider for further research:

- Collecting additional diverse samples, including a wider range of ALGSL signs, gestures, and expressions to provide a complete basic working vocabulary for ALGSL.
- Extending SignAtlas for Continuous Sign Language Recognition where the system can interpret complete sign language sentences and conversations would enable more natural and fluid communication.

REFERENCES

Abdelouafi, H. (2019). Teaching Sign Language to the Deaf Children in Adrar, Algeria.

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

AL-Qurishi, M., Khalid, T., & Souissi, R. (2021). Deep Learning for Sign Language Recognition: Current Techniques, Benchmarks, and Open Issues. *IEEE Access, 9*. https://doi.org/10.1109/ACCESS.2021.3110912

Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A. Q., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M., & Farhan, L. (2021). Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data, 8*.

Amal, D. (2016). Les points communs entre la Algerian Sign Language (LSA) - dialecte de Laghouat, Sud de l'Algérie - et la Langue des Signes Française (LSF). Licence thesis, « Acquisition et dysfonctionnement », Faculté ALLSHS d'Aix-en-Provence.

Chang, J., & Jin, S. (2017). An efficient implementation of 2D convolution in CNN. *IEICE Electronics Express, 14*, 4299-4308. https://doi.org/10.1587/elex.13.20161134

Helen, C., Brian, H., & Richard, B. (2011). Sign Language Recognition. Dans *Visual Analysis of Humans: Looking at People* (pp. 539 - 562). Springer.

Hicham, A. (2021, October). Deaf Education in Algeria: Is it a Sustainable Approach? *Sociology Review, 5*, 417–429.

Lanesman, S. (2016). *Algerian Jewish Sign language: its emergence and survival*. Masters thesis, University of Central Lancashire.

Lanesman, S., & Meir, I. (2012). Algerian Jewish Sign Language: A sociolinguistic sketch. Dans *Sign Languages in Village Communities: Anthropological and Linguistic Insights* (éd. 1, pp. 361–364). De Gruyter. http://www.jstor.org/stable/j.ctvbkjwzx.17

LeCun, Y., Bengio, Y., & Hinton, G. (2015, May). Deep Learning. *Nature, 521*, 436-44. https://doi.org/10.1038/nature14539

Lindemann, B., Müller, T., Vietz, H., Jazdi, N., & Weyrich, M. (2021). A survey on long short-term memory networks for time series prediction. *Procedia CIRP, 99*, 650-655. https://doi.org/https://doi.org/10.1016/j.procir.2021.03.088

Lugaresi, C., Tang, J., Nash, H., McClanahan, C., Uboweja, E., Hays, M., Zhang, F., Chang, C.-L., Yong, M. G., Lee, J., Chang, W.-T., Hua, W., Georg, M., & Grundmann, M. (2019). MediaPipe: A Framework for Building Perception Pipelines. *MediaPipe: A Framework for Building Perception Pipelines*.

Neidle, C., Thangali, A., & Sclaroff, S. (2012). Challenges in development of the American Sign Language Lexicon Video Dataset (ASLLVD) corpus. *Proceedings of the LREC2012 5th*

Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon. 143-150.

Nekkaa, F. (2015). *Détection automatique de la main : Application à la reconnaissance de la langue des signes arabe.* Master thesis, Systèmes Distribués et Méthodes Formelles (SDMF), Université Abdelhamid Mehri-Constantine 2.

Obi, Y., Claudio, K. S., Budiman, V. M., Achmad, S., & Kurniawan, A. (2022). Sign language recognition system for communicating to people with disabilities. *Procedia Computer Science, 216*, 13-20. https://doi.org/https://doi.org/10.1016/j.procs.2022.12.106

Rosete, F. R., Quiroga, F. M., Estrebou, C., Lanzarini, L., & Rosete, A. (2016). LSA64: An Argentinian Sign Language Dataset. *ArXiv, abs/2310.17429*, 794-803.

Ruben, R. (2005). Sign language: Its history and contribution to the understanding of the biological nature of language. *Acta oto-laryngologica, 125*, 464-467. https://doi.org/10.1080/00016480510026287

Sghier, M. M. (2007). Langage et surdité, descriptive de la langue des signes des sourds Oranais. Magistère thesis, Université d'Oran Es-Sénia.

Siriak, R., Skarga-Bandurova, I., & Boltov, Y. (2019). Deep Convolutional Network with Long Short-Term Memory Layers for Dynamic Gesture Recognition. *2019 10th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), 1*, pp. 158-162. https://doi.org/10.1109/IDAACS.2019.8924381

Srivastava, S., Gangwar, A., Mishra, R., & Singh, S. (2022). Sign Language Recognition System Using TensorFlow Object Detection API. Dans *Communications in Computer and Information Science* (pp. 634–646). Springer International Publishing. https://doi.org/10.1007/978-3-030-96040-7_48

Tomasz, K., Mariusz, O., Marian, W., & Dawid, W. (2015). Recognition of Hand Gestures Observed by Depth Cameras. *International Journal of Advanced Robotic Systems, 12*(4).

Vargas, L. P., Barba, L., Torres, C. O., & Mattos, L. (2011, January). Sign Language Recognition System using Neural Network for Digital Hardware Implementation. *Journal of Physics: Conference Series, 274*, 012051. https://doi.org/10.1088/1742-6596/274/1/012051

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention Is All You Need. *CoRR*.

Appendix: Detailed summary of the proposed deep learning models.

We present the architectures of the developed models used in this paper. These models were designed to address the specific objectives and requirements of our study, as well as to facilitate a comprehensive comparison between different approaches. The following figures provide detailed architectures of each model, including the Autoencoder (Figure 11), ConvLSTM (Figure 12), 2DCNN-LSTM (Figure 13(a)), and LSTM Figure 13(b) architectures.

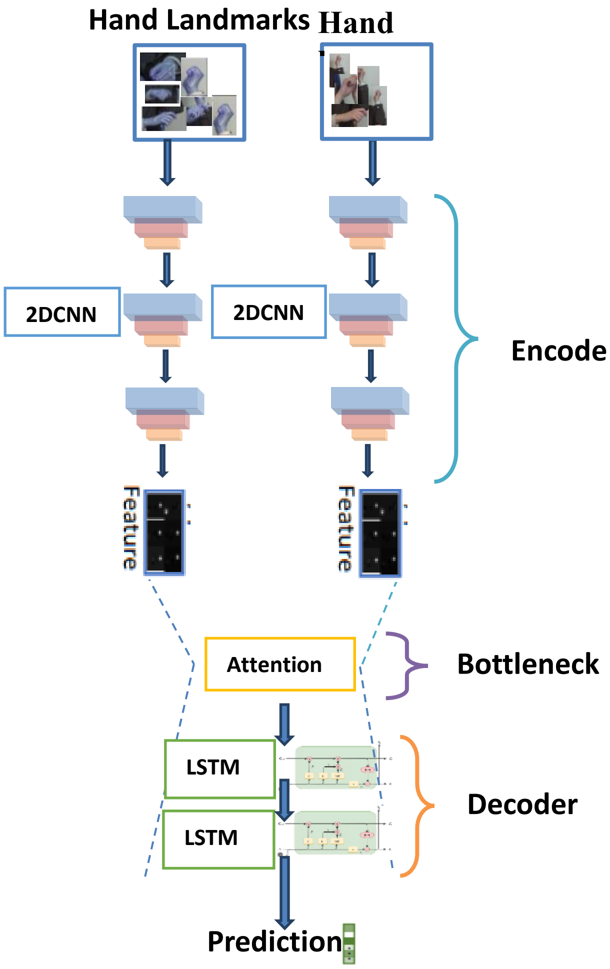


Figure 11. Detailed summary of the Autoencoder model

As shown in Figure 11, the Autoencoder model is composed of an encoder, a bottleneck, and a decoder.

- **Encoder:** The encoder part of our model consists of two branches. Each branch using a series of convolutional layers, max-pooling layers, and dense layers. It applies a 2D convolutional operation in a time-distributed manner to capture spatial information. Both branches extract relevant features from their respective inputs.
- **Bottleneck:** The encoded features from the two branches are concatenated and passed through an attention mechanism, which selectively focuses on important parts of the combined feature representation.
- **Decoder:** The decoder part of our model utilizes an LSTM network to decode the encoded features. The LSTM layer takes the attention-weighted features as input and processes them in a recurrent manner, capturing sequential dependencies.

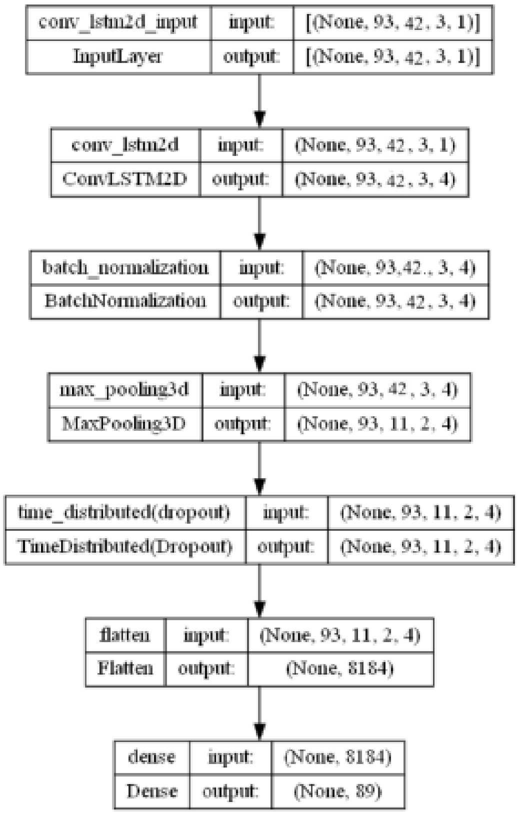
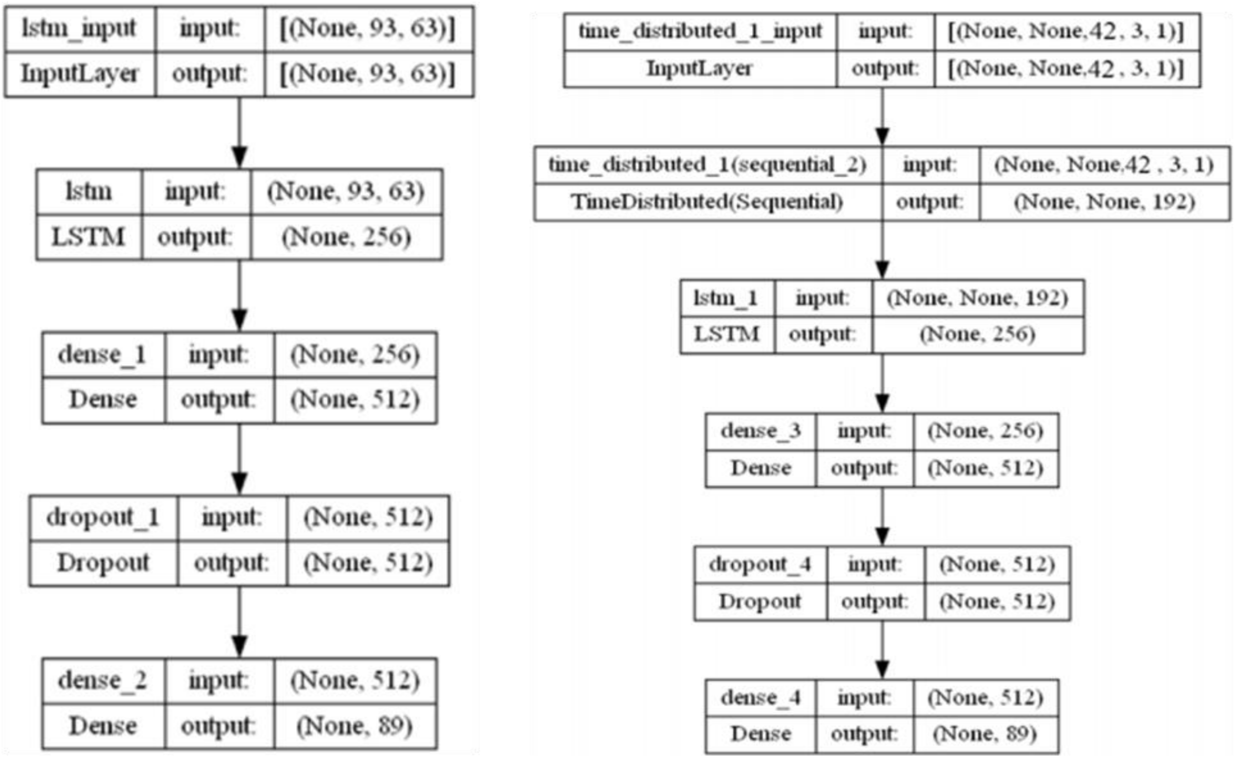


Figure 12. Detailed summary of the ConvLSTM model



(a) 2DCNN-LSTM

(b) LSTM

Figure 13. Detailed summary of the 2DCNN-LSTM and LSTM models