IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received February 16, 2022, accepted March 12, 2022, date of publication March 22, 2022, date of current version April 21, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3161440

Technological Solutions for Sign Language Recognition: A Scoping Review of Research Trends, Challenges, and Opportunities

BOBAN JOKSIMOSKI[1], EFTIM ZDRAVEVSKI[1], PETRE LAMESKI[1], IVAN MIGUEL PIRES[2,3], FRANCISCO JOSÉ MELERO[4,5], TOMÁS PUEBLA MARTINEZ[4], NUNO M. GARCIA[2], MARTIN MIHAJLOV[6], IVAN CHORBEV[1], AND VLADIMIR TRAJKOVIK[1]

[1]Faculty of Computer Science and Engineering, Saints Cyril and Methodius University in Skopje, Skopje 1000, North Macedonia
[2]Instituto de Telecomunicações, Universidade da Beira Interior, 6201-001 Covilhã, Portugal
[3]Escola de Ciências e Tecnologia, Universidade de Trás-os-Montes e Alto Douro, 5001-801 Vila Real, Portugal
[4]Technological Centre of Furniture and Wood of the Region of Murcia (CETEM), 30510 Murcia, Spain
[5]Telecommunication Networks Engineering Group, Polytechnic University of Cartagena, 30202 Cartagena, Spain
[6]Laboratory for Open Systems and Networks, Jožef Stefan Institute, 1000 Ljubljana, Slovenia

Corresponding author: Eftim Zdravevski (eftim.zdravevski@finki.ukim.mk)

This work was supported in part by the Faculty of Computer Science and Engineering, Saints Cyril and Methodius University in Skopje, Skopje, North Macedonia; in part by the FCT/MEC through national funds; in part by the FEDERPT2020 Partnership Agreement under Project UIDB/50008/2020. This article is based upon work from COST Actions CA16226 Indoor living Space Improvement: Smart Habitat for the Elderly, and CA19121 Network on Privacy-Aware Audio- and Video-Based Applications for Active and Assisted Living, supported by COST (European Cooperation in Science and Technology). COST (European Cooperation in Science and Technology) is a funding agency for research and innovation networks. Our Actions help connect research initiatives across Europe and enable scientists to grow their ideas by sharing them with their peers. This boosts their research, career and innovation. www.cost.eu.

ABSTRACT Sign languages are critical in conveying meaning by the use of a visual-manual modality and are the primary means of communication of the deaf and hard of hearing with their family members and with the society. With the advances in computer graphics, computer vision, neural networks, and the introduction of new powerful hardware, the research into sign languages has shown a new potential. Novel technologies can help people learn, communicate, interpret, translate, visualize, document, and develop various sign languages and their related skills. This paper reviews the technological advancements applied in sign language recognition, visualization, and synthesis. We defined multiple research questions to identify the underlying technological drivers that strive to improve the challenges in this domain. This study is designed in accordance with the PRISMA methodology. We searched for articles published between 2010 and 2021 in multiple digital libraries (*i.e.*, Elsevier, Springer, IEEE, PubMed, and MDPI). To automate the initial steps of PRISMA for identifying potentially relevant articles, duplicate removal and basic screening, we utilized a Natural Language Processing toolkit. Then, we performed a synthesis of the existing body of knowledge and identified the different studies that achieved significant advancements in sign language recognition, visualization, and synthesis. The identified trends based on analysis of almost 2000 papers clearly show that technology developments, especially in image processing and deep learning, are driving new applications and tools that improve the various performance metrics in these sign language-related task. Finally, we identified which techniques and devices contribute to such results and what are the common threads and gaps that would open new research directions in the field.

INDEX TERMS Sign language recognition, systematic review, sign language visualization.

I. INTRODUCTION

People are considered to have a hearing loss when they are not able to hear under a hearing threshold of 25dB or less in both ears. Around 430 million people worldwide have disabling hearing loss, and it is estimated that by 2050 over 700 million people will have disabling hearing loss.[1] Hearing

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

[1]World Health Organisation (2022, March 22), Deafness and hearing loss. Retrieved from https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss

loss can be mild, moderate, severe, or profound. ''Hard of hearing'' individuals have a hearing loss ranging from mild to severe, while ''Deaf'' individuals mostly have profound hearing loss.

Hearing loss is one of the most common chronic impairments that appear with age as degeneration of sensory cells. It results from different congenital or acquired causes (e.g., genetic causes, complications at birth, infectious diseases, exposure to excessive noise, among others). It is known as ''Presbycusis'' and affects approximately one-third of people over 65 years of age, and it cannot be reversed [1]. However, it can be effectively treated with common hearing aids and communication devices. Moreover, some disruptive assistive technologies based on Artificial Intelligence are emerging to improve the well-being and quality of life of hard of hearing and deaf individuals.

The problems tackled by assistive technologies are numerous and spread around different fields. Newly developed technologies and Information Technology (IT) driven and supported systems are at the forefront of the assistive technologies field. One of the many fields of application of assistive technologies is improving the life quality of hard of hearing and deaf persons.

Hard-of-hearing persons usually communicate through spoken language and can benefit from hearing aids, cochlear implants, and other assistive devices, as well as captioning. On the other hand, deaf persons face challenges in everyday communication that can significantly impact daily life, causing feelings of loneliness, isolation, and frustration, particularly among older people with hearing loss. One of the challenges for this community is proper integration in everyday ''normal'' society.

Sign languages are classified as natural languages and exhibit all the design features of other natural languages [2], [3]. Sign languages' visual and spatial nature and their variability present an exciting challenge for research in several scientific fields, such as linguistics, medicine, machine learning, computer vision, natural language processing, and computer graphics.

Interpretation and linguistics of sign languages are primarily concerned with the meaning conveyed using the sign language. With the recognition of sign languages as natural languages in the late 1970 and early 1980, linguist research took an in-depth look into this field [4]. Neural aspects are considered for fully grasping the connection between sign and phonetic languages. Natural language processing is also concerned with interpretation, a task similar to the interpretation and comprehension problems.

Sign language synthesis and visualization is an area that tackles visualization issues of sign languages and the creation of signed speech. This field researches and develops means of realistically ''spoken'' sign languages, using video and image sequences or digital characters [5].

Sign Language Recognition (SLR) is the scientific area responsible for capturing and translating sign speech using computer vision and artificial intelligence techniques [6]. The research benefited in the last decade by introducing general-purpose sensors such as *e.g.*, Intel RealSense$^{\text{TM}}$, Microsoft Kinect, LeapMotion Controller, etc. [7], [8].

In this paper, we focus on sign language research, a relatively young field that took off with the breaking research of William Stokoe [9], [10]. When tackling sign languages, we have identified three areas prominent for research: interpretation and linguistics of sign languages, sign language synthesis and visualization, and sign language recognition (SLR). Considering the importance of sign languages for the communication of millions of people across the world and the rapid technological developments, this article performs a scoping review of the most recent technologies applied in sign language recognition. One of the novelties of our work is that we start from a systematic mapping (i.e., a scoping review) where the trends in literature in the last decade are explored so that emerging technologies, sensors, algorithms can be analyzed. Then, we proceed in a more detail-oriented fashion that is more common of systematic reviews where we identify the most significant works in the field. To that end, we use an NLP-based tool to support and simplify the literature review, which, as a methodological approach, is novel and considerably reduces the manual effort usually involved in such activities.

The remainder of the paper is structured as follows. Section II discusses the related systematic reviews, and section III discusses our methodology for this paper. Section IV presents the results of our work, and section V discusses them. Finally, section VI concludes the paper.

II. RELATED WORK

This section covers other literature reviews relevant to the domain of this study. First we analyze most important recent works for sign language recognition. Then, we analyze recent systematic surveys in gesture recognition. Each area of research into sign languages has a vast amount of valuable research that needs to be filtered and analyzed. Surprisingly, few literature reviews have considered gesture and sign language analysis.

A. LITERATURE REVIEWS IN SIGN LANGUAGE RECOGNITION

In 2011, Cooper *et al.* [6] provided an excellent review on sign language recognition approaches and challenges. The publication focuses on all aspects of sign language recognition (SLR): sign language linguistics, data acquisition related to sign languages, and approaches for sign language recognition. The linguistics portion describes some complexity in sign languages, like body posture and non-manual features. The data acquisition section enumerates the approaches for acquiring sign languages available at that time (data gloves, images and video input, depth-based cameras). It also listed the prominent sign language datasets that have been obtained using the methods. Also, the review discusses the main approaches of hand pose acquisition, as well as individual fingerspelling and other non-manual features. As for recognition, the authors describe the state-of-the-art approaches, focusing on individual or continuous sign recognition. Various approaches using various types of Neural Networks, Hidden Markov Models and their variations (like Parallel HMM), decision trees, and self-organizing maps are utilized for various parts of SLR.

Similarly, in 2013, Toiba and Elons [11] discuss the developments in sign language recognition, taking into consideration much of the results mentioned by Cooper *et al.* and focusing on neural network approaches for SLR. They report recent results (as of the time) near 90% recognition for specific and carefully chosen cases. The paper, however, does not present a methodology of the surveyed results and provides a more general overview of the field.

Authors of [12] created a taxonomy to describe the latest research divided into four main categories: development, framework, other hand gesture recognition, and reviews, and surveys. Also, the authors created a roadmap and discussed the limitation of the use of this technology.

Er-Rady *et al.* [13] in 2017 published a survey about automatic sign language recognition. In it, they describe the approaches as well as the best methodologies for sign language recognition. In the paper, the authors explain the complexity of data acquisition as well as the features of the manual features of sign languages as well as other properties that the modality allows (*e.g.* facial expressions). They also describe the most common components that should be present in a system that uses an Automatic Speech Recognition system (ASLR).

A comprehensive review of recent works for sign language recognition is presented in [14], where authors defined a taxonomy to group existing works and discuss their pros and cons. The article also discusses features, modalities, evaluation metrics, applications, and datasets. Even though our work is on a similar topic, we focus on technological solutions, particularly those applying advances in camera technologies and machine learning methods. Likewise, we also analyze the publications in the recent decade at various publishers and identify the emerging trends and promising directions for future work.

A quantitative survey of state of the art in sign language recognition is presented in [15], where authors centered the research for the creation of a framework of the hand recognition system composed by image datasets, preprocessing, feature extraction, image segmentation, and classification with supervised and unsupervised learning strategies. The authors' motivation was to increase the susceptibility, decrease computational complexity, increase classification rate, and reduce the error rate. Similarly, our review is focused on the different methods for the recognition of sign language, but we included more recent studies.

Authors of [16] present a review of hand gesture and sign language recognition techniques. Unlike it, our survey presents more quantitative metrics about the trends and also shows the most relevant works, including novel techniques that have emerged in the recent years not covered by this paper.

In [17] a critical review and analysis of machine learning methods for sign language recognition is presented. It provides a comprehensive review of artificial intelligent methods applied in sign language recognition systems and brief overview of the feature extraction and segmentation methods. Unlike it, our work covers the wider field of visualization and interpretation of sign languages.

In [18], the authors presented the analysis of existing methods related to sign language capturing, recognition, translation, and representation, also showing their pros and cons. In addition, the authors analyzed the applicability of the studies. However, this study is only focused on artificial intelligence techniques and their applicability for sign language recognition.

B. LITERATURE REVIEWS IN GESTURE RECOGNITION

A field that is in close connection to Sign language recognition is gesture recognition. Significant research efforts have been made in that direction. As more notable, Sagayam and Hemanth have compiled a survey about the use of hand posture and gestures recognition [19]. In it, the authors describe the most effective research approaches as well as results from the recognition. Specifically, all the approaches mentioned (decomposition, decision trees, support vector machines, hidden Markov models, among others) can produce a 90+% recognition rate. In the paper, the authors have a special section dedicated to studying hand motion analysis (HMA), mentioning HMM and its variations as the most common way to analyze hand motion.

The gestural language signs differ from different countries, where, in 2020, the authors of [20] explored the recognition of Portuguese Sign Language. Several classifiers, including Deep Neural Network (DNN), Multilayer Perceptron (MLP), Support Vector Machines (SVM), Hidden Markov Models (HMM), and Subspace Gaussian Mixture Model (SGMM), were used for the classification of Portuguese Sign Language between 2021 and 2018, proving its reliability.

In [21], the authors presented a literature review related to gesture recognition in a mobile context as well as facial recognition in sign languages. The classification models used in the literature included SVM, Hierarchical Temporal Memory, Feedforward backpropagation neural network, RandomForest, and MLP. The use of these devices makes feature extraction easier.

III. METHODOLOGY

This study adopted a scoping review methodology to identify and process the literature on sign language recognition, visualization, and interpretation published from January 2010 to September 2021. Using a scoping technique, we aimed to examine the research evidence in the broad field of technological solutions and sign languages, analyzing technology trends, including the resolved and emerging issues. The lack of a qualitative analysis of identified papers, the broad topic range, and the number of studies involved defined our approach as a scoping review and differentiated it from a systematic review [22], [23]. The purpose of this study fully complies with the aims of a scoping review ''to search, select and synthesize the knowledge addressing an exploratory question to map key concepts, types of evidence, and research gaps'', as defined in [24].

The methodological framework for scoping reviews proposed in [22] was adopted in this study, composed of five stages: Identification of a research question; Identification of relevant studies; Study selection; Charting the data; and Collating, summarizing, and reporting the results. The identification of relevant studies and study selection stages in the scoping review methodology corresponds to the ''Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)'' [25] workflow phases: study collection, scanning, and eligibility evaluation (see Fig. 1). Therefore, after the scoping review, our research identifies the relevant works according to the defined research questions and inclusion criteria and discusses them in more detail.

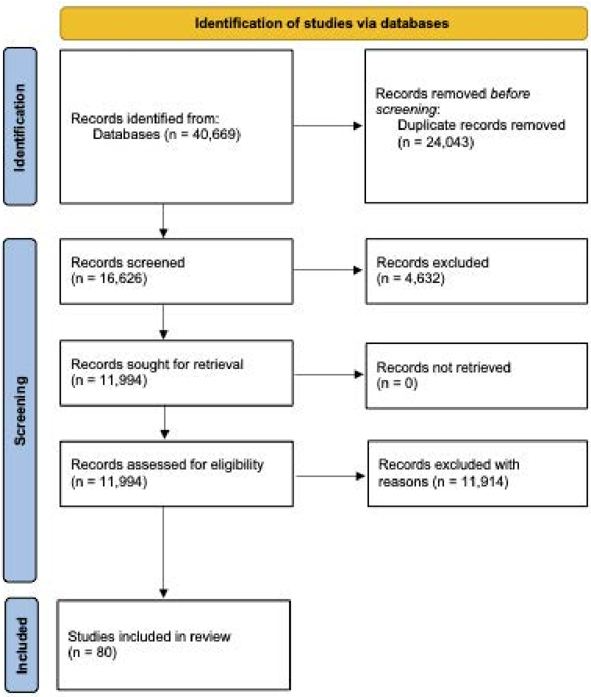

FIGURE 1. Flowchart of the PRISMA-based selection process.

A. RESEARCH QUESTIONS

The primary research questions of this review were defined as follows:

RQ1 What are the prominent research topics and publication trends in literature?

RQ2 In which way machine learning and deep learning algorithms can improve sign language recognition and interpretation?

RQ3 Which approaches are most suitable for recognizing sequences of signs in sign language?

RQ4 Is there any considerable difference in the state-of-the-art in sign language recognition, visualization, and synthesis performance between different sign languages?

RQ5 What are future research directions for computer translation between sign languages and written/spoken languages?

To answer these research questions, we performed an initial search and screening of papers and then selected the most relevant works and analyzed them in more detail. The following subsections describe the scoping review parameters, the search strategy, and the inclusion criteria.

B. SCOPING REVIEW PARAMETERS

This stage employs an NLP toolkit described in [26] enabling both automated search, scanning, and processing procedures. Its input parameters are a collection of search terms used to identify potentially relevant articles, the target digital libraries (i.e., Elsevier, Springer, IEEE, PubMed, and MDPI), and a set of properties that should be satisfied by the identified articles. To determine the initial pool of articles that are then evaluated for eligibility, we performed the search in the aforementioned digital libraries with the search phrases:

- ''sign language recognition'' OR
- ''sign language synthesis'' OR
- ''sign language interpretation'' OR
- ''sign language linguistics'' OR
- ''sign language processing''

Keywords are search terms or phrases used to query a digital library (e.g., ''sign language recognition ''). First, eventual duplicates in the results are removed in a later phase. Then to evaluate whether an article is relevant to the research question, we define properties (see Fig. 2). Properties are words or phrases that are being searched in the title (i.e., text in italics in Fig. 2, such as ''Sign language production''), abstract or keywords section of the articles identified with the keywords. Properties can also have synonyms (i.e., the comma-separated terms after the ones in italics, such as ''sign language generation'', which is a synonym of ''sign language production''). Property groups are thematically, semantically, or otherwise grouped properties for a more comprehensive presentation of results in charts (e.g., text in bold in Fig. 2, such as ''Domain-specific topics'').

The start year indicates the starting year of publishing (inclusive) for the papers included in the study. The end year is the last year of publishing (inclusive) to be considered in the study. This review encompasses studies published from January 2010 to September 2021. The minimum of the relevant properties denotes the minimum number of properties that an article must contain to be considered suitable. This study set this value to 4, providing the right balance between falsely identifying relevant papers and potentially missing a relevant one.

When researchers perform a scoping review according to the methodology mentioned above, the actual tasks involve searching digital libraries with different search phrases, often involving complex Boolean conditions. The NLP toolkit

AI Topics:
- *NLP*, Natural language processing, Natural language understanding
- *NLG*, Natural language generation, language production
- *NLU*, Natural language understanding
- *AI*, Artificial Intelligence
- *ML*, Machine Learning
- *DL*, Deep Learning
- *SVM*, Support Vector Machines
- *Ensemble Classifiers*, XaBoost, XGBoost, GBT, GBTs, Gradient Boosted, random forest, extra trees, extremely randomized trees
- *CNNs*, Convolutional networks, Convolutional nets, Convolutional neural, CNN
- *RNNs*, Recurrent neural network, Recurrent nets, Recurrent neural, RNN
- *HMMs*, Hidden Markov Models, HMM
- *NNs*, Neural networks, Neural net, Neural nets, ANN, NNs, ANNs

Domain-specific topics:
- *Visualization*
- *Recognition*
- *Depth Camera*, Real Sense, Kinect, RGBD
- *Spatial Analysis*
- *Spatial Synthesis*
- *Spatial Visualization*, 3D Visualization, Three-dimensional visualization
- *Sign Language Modelling*
- *Sign Language Production*, Sign Language Generation
- *Sign Language Recognition*
- *Image Analysis*
- *Video Processing*, Image Processing

General Topics:
- *Assistive Technologies*
- *3D*, Three-dimensional, Avatar
- *Sign Language*
- *Deaf*, Hearing Impaired

Study Type:
- *Review*
- *Survey*
- *Randomized trial*, Randomized control trial, Randomized controlled trial, RCT
- *Clinical trial*
- *Observational study*
- *Cohort*
- *Experimental study*

FIGURE 2. Property groups, properties and synonyms used as search input.

counterpart to these phrases are the keywords described above. By screening the title and abstract, a reviewer determines whether the article is relevant to the study. The NLP toolkit is automated using the properties and synonyms to define what we are looking for in a report. Articles that contain more properties are considered more relevant. Undoubtedly, a human reader might better understand the context and better assess the relevance of an article. However, the NLP tool kit mimics these tasks in an automated and more thorough way, providing a higher efficiency in the scoping review process in far less time. For more information about the actual implementation, we refer the readers to [26].

C. SEARCH STRATEGY

The initial screening of duplicates and eligibility analysis was performed automatically using the aforementioned NLP-based toolkit utilizing Natural Language Processing (NLP) techniques, as described in [26]–[28].

The PRISMA methodology is described with details for our case in fig. 1 and it was used for the selection and processing of the collected research articles. Upon merging the results of multiple independent keyword-based searches, the collected papers were screened to remove duplicates. In addition, the screening process discarded articles that were not published in the required period or for which the title or abstract could not be analyzed due to parsing errors, unavailability, or other reasons.

The qualitative analysis of the remaining articles relied on the NLP toolkit that automates the initial steps in the PRISMA approach, significantly reducing the number of articles for manual evaluation. The identified articles were labeled as potentially relevant if they contained at least four predefined properties in their title or abstract (considering the above NLP-enhanced searching capabilities, thus performing a rough screening).

The NLP toolkit generates a spreadsheet file with the following fields: DOI (Digital Object Identifier), link, article title, authors, publication date, number of citations, abstract, keyword, source, publication title, affiliations, number of different affiliations, countries, number of different countries, number of authors, number of found property groups, and number of found properties. These additional properties aided the manual search of the articles with specific filtering criteria. Each of the articles in this reduced set was subsequently manually retrieved from their publisher and analyzed for potential inclusion in the qualitative and quantitative synthesis.

Finally, we used a more traditional systematic survey approach to provide a detailed analysis of the most relevant works per the previously mentioned topics. Five reviewers independently evaluated the remaining studies, and their suitability was determined with the agreement of all parties. Thus, the studies were evaluated per the previously defined eligibility criteria and the reviewers' evaluation of the study's relevance, quality, and impact.

D. INCLUSION AND EXCLUSION CRITERIA

Considering the abundance of research available on this topic, we defined multiple inclusion criteria related to some qualitative properties of the articles, as follows:

IC1 The articles should be written in English and should be published between 2010 and 2021, both inclusive;

IC2 The sign languages treated in the article should be explicitly mentioned. It should contain at least one of the most common sign languages per Ethnologue [29] estimates (40 000 speakers or more). The list of languages includes, but it's not limited to: Indian, Pakistani, Brazilian, American, British, Mexican, Japanese, Russian, French and Chinese Sign language.[2] Publications that do not explicitly mention the sign language are excluded; Alternatively, if the research is sign language-independent, it should provide significant value in respect to IC3 and IC4.

[2]Ethnologue does not give estimates of the speaker population of Chinese Sign Language, but it is included because of its perceived size.

IC3 Articles related to sign language recognition are included if:

- IC3a They describe the algorithms used (*e.g.* Hidden Markov Models, deep learning, convoluted neural networks, among others);
- IC3b It has been tested on a publicly available data set, which is explicitly referenced;
- IC3c The results are quantitatively described (*i.e.*, in terms of accuracy, F-score, among others);

IC4 Articles related to sign language synthesis and visualization are included only if:

- IC4a A successful visualization has been made, and the output has been qualitatively or quantitatively assessed by an expert or proficient sign language speaker (in that language);
- IC4b The article discloses the used technology or device (*e.g.*, depth-enabled cameras, smartphones, Kinect, among others).

Considering that the screening part of the methodology was automated, there were cases when the potentially relevant articles were in fact not relevant based on the defined inclusion criteria above. Therefore, further manual inspection of the remaining articles (1856 as described below in the Results section), was key to identify whether the article should be included per the defined criteria above. Some of the works were included in the in the quantitative trends, but a minor subset was chosen for a more in-depth qualitative analysis.

IV. RESULTS

The funnel of article selection and screening is presented in Fig. 1. In the process of identification, 40,669 studies were deemed potentially relevant to the research field. Duplicates that emerged in the independent searches were removed, thus reducing the total number to 16,626 potentially relevant studies. The first screening process further eliminated 4632 studies outs, or if errors occurred while parsing or analyzing the title and abstract of the studies.

The remaining 11,994 studies underwent an automated eligibility assessment using advanced methods provided by the NLP toolkit [26]–[28]. After preprocessing, the articles were tagged with identified properties, and articles containing less than four properties were removed (i.e., 10,138 articles).

Of the remaining research publications, 1856) were deemed eligible for further manual inspection and inclusion in qualitative analysis. The statistics on the number of the collected articles, duplicates, articles with invalid time or the articles with incomplete data, and relevant articles are presented in Figure 3, for each digital library.

A. STATISTICS PER DIGITAL LIBRARY

Regarding Fig. 3, the Springer digital library provides by far most of the potential candidates. However, after applying the additional PRISMA-defined filters based on the defined properties, the number of relevant articles is severely reduced, as visible in Fig. 4. This figure shows that most of the relevant articles for the defined research questions originate from IEEE Xplore.

The distribution of the number of collected and relevant articles per year is presented in Figure 5. From the chart, we can identify increasing research in the field, with a noticeable increase in the relevant research in time, with a small peak in 2014. However, the most relevant number of articles per the defined properties is in 2020.

Depending on the digital library, the ratio of the relevant papers containing specific keywords changes (as seen in Figure 6). For example, the Springer and IEEE digital library focus on technology-driven research into sign languages, with most studies published for sign language recognition and sign language visualization. Thus, it is a logical outcome of an engineering-based organization and publication. Springer has a similar tendency, although it does have a more general-purpose nature. PubMed shows a difference, recognizing as the most identified keyword, with a significant percentage of publications identified as sign language processing, linguistics, and interpretation. MDPI seems to have minor publications focused on this subject.

B. STATISTICS OF KEYWORD BASED QUERIES

Selected keywords were used to map the literature corpus on sign language research. It is done concerning the set research questions that appear in the relevant articles with different distributions.

Regarding the source of relevant articles in relation to the search keywords, the trend is displayed in Fig. 6.

In figure 7 we have a distribution of the number of research publications that were identified with each keyword for the period January 2010 until September 2021. Here we present the annual number of research papers identified by the search engine of the three libraries with the defined keywords and additionally filtered manually based on their relevance to the specified properties. Please note that the internals of their search engines are not known, and the libraries might differ in the way they look for these keywords: only in the title, keywords section, abstract, or the whole article.

C. EXTRACTION OF STUDY CHARACTERISTICS

The evaluated research studies are segmented into two tables. In Table 1, we have included the studies that have sign language recognition as the main research question. In the table, we present the research data in the following columns: *reference*, *year*, *key findings*, *test protocol*, *dataset*, *sign language*, *problem sub-type* and the *data input type* used for the publications. The column *key findings* summarises the most important parts of the publication, as well as the method that is being used to achieve the results. The *testing protocol* describes the training and testing phases (if present) in the paper. Next, the *dataset* column describes or references the data that is used for training or validating the results. The column *sign language* denotes the SL that is being used (if applicable). Next, the *problem sub-type* column describes whether the recognition is static, continuous,

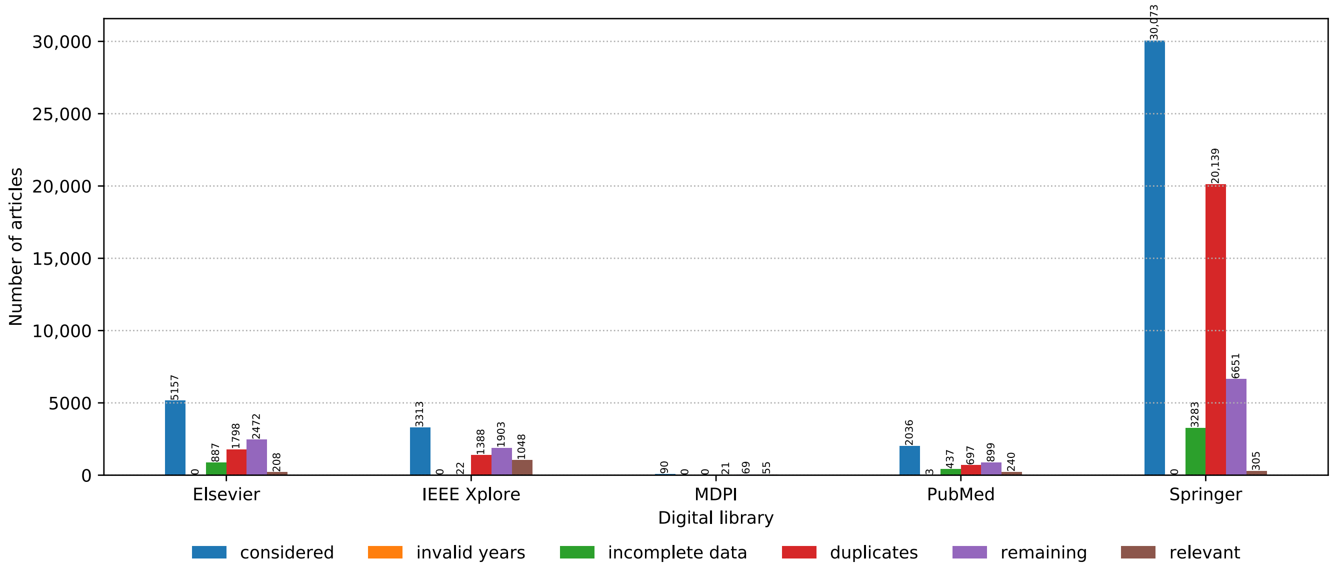

FIGURE 3. **Number of articles (considered, rejected for various selection and eligibility issues, remaining, and relevant) per digital library.**

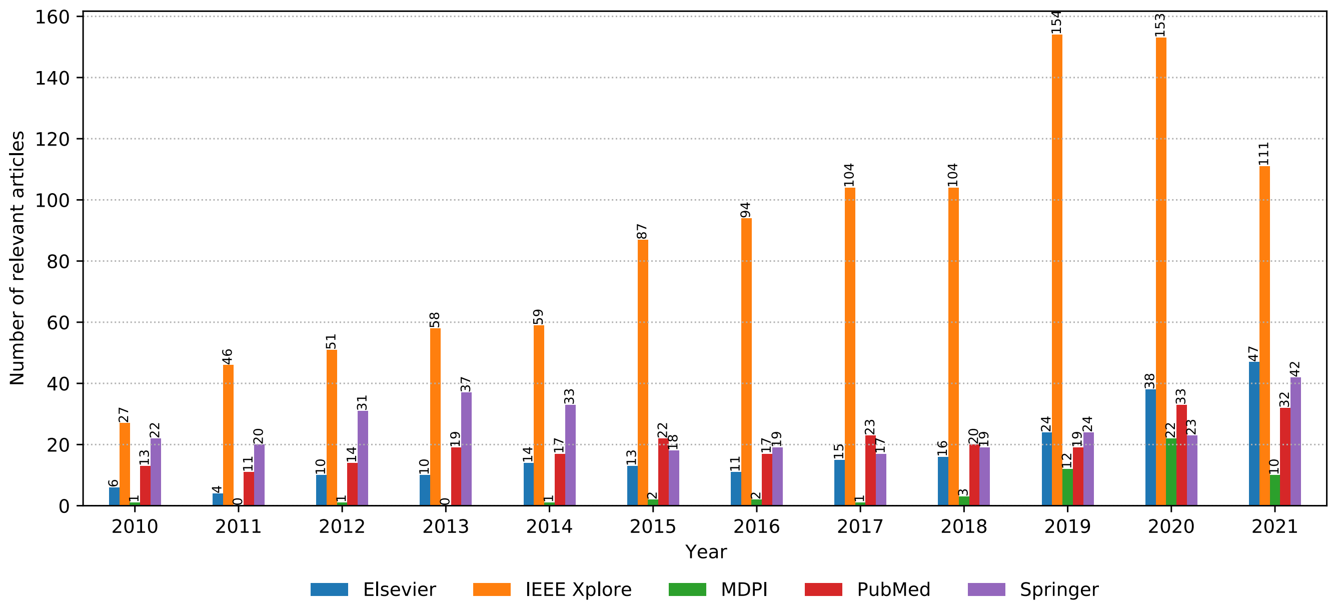

FIGURE 4. **Number of identified relevant articles per year (2010–2021) per digital library.**

isolated, etc. Finally, the *data input type* column shows the device used to capture input data.

V. DISCUSSION

From Fig. 6 and Fig. 7 we can infer some interesting trends. The most active field for sign language research, indexed in the specified libraries, is sign language recognition, while sign language visualization and sign language processing are trailing behind.

This disparity in the number of publications can be attributed to multiple factors: the nature of the area of research (e.g., mostly engineering vs. medical), the availability of new devices and methods (e.g., the introduction of depth-based cameras, development of the machine learning field), etc.

A. EVOLUTION OF RESEARCH INTO SIGN LANGUAGES

One of the contributions of this study is categorizing the research around sign languages through the years. In the 1990s, most research related to sign languages was based on linguistic aspects (e.g., acquisition, evolution, teaching, analysis, etc.).

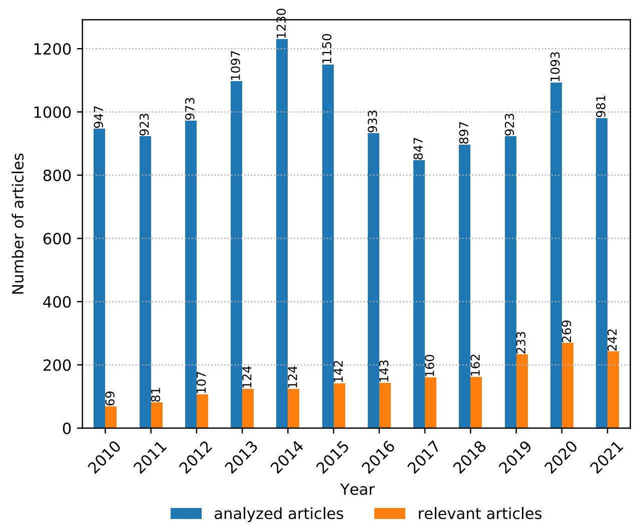

FIGURE 5. Number of analyzed and relevant articles in the span of 2010–2021.

1) SIGN LANGUAGE ANALYSIS

When talking about the acquisition of sign languages, Volterra and Erting [30] published a book that compiled the most notable research studies about sign language acquisition for children of different ages and backgrounds. The amount of the presented research (21 research studies) enables the authors to establish correlations between sign languages and spoken languages. The most important conclusion is that children's stages when developing verbosity in sign and spoken language correspond to one another, and the acquisition phases are the same. Similar research was published by Boyes Braem [31] focusing on one deaf participant from deaf parents fluent in American sign language. Acquisition of sign languages by primates is also studied [32].

2) CHALLENGES IN SIGN LANGUAGE RESEARCH

Due to the visual modality present in sign language research, several considerations have to be taken into account. When discussing sign language recognition, there is a distinction between isolated, single-sign recognition and recognising and interpreting sign language speech (continuous integration). Isolated recognition focuses on providing a specific term (e.g. noun, verb, adjective, etc.) to a visual image or sequence of images (in case of dynamic, motion based sign).

Sign language data can come from different sources, but usually take the form of:

- **image/sequence of images/videos** that have some form of color data (e.g. standard cameras).
- **transformation data** that can be acquired from specialized hardware like data gloves or through usage of motion-tracking techniques. Usually this data gives information for the movement (rotation and/or location of joints) of the arm and hands. The data is presented either as hierarchical transformations in respect to the anatomy, or as raw transformation/joint pairings that need to be further processed.
- **depth data** that comes from specialized hardware that includes depth sensors. The most common input devices are Microsoft® Kinect™ (discontinued) and Intel® RealSense™ series of sensors. Kinect is better suited for whole-body movements and works best at larger distances (1-4 meters) while RealSense devices usually require closer distances (0.5 - 1.5 meters) from the sensors.

Some datasets can come with both image and depth data.

Also, the process of recognition for sign languages follows the standard pipeline for computer-vision based methodology. Ideally, when discussing any problem that needs to be processed from computer-vision perspective, it needs to follow the standard steps [17]:

- **Acquisition.** Acquisition of the data that is needed. In the case for sign language recognition, this includes gathering a specific set of images showing different signs under different conditions (e.g., lighting, environment, different people, etc.)
- **Preprocessing.** This can include various preprocessing tasks ranging from standardization of images taken from different devices to be in the same format, to different data augmentation techniques. Namely, to improve the robustness of the methods, this can include subtracting the per-channel mean pixel values calculated on the training dataset; random rescaling, horizontal flips, perturbations to brightness, contrast, and color, random cropping, etc.
- **Segmentation.** This involves the process of identifying and segmenting the relevant parts of the data (image), in this case, it would be image or video segments that are focusing on the hand gestures.
- **Feature extraction.** This process aims to create manageable sets of data that are derived from the initial raw data. This is usually done by reducing the raw data (a.k.a. dimensionality reduction), transformation to other color space, or extracting some custom properties that better describe certain aspects of the objects in the image, so that the classification is more robust and accurate.
- **Classification.** Categorize the input into a specific classes, in this case, recognizing different language symbols from the sign language gestures.

A driver to boost research into sign and gesture recognition was the ChaLearn competition held in 2012 [33] because of the dataset that was made available with it. The competition data set consisted of 50 000 gestures containing RGBD data, using the then-novel Microsoft Kinect sensor. The gestures recorded varied significantly, including signaling for drivers, aircraft, representing numerals, and sports referees. Here, the accompanying depth channel contributes considerably to the approaches taken by the competitors, although HMM-based methods dominate for the recognition of gestures.

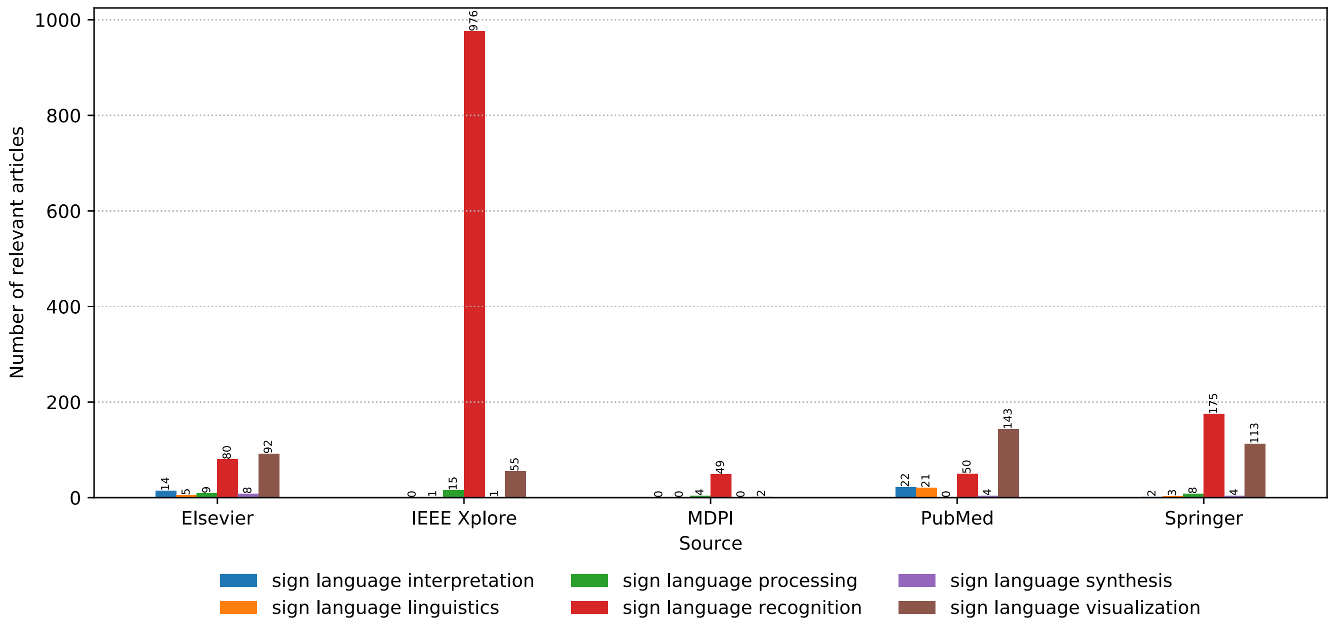

FIGURE 6. Distribution of the number of relevant articles with each of the defined keywords from each of the source publication databases.

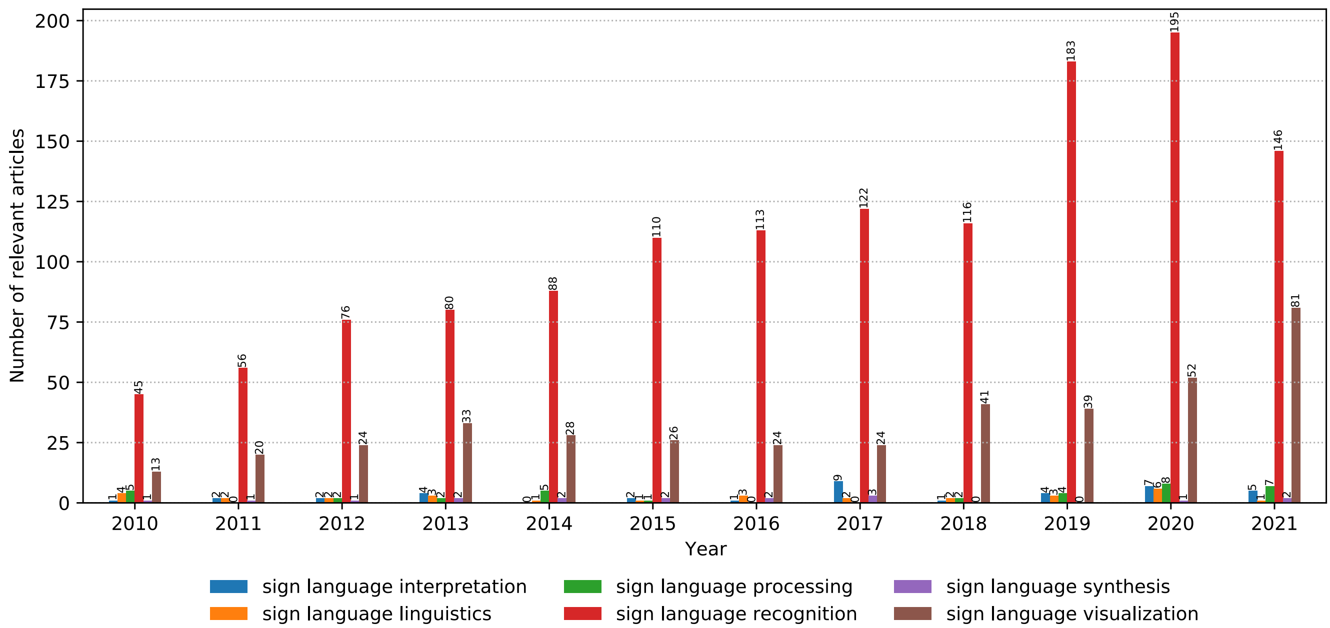

FIGURE 7. Distribution of the number of relevant articles with each of the defined keywords on the annual basis.

3) EDUCATION

The methodologies and results of education of Deaf people in different environments are studied and tested. It includes proposing training models for the education of deaf people, as well as sign language translators [34], [35], and the challenges of learning sign languages for in bilingual environment [36], as well as challenges in learning sign language [37].

4) MEDICAL ASPECTS

From a medical aspect, a lot of research has been conducted on the effect of various medical conditions on sign speaking and analyzing the neuro-biological aspects of sign languages. Most notable, cases of deaf patients that have acquired significant brain injuries and the effects on their communication have been investigated, with specifics on different conditions, like left [38], [39] and right hemisphere [40] damage.

The research gave insight into the parts of the brain that are responsible for speech and signing. Further research, using functional MRI, has been used to figure out the parts of the brain that specialize in sign language communication [41], [42]. To summarize the research, a strong indication of independence of modality and language organization has been noticed.

5) TECHNOLOGY ASPECT OF SIGN LANGUAGE STUDIES

The technology aspect of the research of sign languages can also be discussed as part of this analysis. Generally, computer science has more impact on sign language studies, providing mechanisms that vary from data-gathering to visualization.

a: HCI TECHNOLOGIES

With the emergence of the human-computer interface (HCI) technology, some research into closely related areas to sign languages has been published. Schmauks and Wille [43] have hypothesized that in HCI, the motor responses, and especially facial expressions and hand gestures provide additional non-vocal information that needs to be taken into account when interfacing computer. Research into gesture recognition paved the road for developing sign language recognition as an emerging topic. Significant research of hand gesture recognition [44], [45] have foreseen as having an impact on sign language recognition. Bordegoni and Faconti [46] compared and outlined architectures for gesture recognition system, categorizing into two categories: data-based and pictorial-based, while the algorithms for recognition have been specified as based on pattern recognition, neural networks or statistical classification.

b: IMPLEMENTED SYSTEMS

Authors of [47] did a similar analysis while additionally taking into account other modalities, such as head and eye movement. Gilbert *et al.* [47] also discuss and present implementation of the recognition system using Hidden Markov Models (HMMs), particularly ARGo and CADRE. This work also describes an implementation of a system for visualization (SAGA).

In 2001, a tool designed explicitly for linguistic, and computer vision research for American Sign Language was developed. The freely available tool, called SignStream [48], primarily was developed as annotation of multimedia content. Later versions of the tool utilize computer vision algorithms for the annotation of facial expressions and hand shapes. GesRec3D [49] is a system developed primarily for augmentative and alternative communication for people that have a motor or speech disability. The system was designed to recognize arm-gestures using an electro-magnetic tracker (Polhemus).

Other systems such as the one presented in [50], [51], [52], [53] use gloves as sensors and machine learning approaches for detection and recognition of the sign letters. Some approaches even use Radio Frequency (RF) sensors to detect and recognise sign languages [54].

c: HMM-BASED METHODOLOGIES

In this phase (late 1990 - early 2000), various HMM recognition solutions are proposed. The most cited of all is the implementation of Waldherr *et al.* [55], where they create an autonomous helper robot that is controlled using gesture recognition based on implementing modified HMM based on the ''Viterbi'' algorithm. In 2003, Kai and Stiefelhagen [56] present another system for Human-Machine interaction using HMMs for detection of pointing gestures. In this system, the authors use a stereo camera to consider the position and orientation of the arm relative to the body. In their later work [57], Kai and Stiefelhagen additionally incorporate head motions in their recognition system. Kingsley and his coworkers [58] explored the possibility of developing context-sensitive HMM that would utilize the HMM for gesture learning and recognition.

d: DEEP LEARNING BASED METHODS: TRANSFORMERS, 3D CNN AND OTHERS

The authors of [59] proposed an interactive alignment network with iterative optimization for weakly supervised continuous sign language recognition. It is composed of a 3D convolutional residual network for feature learning and classification with sequence modeling. However, it only focuses on using two decoders, such as Long short-term memory (LSTM) and connectionist temporal classification (CTC). The results demonstrated the effectiveness of the implemented decoders.

In [60], a deep learning-based pipeline architecture was proposed for automatic hand sign language recognition using Single Shot Detector (SSD), 2D Convolutional Neural Network (2DCNN), 3D Convolutional Neural Network (3DCNN), and LSTM from RGB input videos. The proposed model outperforms the state-of-the-art. It is an important study for the review presented in this paper because an automatic system was developed to recognize sign language.

The authors of [61] proposed a solution to speech sign language without previous knowledge of sign language. The proposed system converts these rotations through programming to vowels and consonants. Thus, computer vision is an important subject for the dissemination of sign language.

Transformers are superior in learning long-term dependency, hence the sign language Transformer model achieves remarkable progress in Sign Language Recognition (SLR) and Translation (SLT). The following works all report improvements over the state-of-the-art methods and they use different alternatives of transformer architectures.

Authors in [62] describe a method that improves state of the art results in gloss-to-text and video-to-text translation using STMC-Transformer that outperforms translation of GT glosses. Similarly, [63] proposes a Connectionist Temporal Classification (CTC) loss to bind the recognition and translation problems into a single unified architecture that does not require any ground-truth timing information, simultaneously solving two co-dependant sequence-to-sequence learning problems.

TABLE 1. Detailed analysis of SL synthesis, interpretation and visualization research.

Paper	Year	Key findings	Test protocol	Dataset	Sign Languages	Problem sub-type	Data input type
[65]	2022	Word-level sign language recognition based on the Transformer model based on the estimation of the pose of the human body in the form of 2D landmark locations. WLASL, we are able to successfully recognize 63.18% of sign recordings in the 100-gloss subset. For the 300-gloss subset, we achieve recognition rate of 43.78%. With the LSA64 dataset, we report test recognition accuracy of 100%.	Introduce a robust pose normalization scheme which takes the signing space in consideration and processes the hand poses in a separate local coordinate system, independent on the body pose.	WLASL and LSA64	American Sign Language	Isolated recognition	Video
[66]	2021	Propose a new model architecture (PiSLTRc) with content-aware and position-aware convolution layers that explicitly select relevant features using a neighborhood gathering method. Then these features are aggregated with position-informed temporal convolution layers. Reports state-of-the-art performance improvement on translation quality with +1.6 BLEU.	The authors propose injecting the relative position information to the attention mechanism in the encoder, decoder, and even encoder-decoder cross attention.	PHOENIX-2014, PHOENIX-2014-T and CSL	German Sign Language and Chinese Sign Language	Continuous recognition	Video
[50]	2021	Authors use wearable capacitive sensor to obtain signals. Authors use Error Correction Output Code Support Vector Machines (ECOC-SVM) and K -Nearest Neighbour (KNN) classifiers and obtain 99.50% and 97.94% average cross-validation accuracy	800 data points consisting 15 features used for training and the rest of the data for test	Special dataset created by authors using the sensor	American Sign Language	Static recognition	Glove
[51]	2021	A new real-time sign recognition system based on a wearable and low cost sensory glove or data glove, which has 17 sensors with 65 channels, Results show gesture recognition accuracies of 99%, 96% and 93.4% for numbers, alphabet letters and words.	Validation on data collected with the data glove for numbers, alphabet letters and words	Own dataset	Malaysian Sign Language	Static recognition	Glove
[64]	2020	Applies a combination of feature extraction using OpenPose for human keypoint estimation and end-to-end feature learning with Convolutional Neural Networks. Authors obtain an accuracy of 74.7% on a vocabulary of 100 classes on this dataset.	The proven multi-head attention mechanism used in transformers is applied to recognize isolated signs in the Flemish Sign Language corpus.	Flemish sign language (VGT) corpus	Flemish Sign Language	Isolated recognition	Video
[76]	2020	A novel deep learning method for continuous sign language recognition that applies a cross-modal alignment between video and text embeddings to better model the intra-gloss dependencies in sign language recognition. The proposed method reports 24% WER on the German and 2.4% WER on the Chinese dataset.	Train and test splits of the datasets.	RWTH-Phoenix-Weather-2014T, RWTH-Phoenix-Weather-2014 and The Chinese Sign Language (CSL) dataset	German and Chinese Sign Language	Continuous recognition	Video
[60]	2020	A model for hand sign recognition using SSD, 3DCNN, and LSTM from RGB. A hand skeleton is built using multi-view projections of 3D hand keypoints. A new dataset is also presented. Results: 99.8% accuracy on RKS-PERSIANSIGN and 91.12% accuracy on First-Person dataset 4.64% hand pose estimation error on NYU dataset.	Evaluation of the proposed model performed on two datasets First-Person hand action benchmark and RKS-PERSIANSIGN, for hand action recognition and hand sign recognition. Also hand pose estimation error on NYU dataset. Five fold cross-validation is used for evaluation.	RKS-PERSIANSIGN dataset presented in the paper with 10,000 RGB videos of 100 Persian sign words	Persian Sign Language	Isolated recognition	Video

TABLE 1. *(Continued.)* **Detailed analysis of SL synthesis, interpretation and visualization research.**

[62]	2020	Improves state of the art results in gloss-to-text and video-to-text translation using STMC-Transformer that outperforms translation of GT glosses.	Translate GT gloss annotations to simulate perfect tokenization on both datasets. Then, perform video-to-text translation on PHOENIX-Weather 2014T with the STMC-Transformer.	PHOENIX-Weather 2014T and ASLG-PC12 corpus	German Sign Language and American Sign Language	Continuous recognition and translation	Videos
[63]	2020	Uses a Connectionist Temporal Classification (CTC) loss to bind the recognition and translation problems into a single unified architecture.	A joint approach that does not require any ground-truth timing information, simultaneously solving two co-dependant sequence-to-sequence learning problems and leads to significant performance gains.	PHOENIX-Weather 2014T	German Sign Language	Continuous recognition and translation	Video
[54]	2020	Using RF sensors to recognise signs. 72.5% accuracy in classification of 20 native ASL signs.	3 deaf and 10 imitation people included in the experiment using predefined scenario. Features extracted from sensors for ADL and sign classification and for ASL sign classification. Four 10GHz, One 24GHz and one 77GHz sensors used in the room. Both single features, and fusion of features is used with various classification algorithms. 75%, 25% train-test split.	Special dataset for the purposes of the experiment, measurements made with RF sensors and Kinect sensor used for annotation	American Sign Language	Continuous recognition	RF sensor
[77]	2020	Sign Language Interpretation	The authors developed a prototype to recognize sign language symbols.	The authors demonstrate the use of 8 surface Electromyography sensors to identify the sign language.	Chinese Sign Language	Isolated recognition	Armband
[59]	2019	A novel deep architecture based on 3D-ResNet and encoder-decoder network with connectionist temporal classification by iterative optimization for continuous SLR.	In one setup 40 signers for training and 10 for testing, and in another setup 94 sentences from different signers for training and 6 for testing	RWTH-Phoenix-Weather-2014 dataset and The Chinese Sign Language (CSL) dataset which contains 100 sentences performed 5 times from 50 signers with 25000 videos in total. It also provides 500 word isolated version	German and Chinese Sign Language	Continuous recognition	Video
[78]	2018	Proposing an AI-based tool using convolutional neural networks for language recognition with a 92.88% average recognition rate.	Five subjects performed 200 hand signs recorded in 5 different viewing angles under various background environments.	Custom dataset.	Indian Sign Language	Isolated recognition	Video
[79]	2018	Incorporating 18 new ASCII printable characters and three new features, normalized distance between direction extreme, close figure test, and direction change ratio, of 15 dimensions to enhance the performance of a system. The maximum accuracy achieved using the combination of existing and proposed features is 96.95%.	Classifying the gestures through a comparative study.	58 gestures	Gestures	Isolated recognition	Video
[80]	2018	Sign Language Interpretation	Describes a system for acquisition and interpretation of Indian sign language.	The authors use portable data gloves and an embedded chip to capture and interpret Indian sign language.	Indian Sign Language	Continuous recognition	Gloves
[81]	2017	Using Convoluted Neural networks and Stacked de-noising auto-encoder for audio recognition. Recognition CNN: up to 98%, SDAE: up to 99.4%.	Training: 1440 samples (images), Testing: 600 samples.	PRIMA Moesland gesture recognition DB	American Sign Language	Static recognition	Images

TABLE 1. *(Continued.)* **Detailed analysis of SL synthesis, interpretation and visualization research.**

[82]	2017	Presenting an SL recognition system using portable and cost-affordable Leap Motion sensor and applying kth-Nearest Neighbor (k-NN) with high accuracy in static sign language interpretation.	Collecting and labeling 1000 sign gesture data points for each gesture before the accuracy test.	n/a	Chinese Sign Language	Isolated recognition	Hand-tracking device
[83]	2017	Proposing an MV3D-CNN process applied to both real-world moving vehicle recognition and sign language recognition tasks has a good performance.	Labeling 22,300 two-view videos.	Custom dataset of 50 signs.	N/A	Isolated recognition	Video
[8]	2017	Presenting an SL recognition system based on hand tracking devices that utilize SVM for sign classification and achieving up to 100% accuracy	Performing different evaluations with over 50 individuals.	American Sign Language finger-spelling alphabet	American Sign Language	Static recognition	Hand-tracking device
[84]	2017	Sign Language Visualization	The paper describes a method for generating intermediate frames from 2D image(s) to create facial expressions suitable for sign language visualization and synthesis.	The research contributes to creating intermediate frames from the single input image to generate fluid facial gestures by dividing the face model into little triangles and deforming them with affine transformation.	Sign language-agnostic	Visualization	Video
[85]	2017	Sign Language Visualization	The authors successfully demonstrate the generation of sign language for weather forecast programs.	This research discusses the implementation of a sign language caption system using a 3D computer-generated avatar.	Korean Sign Language	Visualization	Video
[86]	2016	Identifying a new classifier achieves an F1 score of 98% when discriminating between BSL and LSF videos with static backgrounds and a 70% F1 score when distinguishing between ASL and BSL videos.	Using half of each language corpus for training and the other half for testing.	Dicta Sign corpus	American & British Sign Language	Isolated recognition	Video
[87]	2016	Presenting an SL recognition application with Microsoft Kinect with a 90% recognition rate for 13 signs and 100% for 3 signs with an average accuracy rate of 90.68%.	Using a vocabulary of 140 symbols collected by 18 subjects, totaling 5041 images.	140 symbols	Indian Sign Language	Static recognition	Hand-tracking device and Video
[88]	2016	Building a noise-free method for classification during training, Selective Temporal Filtering identifies key feature vectors from a noise-filtered input sequence.	the proposed method not only outperforms Conditional Random Fields and Hidden Markov Models in noisy environments but also in a well-controlled environment where we assume no significant noise vectors exist	Synthetic-dataset and a database of American Sign Language	American Sign Language	Isolated recognition	Video
[89]	2016	Extension on the research presented in [90] focused on embedding a CNN into an HMM for training and utilizing CNNs in a Bayesian fashion for interpretation. Improved upon previous research for sign language recognition tasks by between 15% and 38% relative and up to 13.3% absolute.	The test protocol follows the guidelines of SIGMA and PHOENIX databases	SIGNUM and PHOENIX databases	German Sign Language	Continuous recognition	Video
[91]	2016	Using LTSM and RNN for sign language recognition. Depending on the dataset in use, recognition varies from 62% up to 86%.	25,000 samples from 50 different signers.	Custom RGBD Kinect dataset.	Chinese Sign Language	Isolated recognition	Hand-tracking device
[92]	2016	Testing three approaches on a data gloves-based system for wearable computing with the results showing the following accuracy: SVM 93%, Naive-Bayes 82.5%, and Decision trees 72.3%.	Analyzing 56 data samples from the custom glove	Custom data	Indian Sign Language	Isolated recognition	Glove
[93]	2016	Showcasing an SL recognition system using the Leap motion controller and recognition based on NN multi-layer perception with an average accuracy of 96.5%.	Using a multi-layer perception neural network to analyze 520 samples of recordings	Custom dataset	American Sign Language	Isolated recognition	Hand-tracking device

TABLE 1. ***(Continued.)*** **Detailed analysis of SL synthesis, interpretation and visualization research.**

[90]	2015	Perform continuous sign language recognition across different signers on lab and real-life data. Evaluation on SIGNUM database (25 signers, 455 sign vocabulary, 19k sentences) and unconstrained "real-life" sign language (RWTH-PHOENIX-Weather database: 9 signers, 1081 sign vocabulary, 7k sentences) and achieve up to 10.0%/16.4% and respectively up to 34.3%/53.0% word error rate for single signer/multi-signer setups.	The test protocol follows the guidelines of SIGMA and PHOENIX databases	SIGNUM and PHOENIX databases	German Sign Language	Continuous recognition	Video
[94]	2013	Using an early combination of features, the late fusion of decisions, and synchronous combination on the hidden Markov model state level and asynchronous combination on the gloss level to achieve the best-published word error rate on the SIGNUM database, 10.7% and the RWTH-PHOENIX database, 41.9%.	The test protocol follows the guidelines of SIGMA and PHOENIX databases	SIGNUM and PHOENIX databases	German Sign Language	Isolated recognition	Video
[95]	2013	Using Hu-moment pattern matching for classification of signs via image recognition.		ALS alphabet images	American Sign Language	Static recognition	Images
[96]	2013	Presenting a novel algorithm for hand recognition using image processing and exploring its application in security-based systems by testing for different gestures with an accuracy rate of 95.2%.	Implementing the algorithm in MATLAB programming language and testing on over 50 samples.	Custom dataset.	Gestures	Static recognition	Images
[97]	2013	Proposing a real-time system that recognizes emotions from body movements data streams with an accuracy of 72%.	Using the MoCap UCLIC affective gesture database for training the classifier, then testing the trained classifier with different subjects using continuous Kinect data.	27 postural features	Gestures	Continuous recognition	Video
[98]	2012	Comparing the performance of DTW and HMM methods using different criteria shows that DTW has higher performance.	Analyzing a set of recorded gestures.	A custom dataset that includes a multitude of samples from ten different gestures.	N/A	Isolated recognition	Video
[99]	2012	Proposing a cost-effective and non-intrusive web-cam SL recognition system that allows for multi-language training and recognition. Experimental results show the system works well for dialect-free sign language translation with an average recognition accuracy between 55% and 65%.	Using a sample set of 90 images to train the numeric neural network and a collection of 240 pictures to train the neural network for alphabet characters.	Custom dataset.	English to Spoken	Static recognition	Images
[100]	2012	Comparing the performance of Local Binary Patterns with geometric moments describing the trajectory and shape of hands. When features are combined, the recognition rate increases to 99.75% for signer dependent tests and 57.54% for independent signer tests.	Testing the recognition performance of individual features and their combinations with several repetitions.	23 signs	Gestures	Static recognition	Images
[101]	2012	Sign Language Visualization	Introduces a computing model that is trained to provide better interpolation between words.	The research contributes to creating more accurate and intelligible synthesized sign language animation.	Chinese Sign Language	Visualization	N/A
[102]	2011	Presenting a simple SL recognition system that utilizes skin color segmentation and neural networks to achieve a recognition accuracy rate of 92.58%.	Training the network with 60 % of samples and testing with the remaining 40 % to classify the right and left-hand signs.	A dataset with 3080 feature vectors.	N/A	Static recognition	Images

TABLE 1. *(Continued.)* **Detailed analysis of SL synthesis, interpretation and visualization research.**

[103]	2011	They analyze a system consisting of three components: real-time hand tracking, hand-tree construction, and hand gesture recognition. The results show that the classification ability of single-layered CVNN on unseen data is comparable to the conventional real-valued neural network (RVNN) having one hidden layer.	Testing the system with 26 different gestures to evaluate the effectiveness of the proposed approach.	Custom dataset.	N/A	Static recognition	Images
[104]	2011	Presenting a human-robot interaction system that recognizes meaningful gestures composed of continuous hand motions in real-time based on hidden Markov models achieving an average gesture recognition rate of 96%.	Eight compound gestures are employed in the system, each assigned to a motion or functional control command.to allow users to operate an autonomous robot efficiently.	N/A	Gestures	Continuous recognition	Video
[105]	2010	Presenting an application for learning ASL that uses data gloves and video for sign acquisition and HMM for verification with a recognition rate up to 81.7%.	Using the assigned protocol for Georgian Tech Gesture Recognition Toolkit for HMM training	Georgia Tech Gesture Recognition Toolkit	American Sign Language	Isolated recognition	gloves and video
[106]	2010	Using two approaches, Euclidean distance, and K-nearest neighbor metrics, SL recognizes many signs classified with a direction histogram.	Using a repository of many images to train and test samples under various light illumination conditions.	22 specific kinds of ISL class/word	Indian Sign Language	Static recognition	Images
[107]	2010	Sign Language Visualization	Prototype design of an avatar for sign language visualization.	The authors demonstrate a prototype 3D virtual character for visualizing sign languages.	Italian Sign Language	Visualization	N/A

Another transformer-based approach is presented in [64], which proposes a combination of feature extraction using OpenPose for human keypoint estimation and end-to-end feature learning with Convolutional Neural Networks, using the proven multi-head attention mechanism used in transformers. Another transformer-based approach applied in word-level recognition is presented in [65].

The authors of [66] propose a transformer model with self-attention mechanism. It proposes the use of content-aware and position-aware convolution layers that explicitly select relevant features using a neighborhood gathering method.

e: RECOGNITION USING DECOMPOSITION

In-Cheol and Sung [67] utilized HMMs to create a data-glove-based sign language recognition system using the Polhemus gloves. The algorithm focused on decomposing gestures into primitive strokes, using different HMMs for gesture decomposition and stroke recognition. In its limited experiments, the algorithm achieved a high level of recognition rate (96.88%). Decomposition of gestures, along with linguistic knowledge system, was also used by Konstantinos *et al.* [68]. Here, the authors decompose the image sequences into primitives: handshape, location, and movement. Later, they use linguistic knowledge to recognize the ASL speech. They used this technique on a large set of signs and achieved between 86 and 97.13 percent accuracy.

When directly speaking about sign languages, proto-systems based on neural networks for sign language recognition also emerge. For example, Kim *et al.* [69] are credited for creating a recognition system for Korean sign language based on input generated from data-hand gloves. In 2000, Marcus and his colleagues [70] presented a Neural Network (NN), called T-CombNET, that has been specialized in hand recognition. They tested the network using signs from Japanese fingerspelling and achieved up to 96.5% recognition when using dynamic gestures. Ming and Naendra [71] created a sequential image-based neural network that is capable of recognition of hand-trajectory gestures by utilizing and decomposing the image sequences. The NN was tested for recognition of 40 gestures from American Sign Language.

Fuzzy-rule approaches were also considered and were tested on Taiwanese sign language [72]. Similar fuzzy-based methods were considered and tested for Brazilian sign language [73].

B. VISUALIZATION

From the sign language visualization aspect, efforts were made to simplify the description of the models used for gesture/sign language visualization. Aside from the proposed system by Gibert *et al.* [47], general improvements of the more subtle aspects of sign language visualization. John McDonald *et al.* [74] purposefully created an anatomically correct 3D model and skeleton of a human hand for use with American Sign Language animation and visualization.

C. NOTATION SYSTEMS

The computer representation of sign language notation was discussed in [75]. The paper details the then-known sign language notation systems: Stokoe Notation and its derivative, the BSL notation; SignFont and HamNoSys. The authors conclude that the then-current notation systems are not sufficient for sign language and propose a new type of notation that considers three-layered notation based on: "idea", "word", and " deed".

D. LIMITATIONS OF THE APPROACH

This study considered only three digital libraries, and some relevant articles from other indexed publishers were not considered. However, keeping in mind the size of the evaluated digital libraries, we believe that the obtained results are representative and in line with the study's goals. All digital libraries used in this work have different internal search engines with different rules for the maximum number of papers retrieved and another formatting of search results. The papers obtained for this study were obtained from the various search engines using the same search queries. However, keeping in mind the number of papers analyzed within this review, we believe that specificities of the publishers' search engines have limited impact and have not influenced the findings of this work. In the future, the NLP tool kit needs to be extended to process more digital libraries. In addition, there is an apparent need for a Web app that will make it available to a broader audience. Until then, readers are encouraged to contact the authors if they are interested in using the toolkit.

VI. CONCLUSION

The specific problems tackled in this research occupy a wide range of topics connected to sign language research. The main problem when working with sign language is to effectively translate and interpret them into some easily comprehensible form, both for people and machines. To do this the researchers have to adjust to the visual modality of sign languages and extract key features from the visual data. This problem includes recognition of static images, sequences of images (video) as well as other types of information (e.g., depth data).

Based on the thorough analysis of the selected publication, we can conclude that machine learning and deep learning algorithms can improve sign language recognition and interpretation.

This means that sign language recognition systems depend on machine learning approaches that are typical for other problem domains. These approaches include pattern recognition, statistical approaches, and neural network-based approaches. Historically HMMs gave the best results initially since they, as an algorithm, seem to be most suitable for time series data. However, lately, neural network-based approaches and deep learning approaches seem to give better results. Among them, approaches based on 3D-CNN, Transformer and LSTM architectures seem to dominate the continuous sign language recognition regardless of the sign language. We can attribute the 3DCNN approach due to its success with problems with image/video recognition, time series, and natural language processing, areas that we explained that are of interest in sign language research. Similarly, LSTMs are used due to their effectiveness with time-series data.

Additional studies need to be performed to analyze different sign languages' recognition, visualization, and synthesis performance. Most of the studies compare the more novel approaches to the traditional methods for the same language, but there is a lack of studies that compare methods between different languages. This is due to the fact that there are lack of datasets of translated sign language sequences. Nevertheless, based on our review of the existing literature, one can conclude that the state-of-the-art approaches used for sign language recognition perform similarly for various languages. Therefore, the same methods with slight modification are applied to different sign languages with similar success.

With the improvement of the hardware and the machine learning approaches, including the advancements of neural networks and deep learning, one can expect that there will be a significant improvement in all technologies for sign language recognition and visualization. The current limitations are usually related to a lack of data and training examples for all languages. However, this limitation is becoming less and less relevant with the advancements of the transfer learning approaches. Therefore, we expect that within the following five to ten years, there will be public and commercial products that will perform live translation from and to distinct sign languages. This will ideally allow for real-time communication between people that use different modalities (sign language and vocal speech) as well as developing systems that will translate speech from one sign language to another.

REFERENCES

[1] G. A. Gates and J. H. Mills, "Presbycusis," *Lancet*, vol. 366, no. 9491, pp. 1111–1120, Sep. 2005.

[2] C. F. Hockett and C. D. Hockett, "The origin of speech," *Sci. Amer.*, vol. 203, no. 3, pp. 88–97, 1960.

[3] C. F. Hockett and S. A. Altmann, "A note on design features," in *Animal Communication: Techniques of Study and Results of Research*, vol. 6172. Bloomington, IN, USA: Indiana Univ. Press, 1968.

[4] K. Emmorey, *Language, Cognition, and the Brain: Insights From Sign Language Research*. New York, NY, USA: Psychology Press, 2001.

[5] L. Goyal and V. Goyal, "Text to sign language translation system: A review of literature," *Int. J. Synth. Emotions (IJSE)*, vol. 7, no. 2, pp. 62–77, 2016.

[6] H. Cooper, B. Holt, and R. Bowden, "Sign language recognition," in *Visual Analysis of Humans*. London, U.K.: Springer, 2011, pp. 539–562.

[7] Z. Zafrulla, H. Brashear, T. Starner, H. Hamilton, and P. Presti, "American sign language recognition with the Kinect," in *Proc. 13th Int. Conf. Multimodal Interface (ICMI)*, 2011, pp. 279–286.

[8] L. Quesada, G. López, and L. Guerrero, "Automatic recognition of the American sign language fingerspelling alphabet to assist people living with speech or hearing impairments," *J. Ambient Intell. Humanized Comput.*, vol. 8, no. 4, pp. 625–635, 2017. [Online]. Available: https://link.springer.com/article/10.1007/s12652-017-0475-7

[9] W. Stokoe, *Sign Language Structure, An Outline of the Visual Communications Systems of American Deaf*, vol. 8. Buffalo, NY, USA: Univ. Buffalo, 1960.

[10] W. C. Stokoe, *Problems in Sign Language Research*. Amsterdam, The Netherlands: Elsevier, 1978.

[11] M. F. Tolba and A. S. Elons, "Recent developments in sign language recognition systems," in *Proc. 8th Int. Conf. Comput. Eng. Syst. (ICCES)*, Nov. 2013, pp. xxxvi–xlii.

[12] M. A. Ahmed, B. B. Zaidan, A. A. Zaidan, M. M. Salih, and M. M. B. Lakulu, "A review on systems-based sensory gloves for sign language recognition state of the art between 2007 and 2017," *Sensors*, vol. 18, no. 7, p. 2208, 2018.

[13] A. Er-Rady, R. Faizi, R. O. H. Thami, and H. Housni, "Automatic sign language recognition: A survey," in *Proc. Int. Conf. Adv. Technol. Signal Image Process. (ATSIP)*, May 2017, pp. 1–7.

[14] R. Rastgoo, K. Kiani, and S. Escalera, "Sign language recognition: A deep survey," *Expert Syst. Appl.*, vol. 164, Feb. 2021, Art. no. 113794.

[15] O. Koller, "Quantitative survey of the state of the art in sign language recognition," 2020, *arXiv:2008.09918*.

[16] M. J. Cheok, Z. Omar, and M. H. Jaward, "A review of hand gesture and sign language recognition techniques," *Int. J. Mach. Learn. Cybern.*, vol. 10, no. 1, pp. 131–153, Jan. 2017.

[17] I. A. Adeyanju, O. O. Bello, and M. A. Adegboye, "Machine learning methods for sign language recognition: A critical review and analysis," *Intell. Syst. with Appl.*, vol. 12, Nov. 2021, Art. no. 200056. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2667305321000454

[18] I. Papastratis, C. Chatzikonstantinou, D. Konstantinidis, K. Dimitropoulos, and P. Daras, "Artificial intelligence technologies for sign language," *Sensors*, vol. 21, no. 17, p. 5843, Aug. 2021.

[19] K. M. Sagayam and D. J. Hemanth, "Hand posture and gesture recognition techniques for virtual reality applications: A survey," *Virtual Reality*, vol. 21, no. 2, pp. 91–107, Jun. 2017.

[20] T. Aguiar de Lima and M. Da Costa-Abreu, "A survey on automatic speech recognition systems for Portuguese language and its variations," *Comput. Speech Lang.*, vol. 62, Jul. 2020, Art. no. 101055.

[21] D. Hirafuji Neiva and C. Zanchettin, "Gesture recognition: A review focusing on sign language in a mobile context," *Expert Syst. Appl.*, vol. 103, pp. 159–183, Aug. 2018.

[22] H. Arksey and L. O'Malley, "Scoping studies: Towards a methodological framework," *Int. J. Social Res. Methodol.*, vol. 8, no. 1, pp. 19–32, Feb. 2005.

[23] D. Levac, H. Colquhoun, and K. K. O'Brien, "Scoping studies: Advancing the methodology," *Implement. Sci.*, vol. 5, no. 1, pp. 1–9, Dec. 2010.

[24] H. L. Colquhoun, D. Levac, K. K. O'Brien, S. Straus, A. C. Tricco, L. Perrier, M. Kastner, and D. Moher, "Scoping reviews: Time for clarity in definition, methods, and reporting," *J. Clin. Epidemiol.*, vol. 67, no. 12, pp. 1291–1294, Dec. 2014.

[25] D. Moher, P.-P. Group, L. Shamseer, M. Clarke, D. Ghersi, A. Liberati, M. Petticrew, P. Shekelle, and L. A. Stewart, "Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement," *Systematic Rev.*, vol. 4, no. 1, Dec. 2015.

[26] E. Zdravevski, P. Lameski, V. Trajkovik, I. Chorbev, R. Goleva, N. Pombo, and N. M. Garcia, *Automation in Systematic, Scoping and Rapid Reviews by an NLP Toolkit: A Case Study in Enhanced Living Environments*. Cham, Switzerland: Springer, 2019, pp. 1–18, doi: 10.1007/978-3-030-10752-9_1.

[27] T. Loncar-Turukalo, E. Zdravevski, J. Machado da Silva, I. Chouvarda, and V. Trajkovik, "Literature on wearable technology for connected health: Scoping review of research trends, advances, and barriers," *J. Med. Internet Res.*, vol. 21, no. 9, Sep. 2019, Art. no. e14017, doi: 10.2196/14017.

[28] P. Tonkovic, S. Kalajdziski, E. Zdravevski, P. Lameski, R. Corizzo, I. M. Pires, N. M. Garcia, T. Loncar-Turukalo, and V. Trajkovik, "Literature on applied machine learning in metagenomic classification: A scoping review," *Biology*, vol. 9, no. 12, p. 453, Dec. 2020. [Online]. Available: https://www.mdpi.com/2079-7737/9/12/453

[29] J. Fenlon and E. Wilkinson, *Sign Languages in the World*, A. Shembry and C. Lucas, Eds. Cambridge, U.K.: Cambridge Univ. Press, 2015.

[30] M. C. Caselli and V. Volterra, "From communication to language in hearing and deaf children," in *From Gesture to Language in Hearing and Deaf Children*, V. Volterra and C. J. Erting, Eds. Berlin, Germany: Springer, 1990, pp. 263–277, doi: 10.1007/978-3-642-74859-2_21.

[31] P. B. Braem, "Acquisition of the handshape in American sign language: A preliminary analysis," in *From Gesture to Language in Hearing and Deaf Children*, V. Volterra and C. J. Erting, Eds. Berlin, Germany: Springer, 1990, pp. 107–127, doi: 10.1007/978-3-642-74859-2_10.

[32] R. A. Gardner and B. T. Gardner, "Early signs of language in cross-fostered chimpanzees," in *Language Origin: A Multidisciplinary Approach*, J. Wind, B. Chiarelli, B. Bichakjian, A. Nocentini, and A. Jonker, Eds. Dordrecht, The Netherlands: Springer, 1992, pp. 351–381, doi: 10.1007/978-94-017-2039-7_19.

[33] I. Guyon, V. Athitsos, P. Jangyodsuk, H. J. Escalante, and B. Hamner, "Results and analysis of the chalearn gesture challenge 2012," in *Proc. Int. Workshop Depth Image Anal. Appl.* Springer, 2012, pp. 186–204.

[34] P. J. Berkay, "Making adult education accessible to the deaf: A model of direct instructor communication in ASL," *Amer. Ann. Deaf*, vol. 135, no. 5, pp. 396–401, 1990.

[35] C. Dahl and S. Wilcox, "Preparing the educational interpreter: A survey of sign language interpreter training programs," *Amer. Ann. Deaf*, vol. 135, no. 4, pp. 275–279, 1990.

[36] H. Gibson, A. Small, and D. Mason, "Deaf bilingual bicultural education," in *Bilingual Education*, J. Cummins and D. Corson, Eds. Dordrecht, The Netherlands: Springer, 1997, pp. 231–240, doi: 10.1007/978-94-011-4531-2_23.

[37] M. Kemp, "Why is learning American sign language a challenge?" *Amer. Ann. Deaf*, vol. 143, no. 3, pp. 255–259, 1998.

[38] N. Takahashi, M. Kawamura, H. Shinotou, K. Hirayama, K. Kaga, and M. Shindo, "Pure word deafness due to left hemisphere damage," *Cortex*, vol. 28, no. 2, pp. 295–303, Jun. 1992.

[39] G. Hickok, E. Klima, M. Kritchevsky, and U. Bellugi, "A case of 'sign blindness' following left occipital damage in a deaf signer," *Neuropsychologia*, vol. 33, no. 12, pp. 1597–1601, 1995.

[40] G. Hickok, M. Wilson, K. Clark, E. S. Klima, M. Kritchevsky, and U. Bellugi, "Discourse deficits following right hemisphere damage in deaf signers," *Brain Lang.*, vol. 66, no. 2, pp. 233–248, Feb. 1999.

[41] G. Hickok, "The neural organization of language: Evidence from sign language aphasia," *Trends Cognit. Sci.*, vol. 2, no. 4, pp. 129–136, Apr. 1998.

[42] D. Bavelier, D. Corina, P. Jezzard, V. Clark, A. Karni, A. Lalwani, J. P. Rauschecker, A. Braun, R. Turner, and H. J. Neville, "Hemispheric specialization for English and ASL: Left invariance-right variability," *Neuroreport*, vol. 9, no. 7, pp. 1537–1542, 1998.

[43] D. Schmauks and M. Wille, "Integration of communicative hand movements into human-computer-interaction," *Comput. Humanities*, vol. 25, nos. 2–3, pp. 129–140, 1991.

[44] G. J. Sweeney and A. C. Downton, "Towards appearance-based multichannel gesture recognition," in *Progress in Gestural Interaction*, P. A. Harling and A. D. N. Edwards, Eds. London, U.K.: Springer, 1997, pp. 7–16.

[45] R. Bowden, T. Heap, and D. Hogg, "Real time hand tracking and gesture recognition as a 3D input device for graphical applications," in *Progress in Gestural Interaction*, P. A. Harling and A. D. N. Edwards, Eds. London, U.K.: Springer, 1997, pp. 117–129.

[46] M. Bordegoni and G. P. Faconti, "Architectural models of gesture systems," in *Progress in Gestural Interaction*, P. A. Harling and A. D. N. Edwards, Eds. London, U.K.: Springer, 1997, pp. 61–73.

[47] S. Gibet, A. Braffort, C. Collet, F. Forest, R. Gherbi, and T. Lebourque, "Gesture in human-machine Communication: Capture, analysis-synthesis, recognition, semantics," in *Progress in Gestural Interaction*, P. A. Harling and A. D. N. Edwards, Eds. London, U.K.: Springer, 1997, pp. 89–95.

[48] C. Neidle, S. Sclaroff, and V. Athitsos, "SignStream: A tool for linguistic and computer vision research on visual-gestural language data," *Behav. Res. Methods, Instrum., Comput.*, vol. 33, no. 3, pp. 311–320, Aug. 2001.

[49] M. P. Craven and K. M. Curtis, "GesRec3D: A real-time coded gesture-to-speech system with automatic segmentation and recognition thresholding using dissimilarity measures," in *Gesture-Based Communication in Human-Computer Interaction*, A. Camurri and G. Volpe, Eds. Berlin, Germany: Springer, 2004, pp. 231–238.

[50] W. K. Wong, F. H. Juwono, and B. T. T. Khoo, "Multi-features capacitive hand gesture recognition sensor: A machine learning approach," *IEEE Sensors J.*, vol. 21, no. 6, pp. 8441–8450, Mar. 2021.

[51] M. A. Ahmed, B. B. Zaidan, A. A. Zaidan, M. M. Salih, Z. T. Al-qaysi, and A. H. Alamoodi, "Based on wearable sensory device in 3D-printed humanoid: A new real-time sign language recognition system," *Measurement*, vol. 168, Jan. 2021, Art. no. 108431.

[52] M. Ahmed, B. Zaidan, A. Zaidan, A. Alamoodi, O. Albahri, Z. Al-Qaysi, A. Albahri, and M. M. Salih, "Real-time sign language framework based on wearable device: Analysis of MSL, DataGlove, and gesture recognition," *Soft Comput.*, vol. 25, pp. 11101–11122, May 2021.

[53] A. H. Alrubayi, M. A. Ahmed, A. A. Zaidan, A. S. Albahri, B. B. Zaidan, O. S. Albahri, A. H. Alamoodi, and M. Alazab, "A pattern recognition model for static gestures in Malaysian sign language based on machine learning techniques," *Comput. Electr. Eng.*, vol. 95, Oct. 2021, Art. no. 107383.

[54] S. Z. Gurbuz, A. Cafer Gurbuz, E. A. Malaia, D. J. Griffin, C. S. Crawford, M. Mahbubur Rahman, E. Kurtoglu, R. Aksu, T. Macks, and R. Mdrafi, "American sign language recognition using RF sensing," *IEEE Sensors J.*, vol. 21, no. 3, pp. 3763–3775, 2021.

[55] S. Waldherr, R. Romero, and S. Thrun, "A gesture based interface for human-robot interaction," *Auton. Robots*, vol. 9, no. 2, pp. 151–173, 2000.

[56] K. Nickel and R. Stiefelhagen, "Real-time recognition of 3D-pointing gestures for human-machine-interaction," in *Pattern Recognition*, B. Michaelis and G. Krell, Eds. Berlin, Germany: Springer, 2003, pp. 557–565.

[57] K. Nickel and R. Stiefelhagen, "Real-time person tracking and pointing gesture recognition for human-robot interaction," in *Computer Vision in Human-Computer Interaction*, N. Sebe, M. Lew, and T. S. Huang, Eds. Berlin, Germany: Springer, 2004, pp. 28–38.

[58] K. Sage, A. J. Howell, and H. Buxton, "Developing context sensitive HMM gesture recognition," in *Gesture-Based Communication in Human-Computer Interaction*, A. Camurri and G. Volpe, Eds. Berlin, Germany: Springer, 2004, pp. 277–287.

[59] J. Pu, W. Zhou, and H. Li, "Iterative alignment network for continuous sign language recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 4165–4174.

[60] R. Rastgoo, K. Kiani, and S. Escalera, "Hand sign language recognition using multi-view hand skeleton," *Expert Syst. Appl.*, vol. 150, Jul. 2020, Art. no. 113336.

[61] E. Holdengreber, R. Yozevitch, and V. Khavkin, "Intuitive cognition-based method for generating speech using hand gestures," *Sensors*, vol. 21, no. 16, p. 5291, Aug. 2021. [Online]. Available: https://www.mdpi.com/1424-8220/21/16/5291

[62] K. Yin and J. Read, "Better sign language translation with STMC-transformer," in *Proc. 28th Int. Conf. Comput. Linguistics*, 2020, pp. 5975–5989.

[63] N. Cihan Camgoz, O. Koller, S. Hadfield, and R. Bowden, "Sign language transformers: Joint end-to-end sign language recognition and translation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 10023–10033.

[64] M. De Coster, M. Van Herreweghe, and J. Dambre, "Sign language recognition with transformer networks," in *Proc. 12th Int. Conf. Lang. Resour. Eval.*, 2020, pp. 6018–6024.

[65] M. Bohacek and M. Hruz, "Sign pose-based transformer for word-level sign language recognition," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. Workshops (WACVW)*, Jan. 2022, pp. 182–191.

[66] P. Xie, M. Zhao, and X. Hu, "PiSLTRc: Position-informed sign language transformer with content-aware convolution," *IEEE Trans. Multimedia*, early access, Sep. 2, 2021, doi: 10.1109/TMM.2021.3109665.

[67] I.-C. Kim and S.-I. Chien, "Analysis of 3D hand trajectory gestures using stroke-based composite hidden Markov models," *Appl. Intell.*, vol. 15, no. 2, pp. 131–143, 2001.

[68] K. G. Derpanis, R. P. Wildes, and J. K. Tsotsos, "Hand gesture recognition within a linguistics-based framework," in *Computer Vision—ECCV 2004*, T. Pajdla and J. Matas, Eds. Berlin, Germany: Springer, 2004, pp. 282–296.

[69] J.-S. Kim, W. Jang, and Z. Bien, "A dynamic gesture recognition system for the Korean sign language (KSL)," *IEEE Trans. Syst., Man, Cybern., B (Cybern.)*, vol. 26, no. 2, pp. 354–359, Apr. 1996.

[70] M. V. Lamar, M. S. Bhuiyan, and A. Iwata, "T-CombNET—A neural network dedicated to hand gesture recognition," in *Biologically Motivated Computer Vision*, S.-W. Lee, H. H. Bülthoff, and T. Poggio, Eds. Berlin, Germany: Springer, 2000, pp. 613–622.

[71] M.-H. Yang and N. Ahuja, "Recognizing hand gestures using motion trajectories," in *Face Detection and Gesture Recognition for Human-Computer Interaction*. Boston, MA, USA: Springer, 2001, pp. 53–81, doi: 10.1007/978-1-4615-1423-7_3.

[72] M.-C. Su, Y.-X. Zhao, and H.-F. Chen, "A fuzzy rule-based approach to recognizing 3-D arm movements," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 9, no. 2, pp. 191–201, Jun. 2001.

[73] B. C. Bedregal, A. C. R. Costa, and G. P. Dimuro, "Fuzzy rule-based hand gesture recognition," in *Artificial Intelligence in Theory and Practice*, M. Bramer, Ed. Boston, MA, USA: Springer, 2006, pp. 285–294.

[74] J. McDonald, J. Toro, K. Alkoby, A. Berthiaume, R. Carter, P. Chomwong, J. Christopher, M. J. Davidson, J. Furst, B. Konie, G. Lancaster, L. Roychoudhuri, E. Sedgwick, N. Tomuro, and R. Wolfe, "An improved articulated model of the human hand," *Vis. Comput.*, vol. 17, no. 3, pp. 158–166, May 2001.

[75] K. C. Smith and W. Edmondson, "The development of a computational notation for synthesis of sign and gesture," in *Gesture-Based Communication in Human-Computer Interaction*, A. Camurri and G. Volpe, Eds. Berlin, Germany: Springer, 2004, pp. 312–323.

[76] 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.

[77] S. Yuan, Y. Wang, X. Wang, H. Deng, S. Sun, H. Wang, P. Huang, and G. Li, "Chinese sign language alphabet recognition based on random forest algorithm," in *Proc. IEEE Int. Workshop Metrology for Ind. 4.0 IoT*, Jun. 2020, pp. 340–344.

[78] G. A. Rao, K. Syamala, P. V. V. Kishore, and A. S. C. S. Sastry, "Deep convolutional neural networks for sign language recognition," in *Proc. Conf. Signal Process. Commun. Eng. Syst. (SPACES)*, Jan. 2018, pp. 194–197.

[79] S. Misra, J. Singha, and R. H. Laskar, "Vision-based hand gesture recognition of alphabets, numbers, arithmetic operators and ASCII characters in order to develop a virtual text-entry interface system," *Neural Comput. Appl.*, vol. 29, no. 8, pp. 117–135, Apr. 2018.

[80] H. Rewari, V. Dixit, D. Batra, and N. Hema, "Automated sign language interpreter," in *Proc. 11th Int. Conf. Contemp. Comput. (IC3)*, Aug. 2018, pp. 1–5.

[81] O. K. Oyedotun and A. Khashman, "Deep learning in vision-based static hand gesture recognition," *Neural Comput. Appl.*, vol. 28, no. 12, pp. 3941–3951, 2017.

[82] Y. Xue, S. Gao, H. Sun, and W. Qin, "A Chinese sign language recognition system using leap motion," in *Proc. Int. Conf. Virtual Reality Visualizat. (ICVRV)*, Oct. 2017, pp. 180–185.

[83] T. He, H. Mao, and Z. Yi, "Moving object recognition using multi-view three-dimensional convolutional neural networks," *Neural Comput. Appl.*, vol. 28, no. 12, pp. 3827–3835, Dec. 2017. [Online]. Available: https://link.springer.com/article/10.1007/s00521-016-2277-9

[84] I. Talbi, O. E. Ghoul, and M. Jemni, "Towards realistic simulation of facial deformation in sign language," in *Proc. 6th Int. Conf. Inf. Commun. Technol. Accessibility (ICTA)*, Dec. 2017, pp. 1–5.

[85] J. Oh, S. Jeon, B. Kim, M. Kim, S. Kang, H. Kwon, I. Kim, and Y. Song, "Avatar-based sign language interpretations for weather forecast and other TV programs," *SMPTE Motion Imag. J.*, vol. 126, no. 1, pp. 57–62, Jan. 2017.

[86] C. D. D. Monteiro, C. M. Mathew, R. Gutierrez-Osuna, and F. Shipman, "Detecting and identifying sign languages through visual features," in *Proc. IEEE Int. Symp. Multimedia (ISM)*, Dec. 2016, pp. 287–290.

[87] Z. A. Ansari and G. Harit, "Nearest neighbour classification of Indian sign language gestures using Kinect camera," *Sadhana*, vol. 41, no. 2, pp. 161–182, Feb. 2016.

[88] M.-C. Roh, S. Fazli, and S.-W. Lee, "Selective temporal filtering and its application to hand gesture recognition," *Int. J. Speech Technol.*, vol. 45, no. 2, pp. 255–264, Sep. 2016.

[89] O. Koller, S. Zargaran, H. Ney, and R. Bowden, "Deep sign: Hybrid CNN-HMM for continuous sign language recognition," in *Proc. Brit. Mach. Vis. Conf. (BMVC)*, R. C. Wilson, E. R. Hancock, and W. A. P. Smith, Eds. BMVA Press, Sep. 2016, pp. 136.1–136.12, Art. no. 136, doi: 10.5244/C.30.136.

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

[91] T. Liu, W. Zhou, and H. Li, "Sign language recognition with long short-term memory," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 2871–2875.

[92] A. K. Singh, B. P. John, S. R. V. Subramanian, A. S. Kumar, and B. B. Nair, "A low-cost wearable Indian sign language interpretation system," in *Proc. Int. Conf. Robot. Autom. Humanitarian Appl. (RAHA)*, Dec. 2016, pp. 1–6.

[93] D. Naglot and M. Kulkarni, "Real time sign language recognition using the leap motion controller," in *Proc. Int. Conf. Inventive Comput. Technol. (ICICT)*, vol. 3, Aug. 2016, pp. 1–5.

[94] J. Forster, C. Oberdörfer, O. Koller, and H. Ney, "Modality combination techniques for continuous sign language recognition," in *Pattern Recognition and Image Analysis*, J. M. Sanches, L. Micó, and J. S. Cardoso, Eds. Berlin, Germany: Springer, 2013, pp. 89–99.

[95] M. Fernando and J. Wijayanayaka, "Low cost approach for real time sign language recognition," in *Proc. IEEE 8th Int. Conf. Ind. Inf. Syst.*, Dec. 2013, pp. 637–642.

[96] N. Dhruva, S. R. Rupanagudi, and H. N. N. Kashyap, "Novel algorithm for image processing based hand gesture recognition and its application in security," in *Advances in Computing, Communication, and Control*, S. Unnikrishnan, S. Surve, and D. Bhoir, Eds. Berlin, Germany: Springer, 2013, pp. 537–547.
[97] W. Wang, V. Enescu, and H. Sahli, "Towards real-time continuous emotion recognition from body movements," in *Human Behavior Understanding*, A. A. Salah, H. Hung, O. Aran, and H. Gunes, Eds. Cham, Switzerland: Springer, 2013, pp. 235–245.
[98] J. M. Carmona and J. Climent, "A performance evaluation of HMM and DTW for gesture recognition," in *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, L. Alvarez, M. Mejail, L. Gomez, and J. Jacobo, Eds. Berlin, Germany: Springer, 2012, pp. 236–243.
[99] M. Tariq, A. Iqbal, A. Zahid, Z. Iqbal, and J. Akhtar, "Sign language localization: Learning to eliminate language dialects," in *Proc. 15th Int. Multitopic Conf. (INMIC)*, Dec. 2012, pp. 17–22.
[100] M. Hrúz, J. Trojanová, and M. Železný, "Local binary pattern based features for sign language recognition," *Pattern Recognit. Image Anal.*, vol. 22, no. 4, pp. 519–526, Dec. 2012.
[101] L. Jinghua, Y. Baocai, W. Lichun, K. Dehui, and W. Yufei, "Diversified gesture generation for Chinese sign language animation," in *Proc. 4th Int. Conf. Digit. Home*, Nov. 2012, pp. 179–183.
[102] M. P. Paulraj, S. Yaacob, M. S. Zanar Azalan, and R. Palaniappan, "A phoneme based sign language recognition system using 2D moment invariant interleaving feature and neural network," in *Proc. IEEE Student Conf. Res. Develop.*, Dec. 2011, pp. 111–116.
[103] A. R. Hafiz, M. F. Amin, and K. Murase, "Real-time hand gesture recognition using complex-valued neural network (CVNN)," in *Neural Information Processing*, B.-L. Lu, L. Zhang, and J. Kwok, Eds. Berlin, Germany: Springer, 2011, pp. 541–549.
[104] C.-S. Fahn and K.-Y. Chu, "Hidden-Markov-model-based hand gesture recognition techniques used for a human-robot interaction system," in *Human-Computer Interaction. Interaction Techniques and Environments*, J. A. Jacko, Ed. Berlin, Germany: Springer, 2011, pp. 248–258.
[105] Z. Zafrulla, H. Brashear, P. Yin, P. Presti, T. Starner, and H. Hamilton, "American sign language phrase verification in an educational game for deaf children," in *Proc. 20th Int. Conf. Pattern Recognit.*, Aug. 2010, pp. 3846–3849.
[106] A. Nandy, J. S. Prasad, S. Mondal, P. Chakraborty, and G. C. Nandi, "Recognition of isolated Indian sign language gesture in real time," in *Information Processing and Management*, V. V. Das, R. Vijayakumar, N. C. Debnath, J. Stephen, N. Meghanathan, S. Sankaranarayanan, P. M. Thankachan, F. L. Gaol, and N. Thankachan, Eds. Berlin, Germany: Springer, 2010, pp. 102–107.
[107] V. Lombardo, F. Nunnari, and R. Damiano, "A virtual interpreter for the Italian sign language," in *Intelligent Virtual Agents*, J. Allbeck, N. Badler, T. Bickmore, C. Pelachaud, and A. Safonova, Eds. Berlin, Germany: Springer, 2010, pp. 201–207.

BOBAN JOKSIMOSKI received the M.Sc. degree in software engineering, in 2012, and the Ph.D. degree in computer science, in 2020. He is an Assistant Professor with the Faculty of Computer Science and Engineering, Saints Cyril and Methodius University in Skopje, North Macedonia. He has managed and participated in multiple projects around e-learning and enabling assistive and multimedia technologies for learning enhancements. He has published multiple research papers in conferences and journals. He also coauthored a book chapter. He was a Lead Technical Artist for companies in the CGI/VFX industry and had vast experience in the field. His research interests include computer graphics, visualization, computer animation, assistive technologies, and computer-based education.

EFTIM ZDRAVEVSKI received the Dipl. Eng. degree, in 2008, the M.Sc. degree in computer science, in 2010, and the Ph.D. degree in computer science, in 2017. He has been an Assistant Professor with the Faculty of Computer Science and Engineering, Saints Cyril and Methodius University in Skopje, North Macedonia, since 2010. He's also a Chief Architect at CogniTrek, Canada; and the Founder of MAGIX.AI—a company specialized in applied AI in FinTech, AdTech, and E-commerce. Previously, he was a CSO and a Board Member of Viking Analytics, Sweden; and the Director of machine learning of Xogito, USA. During his research career, he has published over 130 papers in peer-reviewed international conferences and top-tier journals, including *Applied Soft Computing*, *Expert Systems With Applications*, *Big Data Research*, *Journal of Medical Internet Research* (JMIR), *PLOS One*, IEEE Access, *Computer Science Review*, *Remote Sensing*, *Biology*, *Sensors*, *Applied Sciences*, *Electronics*, and *Future Internet*. He has participated in or led national and international research projects related to time series analytics, feature engineering and selection, and ML application for activity recognition. His research interests include FinTech, big data, machine learning, data mining, the IoT, cloud computing, time series analysis and forecasting, and parallel algorithms.

PETRE LAMESKI was born in Kavadarci, North Macedonia, in 1985. He received the bachelor's and master's degrees (Hons.) from the Faculty of Electrical Engineering and Information Technologies, Saints Cyril and Methodius University in Skopje, in 2008 and 2010, respectively, with a master's thesis in robotics and intelligent systems, and the Ph.D. degree from the Faculty of Computer Science and Engineering, in 2017. From 2008 to 2011, he worked as an Assistant Professor with the Faculty of Electrical Engineering and Information Technologies. From 2011 to 2017, he worked as a Teaching and Research Assistant and the Vice-Dean for academic affairs with the Faculty of Computer Science and Engineering, Skopje, North Macedonia. He has also worked as a Consultant in several companies, working on machine learning and mobile development projects in the financial, agricultural, and medical industries. He currently works as an Assistant Professor with the Faculty of Computer Science and Engineering, Saints Cyril and Methodius University in Skopje. He has published more than 100 papers in scientific journals and international conferences. He has participated in and led over ten international and national scientific projects. His research interests include machine learning; deep learning; transfer learning; signal processing of financial, medical, and the IoT data; mobile application development; and application of AI in mobile applications.

IVAN MIGUEL PIRES was born in Penha Garcia, Castelo Branco, Portugal, in 1989. He received the B.Sc. degree in computer science and engineering from the Polytechnic Institute of Castelo Branco, Castelo Branco, in 2010, and the M.Sc. degree in computer science and engineering and the European Ph.D. degree in computer science and engineering from the Universidade da Beira Interior (UBI), Covilhã, Portugal, in 2012 and 2018, respectively. Between February 2019 and September 2019, he frequented the C4—-Cloud Computing Competence Centre, Universidade da Beira Interior, related to passing cloud, for the Postdoctoral Fellowship. The subject was conveyed to sensors embedded on mobile devices to monitor traffic warning situations by car. He was an Invited Adjunct Professor at the Polytechnic Institute of Viseu, Viseu, Portugal, between October 2019 and September 2021. He has been an Invited Assistant Professor with the Universidade de Trás-os-Montes e Alto Douro, since October 2021. His main research interests include related to sensors available in off-the-shelf mobile devices for different purposes, including medicine and sports. His research interests also include the application of data fusion and data classification techniques of the data acquired from the different sensors.

FRANCISCO JOSÉ MELERO is the Innovation Manager with the Technological Centre of Furniture and Wood of the Region of Murcia, Spain. Since 2017, he has been coordinating national and international initiatives from different research and innovation programs mainly focused on designing and validating digital solutions for active and healthy aging. He currently acts as the Chair of the Sheldon COST Action CA16226 ''Living Indoor Space Improvement—Smart Habitat for the Elderly'' and is part of the Management Committee of the Actions CA19136 and CA20104. Under the Horizon 2020 Project Pharaon on Pilots for Healthy and Active Ageing (Grant Agreement ID: 857188), he coordinates the deployment of the pilot planned in the Region of Murcia.

TOMÁS PUEBLA MARTINEZ received the degree in psychology from the University of Murcia, the master's degree (Hons.) in personnel and human resources management from ENAE, and the Diploma of Advanced Studies (DEA) degree in health and clinical psychology. He is the Research and Development Project Manager with the Technological Centre of Furniture and Wood of the Region of Murcia (CETEM). He has coordinated the LLP Leonardo Project VETAAL, and he is involved in several European projects: Working TRI21, Woodability, Skills4Succession, SWING, IM-FUTURE, and Supportability. He was mostly related to adapting life environments, job positions, or workplaces to people with some disability according to their skills or needs or enabling senior citizens to understand the characteristics and possibilities using smart home concepts. Also, he is acting as the Quality Assurance and Risk Manager in most of those European projects. He teaches ''innovation and occupational risk prevention project management'' in the university master's degree in design and business management for the furniture and woodworking sectors at the University of Murcia and CETEM (3ECTS). He has collaborated in the review of ESCO skills and occupations in the wood and furniture manufacturing sector. For his credits, he has published four articles and nine books.

NUNO M. GARCIA received the B.Sc. degree (Hons.) in mathematics/informatics and the Ph.D. degree in computer science engineering from the Universidade da Beira Interior (UBI), Covilhã, Portugal, in 2004 and 2008, respectively. Currently, he is an Associate Professor with Habilitation at the Computer Science Department, UBI, and has been a member of the department, since 2012. He has been an Invited Associate Professor with the Universidade Lusófona de Humanidades e Tecnologias, Lisbon, Portugal, since 2010. His main research interests include next-generation networks, predictive algorithms for healthcare and well-being, distributed and cooperative algorithms, and the battle for a free and open Internet.

MARTIN MIHAJLOV currently works as a Researcher with the Laboratory for Open Systems and Networks, Jožef Stefan Institute, Ljubljana, Slovenia. Previously, he worked as an Assistant Professor with the Department of E-business, Faculty of Economics, Saints Cyril and Methodius University in Skopje. His main research interests include human–computer interaction, user experience, and user-centric systems.

IVAN CHORBEV was born in Ohrid, North Macedonia, in 1980. He received the Ph.D. degree from the Faculty of Electrical Engineering and Information Technologies, Saints Cyril and Methodius University in Skopje, Skopje, in 2009. From 2003 to 2005, he was with the Department for Software and Web Development, Makedonski Telekom—T-Home. In 2005, he was with the Faculty for Electrical Engineering, Institute for Computer Science and Engineering, Skopje. His teaching commitments included lectures in programming, web design, and computer animation. He is currently the Dean of the Faculty for Computer Science and Engineering, Saints Cyril and Methodius University in Skopje. He has participated in more than 85 scientific papers in books, journals, and conference proceedings. He has authored one university textbook, and he was a reviewer of another one. His research interests include combinatorial optimization, heuristic algorithms, constraint programming, web development technologies, software testing, computer science application in medicine and telemedicine, medical expert systems, assistive technologies, knowledge extraction, and machine learning.

VLADIMIR TRAJKOVIK was born in Skopje, North Macedonia, in 1971. He received the Ph.D. degree from Saints Cyril and Methodius University in Skopje, Skopje, in 2003. He joined Saints Cyril and Methodius University in Skopje, in December 1997, where he is currently a Professor with the Faculty of Computer Science and Engineering. He has published four books as an author or an editor, 16 chapters in different books published by international publishers, more than 60 respectable journals, and more than 160 conference papers. He has more than 1300 citations, with an H-index of 21. His research interests include information systems analyses and design, distributed systems, ICT-based collaboration systems, and mobile services, focusing on two areas: connected health and ICT in education. He has participated in more than 60 national and international ICT projects to provide novel ICT-based services. He has coordinated 27 of those projects. He is the co-founder or partner in four ICT companies, has done management consultancy in five ICT companies, and has mentored nine ICT companies in the last 15 years. He is the co-founder of two NGOs that promote digital education and the usage of ICT in education. He has organized or participated as an Invited Lecturer, a mentor, or a judge on more than 60 events related to promoting novel ICT services, education, and entrepreneurship in the last five years.