IEEE Access
Multidisciplinary ⁝ Rapid Review ⁝ Open Access Journal

Received 12 September 2023, accepted 9 October 2023, date of publication 12 October 2023, date of current version 18 October 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3324040

RESEARCH ARTICLE

JUMLA-QSL-22: A Novel Qatari Sign Language Continuous Dataset

OUSSAMA EL GHOUL, MARYAM AZIZ, AND ACHRAF OTHMAN, (Senior Member, IEEE)
Mada, Qatar Assistive Technology Center, Doha, Qatar
Corresponding author: Achraf Othman (aothman@mada.org.qa)

ABSTRACT This paper proposes the first large-scale and annotated Qatari sign language dataset for continuous sign language processing. This dataset focuses on phrases and sentences commonly used in healthcare settings and contains 6300 records of 900 sentences. The dataset collection process involves diverse participants, including both hearing-impaired individuals and sign interpreters, to capture variations in signing styles, speeds, and other linguistic nuances. The data collection setup integrates advanced technology, including true depth cameras, to comprehensively record signing movements from various angles. The collected dataset is rich in content, encompassing different signing variations and linguistic intricacies. The dataset is publicly available in IEEE Dataport. The paper also analyzes the data captured to understand the trends and patterns within the data. As the global population with hearing difficulties continues to grow, there is a pressing need for effective sign language recognition systems to bridge the communication gap between the deaf and non-deaf communities and the introduction of the JUMLA-QSL-22 dataset constitutes a significant stride toward addressing this imperative need.

INDEX TERMS Gestures recognition, hearing disorders, Qatari sign language, sign language, sign language processing.

I. INTRODUCTION

According to World Health Organization (WHO) [1], nearly 432 million adults and 32 million children face hearing difficulties and require help adjusting to their hearing disability. WHO further reports that by 2050, over 700 million people will have hearing disabilities. Hearing-impaired individuals face many challenges in their day-to-day lives, including communication and speech challenges. Communication with and between hearing-impaired individuals is mostly carried out using sign language. Sign language is a complete natural language with its own grammar and structure. Sign language uses gestures based on static and dynamic motion to convey meaning when communicating with others [2]. Static signs depend on the rotation and shape of the hands during signing whereas dynamic signs involve the use of both the hands and other features of the body during the signing. Fingerspelling, using the fingers to express the alphabet and numbers in sign language, is a form of static sign. Gestures made during signing can also be classified into two main categories: manual and non-manual gestures [3]. Manual gestures use hand motions for signing while non-manual gestures use other body features such as eyebrows and forehead movements, lip patterns, and head movements to add more meaning to signs.

Sign languages are not universal that is different countries have different sign languages such as American Sign Language (ASL), British Sign Language (BSL) and Chinese Sign Language (CSL) amongst others. To date, there are more than 300 sign languages around the world [4]. In the Arab World, sign language for Arabic community differs from country to country. Some of the sign languages from the Arab World are Saudi, Moroccan, Jordanian, Qatari, Libyan, and Palestinian [5]. In 1999, the League of Arab States (LAS) and the Arab League Educational, Cultural and Scientific Organization (ALECSO) attempted to standardize the Arabic sign language by combining several existing sign languages in the Arab world [6]. As a result, Arabic sign language (ArSL) was introduced and a two-part dictionary containing

The associate editor coordinating the review of this manuscript and approving it for publication was Alba Amato.

3200 sign words was published. However, ArSL is still not widely adopted by the deaf community due to difficulty in understanding and most Arab countries use their native sign language which differs in both grammar and structure from ArSL [7]. Sign language processing is essential to help the deaf community connect and integrate smoothly within society. With sign language processing, systems like recognition and translation of texts, and messages to signs and vice versa can be implemented which would benefit the deaf community in personal, work, and educational aspects [6]. The use of machine learning (ML) and AI could help bridge the communication gap between deaf and non-deaf communities in several ways. Computer vision algorithms based on ML translate sign language into spoken or written words, while other models transcribe spoken language into real-time text for the deaf. ML and AI can be used also for real-time subtitling of sign language video streaming or for predictive gesture analysis.

The development of sign language recognition remains a complex and ongoing endeavor, primarily due to several existing challenges. Firstly, limited studies exist on sign language processing due to the lack of availability of a standard sign language dataset [8], [9]. To the best of our knowledge, no dataset exists on Qatari sign language (QSL). Most researchers tend to create their own datasets specific to their research purposes for ArSL. Other than the absence of standardized and large-scale datasets, one of the primary obstacles lies in accurately detecting and capturing the movements of hands and facial expressions, which are crucial components of sign language [40]. Sign languages rely not only on manual gestures but also on non-manual elements, such as facial expressions and body motions, to convey meaning effectively. Therefore, sign language recognition systems must be capable of comprehensively representing and interpreting these diverse components in real-time for seamless communication. Compounding this challenge is the complexity of sign language representation. The representation must encompass essential linguistic information, movements, speed, and other variables to ensure no loss of semantic information [10]. Even slight changes in timing, movement, or configuration of different gesture components could lead to entirely different meanings [11]. Furthermore, sign language recognition systems also face challenges in with the difference variations in signs due to factors like epenthesis, individual signing styles, and environmental features. Epenthesis is defined as the dependence of the sign on the signs before or after, which creates problems during sign language recognition [9]. Location, background, and clothes are just some of the environmental features that pose problems in sign language processing. Hence, large datasets featuring different participants and environmental features are needed to mitigate dependencies errors. Furthermore, the issue of sign language dependency arises when different individuals sign the same word or phrase differently. The recognition systems must account for these variations in signing styles while ensuring accuracy and consistency in interpretation. These variations can lead to errors in recognition, hindering the accuracy and reliability of the system. Most publicly available datasets do not focus on continuous sentences but on alphabets or isolated signs, failing to capture the fluidity and dynamic nature of sign language communication in real-world scenarios. This limitation hinders the development of recognition systems that can effectively interpret continuous sign language conversations [9].

Addressing these challenges requires advancements in sign language technology, along with the creation of standardized and comprehensive sign language datasets that encompass various signing styles, regional variations, and natural signing interactions. Hence, this paper aims to introduce the first Qatari sign language dataset – the JUMLA-QSL-22 dataset. The JUMLA-QSL-22 dataset focuses on phrases and sentences used in a hospital and primary health center setting. The dataset is part of a larger research project called the JUMLA project. It is a significant contribution towards the development of language resources in the field of sign language processing. The paper is organized as follows. Section II explores the existing literature of sign language datasets and sign language processing as well as evaluation techniques. Section III describes the setup and collection process of the JUMLA-QSL-22 dataset. In section IV, we present the baseline method adopted and describe the data analysis results in section V. Section VI presents our conclusion and future work in QSL.

II. RELATED WORKS

This section presents the publicly available sign language datasets in literature. It also explores the existing sign language processing techniques and the different methods available for evaluating the performance of the sign language models.

A. SIGN LANGUAGE DATASETS

Existing sign language datasets are of three types mainly: fingerspelling, isolated signs, and continuous signs. Isolated sign datasets are the most common type of dataset available [12]. The existing datasets also employ different methods of data collection ranging from sensors for motion detection to using depth cameras such as Kinect to capture more than one modality. Athitsos et al. [13] introduced the ASL Lexicon Video Dataset which is a public dataset containing videos of almost 3000 signs performed by 6 native signers. The videos are captured from four different views including a side view, two frontal views, and a zoomed face view. The videos are annotated with descriptions on the start and end frame of each sign in the video, the gloss of the sign, whether the sign is performed with one hand or two hands and the signer id. Joze and Koller [14] proposed another ASL-based dataset consisting of 25,513 annotated videos of 1000 distinct signs signed by over 200 signers. The videos were mostly collected from the students and teachers who recorded and uploaded

TABLE 1. Sign language datasets CN: Class Number, SN: Subject Number, LL: Language Level.

Year	Dataset	CN	SN	Sample size	LL	Data type	Device used	Modality	Target Country
2008	ASL Lexicon Video Dataset [13]	3,314	6	9,794	words	videos	camera	RGB	USA
2019	MS-ASL [14]	1000	222	25,513	words, phrases	videos	-	multi-modality	USA
2012	DSG-40 [15]	40	15	3000	words	videos	Kinect camera	depth	Germany
2012	MSR Gesture 3D [16]	12	10	360	words	videos	Kinect camera	depth	USA
2013	PSL Kinect 30 [17]	30	1	300	words	videos	Kinect camera	color, depth	Poland
2014	RWTH-PHOENIX-Weather [18]	1200	9	45760	sentences	videos	camera	RGB	Germany
2021	mArSL [2]	50	4	6,748	words	videos	Microsoft Kinect V2	color, depth, joint points, face, and faceHD	Arab Region (Unified Arabic SL)
2018	ArASL2018 [20]	32	40	54,049	letters	images	smart camera	-	Arab Region (Unified Arabic SL)
2013	SignsWorld Atlas [21]	500	10	535+	letters, words, sentences	images and videos	digital camera	RGB	Arab Region (Unified Arabic SL)
2022	[22]	20	72	8,467	words	videos	smartphones	RGB	Arab Region (Unified Arabic SL)
2021	[23]	215	4	215	words, sentences	videos	-	RGB	Arab Region (Unified Arabic SL)
2019	[24]	40	1	400	sentences	videos	DG5-VHand data gloves, two Polhemus G4 motion trackers, camera	RGB	Arab Region (Unified Arabic SL)
2021	KArSL [25]	502	3	75,300	words	videos	Microsoft Kinect V2	multi-modality	Arab Region (Unified Arabic SL)

themselves signing different words and phrases. Cooper et al. [15] collected 40 isolated signs from 15 participants with each participant repeating the sign 5 times to develop the Deutsche Gebardensprache – German Sign Language (DGS) dataset. The data collection was conducted using a Kinect camera and a mobile system was implemented to get varying views of the participants. The use of Kinect camera enabled them to collect the depth in the dataset as well. Kurakin et al. [16] introduced a dataset, MSR Gesture 3D, comprising of 12 ASL gestures collected from 10 participants. The participants repeated each sign 3 times. A Kinect device was used to collect the depth of the ASL signs. Oszust and Wysocki [17] developed a Polish Sign Language (PSL) dataset using Kinect sensor. The dataset contained 30 PSL words signed by a PSL teacher 10 times. With the Kinect sensor, depth and color data was collected from the videos recorded. Koller et al. [18] introduced the RWTH-PHOENIX-Weather dataset that contains a collection of videos gathered from the daily news and weather recordings of the German tv-station PHOENIX over a period of three years from 2009 to 2011. The videos were recorded on a stationary color camera.

The main challenge that the Arabic sign language processing faces is the lack of availability of a standard dataset. Most researchers create their own datasets for ArSL recognition and processing [19]. Luqman and El-Alfy [2] proposed an ArSL dataset called mArSL that consisted of 6,748 samples based on 50 signs from four participants trained on ArSL. The dataset was mainly introduced to recognize signs that require manual and non-manual gestures. The dataset represents each sign with different modalities of color, depth, skeleton joint points, and facial information. Latif et al. [20] introduced the ArASL2018, an Arabic Alphabet sign language dataset. The dataset contained 54,049 grayscale images based on 32 letters and signs in the Arabic sign language. It was collected from 40 participants from different age groups and contained varying numbers of images per alphabet. Another dataset aimed at ArSL was SignsWorld Atlas, which contained more than 535 images and videos [21]. The dataset was collected through 10 participants within the age range 3 to 30 years and included both manual and non-manual signs. The dataset included facial expressions, fingerspelling, hand shapes, numbers, individual signs, lip movements and continued sentences. Balaha et al. [22] introduced an ArSL dataset based on 20 words in the Arabic language. It included 8,467 smartphone videos from 72 participants within the age range of 20 to 24 years. The recorded videos support different resolutions, locations, and backgrounds. Abbas et al. [23] collected an ArSL dataset of 215 videos focused on communication with deaf drivers. The dataset contained 215 words and sentences in different categories such as general words, directions, place, traffic and transportation, common sentences used by deaf drivers, common sentences used by passengers, and amount. Three different datasets were collected by Hassan et al. [24]. The datasets comprised of 40 Arabic sign languages sentences with each dataset collected using a different medium and each sentence repeated ten times. The first dataset was reused from Tubaiz et al. [26] and was collected using DG5-VHand data gloves. The second dataset was collected using two Polhemus G4 motion trackers and the third dataset was collected using a camera alone. Sidig et al. [25] introduced KArSL, an ArSL dataset based on the Saudi dialect. KArSL comprised of 502 sign words taken from the ArSL dictionary and was signed by three participants within the age range of 30 to 40 years. Each participant was asked to sign each word 50 times resulting in the final dataset containing 75,300 samples. Table 1 shows the summary of the datasets discussed in this paper.

B. SIGN LANGUAGE PROCESSING

Research on sign language processing has looked into systems mainly using alphabet ArSL [27], [28], [29], [30] and isolated signs for recognition [31], [32], [33]. Kamruzzaman [19] used CNN to train a model to recognize the 31 letters in ArSL and translate them into Arabic speech. Their model reached an accuracy of 90%. Elpeltagy et al. [34] looked into classifying 150 isolated ArSL signs using hand segmentation, descriptors and classification. Random classifiers and canonical correlation analysis were used for classification of signs. They reported a maximum accuracy of 55.57% for the 150 signs. Hisham and Hamouda [35] studied two machine learning algorithms of KNN and SVM with each algorithm improved using Ada-boosting technique and Dynamic Time Wrapping technique on them separately. The models were applied on 30 hand gestures for isolated signs with 20 single hand gestures and 10 double hand gestures. An accuracy of 92.3% was reported for single hand gestures and 93% for double hand gestures using Ada-boosting technique. Mohandes et al. [36] presents an image based ArSL recognition system based on the Hidden Markov Model (HMM) algorithm. The system used a Gaussian skin model to detect the face region in an image and used region growing technique to detect hand tracking. Their model achieved an accuracy of approximately 93% when applied to 300 words.

Limited research has been done on continuous signs recognition. Suliman et al. [41] applied the deep learning method of CNN for feature extraction and Long Short-Term Memory (LSTM) model for classification in their study on ArSL recognition. They tested their model on a dataset of 150 signs which contained a mix of letters, words and phrases commonly used in the Arabic language. Their model was trained and tested under two different conditions of signer-dependent and signer-independent cases. Their model scored an accuracy of 95.9% in signer-dependent case and 43.62% for signer-independent case. In signer-dependent conditions, the training and testing set comes from the same signer whereas in signer-independent condition, the training set comes from one signer and the testing set comes from another new signer. While most papers focus on signer-dependent cases, the signer-independent conditions are preferred since the resulting model in signer-independent case is more robust and less dependent on the signer features for ArSL recognition. Hassan et al. [9] presented two sensor-based sign language recognition (SLR) models for ArSL using HMMs and Modified KNN. These models were tested on two different datasets consisting of 40 sentences each and both their performances were measured using word classification accuracy and sentence classification accuracy. For word classification accuracy, HMM performed better than MKNN with a maximum accuracy of 97% whereas for sentence classification accuracy, MKNN performed better with a maximum accuracy of 97%. The paper was further extended in Hassan et al. [24]. In the paper, three datasets are used, each consisting of 40 sentences. The datasets are collected from different sensors, that is Polhemus G4 tracker, camera, and glove-based sensors. The study used two different recognition techniques namely HMM and modified KNN. Their results showed sensor-based data was more precise as compared to visual data with motion trackers achieving higher recognition raters than glove-based sensors. Additionally, their results also showed modified KNN as performing better than the

HMMs for sentence recognition with the maximum accuracy achieved around 97%. On the other hand, HMMs toolkit of RASR performed better than modified KNN for word recognition with the maximum accuracy achieved close to 99%. Tubaiz et al. [26] proposed a modified KNN approach for ArSL recognition using two sensor-based gloves. Their model was trained on a dataset containing 40 sentences and scored an accuracy of 98.9%. Assaleh et al. [39] investigated a vision based ArSL recognition system using motion from the signing videos and HMM. Their results achieved an average word recognition rate of 94% and an average sentence recognition rate of 75%.

III. JUMLA-QSL-22 DATASET

A. PARTICIPANTS

The dataset was collected through seven participants invited through the snowballing approach. The participants were informed beforehand on the data collection process and written consent was acquired from them. Participation in the dataset collection was voluntary and participants were free to leave the collection process at any time. Of the seven participants, five were hearing-impaired individuals whereas two participants were sign interpreters. All participants were male with a mean age of 35.85 years ($SD = 7.47$ years). The participants were a mix of Qataris and non-Qataris however, all were fluent in Qatari sign language.

B. SENTENCE SELECTION

The first step of sentence selection involved the establishment of intents. Intents are a systematic framework for organizing and categorizing sentences within the corpus. To achieve this, we collaborated with a primary healthcare center to identify frequently used phrases, questions, and expressions in healthcare facilities, which allowed us to define 22 distinct intents, as outlined in Table 2. Each intent was then expanded upon with multiple sentences, resulting in a total collection of 900 sentences, forming the fundamental basis of the developed corpus.

TABLE 2. List of intents.

Intent Code	Arabic Description	English Translation	Number of sentences
GREETING	التحية	Greetings	47
CORONA_TEST	فحص كورونا	Corona Test	28
DOCTOR_REQUEST	طلب مقابلة طبيب	Request a doctor consultation	86
NEGATION	النفي	Negative	79
AFFIRMATION	التأكيد	Confirmation	110
FAMILY_DOCTOR	طبيب الأسرة	Family doctor	28
EVENING_REQUEST	طلب موعد مسائي	Request an evening appointment	66
FRIDAY	يوم الجمعة	Friday	5
APPOINTMENT_REQUEST	طلب حجز موعد	Request an appointement	34
TEETH_DOCTOR	طبيب الأسنان	Dentist	22
SUNDAY	يوم الأحد	Sunday	35
PSYCHIATRIST	طبيب نفسي	Psychiatrist	32
MONDAY	يوم الاثنين	Monday	30
ANALYSIS_REQUEST	طلب موعد تحاليل	Request an Analysis appointment	58
MORNING	الصباح	In the morning	48
TUESDAY	يوم الثلاثاء	Tuesday	30
WEDNESDAY	يوم الأربعاء	Wednesday	28
THURSDAY	يوم الخميس	Thursday	27
SATURDAY	يوم السبت	Saturday	9
CERTIFICATION_REQUEST	طلب شهادة طبية	Request a medical certificate	35
REPEAT	إعادة	Repetition	51
TODAY	اليوم	Today	13

C. DATA COLLECTION

The data collection process was initiated by inviting the seven participants to a dedicated studio one at a time. The studio was equipped with a green screen and the participants were asked to stand at a point marked 210 cm from the left and right cameras and 297 cm from the front camera (Figure 1). The room was well-lit with adequate lighting and no restriction was placed on the clothing of the participants if it did not interfere with the data collection process. This helped to reduce dependencies errors that may arise in the dataset from environmental factors. The data collection period lasted for five months starting from August 2022 to December 2022.

The data collection setup also included four true depth cameras situated in four different locations from the participants. True depth cameras are specialized cameras that not only capture the visual RGB information of images but also capture the depth information allowing for a more accurate representation of the 3D space. These cameras were strategically placed at the top, front, left and right angle of the participant to capture comprehensive and detailed information regarding the signing movements done. The camera types used were two Intel Realsense cameras (for left and right angles) and two Zed2I cameras (for front and top). The four cameras were synchronized to ensure that data from the different angles would be recorded at the same time. Additionally, the participants were also recorded at the same distance from the camera from all four angles. During the recording process, each participant was given a random sentence from the list of 900 sentences and then signaled by one of the researchers to inform them that the recording has been started. The participants would then sign the sentence and the recording would be stopped. The data captured was evaluated by two expert QSL signers. One of these QSL experts was an internal deaf signer whereas the other was an external non-deaf interpreter. The internal evaluator first evaluated the recordings before passing them on to the external evaluator, ensuring high levels of quality checks and accuracy.

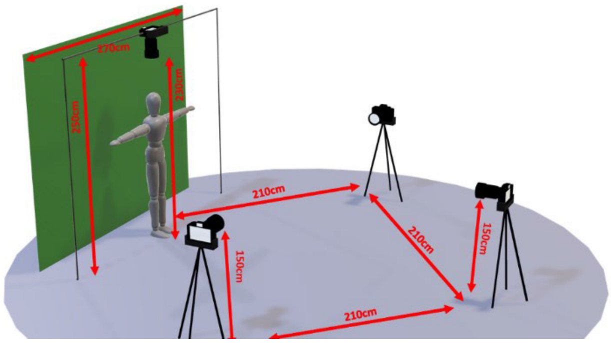

FIGURE 1. Studio setup.

IV. DATA ACQUISITION PROCESS

To optimize the recording process, we developed a specialized software tailored to seamlessly control multiple cameras. This software not only ensures synchronization but also guarantees the simultaneous activation and deactivation of recordings across all four cameras. In a bid to replicate a real-world scenario, we integrated a chatbot. This chatbot is designed to emulate a typical conversation that might occur between a signer and a reception agent at a healthcare center, providing a more authentic and relevant context for our recordings. Our data recording process is meticulous and systematic. It begins with the signer initiating a dialogue with the chatbot. We then capture each of the signer's responses in isolation, ensuring clarity and precision in the data collection. Recognizing the importance of a diverse dataset, we further instruct the signer to vary their answers. This strategy allows us to obtain a rich variety of sentences that, while different in structure and phrasing, still convey the same fundamental intent. Through these methods, we aim to build a comprehensive and robust database that can cater to a wide range of scenarios and applications.

During the recording session, either a deaf individual or a sign language interpreter oversees the process. This supervisor plays a crucial role, ensuring clarity and accuracy in the signer's expressions. They have the authority to intervene at any moment to point out any ambiguities or mistakes to the signer. Their presence is vital in maintaining the integrity and quality of the recorded data, as they ensure that the signer's gestures are clear, correct, and free from potential misinterpretations.

A. DATASET STATISTICS

The JUMLA-QSL-22 dataset comprises of 6300 records of the seven participants signing 900 sentences each. The dataset is publicly available at IEEE Dataport [37]. Table 3 shows the statistics of the dataset.

TABLE 3. Statistics of the dataset.

Parameter	Range
Number of signers	7
Number of sentences	900 per signer
Vocabulary size (intent)	22
Total number of files	25,200
Total duration of recorded videos	≈ 28 hours (average)

The sentences used are primarily focused on the healthcare setting and contain phrases such as فيروس كورونا فحص (English: 'Coronavirus test'), أريد طبيب رؤية (English: 'I want to see a doctor'), and موعد لا مافيه (English: 'No date'). Each participant was giving 900 sentences to sign one at a time however in a random order to reduce order bias in the data collection. Therefore, each participant has a unique sentence coded list in the dataset. Table 4 shows a sample of the coded sentences in the dataset.

The dataset contains videos from four different angles that is front, top, left and right for each signed sentence. Figure 2 shows a sample of the four different angles of a participant signing a given sentence. The videos are stored in two different formats where the front and top-angle videos have a resolution of 2560 × 720 pixels and a frame rate of 60 fps. Further, the left and right-angle videos have a resolution of 640 × 480 pixels and a frame rate of 60 fps.

To account for variations in sign language signing, we invited multiple signers to perform the same sentences. These variations encompassed differences in signing speed, signing order, and the dominant hand used for signing. Some signers exhibited faster signing than others, leading to slight mixing of signs within the sentence. Additionally, the order in which signers performed the same sentence varied among individuals. Moreover, signers had a dominant hand preference for signing, which could be either the right or left hand, adding further diversity. To ensure the dataset captured these variations comprehensively, we annotated the videos with the identified differences and had each sentence repeated three times by the same signer every two weeks. Details on the annotation process are mentioned in [38]. It is important to note that even with the same signer, the same sentence might be signed differently after the two-week interval, contributing to the richness and complexity of the dataset.

V. DATA ANALYSIS

Analyzing the data within the corpus can help identify the different trends and patterns within the corpus records. As such, we conducted data analysis on the JUMLA-QSL-22 corpus. We first collected statistics on the frequency of frames within each record file. Figure 3 shows the bar chart on the frequency of frames with the average frames between 225-250 frames. Using the bar chart, we can determine the cutoff range for upper limits. We can observe from the chart that the shortest videos contain about 125 frames, equating to

TABLE 4. Sample of coded sentences in the dataset.

Code	Intent	Qatari sign language phrases	Arabic language sentences	English translation
3	GREETING	مرحبا	مرحبا	Welcome
4	DOCTOR_REQUEST	أريد طبيب رؤية	أريد رؤية الطبيب	I want to see a doctor
5	NEGATION	موعد لا مافيه	لا يوجد موعد متاح	No date
6	AFFIRMATION	أريد موافق	أنا موافق	I want to accept
7	FAMILY_DOCTOR	أسرة طبيب	طبيب الأسرة	Family doctor
8	EVENING_REQUEST	أريد مساء	أريد تحديد موعد في المساء	I want evening
9	FRIDAY	يوم الجمعة	الجمعة	Friday

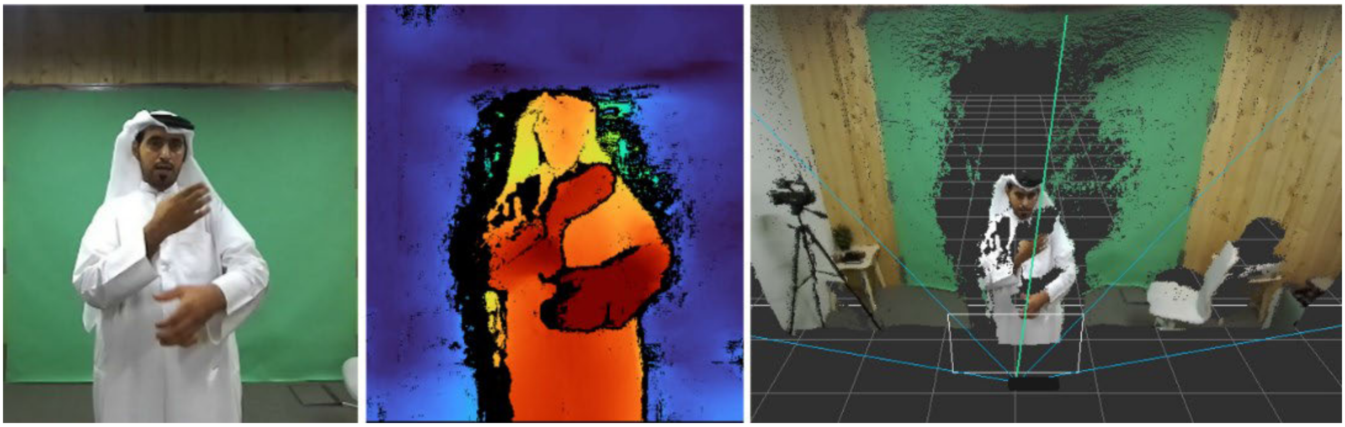

FIGURE 2. Different types of video data available in JUMLA-QSL-22 dataset.

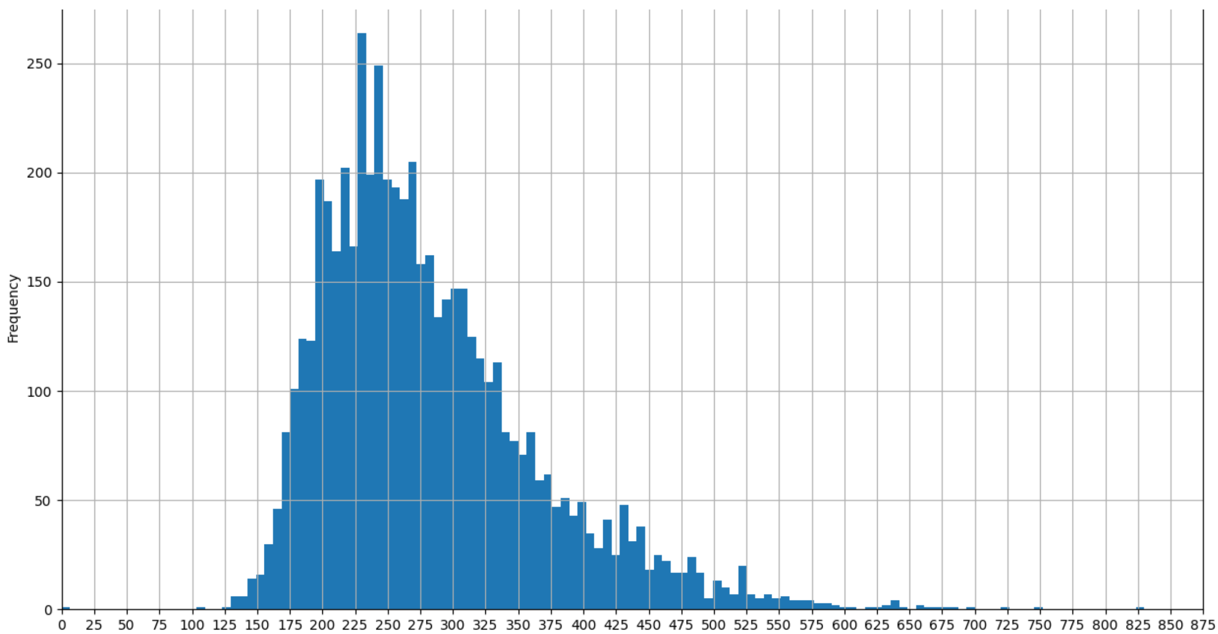

FIGURE 3. Number of frames per record.

approximately 2 seconds, whereas the longest videos exceed 600 frames, or about 10 seconds.

The lips of the participants are represented by 12 landmark coordinates. Figure 4a shows the x-coordinates of the movements of the participants' lips when signing during the recordings while Figure 4b shows the y-coordinates motion. Figure 4c shows the corresponding 12 landmark positions. Furthermore, the x-coordinates motion ranges between

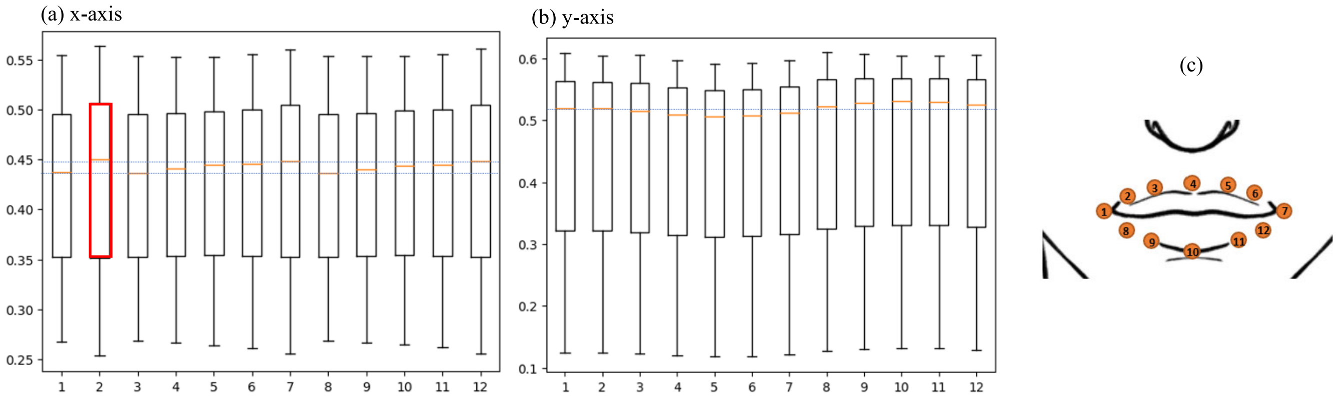

FIGURE 4. **(a) Represents the x-coordinates of the participants' lips movements and (b) represents the y-coordinates of participants' lips movements when signing. (c) Position of the different landmarks on lips.**

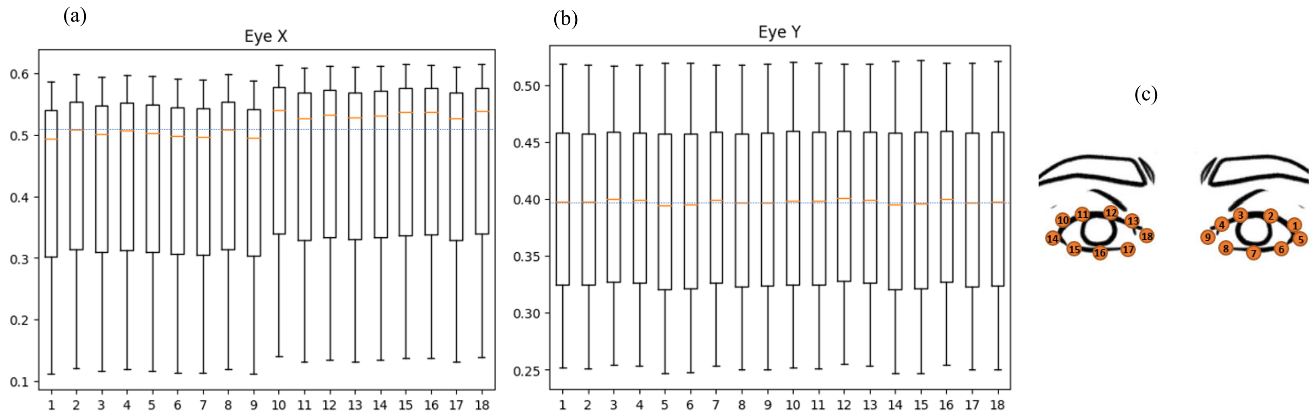

FIGURE 5. **(a) Represents the x-coordinates of the participants' eyes movements and (b) represents the y-coordinates of participants' eyes movements when signing. (c) Position of the different landmarks on both eyes.**

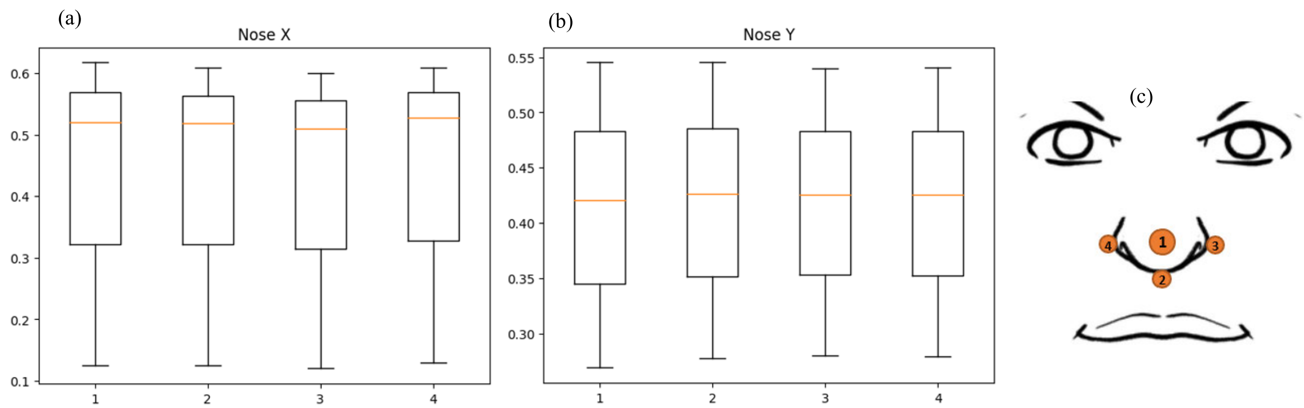

FIGURE 6. **(a) Represents the x-coordinates of the noses of participants and (b) represents the y-coordinates of noses of participants when signing. (c) Position of the different landmarks on the nose.**

0.26 and 0.57 whereas the y-coordinates motion ranges between 0.12 and 0.62. We can observe from Figure 4a that the x-coordinate of leap point 2 is close to point 7, even though it should be positioned on the opposite side of the mouth, closer to point 1. This anomaly can be categorized as noise. Upon examining Figure 4b, we can observe that point 7, which should share the same y-value as point 1, actually has an average value slightly lower than anticipated. This discrepancy can likely be attributed to accuracy reasons related the tracking algorithm.

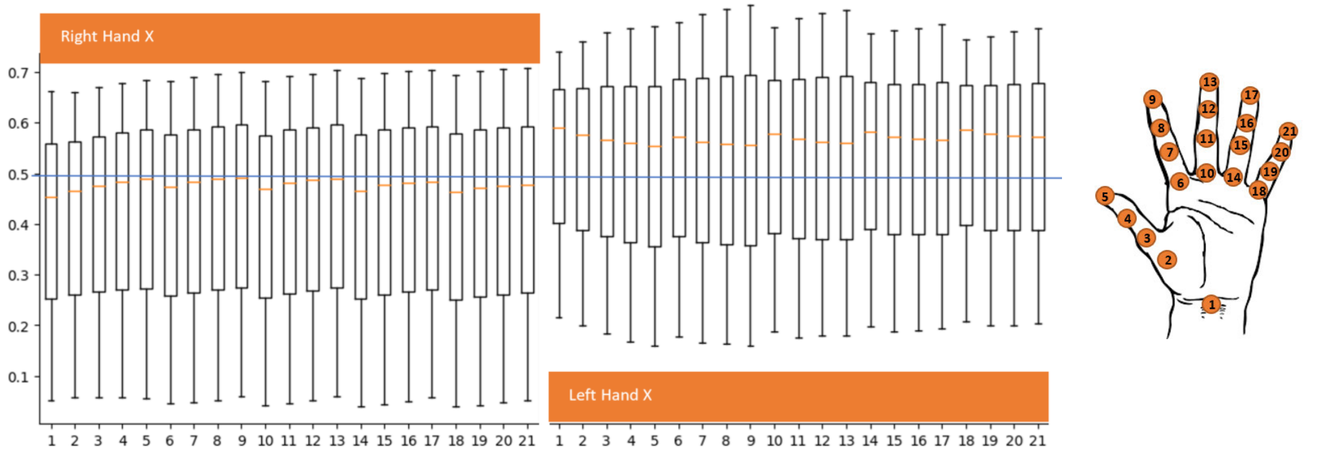

FIGURE 7. (a) Represents the x-coordinates of the participants' left-hand movements and (b) represents the y-coordinates of participants' left-hand movements when signing. (c) Position of the different landmarks on the left hand.

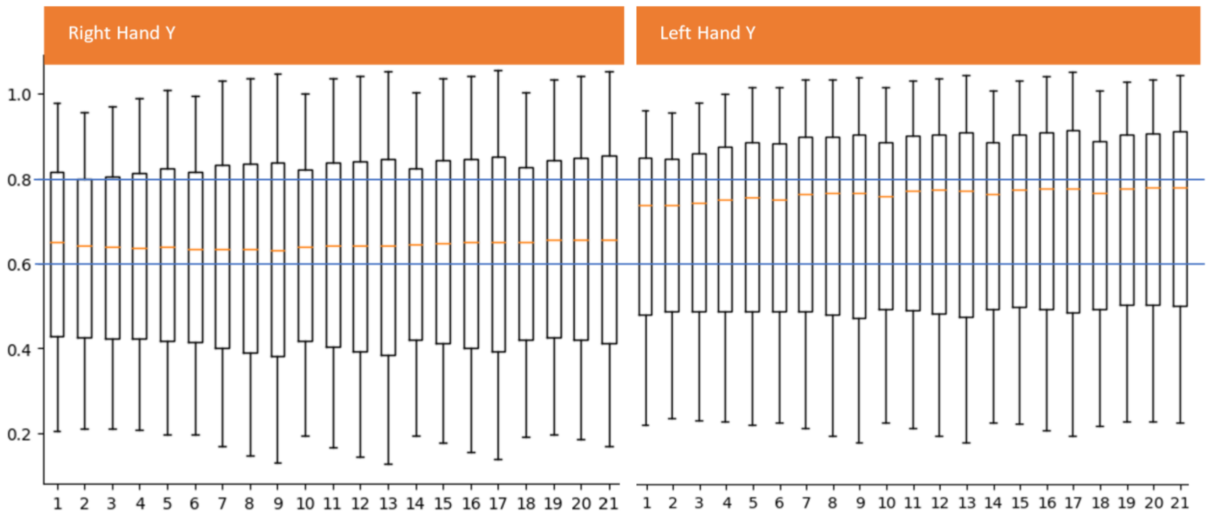

FIGURE 8. (a) Represents the y-coordinates of the participants' right-hand movements and (b) represents the y-coordinates of participants' right-hand movements when signing.

Figure 5a shows the x-coordinates of the participants' eyes when signing. The x-axis represents the landmarks of both eyes with point 1 to point 9 representing the left eye and point 10 to point 18 representing the right eye. The left eye movements range between 0.12 to 0.60 whereas the right eye movements in the x-plane are from 0.14 to 0.63. Figure 5b shows the y-coordinates of the participants' eyes when signing. Similar to the x-place, the x-axis represents the landmarks of both eyes with point 1 to point 9 representing the left eye and point 10 to point 18 representing the right eye. The left and right eye movements range between 0.25 and 0.55 in the y-plane. Figure 5c shows the corresponding landmark positions of both eyes.

Figure 6a shows the x-coordinates of the movement of the noses of the participants when signing. The movement in the x-plane ranges between 0.12 and 0.63. Figure 6b shows the y-coordinates of the participants' nose movements with the y-plane motion ranging between 0.23 and 0.54. Figure 6c shows the corresponding landmark positions on the nose. Nose landmarks are among the most stable features on the face. They can provide insights into head movements. Upon examining Figure 6a, we observe that the nose landmarks are predominantly situated in a central position. However, in some instances, the x-value is close to 0.1. This can be attributed to the signer looking slightly to the left. The reason for this orientation is that during the recording session, deaf individuals were seated in that direction to monitor each recorded sentence. The same observation can be noted in both eye and leap charts.

Figure 7a shows the x-coordinates of the movement of the left hand of the participants when signing. The movement in the x-plane ranges between 0.05 and 0.7. Figure 7b shows the y-coordinates of the movement of the left hand of the participants when signing. The movement in the y-plane ranges between 0.15 and 0.9. Figure 7c shows the corresponding landmark positions on the left hand. The right hand is positioned more towards the center of the x-axis compared to the left hand. This implies that in most cases, the right hand is the dominant one.

Figure 8a shows the x-coordinates of the right-hand motion of the participants when signing. The movement in the x-plane ranges between 0.1 and 1. Figure 8b shows the y-coordinates of the left-hand motion of the participants. The movement in the y-plane ranges between 0.2 and 1. Furthermore, we notice that the left hand has a slightly higher value of motion in the y-plane compared to the right hand as it is located in lower position according to the origin located in the top-left, which may imply that the right hand is the more dominant hand when signing.

VI. CONCLUSION

In conclusion, the development of sign language recognition systems holds immense potential for transforming the lives of individuals with hearing disabilities. As demonstrated through the introduction of the JUMLA-QSL-22 dataset, publicly available in IEEE Dataport [37], creating comprehensive and standardized language resources is a critical step toward advancing sign language processing technology. The presented dataset not only captures variations in signing styles but also accounts for environmental factors and linguistic intricacies, enhancing the authenticity and reliability of the collected data. The successful creation of the JUMLA-QSL-22 dataset serves as a foundation for future research in sign language recognition, facilitating the development of advanced systems that can seamlessly interpret and translate sign language conversations. With the integration of machine learning and artificial intelligence, these systems have the potential to bridge the communication gap between hearing-impaired individuals and the wider society, enabling smoother interactions in personal, work, and educational contexts. As the field of sign language technology continues to evolve, efforts like the JUMLA-QSL-22 dataset contribute significantly to fostering inclusivity and accessibility for the deaf community, paving the way for a more inclusive and interconnected future.

REFERENCES

[1] World Health Organization. *Deafness and Hearing Loss*. Accessed: Sep. 9, 2023. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss

[2] H. Luqman and E.-S.-M. El-Alfy, "Towards hybrid multimodal manual and non-manual Arabic sign language recognition: MArSL database and pilot study," *Electronics*, vol. 10, no. 14, p. 1739, Jul. 2021, doi: 10.3390/electronics10141739.

[3] K. M. Lim, A. W. C. Tan, C. P. Lee, and S. C. Tan, "Isolated sign language recognition using convolutional neural network hand modelling and hand energy image," *Multimedia Tools Appl.*, vol. 78, no. 14, pp. 19917–19944, Jul. 2019, doi: 10.1007/s11042-019-7263-7.

[4] United Nations. (2022). *International Day of Sign Languages | United Nations*. [Online]. Available: https://www.un.org/en/observances/sign-languages-day

[5] Y. O. M. Elhadj, Z. Zemirli, and K. Ayyadi, "Development of a bilingual parallel corpus of Arabic and saudi sign language: Part I," in *Intelligent Informatics* (Advances in Intelligent Systems and Computing), vol. 182, A. Abraham and S. M. Thampi, Eds. Berlin, Germany: Springer, 2013, pp. 285–295, doi: 10.1007/978-3-642-32063-7_30.

[6] H. Luqman and S. A. Mahmoud, "Automatic translation of Arabic text-to-Arabic sign language," *Universal Access Inf. Soc.*, vol. 18, no. 4, pp. 939–951, 2018, doi: 10.1007/s10209-018-0622-8.

[7] K. Al-Fityani and C. Padden, "Sign languages in the Arab world," in *Sign Languages* (Cambridge Language Surveys), D. Brentari, Ed. Cambridge, U.K.: Cambridge Univ. Press, 2010, pp. 433–450, doi: 10.1017/CBO9780511712203.020.

[8] G. Batnasan, M. Gochoo, M.-E. Otgonbold, F. Alnajjar, and T. K. Shih, "ArSL21L: Arabic sign language letter dataset benchmarking and an educational avatar for metaverse applications," in *Proc. IEEE Global Eng. Educ. Conf. (EDUCON)*, Mar. 2022, pp. 1814–1821, doi: 10.1109/educon52537.2022.9766497.

[9] M. Hassan, K. Assaleh, and T. Shanableh, "User-dependent sign language recognition using motion detection," in *Proc. Int. Conf. Comput. Sci. Comput. Intell. (CSCI)*, Dec. 2016, pp. 852–856, doi: 10.1109/csci.2016.0165.

[10] R. Wolfe et al., "The myth of signing avatars," in *Proc. 1st Int. Workshop Autom. Transl. Signed Spoken Lang.*, 2021, pp. 33–42.

[11] R. Wolfe, "Special issue: Sign language translation and avatar technology," *Mach. Transl.*, vol. 35, no. 3, pp. 301–304, 2021, doi: 10.1007/S10590-021-09270-4.

[12] E.-S.-M. El-Alfy and H. Luqman, "A comprehensive survey and taxonomy of sign language research," *Eng. Appl. Artif. Intell.*, vol. 114, Sep. 2022, Art. no. 105198, doi: 10.1016/J.ENGAPPAI.2022.105198.

[13] V. Athitsos, C. Neidle, S. Sclaroff, J. Nash, A. Stefan, Q. Yuan, and A. Thangali, "The American sign language lexicon video dataset," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2008, pp. 1–8, doi: 10.1109/cvprw.2008.4563181.

[14] H. V. Joze, and O. Koller, "MS-ASL: A large-scale data set and benchmark for understanding American sign language," in *Proc. Brit. Mach. Vis. Conf. (BMVC)*, 2019. [Online]. Available: https://www.microsoft.com/en-us/research/publication/ms-asl-a-large-scale-data-set-and-benchmark-for-understanding-american-sign-language/

[15] H. Cooper, E. J. Ong, N. Pugeault, and R. Bowden, "Sign language recognition using sub-units," in *Gesture Recognition* (The Springer Series on Challenges in Machine Learning), S. Escalera, I. Guyon, and V. Athitsos, Eds. Cham, Switzerland: Springer, 2017, doi: 10.1007/978-3-319-57021-1_3.

[16] A. Kurakin, Z. Zhang, and Z. Liu, "A real time system for dynamic hand gesture recognition with a depth sensor," in *Proc. 20th Eur. Signal Process. Conf. (EUSIPCO)*, Aug. 2012, pp. 1975–1979. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/6333871/

[17] M. Oszust and M. Wysocki, "Polish sign language words recognition with Kinect," in *Proc. 6th Int. Conf. Hum. Syst. Interact. (HSI)*, Jun. 2013, pp. 219–226, doi: 10.1109/hsi.2013.6577826.

[18] O. Koller, J. Forster, and H. Ney, "Continuous sign language recognition: Towards large vocabulary statistical recognition systems handling multiple signers," *Comput. Vis. Image Understand.*, vol. 141, pp. 108–125, Dec. 2015, doi: 10.1016/J.CVIU.2015.09.013.

[19] M. M. Kamruzzaman, "Arabic sign language recognition and generating Arabic speech using convolutional neural network," *Wireless Commun. Mobile Comput.*, vol. 2020, pp. 1–9, May 2020, doi: 10.1155/2020/3685614.

[20] G. Latif, N. Mohammad, J. Alghazo, R. AlKhalaf, and R. AlKhalaf, "ArASL: Arabic alphabets sign language dataset," *Data Brief*, vol. 23, Apr. 2019, Art. no. 103777, doi: 10.1016/j.dib.2019.103777.

[21] S. M. Shohieb, H. K. Elminir, and A. M. Riad, "SignsWorld atlas; a benchmark Arabic sign language database," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 27, no. 1, pp. 68–76, Jan. 2015, doi: 10.1016/j.jksuci.2014.03.011.

[22] M. M. Balaha, S. El-Kady, H. M. Balaha, M. Salama, E. Emad, M. Hassan, and M. M. Saafan, "A vision-based deep learning approach for independent-users Arabic sign language interpretation," *Multimedia Tools Appl.*, vol. 82, no. 5, pp. 6807–6826, Feb. 2023, doi: 10.1007/s11042-022-13423-9.

[23] S. Abbas, H. Al-Barhamtoshy, and F. Alotaibi, "Towards an Arabic sign language (ArSL) corpus for deaf drivers," *PeerJ Comput. Sci.*, vol. 7, p. e741, Nov. 2021, doi: 10.7717/peerj-cs.741.

[24] M. Hassan, K. Assaleh, and T. Shanableh, "Multiple proposals for continuous Arabic sign language recognition," *Sens. Imag.*, vol. 20, no. 1, pp. 1–23, Dec. 2019, doi: 10.1007/s11220-019-0225-3.

[25] A. A. I. Sidig, H. Luqman, S. Mahmoud, and M. Mohandes, "KArSL: Arabic sign language database," *ACM Trans. Asian Low-Resource Lang. Inf. Process.*, vol. 20, no. 1, pp. 1–19, Jan. 2021, doi: 10.1145/3423420.
[26] N. Tubaiz, T. Shanableh, and K. Assaleh, "Glove-based continuous Arabic sign language recognition in user-dependent mode," *IEEE Trans. Human-Mach. Syst.*, vol. 45, no. 4, pp. 526–533, Aug. 2015, doi: 10.1109/thms.2015.2406692.
[27] O. Al-Jarrah and A. Halawani, "Recognition of gestures in Arabic sign language using neuro-fuzzy systems," *Artif. Intell.*, vol. 133, nos. 1–2, pp. 117–138, 2001, doi: 10.1016/s0004-3702(01)00141-2.
[28] K. Assaleh and M. Al-Rousan, "Recognition of Arabic sign language alphabet using polynomial classifiers," *EURASIP J. Adv. Signal Process.*, vol. 2005, no. 13, pp. 2136–2145, Dec. 2005, doi: 10.1155/asp.2005.2136.
[29] M. Mohandes, S. Aliyu, and M. Deriche, "Arabic sign language recognition using the leap motion controller," in *Proc. IEEE 23rd Int. Symp. Ind. Electron. (ISIE)*, Jun. 2014, pp. 960–965, doi: 10.1109/isie.2014.6864742.
[30] A. Tharwat, T. Gaber, A. E. Hassanien, M. K. Shahin, and B. Refaat, "Sift-based Arabic sign language recognition system," in *Proc. Afro-Eur. Conf. Ind. Adv.*, in Advances in Intelligent Systems and Computing, vol. 334, 2015, pp. 359–370, doi: 10.1007/978-3-319-13572-4_30.
[31] K. Assaleh, T. Shanableh, and M. Zourob, "Low Complexity Classification system for glove-based Arabic sign language recognition," in *Neural Information Processing*, 2012, pp. 262–268, doi: 10.1007/978-3-642-34487-9_32.
[32] N. B. Ibrahim, M. M. Selim, and H. H. Zayed, "An automatic Arabic sign language recognition system (ArSLRS)," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 30, no. 4, pp. 470–477, Oct. 2018, doi: 10.1016/j.jksuci.2017.09.007.
[33] A. A. I. Sidig, H. Luqman, and S. A. Mahmoud, "Transform-based Arabic sign language recognition," *Proc. Comput. Sci.*, vol. 117, pp. 2–9, 2017, doi: 10.1016/j.procs.2017.10.087.
[34] M. Elpeltagy, M. Abdelwahab, M. E. Hussein, A. Shoukry, A. Shoala, and M. Galal, "Multi-modality-based Arabic sign language recognition," *IET Comput. Vis.*, vol. 12, no. 7, pp. 1031–1039, Oct. 2018, doi: 10.1049/IET-CVI.2017.0598.
[35] B. Hisham, and A. Hamouda, "Arabic sign language recognition using Ada-Boosting based on a leap motion controller," *Int. J. Inf. Technol.*, vol. 13, no. 3, pp. 1221–1234, 2021, doi: 10.1109/AINAW.2007.98.
[36] M. Mohandes, S. I. Quadri, and M. Deriche, "Arabic sign language recognition an image-based approach," in *Proc. 21st Int. Conf. Adv. Inf. Netw. Appl. Workshops (AINAW)*, May 2007, pp. 272–276, doi: 10.1109/ainaw.2007.98.
[37] O. El Ghoul, A. Othman, M. Aziz, and S. Sedrati, "JUMLA-QSL-22: A dataset of Qatari sign language sentences," IEEE Dataport, Oct. 2022, doi: 10.21227/ckzp-3754.
[38] A. Othman, O. El Ghoul, M. Aziz, K. Chemnad, S. Sedrati, and A. Dhouib, "JUMLA-QSL-22: Creation and annotation of a Qatari sign language corpus for sign language processing," in *Proc. 16th Int. Conf. Pervasive Technol. Rel. Assistive Environ.*, Jul. 2023, pp. 686–692, doi: 10.1145/3594806.3596525.
[39] K. Assaleh, T. Shanableh, M. Fanaswala, H. Bajaj, and F. Amin, "Vision-based system for continuous Arabic sign language recognition in user dependent mode," in *Proc. 5th Int. Symp. Mechatronics Appl.*, 2008, doi: 10.1109/ISMA.2008.4648841.
[40] L. Naert, C. Larboulette, and S. Gibet, "A survey on the animation of signing avatars: From sign representation to utterance synthesis," *Comput. Graph.*, vol. 92, pp. 76–98, 2020, doi: 10.1016/j.cag.2020.09.003.
[41] W. Suliman, M. Deriche, H. Luqman, and M. Mohandes, "Arabic sign language recognition using deep machine learning," in *Proc. 4th Int. Symp. Adv. Electr. Commun. Technol. (ISAECT)*, Dec. 2021, pp. 1–4, doi: 10.1109/isaect53699.2021.9668405.

OUSSAMA EL GHOUL was an Assistant Professor with the University of Tunis, from 2008 to 2012. He was also the Project Manager with the Arab League Educational, Cultural and Scientific Organization, from 2013 to 2019. He has been an ICT Access Research Specialist with the Mada, Qatar Assistive Technology Center, Qatar, since 2019. He is a member of the Research Laboratory of Technologies of Information and Communication and Electrical Engineering (LaTICE). At Mada, he is currently leading two research projects related to sign language translation and recognition. He is also a member of the team that received several awards, including the Silver Medal of the International Fair of Inventions, Kuwait, in 2007, and the UNESCO Prize 2008 for the e-learning curriculum they developed for the visually impaired, the "World Summit Award (WSA)–Mobile 2010" in the field of social inclusion. He contributed to the project "ALECSO APPS" that received the Mohammed bin Rashid Arabic Language Award (in the Technology Topic), in May 2017.

MARYAM AZIZ received the B.S. degree in computer engineering from Qatar University, Doha, Qatar, in 2020, and the M.S. degree in data science and engineering from Hamad Bin Khalifa University, Doha, in 2022. She was a Research Assistant with the Mada, Qatar Assistive Technology Center, Qatar. Her research interests include natural language processing and digital health with a focus on well-being, mental health, and behavioral sciences.

ACHRAF OTHMAN (Senior Member, IEEE) received the M.Sc. degree in computer science from Tunis University and the Ph.D. degree in computer science from the University of Sfax. Since 2010, he has been an Assistant Professor of computer science at two universities, where he has taught computer science courses. He is currently the Head of the Innovation and Research Section, Mada, Qatar Assistive Technology Center, Doha, Qatar, where he works on complex IT industry challenges related to assistive technology and digital accessibility, building capabilities, and training, and researching innovative projects that use cutting-edge technologies, NLP, and artificial intelligence to empower persons with disabilities. He is also affiliated as a Researcher with the Research Laboratory LaTICE, Tunis University. He has supervised several bachelor's and master's degree students. He was an international consultant on several global and regional IT projects.