The use of artificial intelligence for sign language recognition in education: from a literature overview to the ISENSE project

Juri Taborri
University of Tuscia
Viterbo, Italy
juri.taborri@unitus.it

Pietro Fornai
University of Tuscia
Viterbo, Italy
pietrofornai@gmail.com

Enrique Yeguas-Bolivar
University of Cordoba
Cordoba, Spain
eyeguas@uco.es

Maria Dolores Redel-Macías
University of Cordoba
Cordoba, Spain
mredel@uco.es

Marlene Hilzensauer
Universität Klagenfurt
Klagenfurt, Austria
marlene.hilzensauer@aau.at

Alexandra Pecher
Universität Klagenfurt
Klagenfurt, Austria
alexandra.pecher@aau.at

Manfred Leisenberg
Fachhochschule des Mittelstands
Bielefeld, Germany
manfred.leisenberg@fh-mittelstand.de

Alessia Melis
Blue Cinema TV srl
Rome, Italy
a.melis@bluecinematv.com

Stefano Rossi
University of Tuscia
Viterbo, Italy
stefano.rossi@unitus.it

Abstract— **Common strategies to guarantee the inclusion of d/Deaf students in university are still absent. Recent studies have demonstrated how innovative technologies, such as artificial intelligence and virtual/augmented reality, could be applied to solve this issue. This paper aims at providing an overview of the AI-based algorithms for sign language recognition, as well as the description of the project ISENSE. As concerning the AI application, the already implemented procedures are generally based on three steps, starting from sign language capture to sign language translation, passing through sign language recognition. The state-of-the art for each step has been reported, highlighting the most widespread solutions and the related pros and cons. Focusing on ISENSE, the project aims at implementing services and tools based on artificial intelligence and immersive technologies for deaf students at university. The project framework has been introduced, reporting both the main outcomes and the operating steps.**

Keywords— artificial intelligence; sign language recognition; sensor system; ISENSE

I. Introduction

According to the World Health Organization, 466 million people worldwide have a disabling hearing loss and the amount of cost due to the social isolation and the challenges of quality education is estimated to be 178 billion euros per year. Several research and reports highlighted the absence of supporting tools for d/Deaf students who decide to begin academic studies. It is well-known that deaf students need vital support in order to undertake academic studies. However, there is a chronic shortage of such services in European universities, leading deaf students to consider abandoning their university studies. For example, the research project conducted by the National Deaf Children's Society in UK shows that 59% of deaf students who need support at university experienced delay in receiving additional support in time for the start of their courses [1]. This situation has been amplified with the pandemic era [2].

Considering the Italian scenario, deaf students are supported by a set of norms on the disability topics, as for example the law n.104/92 with relative modification of law n.17/99; however, a common strategy to guarantee the accessibility of all information is still missing. In particular, this issue is mainly related to the orientation phase, since there is no skilled operator available to assist deaf students in the selection of the most appropriate course, in the bureaucracy of the enrolment phase and in providing supporting tools. In fact, the number of universities that have a specific deaf student orientation guide is still limited and students have to search the needed information only on the university website [3]. Moreover, the communication barriers also exist when deaf sign language users try to communicate with institutions, since almost all people use spoken language. It is evident how there is a need to involve sign language interpreters or to provide sign language training to enable communication with students who are deaf or hard of hearing.

In this context, immersive technologies and artificial intelligence could be applied to facilitate learning for students at any level. Considering the artificial intelligence (AI), one of the main challenges is the implementation of robust AI-based algorithms to automatically recognize the sign languages for the creation of a holographic guide able to interact with students [4]. The implementation of AI-based algorithms for sign language recognition could be structured in three phases, which are: (i) sign language capture; (ii) sign language recognition; and (iii) sign language translation. The sign language capture involves the use of a sensor system for motion capture, ranging from camera-based approaches to wearable gloves [5]. For the recognition phase, convolutional neural networks revealed themselves as the most accurate; whereas the translation phase is typically based on deep learning algorithms, such as long short-term memory [6]. In addition, recent studies demonstrated how augmented reality can be exploited to realize virtual environments to improve the accessibility and inclusion of deaf people in the community. One of the most interesting solutions was proposed by a Japanese researcher that developed a 3D avatar called "I Hear You", a mobile application with sign language keyboard & 3D Character animations to translate sign language to text/audio and from text/audio to sign language. However, such a solution only works for a few sentences. Then, a sign support assistant has been developed to assist people who are deaf or hard of hearing in transaction at a post office, realizing

"TESSA" [7]. TESSA is able to translate from spoken language to British Sign Language only fixed sentences that are frequently used at a post office. A 2D virtual agent was also proposed by McConnell et al. [8], which works on the 26 letters of the English alphabets. However, a similar solution for assistance in university has not been realized.

Even if several studies have been proposed on the use of artificial intelligence and augmented reality for sign language automatic recognition, a common approach to foster the application in education is still missing. This paper aims at presenting a general overview of the AI applications for sign language recognition, as well as to introduce the proof of concept of an innovative project, i,e., ISENSE - Innovative Supporting sErvices for uNiversity Students with dEafness. ISENSE arises from the belief of the importance of deaf students' autonomy and inclusion in mainstream education in order to prepare them for the labour market without isolating them from their peers. Thus, ISENSE also deals with the general need to avoid inequality during ~~the~~ university studies by taking advantage of innovative technologies.

II. ARTIFICIAL INTELLIGENCE RELATED WORKS

A literature overview was conducted by means of Scopus, Web of Science and PubMed, focusing on the studies reporting the application of artificial intelligence for the recognition of the sign language regardless of the nationality. In particular, we sought to provide an overview on both sensor systems used for the data acquisition and the AI-based algorithm for the data processing. The literature overview was performed from January 2022 to January 2023 and the combinations of the following keywords have been applied: *sign language capturing, sign language capturing video analysis; sign language capturing data glove; sign language recognition; sign language translation; artificial intelligence*. Symbols were also taken into account to avoid forgetting important studies. The bibliography of each found paper was then carefully checked to find further papers, which could be missed due to the adopted search strategy. Only papers published from 2005 to 2022 and written in English were included in the studies.

Sign language recognition is divided into two main categories: the one related to recognition of isolated signs (ISLR) and the one related to continuous signs (CSLR). The former refers to the task of detecting with precision single sign gestures from videos, whereas the latter aims to classify annotated videos into the entire phrases [9]. The present analysis focuses almost exclusively on continuous recognition methods since they present greater operational complexities than ISLR and constitute the prevailing technological challenge in practical applications.

The methods using artificial intelligence are generally based on three main steps, reported in Figure 1. Each step is further discussed in the following paragraphs.

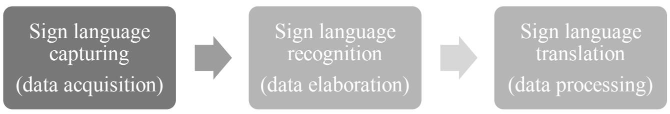

Fig. 1 – Steps for AI-based methodologies.

A. Sign language capturing

The purpose of the data acquisition phase is to capture information from the signs in order to allow their analysis, recognition and translation. It is important to underline that data are not only related to the hand movements, since posture and facial expressions are associated with fundamental information for the translation. The three main categories for accomplishing with this step are: (i) optoelectronic system; (ii) video-based techniques; and (iii) wearable sensors.

Focusing on optoelectronic systems, the study conducted by Sylvie Gibet [10] compared three different configurations of infrared cameras. The first configuration is composed of 12 cameras and 63 reflective markers; the second one is composed of 12 cameras and 84 reflective markers and the last one is composed of 16 cameras and 38 reflective markers on the hands. The study dealt with the realization of a dataset containing pre-established sentences in French sign language related to: (i) city and traffic information (TRAIN); (ii) urban services and places (Sign3D); (iii) recipes (SignCom); and (iv) meteorological forecast (METEO). Similarly, a database for Czech sign language has been realized in [11], using 18 cameras and a total of 109 markers (33 for the hands, 30 for the fingers and 46 for the face).

Considering the video-based techniques, Microsoft Kinect represents the most widespread technology. For instance, Raghuveera *et al.* [12] proposed an Indian sign language recognition based on Kinect, allowing to construct a dataset of 140 unique gestures reproduced by 21 interpreters. The RealSense D435 has been used in [13] for the creation of a Greek sign language dataset.

Finally, moving to the wearable sensors, sensorized gloves represent the most adopted solution for capturing the movements of the hand. For example, Aponno and colleagues demonstrated the feasibility of using a glove sensorized with flex sensors and pressure sensors for the acquisition of the 26 letters of the Indonesian sign language alphabet. Similarly, a glove embedded with an inertial module was used for the acquisition of Japanese sign language [14].

Due to the complexity and plethora of variables to be considered for sign language recognition and translation, it is clear that the best solution is to design a sensor system able to combine data from different sensor types [15].

B. Sign language recognition

This step, named as SLR, often involved the use of a neural network able to automatically recognize the signs, regardless of the sensor systems used for the data acquisition. One of the main problems to be faced in this phase is associated with the sign segmentation, since in CSLR it is not present any division between consecutive signs. As an example, a 3D Convolutional Neural Network (CNN) has been implemented to extract spatio-temporal features from video stream in order to automatically recognize sign language. In particular, the CNN uses information related to color, depth and body joint positions and trajectories [16]. Aiming to improve the accuracy in comparison with the state-of-the-art approaches, Mannan et al. proposed a deep CNN for the recognition of American sign language, achieving an accuracy improvement greater that 5% [17]. A similar approach was also applied on Bengali sign language using video-based techniques, reaching an accuracy of 85% in alphabet recognition [18].

C. *Sign language translation*

The final step, named as SLT, is represented by the sign language translation. This phase aims at translating the sequence of signs recognized by the SLR into the correct sequence, which is coherent with the semantics and grammar of the specific sign language. It is worth underlining the complexity of this phase due to the different linguistic rules not only among different spoken language, but especially due to the intrinsic differences between spoken and sign language of the same country [19]. In fact, sign language does not translate spoken language word by word; for example, the structure of a sentence in Italian Sign Language is subject, complement and verb.

In this context, it is clear how the role of the SLT is to transform the sequence of signs into a correct sequence of words in the spoken language. The most widespread and useful algorithms for achieving such a result are based on Long Short-Term Memory (LSTM). For instance, Mittal et al. validated an LSTM model for continuous Indian Sign Language recognition and translation using data gathered from Leap Motion. In particular, the system has been tested on 942 signed sentences, resulting from the composition of 35 difference words, obtaining an accuracy of 73% for the sentences and 90% for the isolated words [20].

D. *Available dataset*

The applications of artificial intelligence for sign language recognition and translation are constantly growing thanks to the possibility to use open access dataset. Table 1 shows the main datasets available with details related to the language, number of signers, number of data and year of construction. It is worth noticing that the reported information is only related to continuous sign language recognition and to data acquired by means of RGB-cameras.

Table 1 – Information on open access database for continuous sign language recognition in different languages.

Dataset name	Language	Signers	Video instances	Year
Phoenix-2014	German	9	6841	2014
CSL	Chinese	50	25000	2016
Phoenix-2014-T	German	9	8257	2018
GRSL	Greek	15	4000	2020
BSL-1K	British	40	273000	2021
GSL	Greek	7	10295	2021

III. ISENSE FRAMEWORK

A. *Project consortium*

Starting from the literature overview and the needs analysis, a European project has been proposed, named ISENSE - Innovative Supporting sErvices for uNiversity Students with dEafness. The consortium of the project is composed of:

- University of Tuscia (Viterbo, Italy)
- University of Cordoba (Cordoba, Spain)
- Universität Klagenfurt (Klagenfurt, Austria)
- Fachhochschule des Mittelstands (Bielefeld, Germany)
- AUSRU (Italy) – deaf association
- FCNSE (Spain) – deaf association
- Pitagoras (Poland) – deaf association
- Blue Cinema TV srl (Italy) – enterprise
- TUCEP (Italy) – non-profit organization

Figure 2 shows the project partners.

Fig. 2 – Partners of ISENSE project.

B. *Project aims and outcomes*

From this perspective, the project aims at implementing supporting services to be understood as both digital tools and innovative good practices based on higher technologies, such as artificial intelligence, augmented reality, and interactive holograms to assist the students with deafness during their academic life, especially focusing on the university orientation phase. In addition, a set of guidelines to support these students during the stages of enrolment and study, as well as for the entry into employment will be defined. Four tangible outcomes will be achieved within the project:

1. SLAATE – *Sign Language in Augmentative Alternative Technical 3D Environment*. SLAATE is an ICT platform that permits the realization of web-based interactive videos to interact with deaf students during the enrolment phase by using holographic and immersive approaches. The videos will be realized following an innovative digital methodology based on the European patent linked to the trademark OLOS® [21] in order to make the university orientation more inclusive and to help deaf students to select the most appropriate university. A complete interactive immersive installation based on AI whose visual output is a life-size virtual human being [22] [23] (sign interpreter) will be realized at the end of the project at the University of Tuscia as the proof of concept for further steps. At the other universities, the above-mentioned methodology will be used to create web-based online interactive multimedia content with sign interpreters in the five sign languages of the project: Italian, Spanish, German, Polish and Austrian.
2. SHARE SLANG – *SHAring REpository of Sign LANguage Gesture*. This output is a public database

containing data of the signs related to sign language in terms of a movement analysis file acquired by using an optoelectronic system. This step will be performed only considering Italian sign language (LIS) and will be used as a preliminary study to make data available for the research community involved in the analysis of sign language through artificial intelligence. As a further activity, AI-based algorithms for recognition and translation of the activities will be implemented. More generally, the repository will contain all the useful documentation found within the project development and realized during the project.

3. SONAR – *SOcial iNclusion through Augmented Reality*. This outcome concerns the realization of a specific training path for professors, technicians, relatives and friends aiming at learning sign language. The common training path performed by a skilled operator will be combined with learning scenarios based on virtual and augmented reality scenarios. Specifically, virtual reality will be used to raise awareness of the topic of deafness, making it possible to simulate the issues faced by deaf students through emphatic scenarios. In addition, a test in virtual reality will be implemented to test the ability acquired during the training. This outcome can be used in the future by all the universities to train people who want to learn sign language. Courses will be realized in the five countries of the project: Italy, Spain, Germany, Poland and Austria.

4. GUI4DE - *GUIdelines FOR Deaf*. This output will represent the summary of all the findings obtained during the development of the project, leading to the implementation of a memorandum of understanding containing all the guidelines and best practices to guarantee an equal education characterized by a high level of accessibility for students with deafness. Specifically, three round tables will be set up in order to investigate the following topics: (i) good practices for orientation; (ii) good practices for assistance during academic life; and (iii) good practices for entry into employment.

A schematic representation of the four outputs is given in Figure 3.

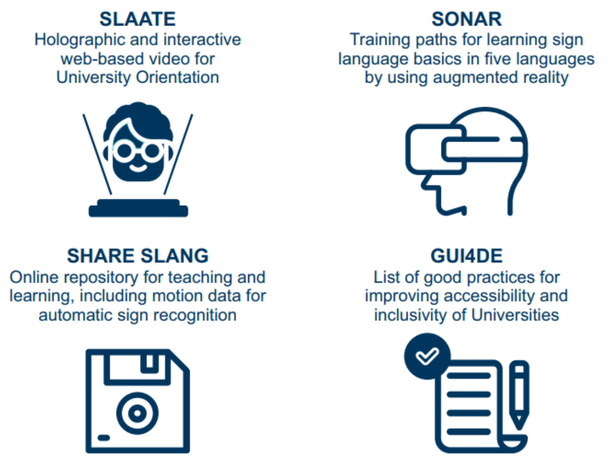

Fig. 3 – Schematic representation of the ISENSE outputs.

C. Project impact

The impact of project results can be analyzed in terms of both social and scientific/technological aspects. Thus, we can affirm that the impact is equally spread among the involved organizations. Considering the involved universities, the main advantages in participating in this project are associated with the possibility to improve the skills of administrative staff working in the orientation office, allowing them to learn the basic of a national sign language to communicate with deaf students, and with the availability at the end of the project of an innovative digital tool to carry out the orientation in a more attractive and inclusive way. Moreover, each university will enlarge the number of deaf students, increasing the inclusion level in education and fulfilling their social and civic duty. Considering the deaf associations, they are the main stakeholders and target group of this project, also considering the involvement of young people, which is one of the priorities of this call. Thanks to the project, deaf associations can create a network of knowledge that can be exploited to guarantee the improvement of social inclusion and well-being of their members.

IV. CONCLUSIONS

Artificial intelligence (AI) together with today's data acquisition systems allow the development of applications useful for overcoming the communication obstacles of deaf or hearing impaired people. This is fundamental in real life context, such as education. An overview of the artificial intelligence applications has been provided, highlighting the future developments that should be conducted to overcome the still existing limitations. Finally, the aims and outcomes of ISENSE have been introduced, as proof of concepts of the still ongoing research project.

ACKNOWLEDGMENT

This research was co-funded by European Union Committee within the Erasmus+ Programme 2022 Key Action 2: Cooperation Partnership in Higher Education (Agreement n. n. 2022-1-IT02-KA220-HED-000089554). EU is not responsible for any contents of the present paper.

REFERENCES

[1] Eleanor Busby, "Lack of support in universities drives deaf students to consider leaving degrees," *Independent*, 2019.

[2] W. Aljedaani, M. Aljedaani, E. A. AlOmar, M. W. Mkaouer, S. Ludi, and Y. B. Khalaf, "I Cannot See You—The Perspectives of Deaf Students to Online Learning during COVID-19 Pandemic: Saudi Arabia Case Study," *Educ Sci (Basel)*, vol. 11, no. 11, p. 712, Nov. 2021, doi: 10.3390/educsci11110712.

[3] O. Oreshkina and A. Gurov, "Solving the Problem of Limited Content Accessibility in Natural Science Disciplines for Students with Hearing Impairments in Technical University," in *2019 IEEE Global Engineering Education Conference (EDUCON)*, IEEE, Apr. 2019, pp. 345–351. doi: 10.1109/EDUCON.2019.8725219.

[4] B. S. Parton, "Sign Language Recognition and Translation: A Multidisciplined Approach From the Field of Artificial Intelligence," *J Deaf Stud Deaf Educ*, vol. 11, no. 1, pp. 94–101, Oct. 2005, doi: 10.1093/deafed/enj003.

[5] L. R. Cerna, E. E. Cardenas, D. G. Miranda, D. Menotti, and G. Camara-Chavez, "A multimodal

LIBRAS-UFOP Brazilian sign language dataset of minimal pairs using a microsoft Kinect sensor," *Expert Syst Appl*, vol. 167, p. 114179, Apr. 2021, doi: 10.1016/j.eswa.2020.114179.

[6] N. Adaloglou *et al.*, "A Comprehensive Study on Deep Learning-Based Methods for Sign Language Recognition," *IEEE Trans Multimedia*, vol. 24, pp. 1750–1762, 2022, doi: 10.1109/TMM.2021.3070438.

[7] S. Cox *et al.*, "Tessa, a system to aid communication with deaf people," in *Proceedings of the fifth international ACM conference on Assistive technologies*, New York, NY, USA: ACM, Jul. 2002, pp. 205–212. doi: 10.1145/638249.638287.

[8] M. McConnell and M. E. Foster, "Two Dimensional Sign Language Agent," in *Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents*, New York, NY, USA: ACM, Oct. 2020, pp. 1–3. doi: 10.1145/3383652.3423898.

[9] I. Papastratis, K. Dimitropoulos, D. Konstantinidis, and P. Daras, "Continuous Sign Language Recognition Through Cross-Modal Alignment of Video and Text Embeddings in a Joint-Latent Space," *IEEE Access*, vol. 8, pp. 91170–91180, 2020, doi: 10.1109/ACCESS.2020.2993650.

[10] S. Gibet, "Building French Sign Language Motion Capture Corpora for Signing Avatars," in *Workshop on the Representation and Processing of Sign Languages*, Miyazaki - Japan, 2018.

[11] P. Jedlička, Z. Krňoul, J. Kanis, and M. Železný, "Sign Language Motion Capture Dataset for Data-driven Synthesis," in *LREC2020 9th Workshop on the Representation and Processing of Sign Languages: Sign Language Resources in the Service of the Language Community, Technological Challenges and Application Perspectives*, Marseille, France, 2020.

[12] T. Raghuveera, R. Deepthi, R. Mangalashri, and R. Akshaya, "A depth-based Indian Sign Language recognition using Microsoft Kinect," *Sādhanā*, vol. 45, no. 1, p. 34, Dec. 2020, doi: 10.1007/s12046-019-1250-6.

[13] N. Adaloglou *et al.*, "A Comprehensive Study on Deep Learning-Based Methods for Sign Language Recognition," *IEEE Trans Multimedia*, vol. 24, pp. 1750–1762, 2022, doi: 10.1109/TMM.2021.3070438.

[14] C. Lu, S. Amino, and L. Jing, "Data Glove with Bending Sensor and Inertial Sensor Based on Weighted DTW Fusion for Sign Language Recognition," *Electronics (Basel)*, vol. 12, no. 3, p. 613, Jan. 2023, doi: 10.3390/electronics12030613.

[15] M. S. Amin, S. T. H. Rizvi, and Md. M. Hossain, "A Comparative Review on Applications of Different Sensors for Sign Language Recognition," *J Imaging*, vol. 8, no. 4, p. 98, Apr. 2022, doi: 10.3390/jimaging8040098.

[16] Jie Huang, Wengang Zhou, Houqiang Li, and Weiping Li, "Sign Language Recognition using 3D convolutional neural networks," in *2015 IEEE International Conference on Multimedia and Expo (ICME)*, IEEE, Jun. 2015, pp. 1–6. doi: 10.1109/ICME.2015.7177428.

[17] A. Mannan, A. Abbasi, A. R. Javed, A. Ahsan, T. R. Gadekallu, and Q. Xin, "Hypertuned Deep Convolutional Neural Network for Sign Language Recognition," *Comput Intell Neurosci*, vol. 2022, pp. 1–10, Apr. 2022, doi: 10.1155/2022/1450822.

[18] M. A. Hossen, A. Govindaiah, S. Sultana, and A. Bhuiyan, "Bengali Sign Language Recognition Using Deep Convolutional Neural Network," in *2018 Joint 7th International Conference on Informatics, Electronics & Vision (ICIEV) and 2018 2nd International Conference on Imaging, Vision & Pattern Recognition (icIVPR)*, IEEE, Jun. 2018, pp. 369–373. doi: 10.1109/ICIEV.2018.8640962.

[19] R. P. Meier, "Sign Language Acquisition," in *Oxford Handbook Topics in Linguistics*, Oxford University Press, 2016. doi: 10.1093/oxfordhb/9780199935345.013.19.

[20] A. Mittal, P. Kumar, P. P. Roy, R. Balasubramanian, and B. B. Chaudhuri, "A Modified LSTM Model for Continuous Sign Language Recognition Using Leap Motion," *IEEE Sens J*, vol. 19, no. 16, pp. 7056–7063, Aug. 2019, doi: 10.1109/JSEN.2019.2909837.

[21] OLOS® (European Patent No. 2965172, Italian Patent MISE UIBM n. 000141612, European trademark UAMI n.011115367, European design UAMI n. 002572685-001).

[22] D. Baldacci, R. Pareschi, "The OLOS® Way to Cultural Heritage: User Interface with Anthropomorphic Characteristics", in *World Academy of Science, Engineering and Technology International Journal of Civil and Architectural Engineering*, Vol. 14, n. 10, 2020, pp. 331-335.

[23] F. Marulli, R., Pareschi, and D. Baldacci, "The internet of speaking things and its applications to cultural heritage", in *Proceedings of the International Conference on Internet of Things and Big Data*, Association for Computing Machinery, New York, 2016, pp. 107-117, doi:10.5220/0005877701070117.