Journal of King Saud University – Computer and Information Sciences 35 (2023) 101642

HOSTED BY

ELSEVIER

Contents lists available at ScienceDirect

Journal of King Saud University – Computer and Information Sciences

journal homepage: www.sciencedirect.com

Facilitating the communication with deaf people: Building a largest Saudi sign language dataset

Mansour Alsulaiman [a,b,*], Mohammed Faisal [a,c,*], Mohamed Mekhtiche [a,b], Mohamed Bencherif [a,b], Tariq Alrayes [d], Ghulam Muhammad [a,b], Hassan Mathkour [a,f], Wadood Abdul [a,b], Yousef Alohali [a,f], Mansour Alqahtani [e], Habib Al-Habib [e], Hassan Alhalafi [e], Mohammed Algabri [a,g], Muneer Al-hammadi [a,h], Hamdi Altaheri [a,b], Taha Alfakih [a]

[a] Center of Smart Robotics Research, College of Computer and Information Science, King Saud University, Saudi Arabia
[b] Computer Engineering Department, College of Computer and Information Science, King Saud University, Riyadh 1543, Saudi Arabia
[c] Center of AI & Robotics, Kuwait College of Science and Technology (KCST), Kuwait
[d] Department of Special Education, College of Education, King Saud University, Riyadh, Saudi Arabia
[e] Higher Education Program for Deaf and Hard-of-Hearing Students, King Saud University, Riyadh, Saudi Arabia
[f] Computer Science Department, College of Computer and Information Science, King Saud University, Saudi Arabia
[g] Computer Science and Information Systems Department, College of Applied Sciences, AlMaarefa University, Riyadh 13713, Saudi Arabia
[h] Department of Civil and Environmental Engineering, Faculty of Engineering, Norwegian University of Science and Technology, Høgskoleringen 1, 7034 Trondheim, Norway

A R T I C L E I N F O

Article history:
Received 28 February 2023
Revised 11 June 2023
Accepted 27 June 2023
Available online 1 July 2023

Keywords:
Saudi Dialect
Saudi sign language
Sign language dataset
Arabic sign language

A B S T R A C T

Recently, several countries have been trying hard to facilitate the integration of disabled people into their societies by ensuring equal opportunities through ease of access to social services, daily human necessities, and the labor market. Deafness is considered one of the major disabilities separating the deaf from their society. To integrate the deaf fully into society, a two-way mode of communication is required: one from the deaf to the hearing people, and the other from the hearing to the deaf. Communication from the hearing person to the deaf is generally easy and can be done through speech recognition and text-to-sign representations, but communication from the deaf to the hearing is somewhat difficult and requires a sign recognition module that recognizes the sign motions from the deaf and translates it to a text; following this, a speech synthesis module will translate this text to speech. To build the sign recognition module, a sign language dataset is required. This paper contributes to the literature by introducing a comprehensive survey of 17 Arabic sign language datasets and by developing a well-organized framework that is used to build a sign language dataset. This paper also contributes to the literature by creating the largest Saudi Sign Language (SSL) database—the King Saud University Saudi-SSL (KSU-SSL database)—with 293 signs, 33 signers, 145,035 samples, and 10 domains (healthcare, common, alpha-bets, verbs, pronouns and adverbs, numbers, days, kings, family, and regions). This paper also contributes to the literature by introducing a convolutional graph neural network (CGCN) architecture for sign language recognition and applying the proposed architecture to the built KSU-SSL database. The architecture is made up of a small number of separable 3DGCN layers, and is augmented with a spatial attention mechanism. This study is a part of the project that aims to develop a two-way communication system for Saudi deaf individuals.

1. Introduction

Over one billion people, globally, experience disability according to the WHO (World Health Organization). This accounts for one of seven people who suffer from disability (Pellegrini et al., 2020). The Demographic Survey 2017 (Nicodemo et al., 2021), issued by the General Authority for Statistics – Kingdom of Saudi Arabia KSA, states that 1,810,358 Saudi residents suffer from disabilities, i.e. 7.1 % of the population are disabled. 360 million people over the world have a hearing disability, 9% of them are

* Corresponding authors at: Computer Engineering Department, College of Computer and Information Science, King Saud University, Riyadh 1543, Saudi Arabia.

E-mail address: msuliman@ksu.edu.sa (M. Alsulaiman).

Peer review under responsibility of King Saud University. Production and hosting by Elsevier.

ELSEVIER Production and hosting by Elsevier

https://doi.org/10.1016/j.jksuci.2023.101642
1319-1578/

children, as stated by the WHO (Pellegrini et al., 2020). In KSA, the number of deaf is about 720,000 (Al-Nafjan et al., 2015). Deaf persons have great difficulty in communicating with other people in the society. Only a small number of them know and use sign language to communicate with others. The lack of sign language interpreters (BaHammam et al., 2006, Alhusseini and Alqahtani, 2020, Al-Musharaf et al., 2021) amplifies the difficulty of the deaf in communicating with the rest of society, especially in the government services, specifically in healthcare as stated by this study (Maschendorf Thomaz et al., 2019).

To help the deaf people get involved in the society and integrate with it, we aim to develop a system for a two-way translation of Saudi sign language based on Avatar. This system will be implemented in a carry-on electronic device (laptop, tablet, or mobile) and integrates four basic functions:

- Recognizes the speech of the hearing person and produces the corresponding text.
- Converts the recognized text of the hearing person to sign language and performs the sign by the Avatar. This allows the deaf person to understand the speech of the hearing person.
- Recognizes the sign performed by the deaf person and produces the corresponding text.
- Converts the text of the recognized sign that the deaf person performed into speech. This allows the hearing person to understand the sign of the deaf person.

This allows the hearing person to understand the sign of the deaf person. This system will help the deaf in two ways. First, the system can be carried anywhere they go and can be used to communicate with the rest of society. Second, it can be used to teach sign language to the deaf, especially to children. Communication from the hearing person to the deaf is easy and can be achieved with speech recognition and text-to-sign representations; however, communication from the deaf to the hearing person is somewhat difficult and requires a sign recognition module that recognizes the sign motions from the deaf and translates them to a text, after which a speech synthesis module will translate this text to speech. To build the sign recognition module, a sign language dataset is required; but due to the lack of a public and rich Arabic sign language (ArSL) dataset, many research groups have developed their own local recordings. These local datasets might contain tens or hundreds of signs, but they are domain-specific or contain few signs and variations or are performed by few signers. Overall, these locally recorded signs do not exceed several hundred in number and have only a few signers. After surveying the databases used in research in recognizing Arabic signs and summarizing their findings in a table, we made three observations. First, the number of the signers does not generally exceed a set of tens of signers and some are for less than ten signers; the number of recorded signs varies inversely, and teams were obliged either to increase signers or signs. The second observation involves the variety of recording devices, which starts from simple cameras such as webcams to RGB + depth cameras to cameras using infrared. The third observation is that most datasets include words and short sentences as a sequence or single images cropped around the hands.

The significance of this work comes from its contributions to the literature: First, a comprehensive survey of 17 ArSL datasets is conducted. This survey describes each dataset and its availability. The datasets are categorized by year, number of signers, number of signs, samples, type of signs, and accessibility. This survey can help readers understand the progress of ArSL datasets and the current limitations. Second, this paper introduces a new framework for constructing a sign language dataset used to build a robust and accurate sign language dataset. This framework consists of four phases: the pre-recording phase, the recording phase, the verification phase, and the labeling and archiving phase. Third, this research creates the largest SSL database, the KSU-SSL database. The KSU-SSL dataset was recorded using three types of recording sources: a fast RGB (Red, Green, and Blue) cam-era, an infrared (IR) camera, and a mobile camera. The use of different types of cameras is in-tended to make the database richer and can be used with different applications. This database includes 10 domains and consists of 293 signs, 33 signers, and 145,035 samples. The healthcare domain consists of 133 signs and 65,835 samples; the common domain consists of 39 signs and 19,305 samples; the alphabets domain consists of 37 signs and 18,315 samples; the verbs domain consists of 20 signs and 9900 samples; the pronouns and adverbs domain consists of 18 signs and 8910 samples; the numerical domain consists of 11 signs and 5445 samples; the day domain consists of 11 signs and 5445 samples; the Saudi Kings domain consists of nine signs and 4455 samples; the family domain consists of eight signs and 3960 samples, and the domain of regions consists of 7 signs and 3465 samples. Fourth, this study proposed CGCN architecture for sign language recognition and applied the proposed architecture to the built KSU-SSL database. The proposed architecture is made up of a small number of separable 3DGCN layers, and is augmented with a spatial attention mechanism. By using a limited number of layers, the proposed architecture is able to overcome the common problem of over-smoothing in deep graph neural networks.

Based on the authors' knowledge, this is the first research in the literature that introduces a novel framework for constructing a sign language dataset. If you are interested in the dataset and need access, please email the corresponding author.

This paper is organized as follows. First, a literature review of relevant ArSL is presented. Second, the proposed framework is described. Third, the pre-recording phase, recording phase, verification phase, and the labeling and archiving phase are described. The KSU-SSL database is presented, and all its statistics are discussed. Finally, the proposed architecture (implementation and experiments) are described.

2. Related work

In the literature on the Arabic Sign Language ArSL dataset, several datasets have been built in the past few years. This section surveys 17 ArSL dataset to summarize and compare their key criteria and motivate the contribution of the proposed one in this paper. In 2007, the ArSL dataset reported in (Shanableh and Assaleh, 2007) is an isolated words dataset. It was created by the College of Engineering at the American University of Sharjah. It contains 23 (Table 1) gesture classes performed by three participants, without restriction on image background or clothing. Each participant repeated the gestures 50 times in three sessions. Therefore, there are a total number of 3450 samples in that dataset, distributed evenly among the 23 classes such that each class has 150 samples. An analog camcorder was used to record that dataset.

Another ArSL was introduced by M. Mohandes, S. Quadri, and M. Deriche in 2007 at King Fahd University of Petroleum (Mohandes et al., 2007). It consists of 300 classes selected from the ArSL dictionary. (The exact dictionary is not identified.) This dataset was recorded by a single hard-of-hearing fluent signer, with each sign repeated 15 times, to produce a dataset containing 4,500 samples in total. The video recording camera (Sony DVD Handycam, model no. DCR-DVD 200E) was placed directly in front of the standing signer, and a frame rate of 25 frames per second was maintained throughout the recording. Moreover, a high-intensity halogen light source was used to avoid any shadow effects in the environment. A whitewashed wall background was also set to match the clothes of the signer, thus making it easy to

Table 1
The 23 words of ArSL dataset reported in (Shanableh and Assaleh, 2007).

No.	Word	No.	Word	No.	Word	No.	Word	No.	Word
1	صديقFriend	6	عليكم السلامPeace upon you	11	اهلا وسهلاWelcome	16	اناI/me	21	امسYesterday
2	جارNeighbor	7	ياكلTo eat	12	شكراThank you	17	يسمعTo listen	22	يذهبTo go
3	ضيفGuest	8	ينامTo sleep	13	تفضلCome in	18	يسكتTo stop talking	23	ياتيTo come
4	هديةGift	9	يشربTo drink	14	عيبShame	19	يشمTo smell		
5	عدوEnemy	10	يستيقظTo wake up	15	بيتHouse	20	يساعدTo help		

isolate the hand of the signer from the background based on the hand color. As the authors mentioned in the study, an ArSL expert was also involved in helping with the dataset recording. In 2014, an ArSL image and video dataset called SignsWorld Atlas was introduced by S. M. Shohieb, H. Elminir, and A. Riad (Shohieb et al., 2015). It was developed in Mansoura and Kafr El-Sheikh Universities in Egypt to evaluate their methods for real-time ArSL gesture and posture recognition. The dataset contained handshapes in isolation and in single signs, Arabic alphabets, the numbers, movement in single signs, movement in continuous sentences, lip movement in Arabic sentences, and facial expressions. The dataset was performed by 10 participants with ages ranging from three to 30 years under controlled lighting conditions. The dataset was acquired by a Canon Power Shot A490 digital camera with an image quality of 1024 $\times$ 768 pixels and video samples of 10 MB each. Another ArSL dataset introduced in 2015 (ElBadawy et al., 2015) contains 20 gesture classes performed by two participants. Each participant was asked to repeat the gestures 100 times. The 20 signs consist of five facial expression signs (Table 2)—happy, sad, normal, surprised, and looking up—as well as body + lips motion. The body movement tracking is mainly concerned with both hand positions against the body and shoulder movements. The hand positions are collected and arranged as 16 different positions (4 positions for each hand). The shoulder movements are five movements (left shoulder up, right shoulder up, left shoulder front, right shoulder front, normal).

Another ArSL dataset build by G. Latif et al. is called ArSL2018 (Latif et al., 2019). ArSL2018 is an image-based dataset for 32 Arabic Alphabets (the basic 28 Arabic alphabets and the extended four alphabets ('ال'،'أل'،'ة'،'ئ'. The ArSL2018 consists of 54,049 images in gray scale with 64 $\times$ 64 dimension. The images were collected by recording 40 participants in different lighting conditions and with different backgrounds. This dataset was made publicly available. In the same way as ArSL2018, in (Al-Jarrah and Halawani, 2001) the authors used a dataset of 1800 Gy-scale images for 30 Arabic alphabets in which each sign was recorded by 60 signers. Moreover, in (Albelwi and Alginahi, 2012) AlBelwi et al. used a dataset of 180 Gy-scale images, in which each sign from a total of six signs was repeated 45 times by two signers. In (SamirElons et al., 2013, Guesmi et al., 2016) the authors also proposed an alphabet dataset consisting of 948 and 840 sign simples successive. In (Assaleh et al., 2008) Assaleh et al. present a dataset of ArSL for 80 sentences, in which each sentence was performed 19 times; in total, this dataset consists of 1520 simple signs. Also in (Tolba et al., 2012), the authors used a dataset of 100 sentences for a continuous sentences recognition system. In (Youssif et al., 2011), Youssif et al. used an ArSL dataset for 20 signs, with each sign repeated 45 times; the dataset consists of 900 simple signs for ArSL words. Furthermore, in (Al-Rousan et al., 2009) Al-Rousan et al. used an ArSL dataset consisting of 540 simple signs for 30 Arabic words, in which each word was recorded by 18 volunteers. In (Amin et al., 2015), Amin et al. present an HMM-based translator for ArSL; the dataset used for this translator contains 60 Arabic words recorded 40 times with a Kinect camera; in total, the dataset used in this work consists of 2,400 simple signs. In (Sidig et al., 2021), Sidig et al. presented an ArSL dataset named KArSL (KFUPM ArSL); this dataset contains 502 sign words from many domains such as numbers, letters, family, common verbs, health, and so on. Each sign from the 502 sign words was recorded by three signers and each signer performed each sign 50 times. This dataset, therefore, consists of 75,300 simple signs which is the largest ArSL until now; however, the number of signers is too small, which means this dataset has too little variance and is not good for training models for signer independence.

Furthermore, the ArSL dataset reported in (Alfonse et al., 2015) was established based on an ArSL dictionary approved by the League of Arab States. The total number of sign classes in this dataset is 1,216 signs, including the Arabic alphabets and numbers. The entire dataset was recorded by four sign language experts and reviewed and validated by the other two experts. The four experts who recorded the dataset are a mix of right-handed and left-handed persons. Three different sensors were used to collect the dataset: ordinary HD camera, Kinect 2, and leap motion. The dataset videos were captured from four different viewing angles (0, 270 then 315, and 225), with four different frame rates (5, 10, 30, and 50 frames per second); hence, the database size is 4 $\times$ 1216 $\times$ 4 = 19,456 samples. The samples of this dataset were recorded in different lighting conditions and involve facial expressions in most of the cases. The samples were presented in the form of RGB videos, infrared, and depth maps, as well as the skeleton indexes produced by the lip-motion sensor.

By contrast, the KSU-ASL dataset reported in (Al-Hammadi et al., 2020) and (Bencherif et al., 2021) consists of 80 gesture classes. It was created by the Center of Smart Robotics Research with collaboration from the Higher Education Program for the Deaf and Hard of Hearing at King Saud University. The dataset comprises selected gestures from common ArSL words and expressions. These expressions contain single- handed actions as well as two-handed actions. Forty non-deaf subjects were recorded, and the recordings were attended by one knowledgeable in sign language. Each subject was asked to perform the gestures five times. Different devices such as RGB cameras and Microsoft Kinect were used for recording this dataset. The dataset recording sessions were performed without restrictions in an uncontrolled environment. There were no constraints on the clothing of the participants, the lighting conditions, or the background color. There was also some variation in the distances between the

Table 2
The 20 words of ArSL dataset introduced in (ElBadawy et al., 2015).

امامFront	بعيدFar	بجانبBeside	ياكلEating	يشربDrinking
ارضGround	رياحWind	سماءSky	انتخاباتElections	جوازMarriage
زرButton	ملفFile	نقلMove	اسدLion	نقودMoney
يساعدHelping	طلاقDivorce	بابDoor	بيتHome	يتسوقShopping

Table 3
Existing ArSL datasets.

Ref	Year	signers	signs	Repetition per signer	Total Samples	Type of sign	Devices	Modality	Recorded by a sign language expert	PubliclyAvailable	Country
(Shanableh and Assaleh, 2007)	2007	3	23	50	3450	WordsPhrases	Video camcorder	Upper body	No	No	UAE
(Mohandes et al., 2007)	2007	1	300	15	4500	WordsPhrases	Sony DVD Handy Camera	Upper body	No	No	Saudi Arabia
(Shohieb et al., 2015)	2014	10	500	–	–	AlphabetWordsSentences	Canon Power Shot A490 Camera	Hands + Upper body + Facial expression + Lips motion	No	No	Egypt
(ElBadawy et al., 2015)	2015	2	20	100	4000	Words	Leap Motion + 2 Cameras	Facial expression + Body + Lips motion	No	No	Egypt
(Latif et al., 2019)	2018	40	32	42	54,049	Alphabet	CameraGray Images	Hands	No	Yes	Saudi Arabia
(Al-Jarrah and Halawani, 2001)	2001	–	30	60	1800	Alphabets	Camera	Hands	No	–	–
(Albelwi and Alginahi, 2012)	2012	2	6	15	180	Alphabets	Camera	Hands	–	–	Saudi Arabia
(SamirElons et al., 2013)	2013	–	–	–	948	Alphabets	Camera	–	–	–	–
(Guesmi et al., 2016)	2016	2	28	15	840	Alphabets	Camera	–	–	–	Tunisia
(Assaleh et al., 2008)	2008	1	80	19	1520	Sentences	Camera	–	–	–	UAE
(Tolba et al., 2012)	2012	–	100	–	100	Sentences	Camera	–	–	–	–
(Youssif et al., 2011)	2011	–	20	45	900	Words	Camera	–	–	–	–
(Al-Rousan et al., 2009)	2009	18	30	–	540	Words	Camera	–	–	–	–
(Amin et al., 2015)	2015	1	60	40	2400	Words	Kinect	–	–	–	Egypt
(Sidig et al., 2021)	2021	3	502	50	75,300	Letters, Numbers, Words	Microsoft Kinect V2	–	–	–	Saudi Arabia
(Alfonse et al., 2015)	2015	4	1216	4	19,456	WordsSentences	Leap MotionKinectV2, Digital Cam	Upper body+Facial expression	No	Yes	Egypt
(Al-Hammadi et al., 2020, Bencherif et al., 2021)	2020	40	80	5	16,000	AlphabetNumbersWordsphrases	-Kinect V1-Kinect V2-Sony Handy Cam	Full Body+ Skeleton	No	No	Saudi Arabia
Our database(KSU-SSL)	2022	33	293	5	145,035	AlphabetNumbersWordsFor daily life and medical field	-RGB color camera-infrared camera- Mobile	Upper body	Yes	No	Saudi Arabia

camcorder device and the signers. Because of this restriction-free recording, KSU-SSL is a challenging dataset. In most cases, the signer's hands are blurred and difficult to detect and track. In Table 3 below, we present the comparison between surveyed databases discussed above and our proposed KSU-ArSL.

3. Proposed framework for KSU-SSL dataset

To build a robust and accurate SSL dataset, we started by designing a well-organized framework (Fig. 1). The proposed KSU-SSL database framework consists of four phases: the pre-recording phase, the recording phase, the verification phase, and the labeling and archiving phase. The pre-recording phase is mainly used to prepare for the recording process and consists of five steps: *selecting the signs* (293 signs including letters, numbers, days, family, pronouns, adverbs, verbs, common, Saudi Kings, regions, medical, and common); *types of signs* (static and dynamic signs); *preparing the recording environment*—the studio and cameras (color camera, infrared camera); and *building the recording system*. The recording phase includes three steps: selecting the volunteers, training the volunteers, and recording. The verification phase is used to verify the correction of the performed signs and has two steps of verification: during the sign recording and after the recording. The last phase is the labeling and archiving phase which consists of the proposed protocol to save and archive the verified signs.

4. Pre-Recording phase

The pre-recording phase is the initial phase, used to prepare for the recording environment, and consists of five steps: selecting the signs (293 signs including letters, numbers, days, family, pronouns, adverbs, verbs, common, Saudi Kings, regions, medical, and common); types of signs (static and dynamic signs); preparing the recording environment—studio and cameras (color camera and Infrared camera); and building the recording system.

4.1. Selection of the signs

In this study, we present SSL dataset for the KSA Dialect; therefore, we adhered to the Saudi sign dictionary produced by the Saudi Association for Hearing Impairment (Senkov et al., 2018). There are other variants of the signs used in KSA, and sometimes the deaf do not follow the dictionary; but it is extremely difficult to include all the variants, and we have to follow the most accepted and agreed upon standard, which is the Saudi sign dictionary. The dictionary is subdivided into sections of varying importance, ranging from the name of towns to daily usual words, words from the medical domain, and so on. The sign selection is an important step in order to produce a pilot system that will convey the importance and usefulness of the dataset and be appealing to the deaf and non-deaf community. Moreover, the Saudi sign dictionary contains more than 3,000 signs, and it will be difficult to cover them in this study. Therefore, we focused on certain important domains. Our team (including sign language experts) considered many domains such as government services, education, healthcare, children's stories, courts, and other law institutions). We realized that the medical field is an important area where many deaf people have difficulties explaining their illness and/or pain, and our system can be used immediately and be beneficial. Therefore, we selected the signs from the medical field. In addition to the medical field, we selected other daily life signs that are needed to build a system in the medical field and are also needed in daily life communications. The team has four sign specialists with good contact with the deaf community; hence, these members suggested the initial draft list of the signs. This list was discussed and refined many times until

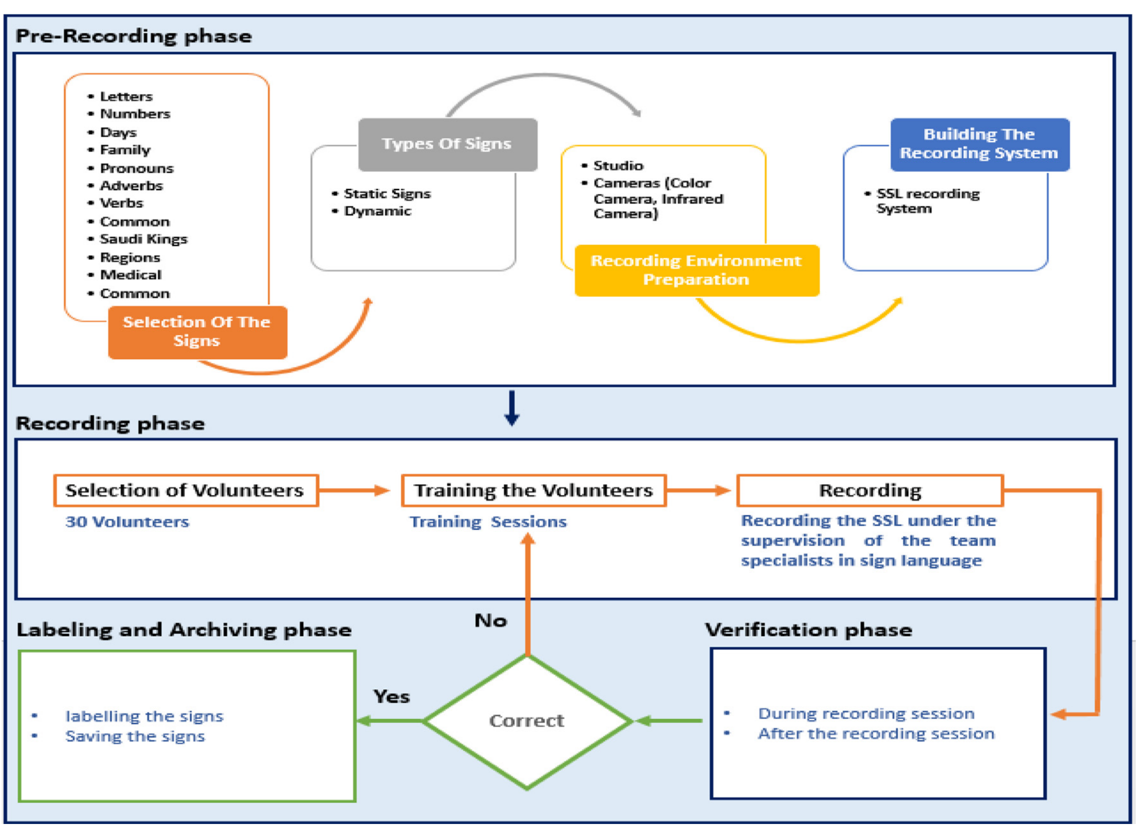


Fig. 1. Proposed KSU-SSL database framework.

Table 4
List of the 293 selected signs.

Alphabets (37 Sign)									
e	ب	ال	آ	ا	ئ	إ	ؤ	أ	ء
	b	al			e	i	o	a	hamza
س	ز	ر	ذ	د	خ	ح	ج	ث	ت
s	z	r	DH	d	KH	h	DJ	th	t
ك	ق	ف	غ	ع	ظ	ط	ض	ص	ش
k	q	f	GH	e	DH	t	DH	s	ch
يy				ى	و	ه	ن	م	ل
				y	w	h	n	m	l
Numbers (11 Sign)									
9	8	7	6	5	4	3	2	1	0
10									
Days (11 Sign)									
الجمعة	الخميس	الأربعاء	الثلاثاء	الاثنين	الاحد	السبت	يوم	سنة	اسبوع
Friday	Thursday	Wednesday	Tuesday	Monday	Sunday	Saturday	Day	Year	Week
أمسyesterday									
Family (8 Signs)									
زواجmarriage		أسرة	ابنة	ابن	أخت	أخ	أم	أب	
		family	daughter	son	sister	brother	Mother	Father	
Pronouns and adverbs (18 Sign)									
ثم	تحت	بعد	بدون	أنت	أنا	أمام	إلى	أعلى	أسفل
Then	Under	after	without	You	Me	ahead	to	upper	Down
منذsince			مع	قبل	في	فوق	على	دائما	خلف
			With	before	in	above	upon	Always	behind
Verbs (20 signs)									
يسمع	يستيقظ	يستطيع	يستحم	يساعد	يدخل	يخرج	يختفي	يجلس	يأكل
To hear	To Wake-up	Can	To shower	To help	To enter	To go out	To disappear	To sit	To eat
ينزل	ينام	يمشي	يقف	يفتح	يغلق	يظهر	يصعد	يشم	يشرب
To descend	To sleep	To walk	To stand up	To open	To close	To appear	To ascend	To smell	To drink
Common (39 Sign)									
ثلاث مرات	بيت	بوابة	بسبب	أين	أهلا وسهلا	النور	السلام عليكم	الخير	اسم
three times	House	Gate	Because	where	Welcome	light	Peace on you	Goodness	Noun
غدا	عفوا	صباح	شكرا	ساعة	رئيس قسم	رئيس	ربع ساعة	دورة مياه	حسنا
tomorrow	Excuse me	morning	Thanks	hour	Head of Department	President	quarter an hour	W.C	Okay
مرة	متى	ماذا	لو سمحت	لماذا	لا	كيف	كم	قسم	في الخارج
Once	when	what	Excuse me	Why	No	How	How many?	Department	out
وعليكم السلامMay peace be upon you		يسار	يمين	هل	هذه	هذا	نهار	مساء	مرتين
		left	right	Does?	This(female)	This (male)	day	evening	twice
Kings (9 Signs)									
الملك فيصلKing Faisal		الملك فهد	الملك عبدالله	الملك عبدالعزيز	الملك سلمان	الملك سعود	الملك خالد	الأمير محمد بن سلمان	آل سعود
		King Fahd	King Abdullah	King Abdulaziz	King Salman	King Saud	King Khaled	Prince Mohammed bin Salman	Al (family) Saud
Regions (7 Signs)									
مكة المكرمةMakkah				جدة	جازان	المدينة المنورة	الرياض	الدمام	أبها
				Jeddah	Jazan	Medina	Riyadh	Dammam	Abha
Healthcare (133 Sign)									
وراثة	وجه	هيكل عظمي	هاتف	نزيف	أزمة قلبية	إزالة	أرق	إدمان	إجهاض
heredity	Face	Skeleton	phone	bleeding	Heart attack	Removal	insomnia	addiction	abortion
اعتذر	إعاقة	إعاقة سمعية	إصابة	أشعة ليزر	إشعاع	إسهال	أسنان	إسعاف	استخدام
Apologize	Obstruction	Impaired hearing	infection	laser beams	radiation	Diarrhea	teeth	Ambulance	Usage
بخير	انتشار	إمساك	النفس	ألم	الطب النفسي	الصحة	الخدمة	التهاب	اكتئاب
Fine	Spread	constipation	self	pain	Psychiatry	Health	In service	inflammation	Depression
تنفس	تقرير	تغيير	تعب	تدليك	تحليل دم	تأمين	بهاق	بنج	بطن
Breathing	report	To change	fatigue	massage	blood analysis	insurance	vitiligo	Anesthesia	Belly
رئتان	ذراع	حساسية	حروق	حادث مروري	جهاز قياس الضغط	جهاز قياس الحرارة	جفن العين	جرح	تورم
lungs	arm	sensitive	burns	Traffic accident	Blood pressure device	Temperature measuring device	Lid	injury	swelling

Table 4 (*continued*)

Alphabets (37 Sign)									
سماعة أذن	سم	دورة شهرية	دوار	دواء	دم	حيض	حمى	حمل	حقنة
headphone	poison	Menstrual period	Vertigo	Medicine	blood	menstruation	Fever	Pregnancy	Injection
طبيب	ضماد طبي	سكر	سكتة قلبية	سرير	سرطان	زكام	رقبة	رضيع	رحم
Doctor	medical bandage	sugar	Heart failure	bed	cancer	common cold	neck	infant	womb
غدة	عينة مختبر	ضغط الدم	صيدلية	صيدلي	صورة أشعة	صداع	شلل نصفي	شعور	شرايين
gland	lab sample	blood pressure	pharmacy	pharmacist	X-ray	headache	hemiplegia	Feeling	arteries
قدم	فيروس	عناية	عمود فقري	عملية جراحية	عظم	عزوبة	عدوى	عدم	ظهر
Foot	virus	Attention	Backbone	Surgery	bone	celibacy	infection	nil	Back
كوع	كعب القدم	فيروس كورونا	فك	فقدان المناعة	فشل كلوي	فرشاة أسنان	فحص	فحص سريري	فحص النظر
elbow	Heel	Corona Virus	untie	Immunodeficiency	Renal failure	Toothbrush	examination	Clinical examination	Eye examination
يدخن	مدير	كرسي	كبد	قياس السمع	قوقعة	قلب	قفص صدري	قطرة	قصبة هوائية
smokes	director	a chair	liver	audiometry	cochlea	heart	Chest	Drop	Trachea
مناعة	ممرضة	مخدرات	مختبر	ليل	لم أفهم	لحظة	لاصق جروح	لا بأس	لا يستطيع
immunity	nurse	drugs	laboratory	night	I do not understand	One Moment	Adhesive wounds	Fine	can't
ولادة	وصفة طبية	ملح	مغص	معجون أسنان	مصعد	مستشفى	مريض	مرهم	مرض السكر
Birth	Prescription	salt	colic	toothpaste	elevator	hospital	Sick	ointment	Diabetes
نتيجةresult								نبضات القلب	ميزان حرارة
								heartbeats	Thermometer

we narrowed it to 293 signs. The selection of the signs concentrated mostly on those that will most probably be used in the daily medical experience of dialogue between a deaf or hard-of-hearing person and a nurse or doctor. This selection took a long time of reflection and consultation as the medical signs within the dictionary consist of about 150 signs; in a practical sense, the more signs there are to video record in multiple sessions, the more the time for each signer increases. This might be a burden for some signers and require more time than is feasible for the project time frame. In this context, we selected 293 signs, as presented in Table 4.

4.2. Types of signs

Some of the signs are just a simple fixed hand position of one or two hands (called *static signs*), while the rest of the signs use a moving action with one or two hands (called *dynamic signs*). In all cases, any sign can be viewed as dynamic if the pre-action (hands up) and post-action (hands down) are included in the entire sign video. Since we hope that our system will allow the deaf to use it live with minimal restrictions, all signs will be considered dynamic, in which we consider that the middle frames of interest of each sign are preceded and followed by some transition frames. A sample sign language performed by the hands is shown in Fig. 2.

4.3. Preparation and testing the recording environment

Recording videos of high quality requires considerable experimenting with people and material, as well as a stable recording machine and excellent programming skills. In this context, our specialist in charge managed to create a video recording studio as shown in Fig. 3.

The designed studio is comprised of the material listed in Table 5. We opted for this design after many trials and recording tests, as well as changing numerous lighting and camera positions.

Different setups and configurations were tried for the recording studio until we reached an optimal setup and configuration as illustrated below.

4.3.1. Environment configuration

The recording of the signs was done inside the arena of the Center for Smart Robotics Research at the College of Computer and Information at King Saud University, where a special computer with high specifications was prepared so that it could record high definition videos at a high speed. The recording setup contained the computer with a recording system, two screens, two cameras, a mobile, and the cameras stand. One of the two screens is facing the team member who is working on the recording system, and the other screen is facing toward the volunteer who performs the signs to display the reference videos to the volunteer.

The signs recording studio contains a place marked with a yellow box, where the volunteer stands to record; to his right, a screen has been placed showing a video of the sign to be recorded so that it is easier for him to perform the signs correctly, as shown in Fig. 3.

A white LED lighting has also been added to the studio to improve the quality of the captured videos, and a green background has been added behind the recording location to facilitate the video processing after recording.

A red and green LED was added to the recording system so that the volunteer can know when the recording is started and stopped.

4.3.2. Cameras configuration

The recording is done using three cameras, the high-speed color camera, mobile camera, and an infrared camera. The cameras are on a stand. The volunteer stands in front of the cameras at a distance of two meters so the upper body of the volunteer will occupy

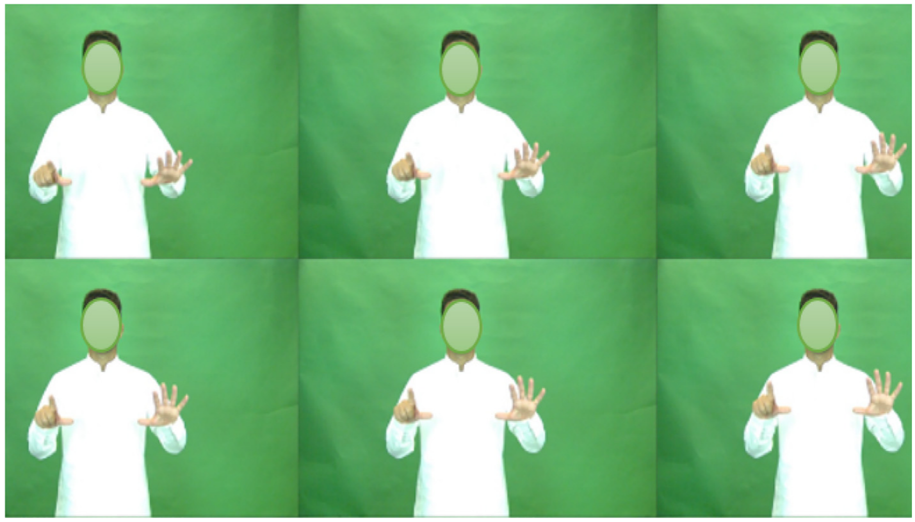


Fig. 2. Example of the sign language gesture –week.

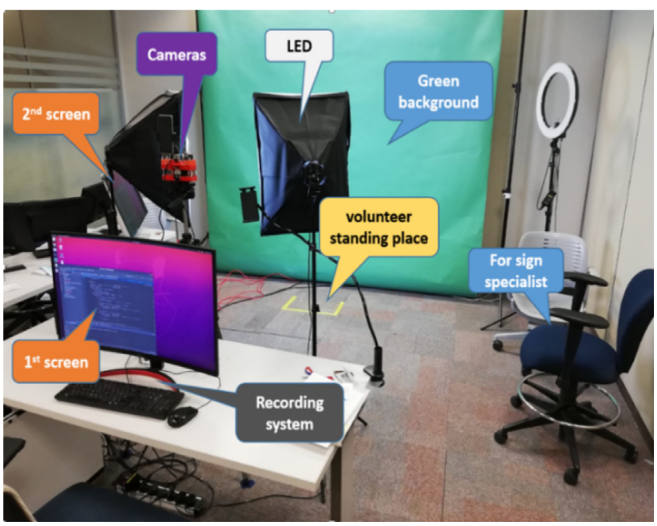


Fig. 3. Recording studio designed by our expert.

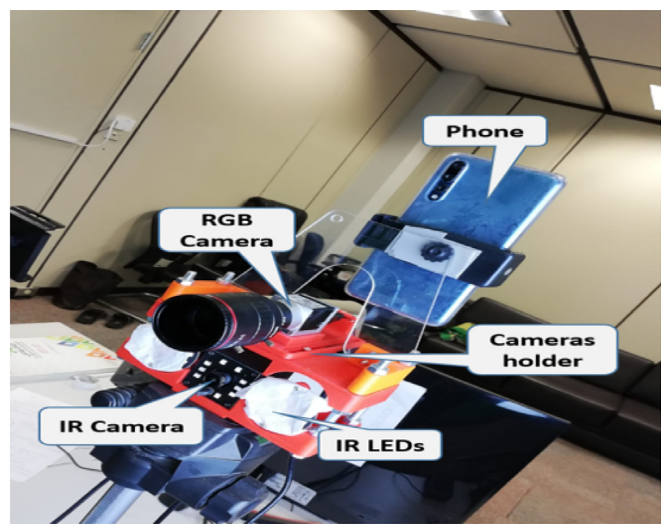


Fig. 4. Support for the cameras and the phone.

most of the camera's field of view. The expert sits on a chair at an angle that allows him to watch the volunteer perform the sign to be recorded and corrects him when needed.

The cameras were connected to the recording computer, where the recording program records videos for each repetition of each sign from the cameras at the same time. As for the *mobile camera*, the recording takes place throughout the session, and then later the signs can be cut manually.

For the *color camera*, in order to avoid the blurring effect generated by low fps cameras where some fast actions are not well recorded and appear as shadowed actions, our expert investigated the use of a very advanced camera with high fps. We choose an

Table 5
Recording studio components.

Item	Description	Remarks
acA1920-150uc - Basler ace	RGB color camera	High speed frame rate up to 150fps
ELP-USBFHD05MT-KL36IR	Infrared camera	Used to record IR videos
Huawei P20 pro	Mobile	To record with a phone camera
LED	Light stand	Used for better light conditions
First Screen	32″ screen	For recording team member
Second Screen	32″ screen	For the volunteer to see reference videos
Camera Holder	Holds RGB and infrared cameras and mobile	Made by the team
Chair	Blue chair	For sign specialist
Marked space in the floor	Marked space where the volunteer will stand	To make sure all volunteers will be facing the camera at the same distance
Paper background	The paper background behind the volunteer	To facilitate the video processing after recording

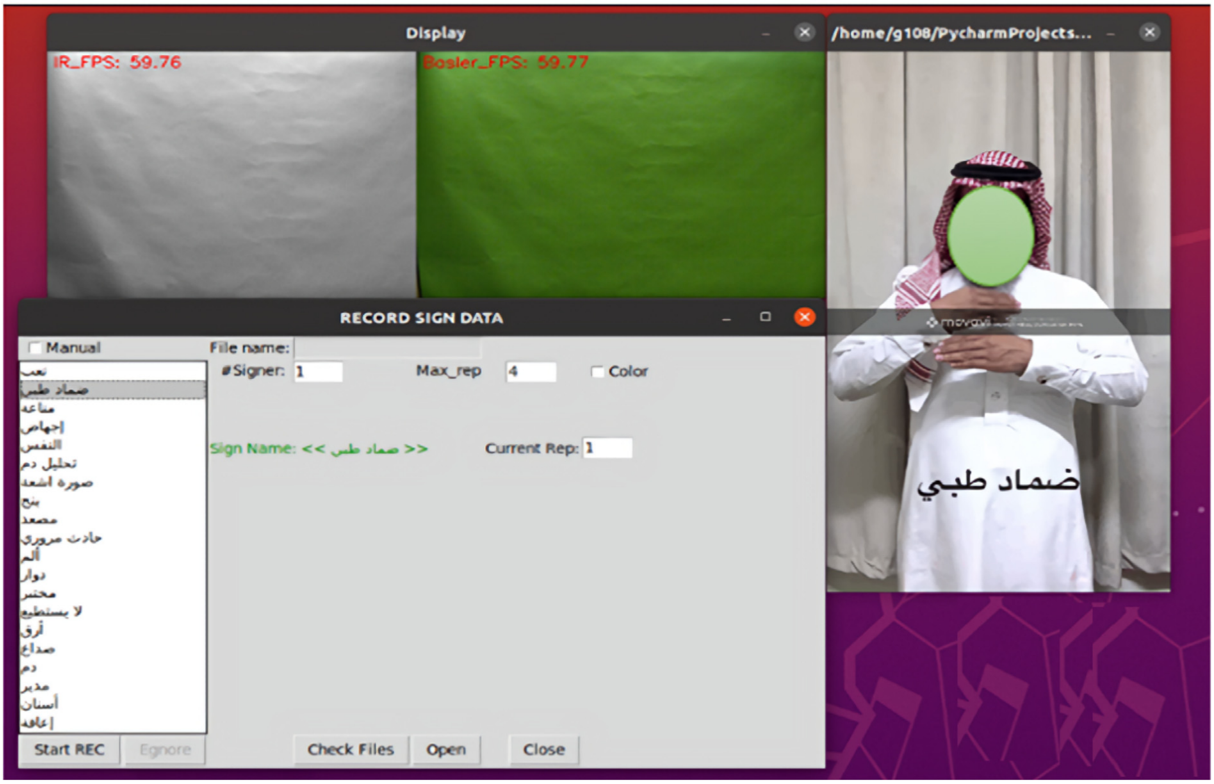


Fig. 5. Recording software interface (sample sign on the right).

industrial camera (acA1920-150uc - Basler ace) with a high frame rate. The camera originally is capable of 120 fps; however, we noticed during testing that 60 fps was quite satisfactory, and the actions or recorded signs could be tracked easily with all the images of the video. This camera can record videos with 1920*1200 px at 150 fps, but at high fps the images will be somewhat dark because the sensor will not have enough time to receive the light from the lens. To solve this issue, we first changed the lens of the camera with a new lens that has a large aperture; then, in the second step, we reduced shooting the frame rate to 120 fps so as to get the best recorded videos with our light conditions. The second issue encountered is that the recording PC was taking too much time to save those videos to the drive and was occupying a large amount of memory; therefore, we decided to change the camera setting to keep shooting videos at 120 fps but send only 60 fps to the recording PC.

The *infrared camera* is a special camera that captures the reflection of infrared rays on objects. This camera was chosen in order to avoid the problem of lighting, as these cameras depend on IR LEDs that are installed around the lens. To further improve performance, the technical team added an additional IR LED to the camera.

As our videos are recorded from three sources: a color camera, an infrared camera, and a mobile phone camera, and in order for all the videos to be captured from the same location, the technical team designed a special stand as shown in Fig. 4 to hold the cameras and the phone, so all recordings will be from the same direction at the same distance.

4.4. Building the recording system

Our technical team developed a system to record the videos of the required signs. The system contains reference video clips of all the signs to be recorded. The reference video clip for each sign is shown to the volunteer before recording the sign, to help him remember the correct way to perform the sign. The recording system has been developed in accordance with the following points:

- The program contains a **"signer#"** field to specify the volunteer's number.
- When the recording member presses the **"Open"** button, as shown in Fig. 5, the program opens the sub-folder of the selected signs and creates folders for saving those signs automatically, in order to avoid errors and standardize the pattern of saving in the database.
- The program records the signs from the cameras (color and infrared) connected to the device at the same time, as shown in Fig. 5.
- The recording member can choose the number of sign repetitions to record by entering the number of repetitions in the field "Max_rep", as shown in Fig. 5.
- The program also contains a "Color" check button, which can be specified by the recording member to record signs with colored hands.
- When recording begins, the program displays a video illustrating the way the sign is performed, as shown in the right display part of Fig. 5.
- The program records each sign several times in succession, according to the number of repetitions entered, then moves to the next sign automatically and displays the reference video for it.
- The recording system contains a list of the signs to be recorded. It also contains a box for the sign number "Current Rep" where the recording coordinator can choose any sign and specify the repetition number to be recorded in case the coordinator wants to re-record a specific sign.
- The program, as shown in Fig. 5.

- contains a "Start REC" button to start and stop recording, and an "Ignore" button to cancel the recording in case something went wrong.
- The red and green LEDs indicate for the volunteer the status of the recording system. First, when the system is ready to record, the green LED will be solid ON and the red LED will be OFF to show the volunteer that the system is ready to record and so that he can prepare himself to record the requested sign; then, when the recording coordinator presses "Start REC" button, the green LED will be OFF and the red LED will be solid ON to show the volunteer that the system is recording so that he will start performing the sign. Finally, when the recording coordinator presses the "STOP REC" button, the system starts saving the recorded videos and the green LED starts blinking until the saving process is finished, showing the volunteer that the system is saving the videos so that he can move freely and relax a little; then, it will turn to solid **ON** to show the volunteer that the system is ready to record again so that he can prepare himself again for the next recording.
- At the end of each recording session, the coordinator presses the **"Check Files**" button so that the program automatically checks that all the signs of the selected section are recorded and that the recordings are intact.

5. Recording KSU-SSL database phase

By the recording time, everything had been prepared; the signs were selected, the recording environment (light, background, distance between signer and cameras) was prepared, as well as the recording equipment (RGB cameras, depth cameras, desktops). As mentioned, we recorded 293 signs; each sign was repeated and recorded five times in different sessions, one with gloves and later with painted hands. Each recording session took an average of three hours; therefore, the total hours where 495 (33 signers * 5 session * 3 h = 495 h).

5.1. Selection and training of volunteers

The project aims to record the signs of volunteers who may be deaf, hard of hearing, and hearing. The volunteers were selected from King Saud University, and they were contacted to explain the project, its goal, and the need for a sign dataset for the project. Having a short time for recording is important; therefore, at the initial stages of the dataset recording and before accepting any volunteer, we sent to volunteers reference videos of the representative sample (10%) of the signs—about 30–35 signs—to practice, performed by a specialist to train them before coming to the studio for recording. When he is ready, we recorded his signs; then we sent their videos of performing these signs to the specialist who

Fig. 6. Colored gloves in sign recording.

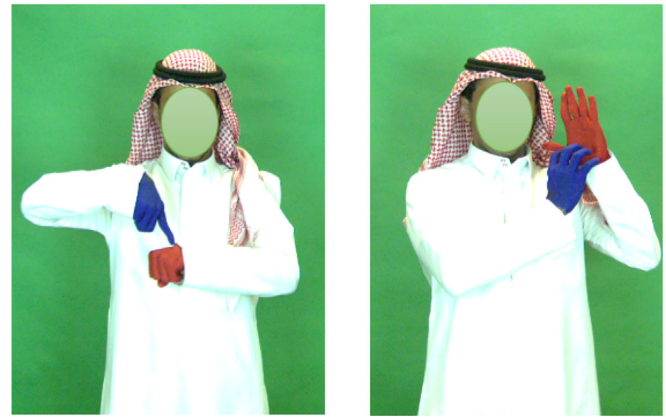


Fig. 7. Colored hands sign recording.

will choose those with near-perfect performance. This step was necessary because it saved us time when recording all signs of the database of the project. After accepting the volunteers, we sent them all pre-recorded reference signs so that the volunteer could train on them for a period ranging from two to five days before recording his signs.

5.2. Recording mechanism

At the beginning of the recording, several recording mechanisms were tried in order to determine the appropriate mechanism that achieved the least effort and the highest quality. It was necessary to record each sign five times, including one time with gloves and later with painted hands. The recording was five days a week, from Sunday to Thursday, under the supervision of the team specialists in sign language. All the signs were recorded in reference videos; then, the videos were sent to the volunteers to train them before coming to the studio for recording. Each volunteer was recorded in five sessions, with an average of 15 h divided over five session.

5.3. Recording with painted hands

After recording some initial volunteers and checking the videos with the volunteers wearing gloves, we found that there were problems with recording with gloves; sometimes there was light reflection. Moreover, hand fingers are not clear, as the light is reflected from the gloves, as shown in Fig. 6.

Therefore, we substituted recording with gloves to recording with the hands painted in colors. We chose blue for the right hand and red for the left hand because these two colors are dark and are the best light-absorbing colors; therefore, the light will not reflect and affect the camera. The green color, which is also dark color and light-absorbing, was used as the background; for this reason, we avoided using green to paint the fingers. We tested our choice of blue and red and this turned out to be a good decision, as shown in Fig. 7.

We first recorded the four repetitions of all the signs without painting (using gloves for the first six volunteers); next, we recorded the signs with the hands painted in the last recording session.

6. Verification phase

This phase is used to insure that the signs have been performed in a correct manner. In order to accomplish this, we have two steps of verification: during the sign recording and after the recording.

		group 1							group 2		
Number of signs		37	39	11	8	11	18	20	133	9	7
gner numb	signer name	Alphabet	common	days	family	numbers	ouns and adv	verbs	hospital and common	kings	regions
Signer 1		370	390	110	80	110	180	200	1330	90	70
Signer 2		370	390	110	80	110	180	200	1330	90	70
Signer 3		370	390	110	80	110	180	200	1330	90	70
Signer 4		370	390	110	80	110	180	200	1330	90	70
Signer 5		370	390	110	80	110	180	200	1330	90	70
Signer 6		370	390	110	80	110	180	200	1330	90	70
Signer 7		370	390	110	80	110	180	200	1330	90	70
Signer 8		370	390	110	80	110	180	200	1330	90	70
Signer 9		370	390	110	80	110	180	200	1330	90	70
Signer 10		370	390	110	80	110	180	200	1330	90	70
Signer 11		370	390	110	80	110	180	200	1330	90	70
Signer 12		370	390	110	80	110	180	200	1330	90	70
Signer 13		370	390	110	80	110	180	200	1330	90	70
Signer 14		296	312	88	64	88	144	160	stoped at "examination"	72	56
Signer 15		296	312	88	64	88	144	160	stoped at "hearing test"	72	56
Signer 16		370	390	110	80	110	180	200	1330	90	70
Signer 17											
Signer 18											
Signer 19											
Signer 20											
Signer 21											

record with painting	(blue)
record 4 repetitions	(yellow)
all recorded	(green)

Fig. 8. Daily follow-up form.

Table 6
KSU-SSL database statistics.

Domain	# signs	Rep.	# signers	# Cameras	Total Samples	%
Healthcare	133	5	33	3	65,835	45.4
Common	39	5	33	3	19,305	13.32
Alphabets	37	5	33	3	18,315	12.63
Verbs	20	5	33	3	9900	6.83
Pronouns & adverbs	18	5	33	3	8910	6.15
Numbers	11	5	33	3	5445	3.76
Days	11	5	33	3	5445	3.76
Kings	9	5	33	3	4455	3.08
Family	8	5	33	3	3960	2.74
Regions	7	5	33	3	3465	2.39
Total	293				145,035	

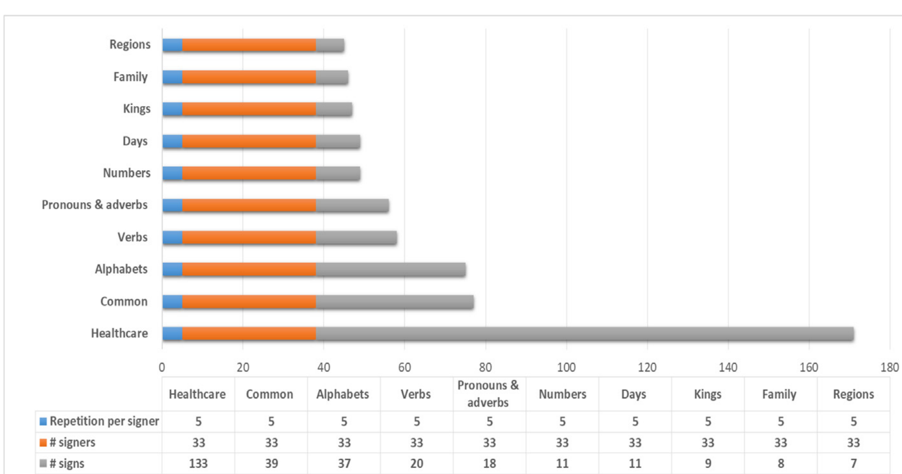
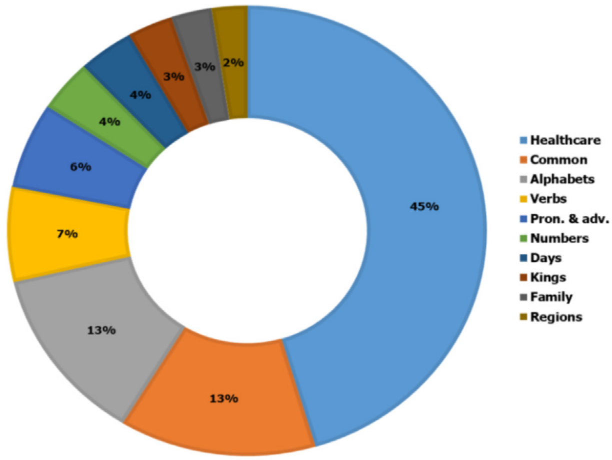
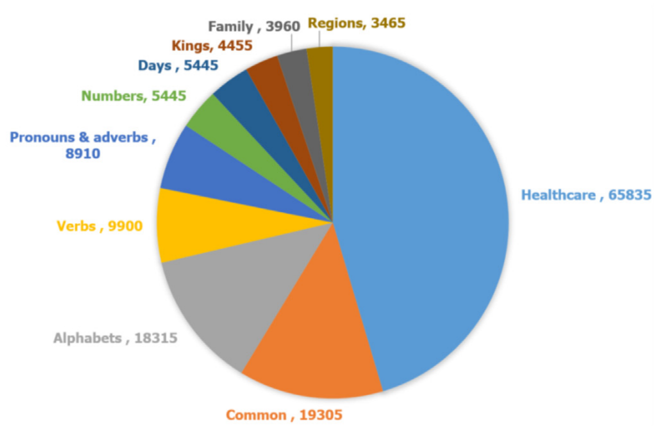


Fig. 9. KSU-SSL database statistics.

As we stated in the previous section, the recording was conducted under the supervision of the team specialists in sign language. This means that verification is done during the recording session with the attendance of a sign specialist. The post-recording verification was performed after completing each recording session per volunteer. In case of any error or mismatching of the performed sign, the specialist retrained the volunteer and the sign was performed again.



a) Percentage of each domain



b) Number of samples per domain

Fig. 10. Domain distribution across the KSU-SSL database.

6.1. Labeling and archiving phase

At the end of each recording session, the recordings were copied and saved in a folder labeled by the day and date on which the recording was made (for example, **Day 12 [1–1–2022]**); within this folder is a folder containing the name of the volunteer for whom the signaling session was recorded (for example, Abdullah Ahmad). With the beginning of the recording of each volunteer, the recording system creates a folder called **"signer_X"**, where **"X"** stands for the volunteer's number; next, under this folder, the program creates two folders (**"group_1**″ and **"group_2**″), and within each of these two folders, subfolders are created for the signs sections according to the following:

- The folder **"group_1**″ contains the following subfolders:
- Alphabet
- Common
- Days
- Family
- Numbers
- Pronouns and adverbs
- Verbs
- The folder **"group_2**″ contains the following subfolders:
- Hospital and common
- Kings
- Regions

Two copies of the recordings were made on different hard drives. In addition, after the daily session, two copies were uploaded into cloud storage: one copy to the **IDRIVE** cloud storage site and another to the **Google Drive** cloud storage site. A follow-up form was created using Excel, which includes the list of participants, to follow up on the volunteers' recordings and the specialists' review of the recorded signs if the signs were recorded in their absence. This form is shown in Fig. 8.

7. KSU-SSL database

As mentioned, this study present a SSL dataset for KSA Dialect; therefore, we adhered to the Saudi sign dictionary produced by the Saudi Association for Hearing Impairment. Our team (including sign language experts) considered many domains, such as government services, education, healthcare, children's stories, courts, and other law institutions). We realized that the medical field is an important area where many deaf people have difficulties explaining their illness and/or pain, and our system can be used immediately and be beneficial. Therefore, we selected the signs from the medical field. In addition to the medical field, we selected other daily life signs that are needed to build a system in the medical field and also needed in daily life communications. Therefore, the KSU-SSL database consists of 293 signs, 33 signers, and 145,035 samples (Table 6 and Fig. 9). The *healthcare domain* consists of 133 signs and 65,835 amples; the *common domain* consists of 39 signs and 19,305 samples; the *alphabets domain* consists of 37 signs and 18,315 samples; the *verbs domain* consists of 20 signs and 9900 samples; the *pronouns and adverbs domain* consists of 18 signs and 8910 samples; the *numbers domain* consists of 11 signs and 5445 samples; the *day domain* consists of 11 signs and 5445 samples; the *Saudi Kings domain* consists of nine signs and 4455 samples; the *family domain* consists of eight signs and 3960 samples; and the *regions domain* consists of 7 signs and 3465 samplesTable 6 shows the statistics of the KSU-SSL database, while Fig. 9 and Fig. 10 illustrates the domain distribution across the KSU-SSL database.

KSU-SSL datasets was recorded using three types of recording source: a fast RGB camera, an IR camera, and a mobile camera. Different types of cameras are used to make the database richer, and it can be used with different applications; moreover, in addition to the recorded repetitions, we add one repetition with painted hands so it can be isolated from the rest of the frame in future work.

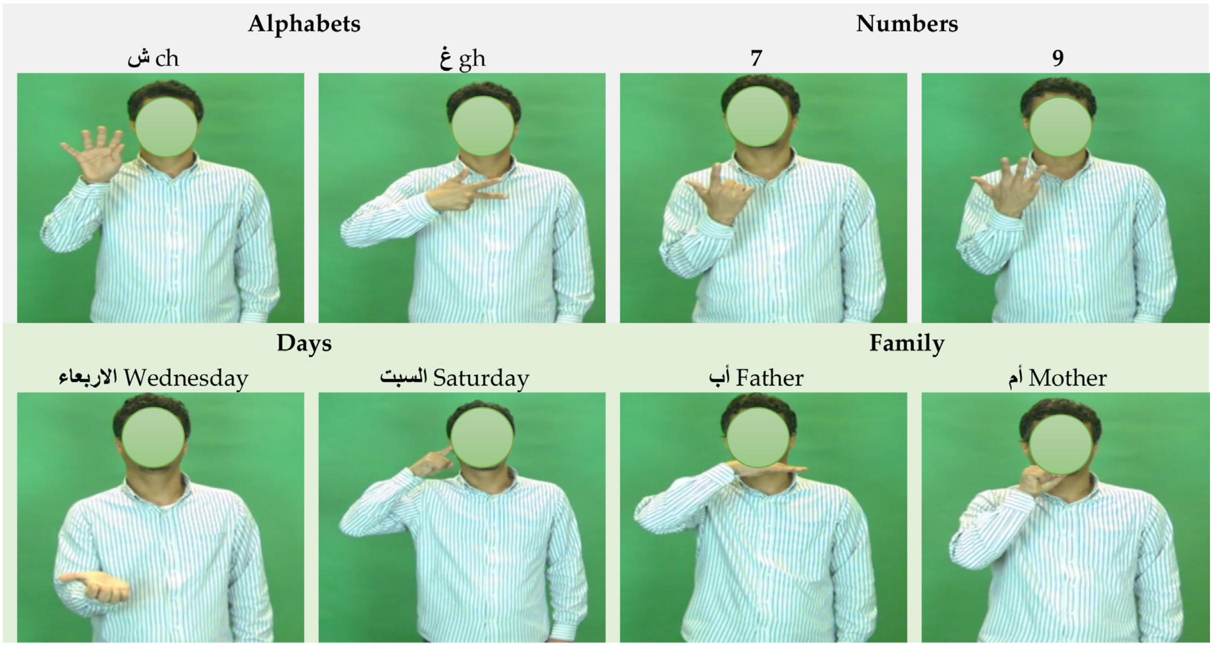
Table 7 illustrates two signs in each domain as an example of the proposed KSU-SSL database. From the healthcare domain, the signs of bone (عظم) and injection (حقنة) are illustrated. From the common domain, the signs of This (male) (هذا) and Noun (إسم) are illustrated. From the alphabets domain, the signs of gh (غ) and ch (ش) are illustrated. From the verbs domain, the signs of To drink (يشرب) and To appear (يظهر) are illustrated. From the pronoun and adverb domain, the signs of Under (تحت) and upper (اعلى) are illustrated. From the numbers domain, the signs of seven (7) and nine (9) are illustrated. From the days domain, the signs of Saturday (السبت) and Wednesday (الاربعاء) are illustrated. From the kings domain, the signs of King Fahd (الملك فهد) and King Saud (الملك سعود) are illustrated. From the family domain, the signs of Mother (أم) and Father (أب) are illustrated. Finally, from the regions domain, the signs of Jazan (جازان) and Abha (أبها) are illustrated in Table 7.

8. Proposed architecture (Implementation and Experiments)

In the proposed approach, we introduced a compact 3DGCN with a small number of trainable parameters to enable efficient representation learning. In addition, we leveraged MediaPipe, the most effective estimator of human landmarks, to extract the

Table 7
KSU-SSL database samples.

necessary graph nodes or recognition. To further enhance learning efficiency, we incorporated techniques such as multi-head self-attention and partitioning of frame nodes. Google has recently introduced MediaPipe (Lugaresi et al., 2019), a framework that offers cross-platform machine learning solutions for streaming media, allowing real-time perception of human pose, hand tracking, and face landmarks on mobile devices. MediaPipe has various modern life applications, including augmented reality, fitness, and sports analysis, and can detect and track 33 pose landmarks, 21 landmarks per hand, and 468 face landmarks in its different solutions. Moreover, MediaPipe integrates models for the pose, hands, and face components, providing a holistic solution that can accurately estimate and track a total of 543 landmarks in ×, y, and z coordinates. A sample frame from the KSU-SSL dataset is shown



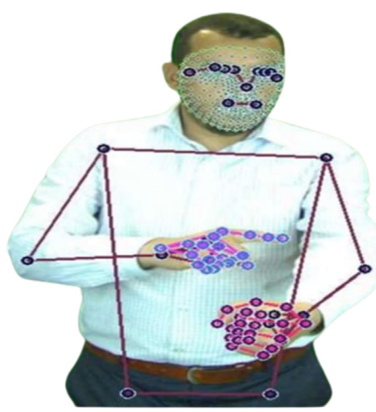


Fig. 11. MediaPipe landmarks (sample from our proposed dataset).

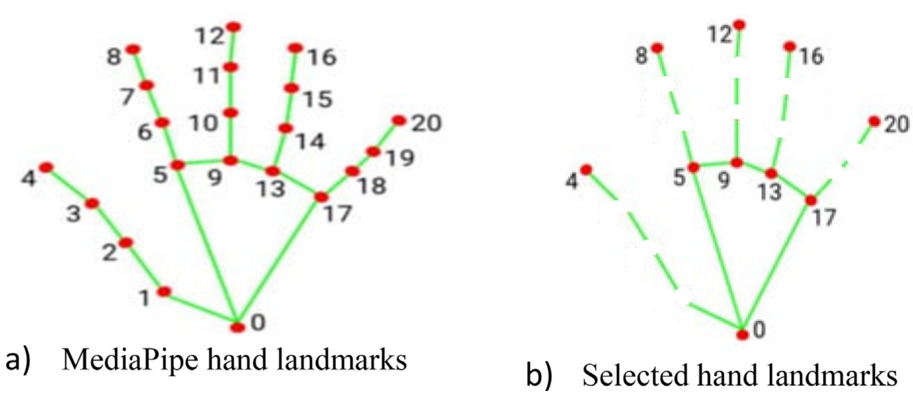


Fig. 12. MediaPipe hand landmarks (Lugaresi et al., 2019) VS selected hand landmarks.

in Fig. 11. In order to achieve a high recognition rate while minimizing computational complexity, we focused on the 25 most relevant landmarks when building the sign graph. From the 21 landmarks shown in Fig. 12 -a, we selected ten landmarks for each hand (0, 4, 5, 8, 9, 12, 13, 16, 17, and 20) Fig. 12 -b, along with five additional landmarks representing the nose, shoulders, and elbows. The nose landmark was utilized as a reference to normalize the landmarks for each frame individually.

The outcome of this process is an undirected spatial–temporal graph denoted as G = (V, E), containing T frames and 10 nodes with both inter-body and intra-frame connections. The node set $V = \{v_{ti} | t = 1, \ldots, T, i = 1, \ldots, 10\}$ consists of the selected 10 landmarks in

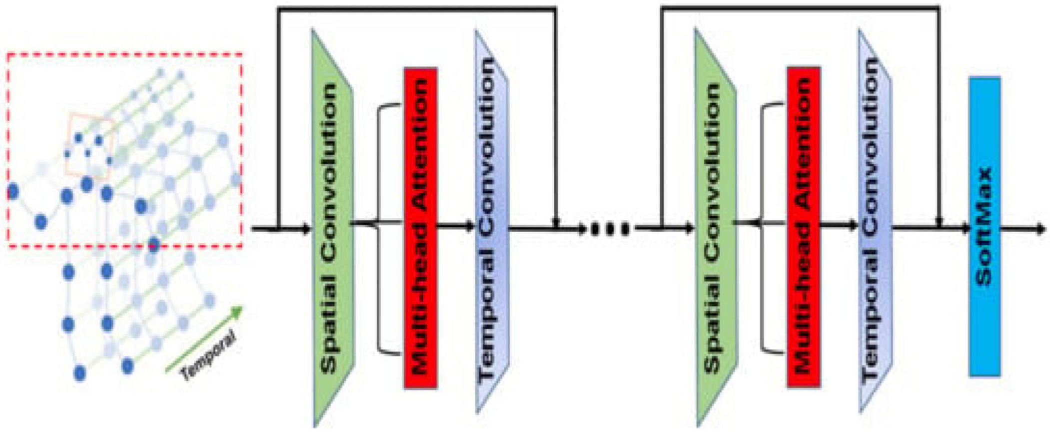


Fig. 13. The proposed 3DGCN architecture.

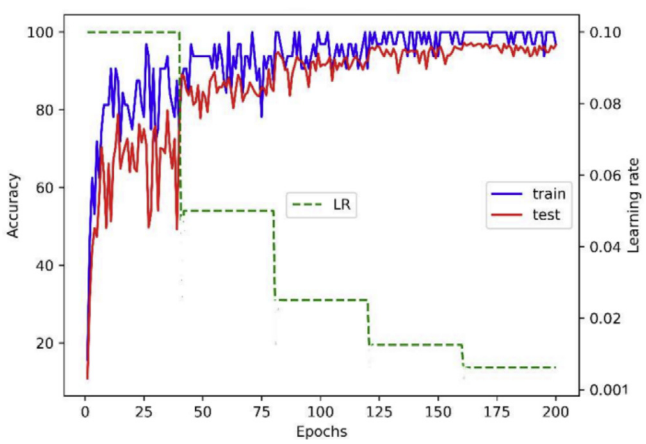


Fig. 14. The behavior of the proposed 3DGCN architecture on the KSU-SSL dataset.

each frame throughout the sequence. Nodes within a frame are connected by edges based on the human body structure, while each node is linked to itself in successive frames. Thus, the edge set E comprises two types of edges: intra-frame edges within each frame and inter-frame edges linking nodes in consecutive frames. The proposed architecture comprises five successive layers of separable 3DGCN, which are separated into spatial and temporal convolution operations through a spatial multi-head self-attention layer, as depicted in Fig. 13. The input layer is connected to the output layer through a residual connection. To perform spatial convolution, this approach considers the spatial neighbourhood of each node to be all the nodes within a single step distance from that node.

In order to exhibit the effectiveness of the suggested basic 3DGCN architecture (Al-Hammadi et al., 2022), which is lightweight, an evaluation was carried out on the KSU-SSL dataset. PyTorch was employed to implement the proposed architecture, and NVIDIA RTX 3090 24 GB GPU was utilized for training. We used 28,021 samples for training and 5,860 for Validation. Training was conducted using a fixed configuration, which utilized mini-batch gradient descent with a batch size of 32 samples and an adaptive learning rate. The initial learning rate was set to 0.1 with an update frequency of 200 epochs. Fig. 14 and Table 8 illustrated the behavior of the proposed 3DGCN architecture on the KSU-SSL dataset.

Table 8
Performance of the proposed 3DGCN architecture on the KSU-SSL dataset.

Avg. accuracy	batch size	Training size	Validation	epochs
97.25	32	28,021	5,860	200

The proposed 3DGCN architecture achieved an average accuracy of 97.25% on the KSU-SSL dataset. The model was trained using a batch size of 32 and a training dataset consisting of 28,021 samples. The validation set used during training contained 5,860 samples. The model was trained for 200 epochs.

9. Conclusion

In this paper, we proposed a framework for constructing a sign language dataset, and we introduced a comprehensive survey of 17 Arabic sign language datasets. The surveyed datasets are categorized by their year, number of signers, number of signs, samples, type of signs, and accessibility. In addition, in this paper, we created and build the largest Saudi Sign Language (SSL) database—the King Saud University Saudi-SSL (KSU-SSL database)—with 293 signs, 33 signers, 145,035 samples, and 10 domains (healthcare, common, alphabets, verbs, pronouns and adverbs, numbers, days, kings, family, and regions). In additional, we introduced the 3CGCN architecture for sign language recognition and applying the proposed architecture to the built KSU-SSL database. In the future, we aim to use the proposed dataset to develop a system for a two-way translation of Saudi sign language based on Avatar and integrates several functions such: 1) recognizes the speech of the hearing person and produces the corresponding text; 2) converts the recognized text of the hearing person to sign language and performs the sign by the Avatar; 3) recognizes the sign performed by the deaf person and produces the corresponding text and; 4) converts the text of the recognized sign that the deaf person performed into speech. Moreover, In the future, we plan to expand the scope of the dataset by adding more signs from different domains.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research project was funded by the Targeted Research Grant Program, The National Transformation Program in King

Abdulaziz City for Science and Technology, Kingdom of Saudi Arabia, Grant No. 5-18-03-001-0003.

References

Albelwi, N.R., Alginahi, Y.M., 2012. Real-time arabic sign language (arsl) recognition. In: International Conference on Communications and Information Technology, pp. 497–501.

Alfonse, M., Ali, A., Elons, A. S., Badr, N. L. & Aboul-Ela, M. Arabic sign language benchmark database for different heterogeneous sensors. 2015 5th International Conference on Information & Communication Technology and Accessibility (ICTA), 2015. IEEE, 1-9.

Al-Hammadi, M., Muhammad, G., Abdul, W., Alsulaiman, M., Bencherif, M.A., Mekhtiche, M.A., 2020. Hand gesture recognition for sign language using 3DCNN. IEEE Access 8, 79491–79509.

Al-Hammadi, M., Bencherif, M.A., Alsulaiman, M., Muhammad, G., Mekhtiche, M.A., Abdul, W., Alohali, Y.A., Alrayes, T.S., Mathkour, H., Faisal, M., 2022. Spatial attention-based 3d graph convolutional neural network for sign language recognition. Sensors 22, 4558.

Alhusseini, N., Alqahtani, A., 2020. COVID-19 pandemic's impact on eating habits in Saudi Arabia. Journal of Public Health Research 9.

Al-Jarrah, O., Halawani, A., 2001. Recognition of gestures in Arabic sign language using neuro-fuzzy systems. Artif. Intell. 133, 117–138.

Al-Musharaf, S., Aljuraiban, G., Bogis, R., Alnafisah, R., Aldhwayan, M., Tahrani, A., 2021. Lifestyle changes associated with COVID-19 quarantine among young Saudi women: A prospective study. PLoS One 16, e0250625.

AL-NAFJAN, A., AL-ARIFI, B. & AL-WABIL, A. Design and development of an educational Arabic sign language mobile application: collective impact with Tawasol. International Conference on Universal Access in Human-Computer Interaction, 2015. Springer, 319-326.

Al-Rousan, M., Assaleh, K., Tala'a, A., 2009. Video-based signer-independent Arabic sign language recognition using hidden Markov models. Appl. Soft Comput. 9, 990–999.

Amin, O., Said, H., Samy, A. & Mohammed, H. K. HMM based automatic Arabic sign language translator using Kinect. 2015 Tenth International Conference on Computer Engineering & Systems (ICCES), 2015. IEEE, 389-392.

Assaleh, K., Shanableh, T., Fanaswala, M., Bajaj, H. & Amin, F. Vision-based system for continuous arabic sign language recognition in user dependent mode. 2008 5th International Symposium on Mechatronics and Its Applications, 2008. IEEE, 1-5.

Bahammam, A., Alfaris, E., Shaikh, S., Saeed, A.B., 2006. Prevalence of sleep problems and habits in a sample of Saudi primary school children. Ann. Saudi Med. 26, 7–13.

Bencherif, M.A., Algabri, M., Mekhtiche, M.A., Faisal, M., Alsulaiman, M., Mathkour, H., Al-Hammadi, M., Ghaleb, H., 2021. Arabic sign language recognition system using 2D hands and body skeleton data. IEEE Access 9, 59612–59627.

Elbadawy, M., Elons, A. S., Sheded, H. & Tolba, M. F. 2015. A proposed hybrid sensor architecture for arabic sign language recognition. *Intelligent Systems' 2014*. Springer.

Guesmi, F., Bouchrika, T., Jemai, O., Zaied, M. & Amar, C. B. Arabic sign language recognition system based on wavelet networks. 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2016. IEEE, 003561-003566.

Latif, G., Mohammad, N., Alghazo, J., Alkhalaf, R., Alkhalaf, R., 2019. Arasl: Arabic alphabets sign language dataset. Data Brief 23, 103777.

Lugaresi, C., Tang, J., Nash, H., McClanahan, C., Uboweja, E., Hays, M., Zhang, F., Chang, C.-L., Yong, M., Lee, J.M., 2019. A framework for perceiving and processing reality. Third Workshop on Computer Vision for AR/VR at IEEE Computer Vision and Pattern Recognition (CVPR).

Maschendorf Thomaz, M., Marten Milbrath, V., Irmgard Bärtschi Gabatz, R., Freitag, V.L., Cardoso Vaz, J., 2019. Accessibility of adolescents with hearing impairment to health services. Revista Eletronica de Enfermagem 21.

Mohandes, M., Quadri, S. & Deriche, M. 2007. Arabic sign language recognition an image-based approach. 21st International Conference on Advanced Information Networking and Applications Workshops (AINAW'07). IEEE, 272-276.

Nicodemo, M., Spreghini, M.R., Manco, M., Wietrzykowska Sforza, R., Morino, G., 2021. Childhood Obesity and COVID-19 Lockdown: Remarks on Eating Habits of Patients Enrolled in a Food-Education Program. Nutrients 13, 383.

Pellegrini, M., Ponzo, V., Rosato, R., Scumaci, E., Goitre, I., Benso, A., Belcastro, S., Crespi, C., de Michieli, F., Ghigo, E., 2020. Changes in weight and nutritional habits in adults with obesity during the "lockdown" period caused by the COVID-19 virus emergency. Nutrients 12, 2016.

Samirelons, A., Abull-Ela, M., Tolba, M.F., 2013. Pulse-coupled neural network feature generation model for Arabic sign language recognition. IET Image Proc. 7, 829–836.

Senkov, O.N., Miracle, D.B., Chaput, K.J., Couzinie, J.-P., 2018. Development and exploration of refractory high entropy alloys—A review. J. Mater. Res. 33, 3092–3128.

Shanableh, T., Assaleh, K., 2007. Telescopic vector composition and polar accumulated motion residuals for feature extraction in Arabic sign language recognition. EURASIP Journal on Image and Video Processing 2007, 1–10.

Shohieb, S.M., Elminir, H.K., Riad, A.M., 2015. Signsworld atlas; a benchmark Arabic sign language database. Journalof King Saud University-Computer and Information Sciences 27, 68–76.

Sidig, A.A.I., Luqman, H., Mahmoud, S., Mohandes, M., 2021. KArSL: Arabic Sign Language Database. ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP) 20, 1–19.

Tolba, M., Samir, A. & Abul-Ela, M. 2012. A proposed graph matching technique for Arabic sign language continuous sentences recognition. 2012 8th International Conference on Informatics and Systems (INFOS). IEEE, MM-14-MM-20.

Youssif, A., Aboutabl, A.E., Ali, H.H., 2011. Arabic sign language (arsl) recognition system using hmm. In: International Journal of Advanced Computer Science and Applications (IJACSA), p. 2.