Check for
updates

# Silent no more: a comprehensive review of artificial intelligence, deep learning, and machine learning in facilitating deaf and mute communication

Hanaa ZainEldin[1] · Samah A. Gamel[2] · Fatma M. Talaat[3,7,8] · Mansourah Aljohani[4] · Nadiah A. Baghdadi[5] · Amer Malki[4] · Mahmoud Badawy[1,6] · Mostafa A. Elhosseini[1,4]

## Abstract
People who often communicate via sign language are essential to our society and significantly contribute. They struggle with communication mostly because other people, who often do not understand sign language, cannot interact with them. It is necessary to develop a dependable system for automatic sign language recognition. This paper aims to provide a comprehensive review of the advancements in artificial intelligence (AI), deep learning (DL), and machine learning (ML) technologies that have been used to facilitate communication for individuals who are deaf and mute (D–M). This study explores various applications of these technologies, including sign language interpretation, speech recognition, and text-to-speech synthesis. By examining the current state of research and development in AI, ML, and DL for the D–M field, the survey sheds light on the potential and challenges faced in utilizing AI, deep learning, and ML to bridge the communication gap for the D–M community. The findings of this survey will contribute to a greater understanding of the potential impact of these technologies in improving access to communication for individuals who are D–M, thereby aiding in the development of more inclusive and accessible solutions.

## 1 Introduction

Artificial intelligence (AI) encompasses various disciplines and aims to advance the development of computer systems with the ability to execute tasks that traditionally necessitate human intelligence. The spectrum of assignments can encompass both rudimentary problem-solving exercises and intricate decision-making procedures. AI comprises a range of sub-disciplines, among which are machine learning (ML) and deep learning (DL), which are considered fundamental components of AI. As a constituent of AI, ML focuses on advancing algorithms and models that facilitate the autonomous acquisition of knowledge and enhancement of computer performance without explicit programming. Machine

---

Extended author information available on the last page of the article

learning algorithms acquire knowledge from data and afterward generate predictions or make judgments by leveraging patterns and conducting statistical analysis. The application of this technology is prevalent throughout several fields, encompassing computer vision, natural language processing (NLP), recommendation systems, and autonomous cars (Sarker 2022; Dwivedi et al. 2021).

Deep learning is a branch of ML that draws inspiration from the structural and functional characteristics of the human brain. It employs artificial neural networks, composed of interconnected layers of artificial neurons, to process information and acquire knowledge from extensive datasets. Through this approach, DL algorithms can automatically acquire hierarchical representations of data, leading to promising performance outcomes in domains such as image and speech recognition, NLP, and generative modeling. The field of AI, encompassing both DL and ML, has a substantial historical background dating back to the mid-20th century. The term "artificial intelligence" was officially coined in 1956 during the Dartmouth Conference, marking a significant milestone in establishing this field. In its early years, AI research focused on developing symbol-based approaches and rule-based systems that mimic human reasoning and intelligence. However, these approaches faced limitations in handling uncertainty and dealing with vast data (Groumpos 2022).

Machine learning algorithms can be categorized into three main types: supervised learning, unsupervised learning, and reinforcement learning. Supervised learning involves training ML models with labeled examples, while unsupervised learning discovers patterns and structures in unlabelled data. On the other hand, reinforcement learning is centered around the process of instructing systems to make decisions by utilizing feedback obtained from their surrounding environment (Sarker 2021a; Janiesch et al. 2021).

In recent years, advancements in computational power, availability of large datasets, and breakthroughs in DL architectures and algorithms have accelerated the progress of AI. DL has achieved remarkable success in various domains, surpassing traditional methods in benchmarks and significantly improving various tasks. AI, DL, and ML have widespread applications across many industries. They are used in healthcare for medical imaging diagnosis, drug discovery, and personalized medicine (Das and Behera 2017). In finance, AI algorithms are employed for fraud detection, algorithmic trading, and risk assessment. AI is also used in autonomous vehicles, robotics, virtual assistants, recommendation systems, and many other areas. As AI continues to evolve, researchers and practitioners are addressing its challenges, exploring ways to guarantee AI technologies' ethical and responsible utilization, and advancing the development of explainable AI and AI systems that align with societal values (Vrontis et al. 2022).

Artificial intelligence has shown immense potential to revolutionize various aspects of society, including communication, accessibility, and assistive technologies. In recent years, AI techniques have gained significant traction in addressing the unique challenges of deaf and mute (D–M) people. This research explores the importance of AI in improving the lives of individuals with hearing and speech impairments and its potential for enhancing their communication capabilities and overall inclusivity (Akbar et al. 2022).

## 1.1 Clarification of research question and objective

- Research Question: The primary research question is: "How can advancements in AI, DL, and ML contribute to enhancing communication for the D–M community, and what are the current challenges and future directions in this field?"

- Objective: The main objective of this paper is to conduct a comprehensive review of the existing literature and advancements in AI, DL, and ML technologies, specifically focusing on their application in facilitating communication for the D–M community.

## 1.2 Detailed scope and contributions

- Scope: The paper extensively covers various applications and technological innovations in AI, DL, and ML that aid in sign language interpretation, speech recognition, and text-to-speech synthesis for the D–M community.
- Contributions to Literature:

   – Bibliometric Analysis: This paper provides a detailed bibliometric study, highlighting the most active journals, countries, and research areas in the field, thereby offering a macroscopic view of the research landscape.
   – Technology Evaluation: A thorough evaluation of different datasets and methodologies for sign language recognition, voice assistants, captioning, and speech recognition is presented, emphasizing the strengths and limitations of various approaches.
   – Sociological and Ethical Considerations: Our survey goes beyond the technological aspects and delves into the sociological and ethical implications of employing AI, DL, and ML in D–M communication, providing a holistic view of the field.
   – Future Directions: We propose potential courses of action and future research directions, emphasizing the need for more inclusive and efficient communication solutions for the D–M community.

The D–M community faces different challenges as D–M individuals encounter significant barriers in communication, education, employment, and social interactions. The inability to hear or speak poses challenges in expressing themselves, understanding spoken language, and engaging in everyday conversations. Traditional methods, such as sign language and text-based communication, have limitations regarding widespread usage, accessibility, and understanding by the general population. As shown in Fig. 1, AI technologies have the potential to significantly impact and bring about disruptive changes in tackling these limitations as follows (Yousaf et al. 2018; Papatsimouli et al. 2023):

- AI-Driven Sign Language Recognition: One of the significant applications of AI in assisting the deaf community in recognizing and interpreting sign language. Sign language recognition systems utilize computer vision techniques, including DL algorithms, to interpret hand gestures and motions. The integration of sophisticated methodologies, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), enables AI systems to effectively comprehend sign language and promote seamless interaction among deaf and hearing individuals (Subramanian et al. 2022).
- Automatic Speech Recognition for Textual Communication: AI-powered automatic speech recognition (ASR) systems have shown remarkable progress in converting spoken language into written text. These systems can transcribe verbal communication into text by leveraging ML and NLP techniques. This technology enables individuals with hearing impairments to communicate effectively by reading the transcriptions in real time, breaking down the communication barrier between deaf and hearing individuals (Proksch et al. 2019).
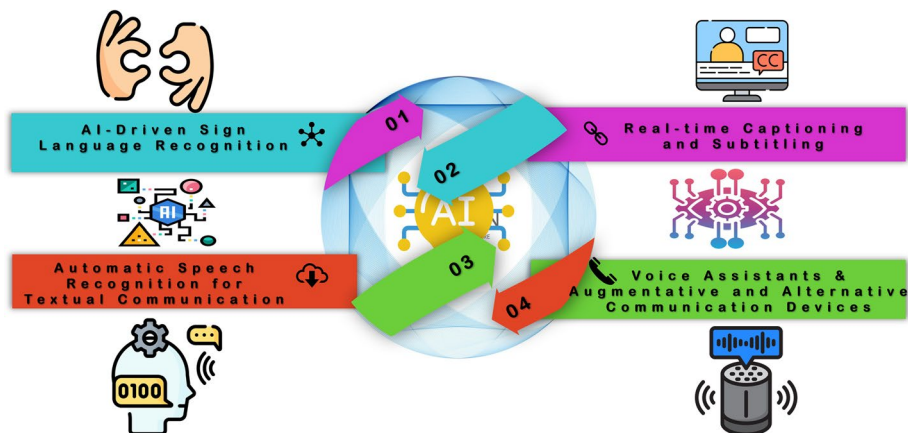
**Fig. 1** The AI technologies for D–M

– Real-time Captioning and Subtitling: AI algorithms also offer real-time captioning and subtitling solutions that can dramatically improve accessibility for the D–M community. By leveraging speech recognition and language processing techniques, AI systems can generate synchronized captions and subtitles for live events, videos, and other multimedia content. This feature enables individuals with hearing disabilities to understand and engage with audiovisual content, making educational materials, entertainment, and public events more inclusive (Masiello-Ruiz et al. 2023).

– Voice Assistants and Augmentative and Alternative Communication (AAC) Devices: AI-enabled voice assistants, including Siri, Google Assistant, and Amazon Alexa, have become increasingly prevalent and can be harnessed to aid individuals with hearing and speech impairments. These voice assistants can provide text-based responses and execute commands, enabling D–M individuals to access information, control smart devices, and perform tasks through voice-activated interactions. Furthermore, AI-driven augmentative and alternative communication devices allow individuals with speech difficulties to express themselves using text, synthesized speech, or image-based communication tools (Jadán-Guerrero et al. 2023).

Integrating AI in D–M communication has tremendous potential to enhance inclusivity, accessibility, and independence for individuals with hearing and speech impairments. The developments in sign language recognition (SLR), automatic speech recognition, real-time captioning, voice assistants, and AAC devices empower the D–M community to overcome communication barriers, promote education and employment opportunities, and foster social interactions. As AI technologies advance, their application in this domain becomes increasingly evident, paving the way for a more inclusive and equitable society (Shahid et al. 2022).

This survey paper aims to offer a comprehensive review of the applications of AI, DL, and ML in facilitating communication for individuals with hearing and speech impairments. Specifically targeting the D–M community, this paper aims to explore the various AI-driven technologies, methodologies, and challenges associated with improving communication accessibility and inclusivity. By examining the existing literature and

advancements in the field, this paper intends to clarify the potential of AI, DL, and ML in empowering D–M individuals to express themselves and engage effectively with the world.

The rest of this paper is organized as follows: Sect. 2 discusses the background of AI, DL, and ML in facilitating communication for individuals with hearing and speech impairments. The Bibliometric analysis is introduced in Sect. 3. Then, Sect. 4 describes the deaf and mute AI, ML, and DL techniques proposed in the literature. Then, in Sect. 5, we detail the comparative Analysis of E-learning and sign-language recognition solutions for assisting individuals with disabilities. After all that, in Sect. 6, the challenges and future research directions are introduced. We conclude the paper in Sect. 9.

## 2 Background

Communication is a fundamental aspect of human interaction, enabling the exchange of information, thoughts, and emotions. For individuals who are deaf or mute, communication barriers pose significant challenges and can hinder their ability to participate fully in social, educational, and professional settings. Traditional methods of communication, such as spoken language or written text, may not be accessible or effective for these individuals. However, recent advancements in AI, DL, and ML have opened up new possibilities for facilitating communication and bridging the gap between individuals with hearing or speech impairments and the broader community (Zhou et al. 2017).

AI, DL, and ML technologies have revolutionized the field of assistive technology, offering innovative solutions to overcome communication barriers. These technologies can potentially empower individuals who are deaf or mute by providing them with effective tools and platforms to communicate more effortlessly, express themselves, and actively engage in various social and professional contexts. This section is divided into seven parts as follows: Fields of Study, Applications, AI, DL, ML, Pattern Recognition and Language Processing, Communication, Accessibility, and Inclusivity.

### 2.1 Fields of study

Computer vision is a field that focuses on enabling computers to understand and interpret visual information from images or videos. It involves tasks such as image recognition, object detection, image segmentation, and scene understanding. Computer vision techniques are widely used in applications like autonomous vehicles, facial recognition, surveillance systems, and augmented reality. Natural Language Processing (NLP) involves the interaction between computers and human language. It encompasses language generation, sentiment analysis, machine translation, and text summarization tasks. NLP techniques enable machines to understand, interpret, and generate human language, leading to advancements in chatbots, virtual assistants, language translation services, and information retrieval systems (Szeliski 2022).

Robotics combines the fields of AI, ML, and DL to design, develop, and program robots capable of performing tasks automatically or with minimal human intervention. Robotic systems can be found in industrial automation, healthcare, agriculture, and exploration (Soori et al. 2023). AI-powered robots can learn from their environment, make decisions, and adapt to different situations, enhancing productivity and efficiency. Reinforcement learning is a subfield of ML that focuses on training agents to make sequential decisions through interactions with an environment. It involves reward-based learning, where an

agent learns to maximize rewards by taking appropriate actions in different states. Reinforcement learning has applications in autonomous systems, game playing, robotics, and resource management.

Data science combines ML, statistical analysis, and domain knowledge to extract insights and knowledge from large datasets. It involves data preprocessing, feature engineering, predictive modeling, and visualization. Data science techniques solve complex problems, make data-driven decisions, and uncover patterns and trends in various industries, including finance, healthcare, marketing, and social sciences. Cybersecurity protects computer systems, networks, and data from unauthorized access, attacks, and breaches. AI and ML techniques are employed in cybersecurity to detect and prevent threats, identify anomalies in network traffic, and enhance security measures. These technologies enable faster and more accurate threat detection, incident response, and vulnerability assessment (Kaur et al. 2021).

These fields continuously evolve, and their applications extend to numerous industries, including healthcare, finance, transportation, entertainment, and more. The advancements in AI, DL, and ML have the potential to shape the future of technology and significantly impact society in various positive ways.

## 2.2 Applications

Pattern identification, language processing, and communication systems are three fundamental applications of AI, DL, and ML (Sarker 2021a). Pattern identification involves teaching AI systems to recognize and categorize specific patterns in vast amounts of data. In fields such as computer vision, AI algorithms are used for image recognition, object detection, and facial recognition, enabling applications like autonomous driving, surveillance systems, and biometric authentication. In natural language processing, AI models can identify patterns in text, enabling sentiment analysis, spam detection, and information extraction. Pattern identification also plays a crucial role in various industries' anomaly detection, fraud detection, and predictive maintenance. Language processing utilizes AI techniques to comprehend and produce human language. AI-powered machine translation systems enable real-time translation between different languages, breaking language barriers and facilitating global communication. AI-based language processing also enables automatic speech recognition, text-to-speech synthesis, and natural language understanding, powering virtual assistants, voice-controlled systems, and chatbots. Additionally, language processing techniques contribute to text summarization, sentiment analysis, and question-answering systems (Zhou et al. 2022). Communication systems encompass how AI technologies enhance accessibility and inclusivity, facilitating effective communication between individuals with different language abilities or hearing and speech impairments. AI-powered systems can convert spoken language into text in real-time for individuals with hearing impairments, enabling better communication through captions or subtitles. SLR systems utilize DL techniques to interpret and translate sign language into spoken or written language, bridging the communication gap between individuals who are deaf or hard of hearing and the wider community. AI-driven communication aids, such as augmentative and alternative communication (AAC) devices, facilitate communication for individuals with speech impairments by converting text or symbols into speech (Robert and Duraisamy 2023).

AI, DL, and ML applications extended beyond pattern identification, language processing, and communication systems. These technologies have found utility in healthcare, finance, recommendation systems, personalized marketing, autonomous robotics, and more. As research and development progress in these fields, new and innovative applications will emerge, further transforming industries and improving the quality of human life.

## 2.3 Deep learning (DL)

Deep learning is an advanced subset of AI that has emerged as a powerful technique for learning and predicting complex data patterns. DL techniques employed neural networks composed of multiple layers of interconnected artificial neurons, enabling the models to learn hierarchical representations of input data. The hierarchical learning approach has proven highly effective in solving challenging tasks such as image recognition, speech recognition, language processing, and pattern recognition. One of the key advantages of DL is its ability to extract meaningful and relevant features from large datasets automatically. Traditional ML approaches often required manual feature engineering, where domain experts identified and defined relevant features for the model. In contrast, DL models could learn and extract significant features directly from raw data, eliminating the need for manual feature engineering. These features allow DL models to effectively capture intricate patterns and correlations that may be difficult for humans to articulate or identify (Wu et al. 2023).

DL has revolutionized the field of computer vision by achieving remarkable accuracy in image recognition tasks. DL models, such as CNNs, have demonstrated unprecedented performance in object detection, image segmentation, and image classification tasks. These advancements have led to applications in autonomous vehicles, medical imaging analysis, surveillance systems, and facial recognition technologies. In natural language understanding, DL techniques have greatly improved language processing capabilities. Recurrent neural networks (RNNs) and transformer models have enabled significant progress in machine translation, sentiment analysis, text generation, and question-answering systems. DL-powered language models, such as GPT (Generative Pre-trained Transformer), have demonstrated impressive language generation capabilities, making substantial contributions to fields like content generation, chatbots, and automated content summarization. DL's impact also extended to data analysis and pattern recognition tasks (Sohail et al. 2023). DL models can uncover complex patterns and correlations in large datasets by leveraging deep neural networks, enabling accurate predictions, classification, and anomaly detection. DL techniques have applications in diverse fields, including finance, healthcare, marketing, recommendation systems, and fraud detection.

The breakthroughs achieved in DL have paved the way for transformative advancements in various industries. DL's ability to learn from vast amounts of data and extract intricate patterns has opened up new possibilities for automation, optimization, and decision-making (Dargazany et al. 2018). However, DL models often require significant computational resources and large amounts of labeled data for training, and they can be susceptible to overfitting. Ongoing research efforts are focused on addressing these challenges, making DL more accessible, interpretable, and robust. As DL continues to evolve, it holds immense potential for driving innovation, improving accuracy, and solving complex problems across industries. Its applications will expand further, empowering advancements in personalized medicine, autonomous systems, intelligent virtual assistants, and more.

## 2.4 Pattern recognition and language processing

Pattern recognition and language processing are two fundamental aspects of AI that have revolutionized numerous domains, enhanced communication systems, and enabled innovative applications. Pattern recognition is a crucial component of AI systems as it involves teaching machines to identify and categorize specific patterns within data. This capability has led to significant advancements in various fields, such as speech-to-text conversion, facial recognition, and object detection (Karras et al. 2022). In speech-to-text conversion, AI systems analyze audio signals and employ pattern recognition algorithms to accurately transcribe spoken words into written text. This technology has greatly facilitated tasks like transcription services, voice-controlled interfaces, and voice commands for smart devices. Facial recognition systems utilize pattern recognition techniques to identify and verify individuals based on their unique facial features, enabling applications in security systems, access control, and personalized user experiences. Additionally, pattern recognition plays a vital role in object detection, enabling AI systems to identify and locate objects within images or videos, which has applications in autonomous vehicles, surveillance systems, and augmented reality (Serey et al. 2023).

Furthermore, pattern recognition has significantly contributed to SLR to bridge the communication gap between deaf or hard-of-hearing individuals and the broader community. AI systems trained in pattern recognition techniques to interpret sign language gestures and convert them into understandable text or speech. This breakthrough has facilitated more inclusive communication and improved accessibility for individuals who rely on sign language as their primary means of communication (Kahlon and Singh 2023). Language processing, on the other hand, focuses on applying AI techniques to comprehend and generate human language. The language processing field has led to remarkable advancements in real-time translation, natural language understanding, and voice assistants. Real-time translation systems employ language processing techniques, such as machine translation and NLP, to automatically translate spoken or written text between different languages. These systems enable seamless communication across language barriers, facilitating international collaboration, tourism, and cross-cultural interactions. Natural language understanding techniques enable AI systems to comprehend and interpret human language, extract meaning, infer intent, and respond appropriately. This capability has led to the development of intelligent chatbots, virtual assistants, and voice-activated systems that can answer questions, provide recommendations, and assist with various tasks. Language processing applications have transformed communication systems by providing efficient, accurate, and interactive interfaces between humans and machines.

The combination of pattern recognition and language processing has the potential to revolutionize numerous industries and improve everyday life. From personalized customer experiences in e-commerce to healthcare diagnostics and personalized medicine, AI systems recognizing patterns and understanding language drive innovation and transform how we interact with technology. However, ongoing research and development are required to address language ambiguity, context understanding, and cultural nuances to enhance further the capabilities and accuracy of pattern recognition and language processing systems (Kumar 2023).

The advancements have opened up new possibilities for individuals with speech impairments, granting them greater autonomy and confidence in their daily interactions. Language processing algorithms have also created seamless communication experiences

for individuals with different language abilities. Real-time captioning and translation technologies, powered by AI, ML, and DL, enable instantaneous conversion of spoken language into written captions or translations in various languages. This approach empowered non-native speakers with limited language proficiency or deaf or hard of hearing to understand and actively participate in conversations, presentations, and multimedia content. By removing language barriers, these technologies foster inclusivity, cross-cultural understanding, and equal access to information and opportunities (Kang 2022). This advancement has the potential to facilitate effective communication between persons proficient in sign language and others who lack proficiency in it. Additionally, AI-powered speech recognition and synthesis technologies can empower individuals with speech impairments to communicate more effectively by converting their input into natural language. Language processing algorithms can also aid in real-time captioning and translation, facilitating smooth communication between individuals with different language abilities. AI in the field of D–M has the potential to revolutionize how communication is facilitated, making it more accessible and inclusive for everyone.

Despite the significant advancements in AI, DL, and ML in facilitating communication for individuals who are deaf or mute, several challenges and considerations need to be addressed. Ethical considerations, privacy concerns, and potential biases in developing and deploying these technologies must be carefully examined. Interdisciplinary collaborations and user-centered design approaches are crucial to effectively addressing the needs and perspectives of individuals with hearing or speech impairments (Sullivan et al. 2018).

The problem addressed in this review is the communication barrier faced by individuals who communicate through sign language. Due to the lack of understanding of sign language, these individuals struggle to interact and communicate effectively with the larger society. The paper aims to explore the development of a reliable system for automatic sign language recognition using artificial intelligence (AI), deep learning, and ML technologies. The goal is to failitate communication for individuals who are D–M by bridging the communication gap and improving access to communication. Although the development of a reliable system for automatic SLR has been widely studied, it involves several issues and challenges as follows:

1. Variability in sign language: Sign language can vary across regions and cultures, leading to differences in hand gestures, facial expressions, and body movements. Developing a system that can recognize and interpret these variations accurately poses a significant challenge.
2. Data availability: Obtaining a large and diverse dataset of sign language gestures can be challenging. Collecting high-quality annotated data that covers different sign language variations and contexts is crucial for training robust recognition models.
3. Real-time recognition: Real-time sign language recognition requires low-latency processing for smooth and natural communication. Achieving high accuracy while maintaining fast inference speeds is challenging when developing practical and efficient systems.
4. Handling occlusions and noise: Sign language recognition systems must be robust to occlusions caused by hands interacting with objects or other body parts. Additionally, environmental noise and variations in lighting conditions can affect the quality of input data, making accurate recognition more challenging.

5. Contextual understanding: Sign language often relies on the surrounding context to convey meaning. Understanding the context and disambiguating signs based on the conversation topic or context-specific gestures adds complexity to the recognition task.
6. User adaptation and personalized recognition: Different individuals may have distinct signing styles, speeds, and idiosyncrasies. Developing adaptive systems that can personalize recognition for individual users and accommodate their unique signing characteristics is a challenge.
7. Ethical considerations: As with any AI system, ensuring fairness, privacy, and unbiased performance in sign language recognition is crucial. Addressing potential biases, protecting user privacy, and promoting inclusivity are important ethical considerations in developing and deploying these technologies.

By addressing these challenges, advancements in AI, deep learning, and ML can revolutionize communication for D–M individuals, enabling more inclusive and accessible solutions.

## 3 Bibliometric study

Pattern recognition in AI systems refers to identifying and classifying distinct patterns, an essential feature for functionalities such as speech-to-text conversion and sign language recognition (SLR). Utilizing AI to process language involves implementing AI techniques to understand and generate human language, facilitating real-time translation and captioning in communication technologies. Consequently, AI, along with deep learning (DL) and ML, has significantly redefined the domains of pattern recognition, language processing, and communication technologies, offering individuals with hearing and speech challenges unprecedented access and inclusivity (Tobore et al. 2019; Mariappan and Krishnan 2023).

The discipline of D–M communication presents specific challenges and possibilities. AI innovations can improve communication, access, and inclusivity for those with hearing and speech impairments. SLR stands as a pivotal area wherein AI could exert a profound influence. By employing deep learning algorithms, AI systems decipher sign language gestures and translate them into intelligible text or spoken words (Papatsimouli et al. 2023; David et al. 2023). Such progress promises to enhance interaction between sign language users and those unfamiliar.

Moreover, AI-facilitated speech recognition and synthesis technologies are poised to assist individuals with speech impairments to achieve clearer communication by transforming their inputs into natural-sounding language. Algorithms for processing language further contribute to real-time captioning and translation, smoothing out interactions among people of varying linguistic capabilities. Integrating AI in D–M communication promises a transformative impact on facilitating interaction, rendering it more accessible and inclusive for all.

1. Inclusion criteria:

   – Time Range: Papers published between January 1, 2015, and December 31, 2023.
   – Publication Type: Peer-reviewed research papers, conference proceedings, and review articles.
   – Content: Publications focusing on applying AI, DL, and ML in communication for the D–M.

– Language: Articles published in English.

2. Exclusion criteria:

   – Non-Peer Reviewed Sources: Grey literature, such as theses, dissertations, and non-peer-reviewed articles.
   – Irrelevant Topics: Papers not directly related to AI, DL, ML, and D–M communication.
   – Duplicate Publications: Articles published in multiple sources.

3. Sources and databases:

   – Primary Database: Scopus, for its comprehensive coverage and advanced analytical tools.
   – Secondary Sources: PubMed, IEEE Xplore, and Google Scholar.
   – Rationale for Selection: These databases were selected due to their extensive coverage of scientific publications in computer science, AI, and communication studies.

## 3.1 Bibliometric analysis

The research on intelligent systems for helping the D–M communicate based on DL and ML techniques published over the past nine years is analyzed in this section using a bibliometric analysis technique. The process is carried out to point out the current knowledge boundaries and identify any prospective research gaps. The selection of nine years of research was made to provide a more comprehensive picture of how research has developed throughout these times and to help researchers better comprehend the activities that makeup research patterns and their features. Using this information is useful since it explains the patterns in scientific publications among institutions, nations, journals, and authors.

### 3.1.1 Search strategy and keywords

- Search Strategy: Employing Boolean operators (AND, OR) to refine search queries.
- Keywords: "Artificial intelligence," "Deep learning," "Machine learning," "Communication," "Deaf," "Mute," "Hearing impaired," and related phrases.
- Combination of Terms: Keywords were used in various combinations to ensure comprehensive coverage.

### 3.1.2 Data analysis methods and tools

- Bibliometric Software: Using the Scopus analysis tool and Vosviewer for bibliometric analysis.
- Analysis Techniques:

  – Keyword Frequency Analysis: To identify the most prevalent themes in literature.
  – Co-Authorship Analysis: To examine collaboration patterns among authors and institutions.
  – Citation Analysis: To assess the impact and influence of the publications in the field.

Numerous keywords were merged throughout the search process to find research articles on helping the D–M communicate based on DL and ML techniques. Figure 2 shows an analysis of term frequency for the search. From 2015 through 2023, a thorough bibliometric study of research trends in helping D–M people communicate based on DL and ML techniques was conducted. Figure 3 depicts the block diagram illustrating the process of gathering and analyzing search results from the Scopus database in the context of publication retrieval. The search terms employed for querying the Scopus database encompassed the following keywords: "D–M communication," "machine learning," "deep learning," "artificial intelligence," and "intelligent systems." We considered the final search containing ML, DL, and AI keywords. Lastly, a bibliometric analysis of 115 published papers was used to discover research trends in ML, DL, and AI techniques for helping the D–M communicate easily with the world.

The main objective is to recognize and comprehend historical trends and publishing patterns by journals, regions, and collaboration across institutions and organizations. As time has passed, intelligent techniques to help D–M people have drawn more and more attention from researchers. Consequently, it will be compelling in this field to look at the general research trend and the new research direction.

We considered the final search containing ML, DL, and AI keywords. Lastly, a bibliometric analysis of 99 published papers was used to discover research trends in ML, DL, and AI techniques for helping the D–M communicate easily with the world.
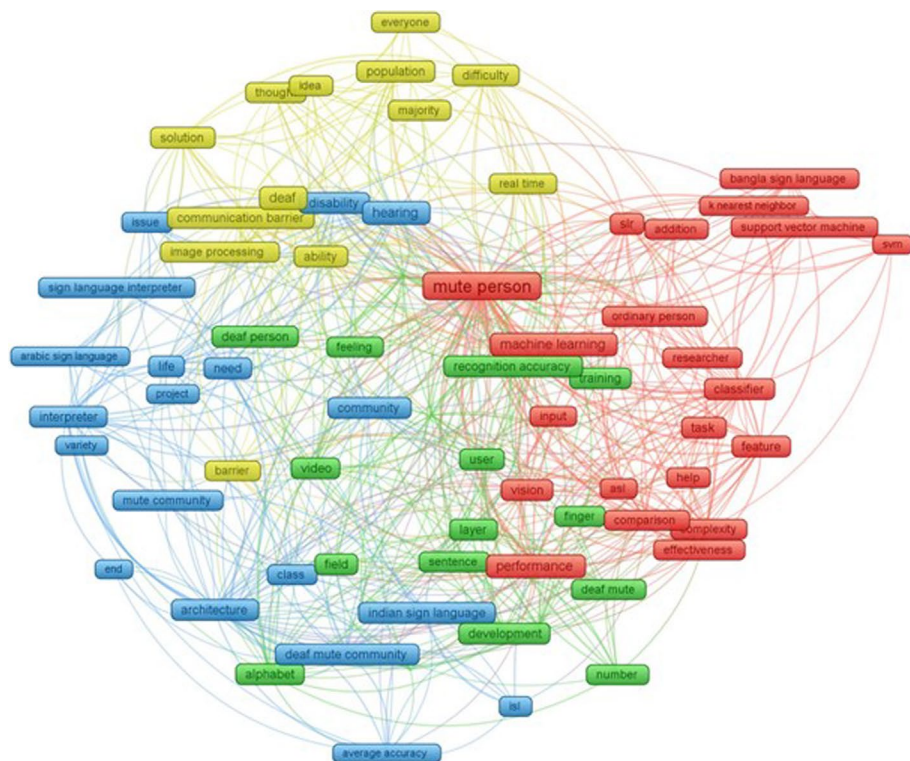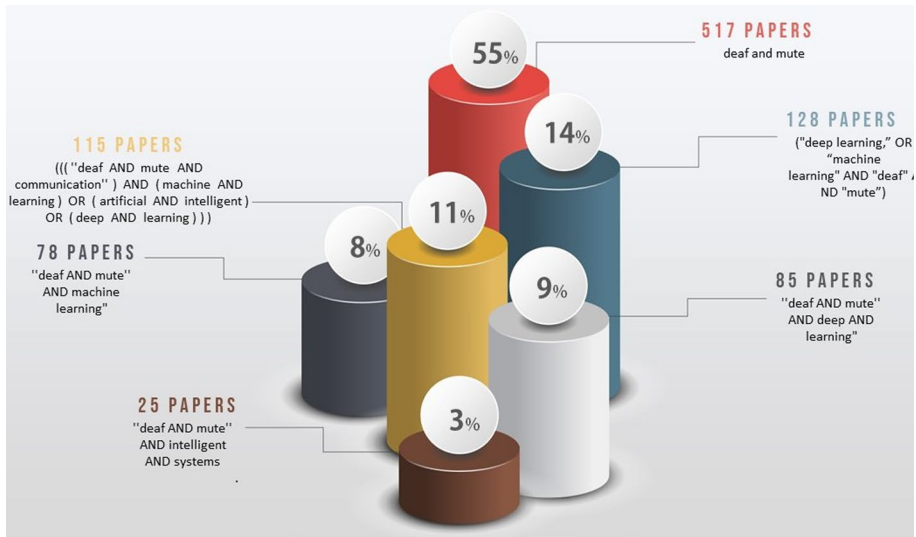


**Fig. 2** keyword frequency analysis

**Fig. 3** Schematic of how publications are collected and analyzed in the Scopus database

### 3.1.3  Findings discussion and analysis

The findings of the bibliometric evaluation of the research on D–M communication based on DL and ML techniques published during the preceding nine years are presented in this section. It is vital to dig deeply into each document and extract the relevant words to verify the main topic area of the study on D–M communication based on DL and ML techniques. The analysis must identify trends in emerging subjects and hotspots that may benefit academic disciplines, community interaction, and innovation. The analysis of keywords linked to D–M communication based on DL and ML techniques yielded 115 results, as mentioned above.

Figure 2 shows keyword frequency analysis for this search. Four groupings of words were produced by using the data. The phrase "mute person" is the one that is highlighted the most throughout the whole network, not only in Group 1. The primary keywords contained in this cluster mostly relate to machine learning, support vector machines (SVM), performance, deaf people, and sign language interpreters. The term "recognition accuracy" stands out in Group 2. Hand gesture recognition, alphabet, real-time, and computer vision are the key phrases in this cluster. The keyword "community" is emphasized in Group 3 and displays the key terms associated with the deaf-mute community: sign language, disability, hearing, life, and challenge. Finally, the words in group 4 that are "communication barriers" include deaf, mute, community, normal people, and real-time recognition. We will discuss and analyze research papers and document results in the Scopus database. The research revealed that 115 publications in the field of intelligent systems for the D–M were published between 2015 and 2023 in 30 different journals, according to the Scopus database. Figure 4 presents the cite score of the most active journals in the study's field by year to avoid a long chart.

Figure 5 shows yearly documents and papers from 2015 to 2023. According to Fig. 4, the year 2022 shows the most publications, with 33 articles released, followed by 2021, with 22 papers. 2016 had the lowest number of publications, with only one article. From
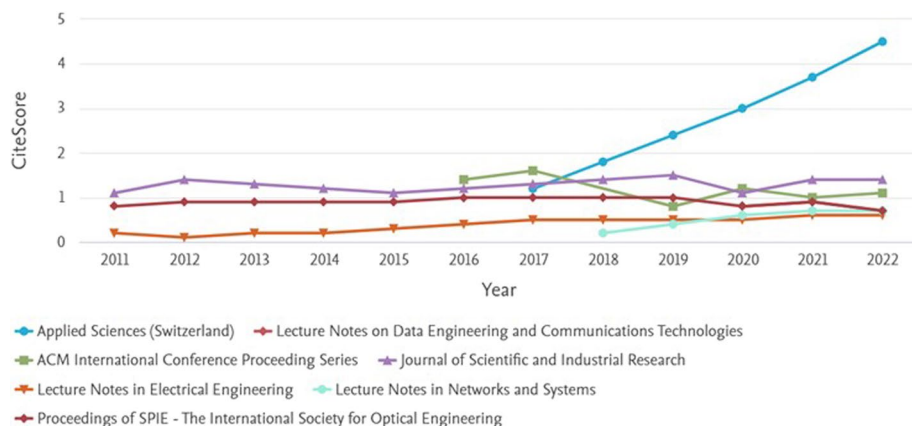
**Fig. 4** Cite score publication by year
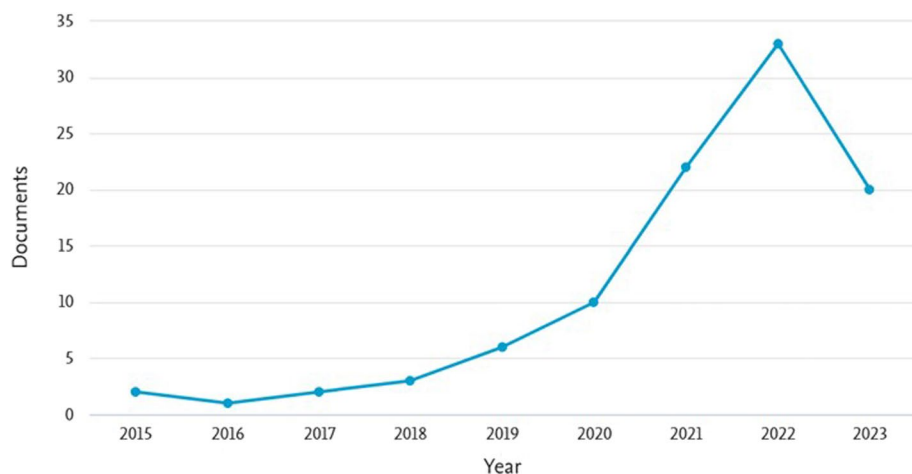


**Fig. 5** Papers and documents by year

2020 to 2023, research production is expected to be more than twice as fast as it has been since 2015.

The number of articles published in every country is calculated based on each country's participation. This search results in 115 research outputs from 25 countries during the past eight years, which have published at least one full-length research publication on intelligent systems in D–M communication-based intelligent systems (AI, ML, DL). The top 10 countries that have contributed most to the expansion of intelligent systems for D–M research over the past eight years are depicted on a map in Fig. 6. About 72% of all publications worldwide came from India, Bangladesh, and China. India is the global leader, topped the list with 49 publications during the previous eight
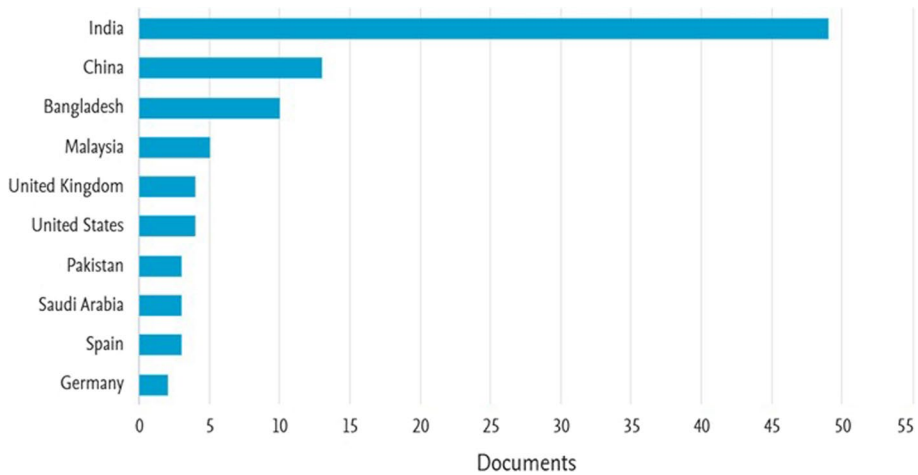
**Fig. 6** The publications count by country

years, accounting for 49.4% of all articles worldwide. Bangladesh and China contributed 13.1% and 10.1%, respectively.

Figure 7 shows publication distribution by document type. It is observed that conference papers have the highest values at 72.7%, followed by article types, which represent 21.2.

Figure 8 shows the distribution of documents by subject area. Computer science and engineering account for approximately 50% of the subject areas. It is shown that computer science is the most popular field for DL and ML techniques for the D–M at 31.1%, followed by engineering at 19%. Also, Mathematics, physics, and decision science subject areas have approximately 30% among different subject areas.
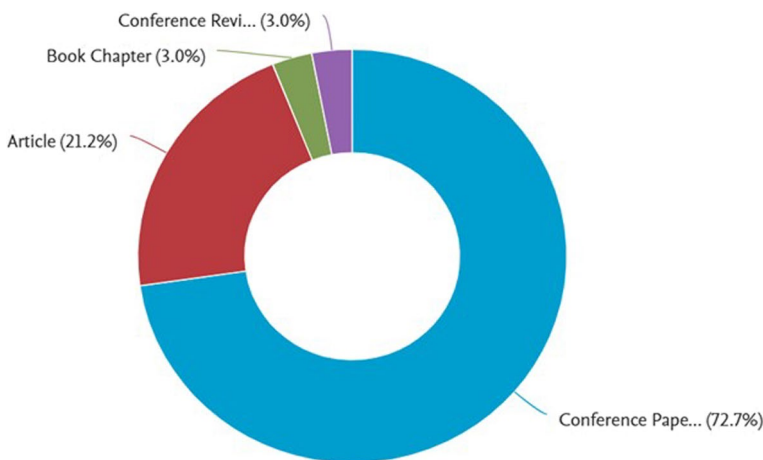


**Fig. 7** The document by type distribution
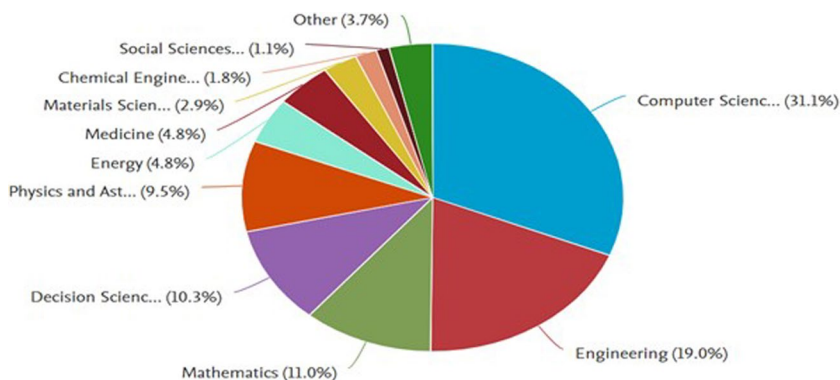
**Fig. 8** The document by type distribution

## 4 Deaf and mute AI, ML, and DL techniques

Numerous sophisticated methodologies facilitate communication between D–M individuals and the broader society. In the subsequent discussion, we will outline contemporary AI, ML, and DL advancements for those with hearing and speech impairments.

### 4.1 Deaf and mute benchmark datasets

Several publicly available datasets may be used to assess how well static, independent, and continuous sign language recognition systems perform. Below are some of the benchmark datasets used by many researchers. citeD1: Alphabet gestures, 0–9 number gestures, and a gesture for space, i.e., whether the deaf or dumb people indicate the gap between two letters or two words—are all included in the dataset, which comprises 37 distinct hand sign gestures. The dataset is divided into two folders. (1) Gesture Image Data comprises colored pictures of hands doing various gestures. Each gesture image is 50 × 50 pixels in size and is in its folder with A–Z files containing photos of A–Z gestures, 0–9 folders containing photographs of 0–9 gestures, and the '_' folder containing images of the space gesture. With an average of 37 gestures and 1,500 photos per gesture, 55,500 images total for all gestures in the first and second folders. (2) Gesture Images Pre-Processed Data has the same folders and images as the first folder, 55,500. These photos were transformed into threshold binary for training and testing. CNNs operate well with the dataset for model training and gesture prediction.

Sign Language MNIST (American Sign Language) (Sign Language MNIST 2017): Hand and face gestures convey meaning in ASL. The collection comprises 27,455 hand-sign pictures, each measuring 28 by 28 and being in grayscale. The dataset format roughly resembles the traditional MNIST. There are no instances of 9 = J or 25 = Z since they entail hand motion; however, images in the collection include labels ranging from 0 to 25 that represent the letters A to Z. Approximately half the size of conventional MNIST, the training data (27,455 examples) and test data (7,172 cases) are comparable to a header row of the label, pixel1, pixel2, pixel3,...,pixel784. The original hand gesture picture data represented multiple users performing movements across various backdrops.

The American Sign Language Lexicon Video Dataset (ASLLVD) by Athitsos et al. (2008) contains 3800 ASL signs in video of excellent quality frames performed by four native signers. ASL is available in isolated and continuous forms in the RWTHBOS-TON-104 (Zahedi et al. 2006) and RWTHBOSTON-400 (Dreuw et al. 2008) datasets. The isolated sign language SL in the RWTHBOSTON-104 dataset has an alphabetical list of 104 signs and 201 terms performed by three signers. The RWTHBOSTON-400 dataset was developed to achieve continuous ASL recognition. It has four signers and consists of 843 phrases with a collection of 406 words.

A novel dataset of dynamic hand gestures for the Indian sign language words commonly used for emergency communication by deaf COVID-19 patients is proposed by Venugopalan and Reghunadhan (2023). It consists of 900 video clips of nine hand gesture classes (100 samples for each), which are distinguished by three distinct forms (flat, spread, and V), as well as three distinct movements (left, right, and contract), and were recorded by two participants using five different types of lighting.

An Indian sign language (ISL) used by deaf farmers (Venugopalan and Reghunadhan 2021) includes 260 original videos totaling $1080 \times 1920$ pixels, with 20 examples for each type of gesture. The dataset consists of 900 video clips, with 100 examples for each of nine classes of hand gestures, each of which has three distinct forms (flat, spread, and V), as well as three distinct movements (left, right, and contract). The video snippets were recorded using ten random actions from two people and five lighting conditions. The dataset was split into training and testing, with 450 sample movies used for testing and the other 450 used for training.

A dataset for recognizing each Urdu alphabet in Pakistani sign language PSL (Imran et al. 2021). The Urdu language has 37 letters, with 40 photos of each alphabet in various hand and finger orientations taken by the webcam. The alphabets of Urdu sign language are divided into groups based on the dimensions and orientations of the hands and fingers. One thousand four hundred eighty images are obtained, and 37 categories are formed using these images. Each category of photographs is assigned a different Urdu alphabet. The created dataset would aid academics working on sign identification and translation to create deaf people in Pakistan. The Bengali Sign Language (BSL) dataset in Islam et al. (2022) is entirely novel and exclusive. It includes pictures of 11 Bengali sign language classes gathered from several volunteer signers. There are 11 classes and 11 sign words in the dataset. The collection has 1105 images overall, with more than 100 images of each kind. Besides, the authors in Siddique et al. (2023) construct a massive collection of 49 classes of Bangla sign language, each with about 80 photos. Several backgrounds and conditions of lighting, as well as both left and right hands, have been used to construct the proposed dataset. At Prince Mohammad Bin Fahd University, Al Khobar, Saudi Arabia, a new, complete, fully labeled collection of Arabic Sign Language pictures, known as ArSL2018 (Latif et al. 2019), was presented. The dataset consists of 54,049 grayscale images with sizes $64 \times 64$ for each. Dataset images were changed to have varied backgrounds and lighting effects. Additionally, there are 36 classes of alphabets (A–Z) and numerals (0–9) in the Massey University dataset (Barczak et al. 2011). There are 2,160 photos for the Arabic Sign Language (ArSL) dataset.

The British Sign Language (BSL) dataset is proposed in Schembri et al. (2013), consisting of recordings of 249 BSL-speaking individuals speaking while 6330 hand gestures were annotated from the discussion. German sign language recognition GSL includes datasets named RWTH-PHOENIX-Weather 2014 (Forster et al. 2014) and SIGNUM (Von Agris and Kraiss 2007). The continuous sign language in RWTH-PHOENIX-Weather 2014 consists of 6861 phrases and 1558 vocabulary words. Meanwhile, 450 fundamental

**Table 1** Deaf and mute datasets with different languages

| References | Dataset | Language | Availability |
|---|---|---|---|
| Sign Language Gesture Images Dataset (2020) | Sign Language Gesture Images Dataset | American | Public |
| Sign Language MNIST (2017) | Sign Language MNIST | American | Public |
| Athitsos et al. (2008) | The American Sign Language Lexicon Video Dataset (ASLLVD) | American | Public |
| Zahedi et al. (2006) and Dreuw et al. (2008) | RWTHBOSTON-104 and RWTHBOSTON-400 | American | Public |
| Venugopalan and Reghunadhan (2023) | Indian Sign Language for positive deaf COVID-19 patients | Indian | Limited access |
| Venugopalan and Reghunadhan (2021) | Indian sign language (ISL) used by deaf farmers | Indian | Limited access |
| Imran et al. (2021) | Urdu alphabet in Pakistani sign language PSL | Pakistani | Public |
| Islam et al. (2022) | Bengali Sign Language (BSL) | Bengali | Limited access |
| Latif et al. (2019) and Barczak et al. (2011) | Arabic Sign Language pictures ArSL2018 and Arabic Sign Language (ArSL) | Arabic | Public |
| Schembri et al. (2013) | British Sign Language (BSL) | British | Limited access |
| Forster et al. (2014) and Von Agris and Kraiss (2007) | German Sign Language (GSL) and RWTH-PHOENIX-Weather 2014, SIGNUM | German | Limited access |

gestures and 780 phrases signed by 25 signers make up the vocabulary size of the SIG-NUM Database. Table 1 summarizes the previously discussed deaf and mute datasets.

## 4.2 Deep learning-based techniques

Deep learning models have been extensively utilized in various applications to perform classification tasks on many data types, including images, text, and audio. This approach has been demonstrated to be highly effective and influential. Deep learning models can extract task-specific information from large datasets. However, due to the deterministic nature of DL models, they cannot effectively process unclear or inaccurate data. Deep learning models often demonstrate high sensitivity to noise in the data, leading to poor performance when faced with ambiguous or uncertain data.

Furthermore, the classification performance of deep learning models is diminished when trained on datasets containing many features or high-dimensional data that include redundant and irrelevant properties. In scenarios characterized by high dimensions, deep learning models necessitate a substantial quantity of data and requirements that frequently grow exponentially in proportion to the number of features. Low performance in classification tasks is caused by DL model issues with irrelevant features and data ambiguity.

An optimized fuzzy deep learning (OFDL) model for data categorization utilizing a Non-Dominated Sorting Genetic Algorithm II (NSGA-II) was developed (Yazdinejad et al. 2023). Utilizing the NSGA-II algorithm in the context of multi-modal learning is a key aspect of optimizing the OFDL framework. The OFDL framework leverages optimization techniques to design DL models and incorporate fuzzy learning methodologies effectively. The OFDL approach initially considers intelligent feature selection to get an efficient classification. The feature selection process entails reducing the number of features while enhancing accuracy, achieved by raising the weights assigned to selected features. By doing so, OFDL aims to establish the optimal balance between two conflicting objective functions. Depending on their objective functions, OFDL employs Pareto optimal solutions for multi-objective optimization. The optimization process is achieved by applying the NSGA-II algorithm, which facilitates the attainment of optimized backpropagation and fuzzy membership functions. According to a neuroscientific study, synced audiovisual stimuli will produce a larger response from visual perception than an independent stimulus. Numerous studies have shown that audio cues can change how people view natural video situations. As a result, authors in Chen et al. (2021) presented a multi-modal framework using auditory and visual data to predict video saliency. It consists of four modules: feature fusion, auditory feature extraction, visual feature extraction, and semantic interaction between auditory and visual features. A network architecture of DL was outlined to fulfill the functions of these four modules by utilizing auditory and visual signals as inputs. The end-to-end approach was designed to interact meaningfully with audible and visual stimuli. The numerical and visual findings demonstrate that our approach significantly outperforms eleven current saliency models, some of which are state-of-the-art DL models, regardless of the auditory stimuli.

For those with hearing loss, sign language, often called silent conversation, serves as their main form of gesture-based visual communication. Static and dynamic sign languages are employed by those who cannot speak or hear and the general people. Adopting sign language into our culture and developing a straightforward and cost-effective method for its detection is of heightened significance. These measures are vital to mitigate the exclusion

experienced by individuals who are deaf and unable to communicate, thereby ensuring their inclusion in the rapidly evolving societal landscape.

An automatic Bangla sign language (BSL) identification approach was proposed using DL techniques and a Jetson Nano edge device (Siddique et al. 2023). These methods have been validated using a bespoke author-curated dataset of 3,760 photos and 49 categories, which was utilized for training the deep-learning models from the open-source database Okkhornama. Within the Roboflow framework, the provided images undergo preprocessing procedures, specifically auto-orientation, and scaling, resulting in their dimensions being adjusted to 416 by 416 pixels. Then, they deployed the Detectron2, EfficientDet-D0 with TensorFlow, and PyTorch-built YOLOv7 models. Also, they used the Jetson Nano, a portable and very versatile NVIDIA computer, to develop an item recognition model and subsequently infer it from test images in real-time. The Dectectron2 model exhibited superior performance in terms of different performance metrics.

Authors in Islam et al. (2022) used Transfer Learning (TL) techniques to recognize eleven common Bengali Sign Words. They modified four established and highly regarded transfer learning algorithms and subsequently evaluated the performance of each system using their dataset. Besides, the dataset underwent preprocessing before division into training and test sets to enhance performance. The proposed methodology demonstrated notable levels of accuracy, with the VGG16 model exhibiting superior performance in terms of both time and accuracy when compared to the other models. The authors have provided a comparative analysis with pre-existing models to demonstrate the efficacy of their proposed models.

Prasath (2022) discusses modern voice recognition, which uses AI techniques to predict human speech. It has various applications such as identity verification, assisting individuals with hearing impairments or speech disabilities, electronic voice eavesdropping, etc. The main challenge in voice recognition is predicting the important steps in the process. The paragraph focuses on proposing an efficient approach using DL. It involves four phases: data acquisition, voice preprocessing, word segmentation, and classification. An online dataset is used for experimentation. The input voice is preprocessed using A refined mechanism for voice activity detection. Word segmentation is performed using Grab Cut segmentation. Finally, the hybrid model of RNN and CNN is used for classification to improve prediction accuracy compared to a single classifier model.

The work in Nahar et al. (2023) proposed a fresh and creative method for automatically converting Arabic speech into Arabic Sign Language (ATSL). The suggested method retrains 12 image identification models using a DL-based classification methodology and the TL method. The image-based translation methodology uses classification as a machine learning technique to map sign language motions to matching letters or phrases. The results demonstrate that the suggested model, with a translation accuracy of 93.7%, classifies Arabic-language signs more precisely and quickly than conventional image-based models.

To translate Indian sign language digits effectively, Sabharwal and Singla (2022) suggested a seven-layered, two-dimensional CNN built on DL and uses the swish activation function. The suggested approach incorporates several techniques commonly employed in DL models, including max pooling, batch normalization, dropout regularization, and the Adam optimizer. A publicly available "India Sign Language" dataset consisting of numerical data with a size of around 12 kilobytes was used. The average validation accuracy for our model is 99.22%, while the maximum validating accuracy is 99.55%. Swish activation functions outperform standard ReLU and Leaky ReLU activation functions.

A MobileNetV2-based TL model (Lum et al. 2020) is added to the previously published CNN model for detecting American Sign Language. The latter model was successfully

adapted to a dataset about 18 times bigger and contained five more types of hand signals. The claimed recognition accuracy was over 98%. The algorithm was also perfect for mobile device implementation since it had comparatively fewer parameters and less demanding computational operations than other deep learning architectures. The concept will be essential for implementing sign language translation applications to improve deaf-mute people's ability to communicate with hearing people. In Huang et al. (2015), a novel 3D CNN was introduced, which can independently extract distinctive spatial-temporal attributes from the unprocessed video data, hence avoiding the need for feature construction. The 3D CNN provides several video input channels to enhance efficiency, encompassing color, depth, trajectory data, body joint positions, and depth cues. The superiority of the proposed model over standard methods that rely on hand-crafted features is demonstrated through its validation on a real dataset obtained using Microsoft Kinect. Furthermore, the studies explored a user-centered approach, tailoring their solutions to different user groups' specific needs and challenges. For instance, recognizing depression using audiovisual cues (He et al. 2022) demonstrated the potential of AI and deep learning in advancing mental health diagnosis and care for individuals who may struggle to express their emotions verbally. This personalized approach highlights the transformative impact of technology in addressing a wide range of communication barriers. Creating specialized datasets specific to different sign languages, such as PSL (Imran et al. 2021) and BSL (Siddique et al. 2023), underscores the importance of accurate and contextually relevant training data. These datasets serve as valuable resources for training and evaluating machine learning models, ensuring that the technology accurately recognizes and translates unique signs and gestures. Table 2 summarizes different Deep learning-based techniques for the deaf and mute, describing the advantages and disadvantages of each technique. Understanding the strengths and weaknesses of these techniques is crucial for selecting the most suitable approach based on the specific requirements and constraints of the deaf and mute communication context. A seven-layered, 2D-CNN built on DL and uses the swish activation function proposed in Sabharwal and Singla (2022) and a MobileNetV2-based TL model proposed in Zhang et al. (2022) achieved high accuracy compared with other state-of-the-art techniques.

## 4.3 Machine learning-based techniques

The Urdu sign language, a visual-gestural system, is employed by those with hearing impairments in Pakistan as a means of communication. Nevertheless, the Pakistan Sign Language (PSL) datasets remain inaccessible. Photos of various hand configurations were collected using a webcam to create the PSL dataset (Imran et al. 2021). Each hand arrangement has 40 photos in this piece, each with different orientations. Furthermore, a mobile application for Android was developed using the PSL dataset, which uses machine learning techniques to bridge the communication gap between individuals who are deaf and those who are not effectively. The Android application can identify the Urdu alphabet by analyzing input hand configuration. Potential communication information may be contained in throat vibration signals. However, a thorough study of throat vibration signals is currently lacking.

Authors in Fang et al. (2023) suggested a novel throat-language decoding system (TLDS) that uses adaptable, inexpensive, self-powered sensors to capture throat vibration signals and machine learning classifiers for semantic interpretation. A flexible piezoelectric polyvinylidene fluoride (PVDF) sensor was constructed to gather the signals from the throat vibrations. High softness, great response repeatability, outstanding linear sensitivity,

**Table 2** Deep learning-based techniques for the deaf and mute

| Study | Technique | Dataset | Accuracy | Pros | Cons |
|---|---|---|---|---|---|
| Yazdinejad et al. (2023) | A Non-Dominated Sorting GA II (NSGA-II)–based OFDL model for data classification | Dirty-MNIST dataset, Scene-15 dataset, and High-frequency trading (HFT) dataset | 95.20% | Integrating fuzzy logic and DL techniques improves classification performance. The proposed optimization technique led to better model configurations and higher accuracy | Complexity and Computational Cost |
| Chen et al. (2021) | A multimodal framework using auditory and visual data. It consists of four modules | Dataset 1: 60 movies total. Dataset 2: 15 movies chosen from the AMI Meeting Corpus | N/A | Understanding that saliency in audiovisual data has practical implications in several domains, such as multimedia content analysis and human-computer interaction | The performance of video saliency prediction needs to be improved |
| Siddique et al. (2023) | An automatic BSL identification system utilizing DL techniques and a Jetson Nano edge device is created | Construct a massive collection of 49 classes of BSL, each with about 80 photos. An immense open-source database of Bangla sign letters is called Okkhornama | 94.9 mAP@ 0.5 | The proposed technique leverages the power of neural networks to detect and recognize sign language gestures accurately. A new BSL dataset is proposed | The limited dataset may hinder the generalization ability of the proposed system, potentially leading to reduced performance on unseen sign language gestures |
| Islam et al. (2022) | A TL technique is used to recognize 11 common Bengali Sign Words. The authors modified four well-performed and renowned TL algorithms and tested their dataset with each | The dataset utilized is entirely novel and exclusive. The collection has 1105 images overall, with more than 100 images of each kind | The best validation accuracy is 92.41% | SLR has significant practical implications, as it can facilitate communication and inclusion for individuals with hearing and speech impairments. High detection accuracy | limited generalization ability |

**Table 2** (continued)

| Study | Technique | Dataset | Accuracy | Pros | Cons |
|---|---|---|---|---|---|
| Prasath (2022) | A combination of RNN and CNN is used for classification to improve prediction accuracy compared to a single classifier model | An available online dataset | N/A | Integrating multiple learning methods potentially improves accuracy in real-time deaf applications | Insufficient training data. Issues such as difficulty in effectively identifying accents, dialects, or speech patterns might still arise |
| Nahar et al. (2023) | The suggested model retrains 12 image identification models using a DL-based classification methodology and the TL method | A fresh and creative method for automatically converting Arabic speech into Arabic Sign Language (ATSL) is proposed | 93.70% | The model can have significant practical implications by providing real-time support and accessibility to deaf individuals | Limited scalability and robustness. Accuracy needs to be improved |
| Sabharwal and Singla (2022) | A seven-layered, 2D-CNN built on DL and uses the swish activation function is proposed to translate ISL digits effectively | 12 K image open-access numeric tailored dataset of India Sign language | 99.55% | High detection accuracy. The proposed model benefits from improved model convergence and accuracy in SL digit recognition by incorporating Swish activation | Need to generalize and test on more sign and alphanumeric datasets |
| Zhang et al. (2022) | A MobileNetV2-based TL model is added to the previously implemented CNN model for detecting ASL | There are 29 categories and a total of 87,000 images. There are 26 categories for the letters A through Z, plus three additional categories for SPACE, DELETE, and NOTHING | 98% | MobileNetV2 is a lightweight CNN architecture designed for efficient computation and deployment on resource-constrained devices. Detecting ASL with high accuracy | Real-time video processing is required |

and long-term durability are all features that make the sensor safe to use on human skin. The neck vibration signals were denoised, and the Grid Search-Support Vector Machine (GS-SVM) was used to extract the characteristics and nonlinear dynamics features of time-frequency dynamics. Letter recognition accuracy ranged from 87.26% for multi-person recognition to 90.55% on average for single-person recognition. Additionally, speaker recognition and straightforward semantic recognition had accuracy rates of 95.97% and 97.50%, respectively.

In their study, Aazam et al. (2021) conducted case studies to examine the implementation of smart healthcare, safety, and emergency response in indoor and outdoor settings. They demonstrated the feasibility and effectiveness of providing opportunistic healthcare and safety measures in these scenarios. Furthermore, the study presented an analysis of the function of ML in three specific use scenarios of smart and opportunistic healthcare, as well as its application in task offloading. Besides, this study examines the assessment of three distinct ML algorithms, namely KNN, SVM, and naive Bayes (NB), as applied in the case studies. The preliminary performance outcomes presented in this research shed light on the influence of executing machine learning on individual edge nodes and femto-cloud regarding energy consumption.

A hand gesture recognition method for sign language communication, demonstrating promising results with a high recognition rate, was proposed in Sahana et al. (2020). The research advances the interface between individuals with speech and hearing disabilities and computer systems, facilitating improved communication and accessibility. The gesture recognition system described in this study operates through three main steps: preprocessing the publicly available raw data, normalizing orientation, and feature extraction. The raw data is processed to enhance its quality and prepare it for further analysis. Orientation normalization is a crucial step in feature computation as it guarantees the invariance of the features to scale, rotation, and translation. This invariance allows for consistent recognition of hand movements across various instances.

Luo et al. (2021)devised a genetic diagnosis model utilizing ML techniques to identify variations associated with hereditary hearing loss (HHL) in three specific genes, namely GJB2, SLC26A4, and MT-RNR1. The researchers conducted a case–control study involving 1898 subjects, including HHL patients and carriers. Risk assessment models were constructed by employing variations at 144 locations within the three genes and constructing six ML models afterward. The SVM demonstrated superior performance compared to the other five models. The SVM model demonstrated a performance of 0.803 regarding the area under the receiver operating characteristic (AUC) during the 10-fold stratified cross-validation process for predicting gene sites associated with HHL and variations. The model attained an AUC of 0.751 during external validation. The SVM model's predicted accuracy outperformed expert interpretation and genetic risk score (GRS) techniques. They have successfully discovered a total of 11 sites that constitute the minimum feature set capable of being reliably predicted.

A full-duplex communication system for individuals who are D–M utilizes ML techniques was proposed in Saleem et al. (2023). The proposed system enables non-deaf and mute (ND-M) people to interact with D-M people without learning sign language. The system is inexpensive, dependable, simple to use, and relies on a Leap Motion Device (LMD) that is available commercially off-the-shelf (COTS). This publication creates and presents a new data set for the ML-based method. The dataset consists of three signed language datasets: American Sign Language (ASL), Pakistani Sign Language (PSL), and Spanish Sign Language (SSL). A static hand gesture detection system for recognizing signs was created using real-time vision, as described in Jiang and Ahmad (2019). The data collection

process involved a USB camera connected to a computer, without additional equipment like gloves. The proposed methodology employs a skin color technique to identify the hand gesture's Region of Interest (ROI) in the HSV color space. After completing all preprocessing tasks, eight characteristics were retrieved from each sample using Principal Component Analysis (PCA). The SVM formed the foundation for the machine learning method utilized for recognition. According to the research results, the method successfully distinguishes B, D, F, L, and U, the five hand movements used in ASL, with a success rate of roughly 99.4%.

Authors in Sahoo (2001) described a method for automated identification of Indian Sign Language (ISL) immobile numeric signs, where the signs were solely acquired using a regular digital camera, and electrical signals were captured without wearable equipment. Since the technology is designed to translate isolated digit signs into text, each sign picture supplied must include exactly one numerical sign. A sign database including ISL digits is developed and comprises 5000 photos and 500 images for each numerical sign (0–9) to detect ISL sign images in real-time. Regarding classification accuracy, the k-Nearest Neighbor classifier performs better than the Naive Bayes classifier.

The study (Lee and Lee 2017) recommended a smart sign language interpreting system based on a wearable hand device to help the deaf and mute incorporate into society without hindrance. The ASL alphabet is characterized by wearable technology comprising five flex sensors, two pressure sensors, and a three-axis inertial motion sensor. The system has three distinct modules: a wearable device equipped with a sensor module, a processor module, and a mobile application module designed for a display unit. The system incorporates an integrated SVM classifier to collect and assess sensor data. Table 3 summarizes different machine learning-based techniques for the deaf and mute, describing the advantages and disadvantages of each technique. An interactive sign language avatar is a computer-generated model miming human movements and expressions. These models, powered by machine learning algorithms, provide deaf and mute people with an engaging and aesthetically pleasing way to communicate, particularly in virtual or digital contexts. From Table 3, a static hand gesture detection system for recognizing signs proposed in Jiang and Ahmad (2019) and automated identification of ISL immobile numeric signs proposed in Lum et al. (2020) achieved high accuracy compared with other state-of-the-art techniques.

## 4.4 Hybrid techniques

Motivated by understanding the sign language used by deaf farmers for agricultural workers' productivity and efficiency, hand movements for words in the common Indian sign language (ISL) used by deaf farmers were identified (Venugopalan and Reghunadhan 2021). The utilization of a convolutional long short-term memory (LSTM) network has been implemented within a hybrid DL framework for gesture categorization. The model demonstrated an average classification accuracy of 76.21% on the provided dataset of ISL words related to the agriculture domain.

In the study by Mete (Yağanoğlu 2021), a proposition is made for creating a real-time wearable speech recognition system designed specifically for hearing impairments. The proposed system utilizes signal processing techniques and ML algorithms for real-time detection and recognition of speech. Then, it displays the recognized text on a small screen attached to the device. The system was tested on a group of deaf individuals and showed promising results regarding accuracy and usability.

**Table 3** Machine learning-based techniques for the deaf and mute

| Study | Technique | Dataset | Accuracy | Pros | Cons |
|---|---|---|---|---|---|
| Imran et al. (2021) | An interactive Android mobile application based on ML that reduced the communication gap between the deaf and non-deaf communities | A dataset for recognizing each Urdu alphabet in PSL. Urdu has 37 alphabets, with 40 photos for each alphabet | 90% | Developing accurate recognition models addresses the unique needs and challenges of the Pakistani deaf community | The generalizability and performance of the model in real-world situations. Accuracy needs to be improved |
| Fang et al. (2023) | A throat-language decoding system (TLDS) uses adaptable, inexpensive sensors to capture throat vibration signals and ML classifiers for semantic interpretation | A flexible polyvinylidene fluoride (PVDF) piezoelectric sensor was made to gather the data from the neck vibrations | 95.97% (speaker recognition) and 97.50% (straightforward semantic recognition). | The suggested method enhances throat-language decoding's precision and effectiveness by employing ML | Elements like comfort, usability, and user experience may influence the adoption and applicability of the technology in real-world circumstances |
| Sahana et al. (2020) | A hand gesture recognition method for sign language communication is proposed | 1000 distinctive signs representing ten distinct subjects are included in the 10-gesture dataset | 98.20% | The suggested method could capture and examine hand form and motion features more successfully, achieving high accuracy | The system must be modified to identify and understand hand signs in different sign languages |
| Luo et al. (2021) | Develop an ML-based genetic diagnosis model for hereditary hearing loss (HHL) related variants in GJB2, SLC26A4, and MT-RNR1 genes | 1898 samples were divided into two groups of information | AUC (s 0.803) | The analysis and interpretation of genetic data can focus more precisely on certain gene variations | Expanding models to include more genetic variations. More data, feature engineering, and model modifications could be necessary |
| Saleem et al. (2023) | An ML-based full-duplex communication system for the D–M is proposed | ASL, PSL, and SSL | 95% | The proposed approach improves the accuracy and reliability of SLR and translation | The dataset size needs to be increased. Videos and other data can be added to datasets to enhance them further |

**Table 3** (continued)

| Study | Technique | Dataset | Accuracy | Pros | Cons |
|---|---|---|---|---|---|
| Jiang and Ahmad (2019) | A static hand gesture detection system for recognizing signs was created using real-time vision | A database with five movements was created, and for each gesture, 1,400 images were presented by four individuals | 99.40% | Real-time detection enables rapid gesture interpretation and identification with high recognition accuracy | A small dataset or a constrained representation restricts the generalizability and performance of the system |
| Lum et al. (2020) | Automated identification of ISL immobile numeric signs is proposed, where the signs were solely acquired using a regular digital camera and electrical signals | A sign database including ISL digits is developed and comprises 5000 photos and 500 images for each numerical sign (0–9) | KNN is 98.36%, whereas Naive Bayes is 97.79% | The proposed system provides more relevant and accurate recognition results for the Indian deaf community | Limited generalization ability due to the use of a specific Indian dataset |
| Lee and Lee (2017) | The ASL alphabet is distinguished by this wearable technology using five flex sensors, two pressure sensors, and a three-axis inertial motion sensor | American Sign Language alphabet dataset | 65.70% | Wearable technology has the benefit of portability and convenience. For those who are deaf, it can improve accessibility and communication | accuracy needs to be improved. Strategies to deal with recognition faults or confusing gestures must be addressed |

In Venugopalan and Reghunadhan (2023), the authors proposed a novel dataset of dynamic hand gestures representing commonly used emergency communication words in the ISL. The proposed dataset is specifically designed for deaf individuals who have tested positive for COVID-19. They also suggested applying a hybrid model, combining deep CNN and long short-term memory networks, to recognize the hand gestures in the proposed dataset accurately. The model demonstrated an average accuracy of 83.36% on the designated dataset. Additionally, it attained accuracy rates of 97% and 99.34 + 0.66% on an alternative ISL word dataset and a benchmarking dataset, respectively. The experimental study serves as a reference point for future advancements in SLR, intending to overcome the current obstacles in communication for the deaf community.

In their work, Wadhawan and Kumar (2021) presented the initial literature review for SLR systems. Their study encompasses a thorough review of academic literature on SLR systems spanning 2007 to 2017. Additionally, the authors proposed a systematic classification framework to categorize the research articles in this field. Furthermore, they compared the selected research articles, considering twenty-five distinct sign languages and six dimensions. The comprehensive investigation sheds light on the prevailing research trends and key findings in the domain of SLR. The authors' ultimate objective is to provide guidance for future research endeavors and to facilitate the accumulation and generation of knowledge in the field of SLR.

The authors in Mandal et al. (2023) focused on developing an application-based system to facilitate communication between hearing-impaired individuals and those who are not. The proposed system aims to translate sign language, incorporating dynamic gestures and a centralized system. To do this, the authors proposed utilizing two modules: the first is for speech-to-3-D animated sign motions, and the second is for speech-to-sign gestures. The dataset used in this study consisted of around 1800 words and was obtained from the D-M School located in Dhule, Maharashtra, as well as the ISL's official website. The CNN+LSTM DL approach converts sign motions to speech, while the voice-to-3D sign gestures technique uses live speech to extract keywords, which are then used to create 3D animated gestures. The proposed method will help the deaf, the mute, and the normal communication. In their study, Zhang et al. (2022) proposed an innovative real-time translation system that integrates optical data, data from a flexible strain sensor, camera sensory data, and DL techniques. The training structure involves the utilization of a CNN to handle the visual input. Then, the feature layer implements a sparse neural network to fuse and recognize sensor data. They proposed implementing a module that facilitates knowledge acquisition regarding the correlation between fundamental manual motions and the alphabet, encompassing the letters A through Z. The proposed approach is evaluated using data derived from signs and descriptive text. The suggested system demonstrates a high identification rate of up to 99.50% and a detection time of less than one second. The proposed system can accurately identify the 26 letters, facilitating the comprehension of sign observations as words.

A proposed approach for vision-based hand gesture identification involves utilizing a hybrid network architecture called Lightweight VGG16, and Random Forest (Lightweight VGG16-RF) was proposed in Ewe et al. (2022). The CNNs are used in the proposed model to implement feature extraction techniques, and machine learning is used to accomplish categorization. They used publicly accessible datasets to conduct tests, including the NUS Hand Posture dataset, ASL Digits, and ASL. The experimental findings show that the suggested model, the lightweight VGG16 and random forest combination, surpasses alternative approaches. Table 4 summarizes different hybrid techniques for the deaf and mute, describing the advantages and disadvantages of each technique.

**Table 4** Deep learning-based techniques for hybrid techniques for the deaf and mute

| Study | Technique | Dataset | Accuracy | Pros | Cons |
|---|---|---|---|---|---|
| Venugopalan and Reghunadhan (2021) | An LSTM network has been employed in a hybrid DL model for gesture categorization | ISL used by deaf farmers includes 260 original videos with 20 examples for each type of gesture | 76.21% | The proposed technique significantly advances ISL identification in agriculture | High computational requirements and processing time. Accuracy needs to be improved |
| Yağanoğlu (2021) | Development of a real-time wearable speech recognition system for deaf individuals is proposed | A dataset of 4,000 audio blocks is set aside for training | 95% | The mobility and simplicity of wearable devices provide deaf people with a small, hands-free voice recognition tool | The system is under various acoustic situations, including variable background noise, accents, and speech patterns |
| Venugopalan and Reghunadhan (2023) | Applying a hybrid deep convolutional LSTM network model for ISL recognition | The proposed dataset consists of 900 video clips of nine hand gesture classes (100 samples for each | 97 % | Proposing a new dataset for deaf COVID-19-positive patients | Complexity and Computational Cost |
| Mandal et al. (2023) | An application-based platform that translates sign language, allowing hearing-impaired and hearing persons to converse with each other while developing adaptive gestures | The approximately 1800-word dataset was gathered from the deaf and mute institution in Dhule, Maharashtra, and the official ISL website | N/A | The dual-mode method enables reliable and precise detection of sign language motions. Real-time prediction for enabling immediate feedback and interaction | A large and diverse dataset is essential for training a robust and generalizable model |
| Zhang et al. (2022) | A real-time translation system that combines optical data with data from a flexible strain sensor and the camera, plus DL, is proposed | A collection of data from signs and descriptive letters is used | 99.50% | Enhanced Accuracy and Reliability. Integrating visual and sensorimotor data from a flexible strain sensor-based data glove is suggested | A large and diverse dataset is essential |
| Ewe et al. (2022) | A hybrid network architecture known as Lightweight VGG16 and Random Forest (Lightweight VGG16-RF) is suggested | American Sign Language Alphabet, ASL Digits Dataset, NUS Hand Posture Dataset | 99.98% | The suggested model achieves a reasonable compromise between accuracy and computational cost | Strategies to deal with recognition faults must be addressed |

To conclude, the comparative analysis of the selected studies reveals the transformative potential of AI, deep learning, and machine learning in facilitating communication for D–M individuals. These technologies offer innovative approaches to recognizing, translating, and enhancing sign language and non-verbal communication. Through user-centered design, specialized datasets, real-time interaction, and ongoing research, AI-based solutions can create inclusive and accessible communication environments that empower individuals with different abilities to connect and communicate effectively with the world around them. Tables 2,3, and 4 provide a comprehensive comparative study of the scholarly articles on the domains of AI, DL, and ML, specifically in facilitating communication for DM Individuals. Each study contributes to the overarching goal of enhancing accessibility and inclusivity through innovative approaches. The papers cover a range of methodologies, including SLR, optimized fuzzy deep learning, convolutional LSTM networks, multi-modal deep learning, machine learning for SLR, throat vibration signals, deep learning with edge devices, and depression recognition through audiovisual cues. These diverse approaches highlight the versatility of AI-based solutions in addressing various communication challenges the D–M community faces. The table effectively summarizes the key contributions of each paper, showcasing their unique focus, technology, and application, and demonstrates the collective effort to create inclusive communication environments for individuals with different abilities. With the previously discussed state-of-the-art techniques and an understanding of the strengths, weaknesses, challenges, and opportunities highlighted in this analysis, researchers and practitioners in D–M communication can make informed decisions when designing and implementing AI, DL, and ML-based solutions. Collaborative efforts between the technical community and end-users are crucial for advancing the state-of-the-art and ensuring the practicality and inclusivity of these technologies.

Current research Challenges and gaps are: (1) restricted access to vast and varied datasets for D–M communication. The model generalization may be limited by the inability of existing datasets to sufficiently capture the range of verbal expressions, cultural subtleties, and signing techniques. (2) The achievement of real-time processing, a crucial need for perfect communication, presents difficulties for several current methods. The practical use of these methods may be limited by the delay caused by processing, which might affect the normal progression of discussions. (3) Recognizing gestures is one aspect of effective D–M communication; other skills include reading body language, facial emotions, and context. There is still a research gap in addressing these complexities because gesture detection is frequently the primary emphasis of models.

# 5 E-learning and sign-language recognition solutions

This section will examine E-learning and sign-language recognition solutions for assisting individuals with disabilities. Both technologies can potentially improve education and communication for individuals with different abilities. E-learning is a method of education that utilizes the Internet as a distribution medium, making it accessible to individuals regardless of their geographical location and time constraints. It places a larger emphasis on written materials than spoken ones. While E-learning removes many access barriers, it fails to adequately consider enhancing students' inclusivity irrespective of their sensory, cognitive, or functional abilities (Pivetta et al. 2014).

On the other hand, SLR solutions focus on enabling communication for individuals with disabilities through gestures and facial expressions. There are two main approaches: sensory gloves and vision-based methods. Sensory gloves use sensors and electromechanical integration to track hand movements, while vision-based methods use cameras to capture gestures. Both approaches aim to accurately interpret sign language gestures, with the vision-based methods having an advantage in recognizing non-manual variations such as facial expressions (Wang et al. 2006).

Regarding recognition accuracy, vision-based methods have an edge over sensory gloves in capturing the complexity of sign language. Microsoft Kinect and Leap Motion offer 3D motion capture systems that accurately track hand and arm movements, enhancing human-computer interaction (Microsoft Kinect 2023; Leap Motion 2023). Additionally, artificial intelligence and digital image processing methods are employed to analyze captured data, improving recognition accuracy. Furthermore, DL models have been applied to both SLR and augmented reality applications. These models have shown potential in extracting task-driven features from data, improving classification accuracy (Yazdinejad et al. 2023). In augmented reality, DL integrates virtual content with the real environment, enhancing user interaction and engagement (Burns 2023).

Regarding applications, E-learning and SLR solutions have different focuses. E-learning aims to provide education through online platforms, making it accessible to many individuals. SLR solutions target individuals with disabilities, specifically those who use sign language as their primary mode of communication. These technologies aim to bridge the communication gap and enhance inclusion for this group. Table 5 comprehensively comparatively examines E-learning and SLR solutions. The next paragraphs introduce e-learning techniques used to facilitate communication for D–M individuals.

In conclusion, both E-learning and SLR solutions can potentially improve education and communication for individuals with disabilities. While E-learning focuses on providing education through online platforms, SLR solutions aim to enable communication through gestures and facial expressions. Both technologies utilize advanced methods, such as deep learning, to enhance recognition accuracy and user engagement. However, their applications and target audiences differ, with E-learning targeting a broader audience and SLR solutions focusing on individuals with specific communication needs. The next paragraphs introduce a comparative analysis of the papers on AI, DL, and ML in facilitating communication for D–M Individuals.

Recently, there has been extensive research on using AI, DL, and ML techniques to facilitate communication for D–M individuals. The overarching goal is to bridge the communication gap between these individuals and society by recognizing and translating sign language, enhancing their learning experiences, and aiding their social interactions. The research community utilized various technological methods and datasets. Authors in Martins et al. (2015) discussed the necessity of developing E-learning platforms that are accessible to individuals who are deaf using Sign Language (SL) translation. This study investigated various assistive technologies, methodologies, and approaches to examine prospective choices for SLR, translation, and presentation. The analysis highlights promising technological solutions for recognizing and translating SL but acknowledges existing challenges in fully integrating these technologies into e-learning platforms.

Authors in Alshawabkeh et al. (2021)examine how deaf students were seen in online distance learning during the COVID-19 epidemic regarding technology training and adaptations. Semi-structured, one-on-one interviews that were conducted in confidence were employed in this qualitative study. They conducted a convenience sample interview with 15 deaf kids and three of their teachers in June 2020 and systematically evaluated the results.

**Table 5** A comparative analysis of E-learning and SLR solutions

| Aspect | E-Learning | SLR Solutions |
| --- | --- | --- |
| Technology | Utilizes the Internet for online education | Focuses on recognizing and translating sign language gestures |
| Accessibility | Accessible regardless of location and time constraints | Enhances communication for individuals with disabilities |
| Emphasis | Written materials | Gestures and facial expressions |
| Inclusion | It does not address inclusion for individuals with sensory, cognitive, or functional capacities | It aims to bridge the communication gap and enhance inclusion |
| Approaches | Various methodologies for online education | Sensory gloves and vision-based methods |
| Recognition Accuracy | Not applicable | Vision-based methods have an edge in recognizing complex gestures |
| Technologies | Not applicable | Microsoft Kinect, Leap Motion, AI, digital image processing |
| Deep Learning Applications | Not applicable | Applied to SLR and augmented reality |
| Applications | education through online platforms | Targeted communication for individuals with disabilities |
| Target Audience | Broad audience | Individuals with specific communication needs |
| User-Centered Approach | Not applicable | Tailored solutions for specific user groups |
| Dataset Creation | Not applicable | Creation of specialized sign language datasets |
| Real-Time Interaction | Not applicable | Utilization of edge devices for real-time communication |
| Multi-modal Approach | Not applicable | Utilization of audiovisual cues for enhanced communication |
| Challenges and Future Research | Not applicable | Addressing noise reduction, dataset limitations, model accuracy |
| Diversity of Applications | Not applicable | Diverse applications in education, mental health, and more |

After obtaining theme saturation, the study of the thematic structure was completed. Five key themes emerged from the findings about deaf students' online distance learning experiences during COVID-19. The themes included social interactions, technological usage, delivery strategy, and delivery of the course material. It is examined and contrasted with the pertinent literature to scientifically encompass each theme's proposed aspects.

Because of a shortage of resources, a lack of educational facilities, and certain social problems, the deaf and mute society no longer tries to enroll in traditional organizations without many facilities. It is the main issue the community faces in the e-learning environment. The system proposed in Ranasinghe et al. (2022) will solve that issue by transcribing the lecturer's voice into text, mapping words to pre-made sign language movies, creating subtitles for lecture videos, clearly identifying the lecturer's face position, detecting difficult words, tracking hand gestures, and practicing sign language. It will also increase educational assets, facilities, and usability while assisting teachers in teaching methods through this platform. Thus, regular institutions may use this system as an educational management system.

The learning strategy (Haron et al. 2019) produces a mobile application (apps) to aid the learning of dumb and deaf individuals. This study explored the development and use of mobile apps targeted at dumb and deaf individuals. The software, commonly known as Malaysian Sign Language (MSL), was created to help dumb and deaf people learn sign language whenever and wherever it is most convenient. In addition, the application may teach fresh information to anybody who wants to learn sign language, especially in the outside world.

To improve the learning experience for deaf and mute students, Mubin et al. (2022) provides a framework that makes use of Extended reality (XR) technologies, such as augmented reality (AR), virtual reality (VR), and mixed reality (MR). A working model is constructed based on a study of the user requirements, and an assessment phase will follow. Most of the framework concentrates on a particular technological method. Therefore, their focus is on examining how XR might be able to help and aid deaf and mute learners during their educational process. Creating an efficient real-time vision-based American sign language identification system for Deaf and Dumb people utilizing ASL letters is described in Saleem et al. (2023). They ultimately reached a final accuracy of 92.0% on the dataset used. They improve prediction accuracy by employing two layers of techniques that check and foresee increasingly similar symbols.

Arab nations use Arabic Sign Language (ArSL) to communicate with the deaf and mute. Mubin et al. (2022) emphasizes developing a real-time voice-to-ArSL translation service using 3D animated films to aid learning in online learning environments. The research aided the COVID-19 pandemic-affected deaf and mute students in distance learning. This study was implemented on the TEAMS platform. Using the Al-tarjuman application, a database comprising more than 550 ArSL movies was built and linked to the TEAMS platform.

In Mubin et al. (2022), researchers focused on the benefits of online learning for local students who are deaf and mute. The obstacles, technical solutions, and comparison of a novel teaching strategy for deaf and mute pupils against typical students are all discussed in the paper. We adopted a distributed system to lessen the workload of performing sign interpreting and recognition. The accuracy and effectiveness of the e-learning system for the deaf and mute are increased because of division. Deaf and mute students in remote areas may use the method for online education.

Focusing on augmented reality (AR), the study (Deb and Bhattacharya 2018) introduced an ASLM application for translating Hindi Varnamala into sign gestures using 3D

animated hand movements. The objective of this study is to provide a self-directed educational and communicative resource for individuals who are deaf and unable to speak. The study demonstrates the potential of AR to bridge communication gaps and enhance accessibility in inclusive classrooms.

These research papers collectively contribute to E-learning advancing communication and accessibility for D–M individuals through AI, deep learning, and machine learning. While each study approaches the topic from a unique perspective, they share common themes:

– Technological Innovation: The recent research explores the innovative use of deep learning, augmented reality, sensors, and edge devices to enhance communication, recognition, and translation of sign language and non-verbal cues.
– User-Centered Approach: Many research papers focus on specific user groups, such as deaf agriculturists or individuals with depression, to tailor their solutions to the needs and challenges of those groups. The user-centered approach highlights the importance of addressing specific communication barriers.
– Dataset Creation: Several studies emphasize the creation of datasets specific to the target sign language, such as Pakistan Sign Language or Bangla Sign Language. These datasets serve as valuable resources for training and evaluating machine learning models.
– Real-time Interaction: Several studies emphasize the importance of real-time communication and interaction, particularly relevant in scenarios where timely communication is essential.
– Multi-modal Approach: Some research papers explore the combination of multiple modalities, such as audio and visual cues, to enhance communication and recognition accuracy.
– Application Diversity: The studies introduced diverse applications, ranging from education and learning aids to mental health diagnosis, highlighting the versatility of AI-based solutions in addressing various communication challenges.
– Challenges and Future Directions: While the recent studies demonstrate significant advancements, they also acknowledge challenges such as dataset limitations, noise reduction, and model accuracy, potential areas for future research and improvement.

## 6 Challenges and applications of SLR in deaf–mute communication

This section examines the complexities of sign language recognition (SLR) in deaf-mute communication, addressing the challenges, applications, and ethical considerations integral to its advancement. Despite SLR's progress, issues such as language variability, dataset scarcity, the need for real-time processing, and environmental robustness persist, necessitating ongoing innovation to enhance its inclusivity and effectiveness.

Furthermore, the section showcases the innovative applications leveraging SLR and automated speech recognition (ASR) to aid individuals with hearing and speech impairments. These technologies, ranging from real-time captioning to digital sign language platforms, demonstrate SLR's potential to bridge communication divides. Additionally, a focused discussion on ethical considerations underscores the imperative of developing SLR technologies responsibly. This encompasses ensuring inclusivity, protecting privacy,

maintaining cultural sensitivity, and promoting equitable access, reinforcing ethical foresight's significance in expanding communication access for the deaf and mute community.

## 6.1 Addressing key challenges in sign language recognition

Despite considerable advancements, sign language recognition (SLR) encounters significant obstacles. This narrative examines these challenges and proposes potential avenues for addressing them. Figure 9 illustrates various issues and challenges inherent in SLR, detailed as follows:

– Variability in Sign Language: Sign languages are characterized by significant variability. Future systems must manage this diversity by applying domain adaptation or transfer learning techniques across multiple sign language datasets.
– Data Availability: The scarcity of comprehensive, diversified sign language datasets with precise annotations presents a substantial barrier to developing effective models. Future endeavors should broaden data collection to cover many sign languages, signing styles, and contexts. Exploring data augmentation and active learning methods may also enhance data utility.
– Enhancing Real-Time Performance: The goal of achieving high accuracy alongside rapid processing speeds is vital for facilitating seamless communication. Prospective research should investigate the development of compact deep-learning models and efficient inference strategies tailored for real-time execution on devices with limited resources.
– Building Robustness to Noise and Occlusions: In real-world settings, occlusions and background noise can significantly impair data quality. Future systems must employ strategies such as pose estimation, background subtraction, and resilient deep learning frameworks to maintain performance in adverse conditions.
– Incorporating Contextual Understanding: Sign language interpretation can be highly context-dependent. Integrating contextual data from adjacent signs, facial expressions, and body language could significantly enhance recognition accuracy, possibly by blending NLP with computer vision techniques.
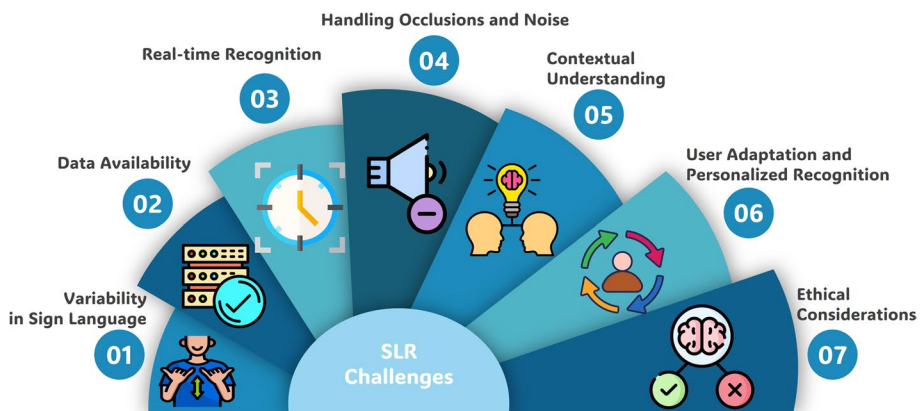


**Fig. 9** The sign language recognition challenges

– Personalization and User Adaptation: Given the individualized nature of sign language use, future systems must be designed to adapt and personalize recognition algorithms to the unique signing styles of individual users.
– Ethical Considerations in SLR Development: Addressing ethical concerns, including fairness, privacy, and eliminating biases in training datasets, is essential. Upcoming research should aim to develop frameworks that ensure user privacy and foster inclusivity within SLR applications.

These challenges span various applications, from traditional sign-language recognition to innovative throat-language interpretation systems. Herein, we detail the primary challenges identified in different studies as shown in Table 6. Addressing these challenges is paramount for the evolution and widespread adoption of sign language and gesture recognition systems.

## 6.2 Ethical considerations

Integrating AI into communication technologies for the deaf and mute communities heralds a potentially transformative era. However, this innovation brings to the forefront several ethical challenges that necessitate careful consideration. Addressing these challenges is crucial for ensuring that AI aids in creating a more inclusive and equitable society without inadvertently introducing new forms of disparity or dependency. Key ethical considerations include:

– Inclusivity and Accessibility: The goal is to ensure that AI communication technologies serve the deaf and mute community, avoiding exclusionary practices that could exacerbate existing inequalities. This requires designing solutions that cater to a broad spectrum of needs within these communities, including varying levels of hearing loss and communication preferences.
– Privacy and Data Security: Given the personal nature of the data processed by AI systems, including speech and sign language patterns, robust measures must be adopted to protect user privacy. This entails implementing stringent data protection protocols and ensuring users are informed and retain control over their data.
– Balancing Empowerment and Dependency: While AI technologies offer new avenues for empowerment through enhanced communication tools, there is also the potential for increased dependency on these technologies. It is important to strike a balance where AI supplements rather than replaces human interaction, preserving the value of direct, personal communication.
– Cultural Sensitivity: Recognizing and respecting the cultural and linguistic diversity within the deaf and mute communities is paramount. AI technologies should be developed with an awareness of cultural nuances, ensuring they support and enhance these unique aspects rather than overlooking or diluting them.
– Equitable Access to Technology: Bridging the digital divide is essential in deploying AI communication technologies. Efforts must be made to ensure these technologies are affordable, accessible, and adaptable across diverse regions and communities, promoting equal opportunities for all to benefit from these advancements.

**Table 6** Challenges in Sign Language and Gesture Recognition Systems

| Application | Challenges |
| --- | --- |
| Sign Language Recognition Systems (Alsulaiman et al. 2023; Sundar and Bagyammal 2022) | The precision of recognition algorithms remains suboptimal, necessitating further refinement |
| | Ambient lighting, camera resolution, occlusions, or external visual interferences can constrain the system's efficacy |
| | variability in gestures, differing among signers and contexts, poses recognition challenges |
| | Relying solely on still images for recognition may not capture the full motion spectrum of sign language gestures |
| | There's a pressing need for extensive and varied datasets for training machine learning models, given the limited availability of sign language data |
| Throat-Language Interpretation Systems (Fang et al. 2023) | Procuring superior-quality throat vibration signals is challenging due to the intricate characteristics of vocal cords |
| | The precision of machine learning classifiers heavily depends on the quality and volume of training data |
| Gesture Recognition Systems (Sundar and Bagyammal 2022; Fang et al. 2023) | Systems relying on still images might compromise accuracy compared to motion-based tracking |
| | The scarcity of sign language data necessitates the development of extensive and varied datasets for training |
| General Recognition Systems (Prasath 2022; Li et al. 2023) | Limited dataset size is a recurring constraint in various methods |
| | Reliance on a single camera might hinder performance in noisy settings or if the user is outside the camera's view |
| | Disease prediction from medical images faces challenges like class imbalance, limited data, and label inaccuracies |
| | Training multimodal deep learning models can be complicated due to conflicts between sub-models and fusion modules |
| Leap Motion Controller (LMC) for Sign Language Recognition (Katılmış and Karakuzu 2023) | sensitivity and noise in sensor signals pose challenges |
| | Elevated costs, unnatural usage, restricted movement flexibility, and calibration adjustments hinder the effective application of this technology |
| Hand Gesture Recognition (HGR) Systems (Bhiri et al. 2023) | Precise detection and tracking of hand movements in complex environments is intricate due to hand variations, occlusions, and visual interferences |
| | Devices like the Leap Motion Controller may have accuracy, resolution, or scope constraints |
| | Functionality under diverse lighting, backgrounds, and camera perspectives can compromise hand-tracking data quality |
| | Amassing and annotating extensive training data for HGR systems is resource-intensive and costly |

**Table 6** (continued)

| Application | Challenges |
|---|---|
| Automatic Sign Language Recognition (Hameed et al. 2022) | The nascent stage of development compared to speech recognition poses challenges |
| | Wearable and vision-based systems face issues like bulkiness, restricted range, and light sensitivity |
| ASL Recognition through RF Sensing (Gurbuz et al. 2020) | Extensive training data is required for machine learning, making data acquisition from native signers challenging |
| | Variations in signing styles and individual differences complicate creating universally applicable systems |
| | RF sensing might be unsuitable in situations with thick attire or multiple nearby signers |

## 6.3 Transformative applications: bridging gaps with SLR technology

Individuals with hearing loss or deafness utilize a diverse range of communication techniques. Recent years have witnessed the emergence of innovative alternatives thanks to advancements in automated speech recognition (ASR). This subsection highlights several key applications developed to facilitate communication for the deaf and mute community:

– Ava (Captions for All 2024): Ava offers real-time captioning for individuals who are hard of hearing or deaf, transcribing conversations with voice recognition technology. This freemium software provides basic features for free, with the option to access more advanced features through a subscription.
– Google Live Transcribe (Live Transcribe & Notification 2024): Designed for the deaf and hard of hearing, Google Live Transcribe delivers real-time voice-to-text transcription, supporting over 80 languages and dialects. It is an open-source application, allowing for widespread use and adaptation.
– Microsoft Translator (Microsoft Translator 2024): A complimentary tool enabling real-time translation of text to speech and vice versa, Microsoft Translator serves as a bridge for communication between hearing and deaf individuals.
– Signily Keyboard (SIGNILY 2024): Signily offers a digital platform for typing in American Sign Language (ASL) symbols, addressing the communication needs of deaf individuals. It underscores the importance of written communication, acknowledging the challenges in learning to read for those born deaf.
– Hand Talk (Handtalk 2024): An open-source app, Hand Talk employs an avatar to convert text and audio into Brazilian Sign Language (LIBRAS), simplifying conversations for the deaf and hard of hearing.
– Ameslan Pro (The ASL App 2024): This paid iOS application provides an extensive library of ASL signs, accompanied by videos demonstrating each sign, catering to the learning and communication needs of ASL users.

– JusTalk (Justalk 2024): JusTalk enhances video calls for the deaf and hard of hearing with real-time text subtitles. This freemium application offers basic services for free, with premium features available for a fee.
– SOS Tap (Tapsos 2024): SOS Tap enables communication with emergency services for those with hearing loss without needing speech or auditory cues. Users select their needs through a visual interface, and the app shares their medical history and location with emergency responders.

# 7 Future directions in sign language and gesture recognition

Sign language and gesture recognition offers exciting opportunities for future exploration. Researchers have proposed several promising directions across various applications and methodologies as follows:

**Sign language recognition systems (Alsulaiman et al. 2023)**

- Data Augmentation and Deep Learning: Expanding datasets with a wider variety of signers signs and addressing regional variations are crucial. Deep learning models hold significant promise for improving recognition accuracy.
- Personalized Recognition: Future systems should explore techniques to adapt to individual signing styles.
- Applications Beyond Research: Sign language recognition can transform education, entertainment, and accessibility in various sectors.

**Throat-language interpretation systems (Fang et al. 2023)**

- Data Collection and Algorithm Development: Building larger databases of throat vibration signals and researching more precise recognition algorithms are essential.
- Integrated Systems: Developing a cohesive throat-language interpretation system requires integrating data collection, signal processing, and recognition components.
- Machine Learning Optimization: Determining optimal training and test data sizes for machine learning in this domain is important.

**Gesture recognition systems (Fang et al. 2023)**

- Real-time Recognition: Real-time gesture recognition is vital for practical applications. Future research should focus on achieving high accuracy with low latency.

**General recognition systems (Prasath 2022; Li et al. 2023)**

- Multi-sensor Fusion: Integrating multiple cameras or sensors can improve system robustness and accuracy (Prasath 2022).
- Expanding DeaF Applications: Exploring applications of DeaF (Deep Ensemble Architecture with Feedback) beyond disease prediction is recommended.
- Transfer Learning for DeaF: Investigating transfer learning techniques to enhance DeaF's performance is promising.

## 8 Leap motion controller for sign language recognition (Katılmış and Karakuzu 2023)

- Alternative Sensors and Meta-learning: Exploring alternative sensors and meta-learning techniques can improve recognition capabilities beyond the Leap Motion Controller.
- Expanded Sign Language Datasets: Training on a broader range of sign language expressions is necessary.

**Hand gesture recognition (HGR) systems (Bhiri et al. 2023)**

- Generalizability and Multi-modality: Future research should improve HGR (Hand Gesture Recognition) generalizability and explore multi-modal approaches that integrate audio or video data.
- Real-time HGR Systems: Developing real-time HGR systems for instantaneous response is crucial for practical applications.
- Deep Learning for Feature Extraction: Deep learning methods hold promise for improved feature discernment in HGR.

**Automatic sign language recognition (Hameed et al. 2022)**

- Sensor Fusion and Deep Learning: Integrating alternative sensing methods like radar with deep learning can enhance recognition accuracy.
- Large and Diverse Datasets: Creating extensive datasets encompassing diverse subjects and signs is essential.

**ASL recognition through RF sensing (Gurbuz et al. 2020)**

- Multi-sensor RF Systems and Deep Learning: Exploring multi-sensor RF systems and integrating deep learning can improve ASL recognition robustness.
- Broader Applications for the Deaf Community: Investigating RF sensing for applications like non-invasive ASL-responsive personal assistants can benefit the Deaf community.

**Assistive technologies for visually impaired users (Talaat et al. 2024b)**

- The field of sign language recognition can be leveraged to develop assistive technologies for visually impaired users. Inspired by works like "SightAid: Empowering the visually impaired in the Kingdom of Saudi Arabia (KSA) with deep learning-based intelligent wearable vision system" (Talaat et al. 2024b), future research can explore sign language recognition systems integrated into wearable devices or smart environments. This could enable real-time translation of sign language gestures into audio or other accessible formats, fostering greater communication and social interaction for visually impaired individuals.

**Facial expressions for sign language disambiguation (Alhussan et al. 2023)**

- While sign language recognition focuses on hand gestures and postures, facial expressions can provide complementary information, aiding in the disambiguation of signs with similar handshapes. Future research, as explored in Alhussan et al. (2023), can investigate incorporating facial expression recognition using optimized Support Vector Machines (SVM) or other machine learning techniques to improve the accuracy and robustness of sign language recognition systems.

   **Emotion recognition in sign language interactions (Talaat et al. 2024a)**

- While significant progress has been made in sign language recognition, incorporating emotion recognition adds another layer of complexity. Future research should explore techniques for recognizing emotions from facial expressions, body language, and signing style, drawing inspiration from recent advancements in facial emotion recognition using deep learning techniques like those presented by Talaat et al. (2024a). This would enable the development of more natural and nuanced communication systems for human-computer interaction.

## 9 Conclusions

The D–M community, while being an integral part of our society, often faces communication barriers due to the predominant lack of understanding and knowledge of sign language among the general populace. This challenge underscores the pressing need for innovative solutions to bridge this communication gap. Motivated by this imperative, the primary aim of this survey was to provide a comprehensive review of the applications of AI, DL, and ML in facilitating communication for the D–M community. An extensive bibliometric study was undertaken, encompassing various fields of study such as Pattern Recognition, Language Processing, Communication, Accessibility, and Inclusivity. This bibliometric analysis served as a foundation to discern the current state-of-the-art techniques and methodologies employed in AI, DL, and ML for the D–M community. Our exploration revealed a plethora of advancements, with deep learning-based techniques, machine learning-based techniques, and hybrid techniques emerging as the forefront methodologies. RNNs, CNNs, and other DL architectures have shown impressive results in sign language recognition, providing the foundation for improved and successful communication. When combined with pre-trained models, transfer learning approaches have made it possible to adapt knowledge across several sign languages, improving performance even in scenarios with a small amount of labeled data. Multimodal techniques that integrate visual and audio information have emerged to provide more inclusive communication systems. Edge computing makes it possible to read sign language in real-time, which helps to minimize processing delays and provide smooth, immediate communication. During our discussion of different machine learning approaches, it became clear that the sector is dedicated to user-centric design as much as it is to technical developments. Involving the deaf and mute population in the co-creation process guarantees that the solutions are user-friendly, socially sensitive, and technically effective. Besides, the integration of e-learning and sign-language recognition solutions showcased the potential of technology in bridging communication gaps. As with any evolving domain, challenges persist. The section on challenges and future work elucidates the existing hurdles and proposes potential avenues for further research and development. Through its meticulous examination and analysis, this survey underscores

the transformative potential of AI, DL, and ML in crafting more inclusive communication platforms for the D–M community, heralding a future where every individual, irrespective of their communication abilities, can effectively express and connect. Although this survey also revealed several problems and directions for further study, it is still essential to have representative and varied datasets, particularly when considering cultural variances and various sign languages. Furthermore, these advances in practical use and scalability call for additional research, necessitating multidisciplinary cooperation and ongoing methodological improvement.

## Declarations

## References

Aazam M, Zeadally S, Flushing EF (2021) Task offloading in edge computing for machine learning-based smart healthcare. Comput Netw 191:108019

Akbar MS, Hussain Z, Sheng QZ, Mukhopadhyay S (2022) 6G survey on challenges, requirements, applications, key enabling technologies, use cases, AI integration issues and security aspects. arXiv preprint. arXiv:2206.00868

Alhussan AA, Talaat FM, El-Kenawy E-SM, Abdelhamid AA, Ibrahim A, Khafaga DS, Alnaggar M (2023) Facial expression recognition model depending on optimized support vector machine. Comput Mater Contin 76(1):499–515

Alshawabkeh AA, Woolsey ML, Kharbat FF (2021) Using online information technology for deaf students during COVID-19: a closer look from experience. Heliyon 7(5):e06915

Alsulaiman M, Faisal M, Mekhtiche M, Bencherif M, Alrayes T, Muhammad G, Mathkour H, Abdul W, Alohali Y, Alqahtani M et al (2023) Facilitating the communication with deaf people: Building a largest Saudi sign language dataset. J King Saud Univ Comput Inf Sci 35(8):101642

Athitsos V, Neidle C, Sclaroff S, Nash J, Stefan A, Yuan Q, Thangali A (2008) The American sign language lexicon video dataset. In: 2008 IEEE Computer Society conference on computer vision and pattern recognition workshops. IEEE, pp 1–8

Barczak A, Reyes N, Abastillas M, Piccio A, Susnjak T (2011) A new 2D static hand gesture colour image dataset for ASL gestures. Res Lett Inf Math Sci 15:12–20

Bhiri NM, Ameur S, Alouani I, Mahjoub MA, Khalifa AB (2023) Hand gesture recognition with focus on leap motion: an overview, real world challenges and future directions. Expert Syst Appl 226:120125

Burns M (2023) Exciting AR apps for student learning (2016). https://www.edutopia.org/blog/ar-apps-for-student-learning-monica-burns. Accessed 26 June 2023

Captions for All (2024) https://www.ava.me/. Accessed 8 Apr 2024

Chen J, Li Q, Ling H, Ren D, Duan P (2021) Audiovisual saliency prediction via deep learning. Neurocomputing 428:248–258

Dargazany AR, Stegagno P, Mankodiya K (2018) WearableDL: wearable Internet-of-Things and deep learning for big data analytics-concept, literature, and future. Mob Inf Syst 2018:1–20

Das K, Behera RN (2017) A survey on machine learning: concept, algorithms and applications. Int J Innov Res Comput Commun Eng 5(2):1301–1309

David D, Alamoodi A, Albahri O, Zaidan B, Zaidan A, Garfan S, Ismail AR, Albahri A, Alsinglawi B, Malik R (2023) Landscape of sign language research based on smartphone apps: coherent literature analysis, motivations, open challenges, recommendations and future directions for app assessment. Univers Access Inf Soc 23(2):1–16

Deb S, Bhattacharya P et al (2018) Augmented sign language modeling (aslm) with interaction design on smartphone-an assistive learning and communication tool for inclusive classroom. Procedia Computer Science 125:492–500

Dreuw P, Neidle C, Athitsos V, Sclaroff S, Ney H (2008) Benchmark databases for video-based automatic sign language recognition. In: LREC

Dwivedi YK, Hughes L, Ismagilova E, Aarts G, Coombs C, Crick T, Duan Y, Dwivedi R, Edwards J, Eirug A et al (2021) Artificial intelligence (AI): multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. Int J Inf Manage 57:101994

Ewe ELR, Lee CP, Kwek LC, Lim KM (2022) Hand gesture recognition via lightweight VGG16 and ensemble classifier. Appl Sci 12(15):7643

Fang H, Li S, Wang D, Bao Z, Xu Y, Jiang W, Deng J, Lin K, Xiao Z, Li X et al (2023) Decoding throat-language using flexibility sensors with machine learning. Sens Actuators A 352:114192

Forster J, Schmidt C, Koller O, Bellgardt M, Ney H (2014) Extensions of the sign language recognition and translation corpus RWTH-PHOENIX-Weather. In: LREC, pp 1911–1916

Groumpos PP (2022) A critical historic overview of artificial intelligence: Issues, challenges, opportunities and threats. Artif Intell Appl. https://doi.org/10.47852/bonviewAIA3202689

Gurbuz SZ, Gurbuz AC, Malaia EA, Griffin DJ, Crawford CS, Rahman MM, Kurtoglu E, Aksu R, Macks T, Mdrafi R (2020) American sign language recognition using RF sensing. IEEE Sens J 21(3):3763–3775

Hameed H, Usman M, Tahir A, Ahmad K, Hussain A, Imran MA, Abbasi QH (2022) Recognizing British sign language using deep learning: a contactless and privacy-preserving approach. IEEE Trans Comput Soc Syst 10(4):2090–2098

Handtalk (2024) https://www.handtalk.me/en/. Accessed 0 Apr 2024

Haron H, Samad H, Diah FM, Yusof H (2019) E-learning approach using mobile apps: Malaysian sign language for dumb and deaf. Int J Adv Res Technol Innov 1(1):1–7

He L, Niu M, Tiwari P, Marttinen P, Su R, Jiang J, Guo C, Wang H, Ding S, Wang Z et al (2022) Deep learning for depression recognition with audiovisual cues: a review. Inf Fusion 80:56–86

Huang J, Zhou W, Li H, Li W (2015) Sign language recognition using 3D convolutional neural networks. In: 2015 IEEE international conference on multimedia and expo (ICME). IEEE, pp 1–6

Imran A, Razzaq A, Baig IA, Hussain A, Shahid S, Rehman T-u (2021) Dataset of Pakistan sign language and automatic recognition of hand configuration of Urdu alphabet through machine learning. Data Brief 36:107021

Islam MM, Uddin MR, Akhtar MN, Alam KR (2022) Recognizing multiclass static sign language words for deaf and dumb people of Bangladesh based on transfer learning techniques. Inf Med Unlocked 33:101077

Jadán-Guerrero J, Arias-Flores H, Cedillo P, Chacón-Castro M (2023) Alternatives for designing augmentative and alternative communication systems for people with disabilities and older adults. In: International conference on human–computer interaction, pp 321–328. Springer, Cham

Janiesch C, Zschech P, Heinrich K (2021) Machine learning and deep learning. Electron Mark 31(3):685–695

Jiang X, Ahmad W (2019) Hand gesture detection based real-time American sign language letters recognition using support vector machine. In: 2019 IEEE international conference on dependable, autonomic and secure computing, international conference on pervasive intelligence and computing, international conference on cloud and big data computing, international conference on cyber science and technology congress (DASC/PiCom/CBDCom/CyberSciTech). IEEE, pp 380–385

Justalk (2024) https://www.justalk.com/. Accessed 08 Apr 2024

Kahlon NK, Singh W (2023) Machine translation from text to sign language: a systematic review. Univ Access Inf Soc 22(1):1–35

Kang H (2022) Effects of artificial intelligence (AI) and native speaker interlocutors on ESL learners' speaking ability and affective aspects. Multimedia Assist Lang Learn 25(2):9

Karras C, Karras A, Sioutas S (2022) Pattern recognition and event detection on iot data-streams. arXiv preprint. arXiv:2203.01114

Katılmış Z, Karakuzu C (2023) Double handed dynamic Turkish sign language recognition using leap motion with meta learning approach. Expert Syst Appl 228:120453

Kaur G, Habibi-Lashkari Z, Habibi-Lashkari A, Kaur G, Habibi-Lashkari Z, Habibi-Lashkari A (2021) Introduction to cybersecurity. In: Understanding cybersecurity management in FinTech: challenges, strategies, and trends. Springer, Cham, pp 17–34

Kumar MN (2023) Artificial intelligence and linguistics: the scientific study of natural language processing, 141***

Latif G, Mohammad N, Alghazo J, AlKhalaf R, AlKhalaf R (2019) ARASL: Arabic alphabets sign language dataset. Data Brief 23:103777

Leap Motion (2023) https://www.leapmotion.com/ Accessed 20 June 2023

Lee BG, Lee SM (2017) Smart wearable hand device for sign language interpretation system with sensors fusion. IEEE Sens J 18(3):1224–1232

Li K, Chen C, Cao W, Wang H, Han S, Wang R, Ye Z, Wu Z, Wang W, Cai L et al (2023) Deaf: a multi-modal deep learning framework for disease prediction. Comput Biol Med 156:106715

Live Transcribe & Notification (2024) https://play.google.com/store/apps/details?id=com.google.audio. hearing.visualization.accessibility.scribe &hl=en &gl=US &pli=1. Accessed 8 Apr 2024

Lum KY, Goh YH, Lee YB (2020) American sign language recognition based on MobileNet-v2. Adv Sci Technol Eng Syst J 5(6):481–488

Luo X, Li F, Xu W, Hong K, Yang T, Chen J, Chen X, Wu H (2021) Machine learning-based genetic diagnosis models for hereditary hearing loss by the GJB2, SLC26A4 and MT-RNR1 variants. EBioMedicine 69:103322

Mandal R, Patil D, Gadhe S, Birari G, Buwa T (2023) Dual mode sign language recognizer-an android based CNN and LSTM prediction model. In: 2023 3rd International conference on artificial intelligence and signal processing (AISP). IEEE, pp 1–5

Mariappan J, Krishnan C (2023) Artificial intelligence: future of advance learning. In: Digital transformation in education: emerging markets and opportunities. Bentham Science Publication, Sharjah, pp 118–135

Martins P, Rodrigues H, Rocha T, Francisco M, Morgado L (2015) Accessible options for deaf people in e-learning platforms: technology solutions for sign language translation. Procedia Comput Sci 67:263–272

Masiello-Ruiz JM, Ruiz-Mezcua B, Martinez P, Gonzalez-Carrasco I (2023) Synchro-sub, an adaptive multi-algorithm framework for real-time subtitling synchronisation of multi-type TV programmes. Computing 105:1467–1495

Microsoft Kinect (2023) http://www.microsoft.com/en-us/kinectforwindows/ Accessed 20 June 2023

Microsoft Translator (2024) https://www.microsoft.com/en-us/translator/. Accessed 8 Apr 2024

Mubin SA, You JST, Samiraj EPR, Jaafar SP (2022) A framework for supporting deaf and mute learning experience through extended reality. In: 2022 IEEE international conference on computing (ICOCO). IEEE, pp 80–83

Nahar KM, Almomani A, Shatnawi N, Alauthman M (2023) A robust model for translating Arabic sign language into spoken Arabic using deep learning. Intell Autom Soft Comput 37(2):2037–2057

Papatsimouli M, Sarigiannidis P, Fragulis GF (2023) A survey of advancements in real-time sign language translators: integration with iot technology. Technologies 11(4):83

Pivetta EM, Saito DS, Ulbricht VR (2014) Deaf and accessibility: analysis of a virtual learning environment. Rev Bras Educ Esp 20:147–162

Prasath A et al (2022) Design of an integrated learning approach to assist real-time deaf application using voice recognition system. Comput Electr Eng 102:108145

Proksch S-O, Wratil C, Wackerle J (2019) Testing the validity of automatic speech recognition for political text analysis. Polit Anal 27(3):339–359

Ranasinghe P, Akash K, Nanayakkara L, Perera H, Chandrasiri S, Kumari S (2022) E-learning assistive system for deaf and mute students. In: 2022 4th International conference on advancements in computing (ICAC). IEEE, pp 13–18

Robert EJ, Duraisamy HJ (2023) A review on computational methods based automated sign language recognition system for hearing and speech impaired community. Concurr Comput Pract Exp 35(9):7653

Sabharwal S, Singla P (2022) Indian sign language digit translation using CNN with swish activation function. In: International conference on information systems and management science. Springer, Cham, pp 245–253

Sahana T, Paul S, Basu S, Mollah AF (2020) Hand sign recognition from depth images with multi-scale density features for deaf mute persons. Procedia Comput Sci 167:2043–2050

Sahoo AK (2021) Indian sign language recognition using machine learning techniques. In: Macromolecular symposia, vol 397. Wiley Online Library, New York, p 2000241

Saleem MI, Siddiqui A, Noor S, Luque-Nieto M-A, Nava-Baro E (2023) A machine learning based full duplex system supporting multiple sign languages for the deaf and mute. Appl Sci 13(5):3114

Sarker IH (2021a) Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions. SN Comput Sci 2(6):420

Sarker IH (2021b) Machine learning: algorithms, real-world applications and research directions. SN Comput Sci 2(3):160

Sarker IH (2022) AI-based modeling: techniques, applications and research issues towards automation, intelligent and smart systems. SN Comput Sci 3(2):158

Schembri A, Fenlon J, Rentelis R, Reynolds S, Cormier K (2013) Building the british sign language corpus. University of Hawaii Press, Honolulu

Serey J, Alfaro M, Fuertes G, Vargas M, Durán C, Ternero R, Rivera R, Sabattin J (2023) Pattern recognition and deep learning technologies, enablers of industry 4.0, and their role in engineering research. Symmetry 15(2):535

Shahid NM, Law EL-C, Verdezoto N (2022) Technology-enhanced support for children with down syndrome: a systematic literature review. Int J Child Comput Interact 31:100340

Siddique S, Islam S, Neon EE, Sabbir T, Naheen IT, Khan R (2023) Deep learning-based Bangla sign language detection with an edge device. Intell Syst Appl 18:200224

Sign Language Gesture Images Dataset (2020) https://www.kaggle.com/datasets/ahmedkhanak1995/sign-language-gesture-images-dataset . Accessed 20 June 2023

Sign Language MNIST (2017) Available: https://www.kaggle.com/datasets/datamunge/sign-language-mnist. Accessed: 2023-06-20

SIGNILY (2024) Available: https://aslized.org/innovations/signily/. Accessed 8 Apr 2024

Sohail SS, Farhat F, Himeur Y, Nadeem M, Madsen DØ, Singh Y, Atalla S, Mansoor W (2023) The future of GPT: a taxonomy of existing ChatGPT research, current challenges, and possible future directions. SSRN. https://doi.org/10.2139/ssrn.4413921

Soori M, Arezoo B, Dastres R (2023) Artificial intelligence, machine learning and deep learning in advanced robotics, a review. Cogn Robot 3:54–70

Subramanian B, Olimov B, Naik SM, Kim S, Park K-H, Kim J (2022) An integrated mediapipe-optimized GRU model for Indian sign language recognition. Sci Rep 12(1):11964

Sullivan LS, Klein E, Brown T, Sample M, Pham M, Tubig P, Folland R, Truitt A, Goering S (2018) Keeping disability in mind: a case study in implantable brain-computer interface research. Sci Eng Ethics 24:479–504

Sundar B, Bagyammal T (2022) American sign language recognition for alphabets using MediaPipe and LSTM. Procedia Comput Sci 215:642–651

Szeliski R (2022) Computer vision: algorithms and applications. Springer, London

Talaat FM, Ali ZH, Mostafa RR, El-Rashidy N (2024a) Real-time facial emotion recognition model based on kernel autoencoder and convolutional neural network for autism children. Soft Comput. https://doi.org/10.1007/s00500-023-09477-y

Talaat FM, Farsi M, Badawy M, Elhosseini M (2024b) SightAid: empowering the visually impaired in the Kingdom of Saudi Arabia (KSA) with deep learning-based intelligent wearable vision system. Neural Comput Appl. https://doi.org/10.1007/s00521-024-09619-9

Tapsos (2024) https://tapsos.com/. Accessed 8 Apr 2024

The ASL App (2024) https://apps.apple.com/us/app/the-asl-app/id921030207. Accessed 8 Apr 2024

Tobore I, Li J, Yuhang L, Al-Handarish Y, Kandwal A, Nie Z, Wang L et al (2019) Deep learning intervention for health care challenges: some biomedical domain considerations. JMIR Mhealth Uhealth 7(8):11966

Venugopalan A, Reghunadhan R (2021) Applying deep neural networks for the automatic recognition of sign language words: a communication aid to deaf agriculturists. Expert Syst Appl 185:115601

Venugopalan A, Reghunadhan R (2023) Applying hybrid deep neural network for the recognition of sign language words used by the deaf covid-19 patients. Arab J Sci Eng 48(2):1349–1362

Von Agris U, Kraiss K-F (2007) Towards a video corpus for signer-independent continuous sign language recognition. In: Gesture in human-computer interaction and simulation, vol 11(2), Lisbon, Portugal

Vrontis D, Christofi M, Pereira V, Tarba S, Makrides A, Trichina E (2022) Artificial intelligence, robotics, advanced technologies and human resource management: a systematic review. Int J Hum Resour Manag 33(6):1237–1266

Wadhawan A, Kumar P (2021) Sign language recognition systems: a decade systematic literature review. Arch Comput Methods Eng 28:785–813

Wang H, Leu MC, Oz C (2006) American sign language recognition using multi-dimensional hidden Markov models. J Inf Sci Eng 22(5):1109–1123

Wu G, Ji X, Yang G, Jia Y, Cao C (2023) Signal-to-image: rolling bearing fault diagnosis using resnet family deep-learning models. Processes 11(5):1527

Yağanoğlu M (2021) Real time wearable speech recognition system for deaf persons. Comput Electr Eng 91:107026

Yazdinejad A, Dehghantanha A, Parizi RM, Epiphaniou G (2023) An optimized fuzzy deep learning model for data classification based on NSGA-II. Neurocomputing 522:116–128

Yousaf K, Mehmood Z, Saba T, Rehman A, Rashid M, Altaf M, Shuguang Z (2018) A novel technique for speech recognition and visualization based mobile application to support two-way communication between deaf-mute and normal peoples. Wirel Commun Mobile Comput. https://doi.org/10.1155/2018/1013234

Zahedi M, Dreuw P, Rybach D, Deselaers T, Ney H (2006) Geometric features for improving continuous appearance-based sign language recognition. BMVC 3:1019–1028

Zhang Y, Xu W, Zhang X, Li L (2022) Sign annotation generation to alphabets via integrating visual data with somatosensory data from flexible strain sensor-based data glove. Measurement 202:111700

Zhou L, Pan S, Wang J, Vasilakos AV (2017) Machine learning on big data: opportunities and challenges. Neurocomputing 237:350–361

Zhou B, Yang G, Shi Z, Ma S (2022) Natural language processing for smart healthcare. IEEE Rev Biomed Eng 17:4–18

## Authors and Affiliations

**Hanaa ZainEldin[1] · Samah A. Gamel[2] · Fatma M. Talaat[3,7,8] · Mansourah Aljohani[4] · Nadiah A. Baghdadi[5] · Amer Malki[4] · Mahmoud Badawy[1,6] · Mostafa A. Elhosseini[1,4]**

✉  Mahmoud Badawy
   engbadawy@mans.edu.eg

[1]  Computers and Control Systems Engineering Department, Faculty of Engineering, Mansoura University, Mansoura 46421, Egypt

[2]  Electronics & Communication Engineering Department, Faculty of Engineering, Horus University, New Damietta, Egypt

[3]  Faculty of Artificial Intelligence, Kafrelsheikh University, Kafrelsheikh 33516, Egypt

[4]  College of Computer Science and Engineering, Taibah University, Yanbu 46421, Saudi Arabia

[5]  Nursing Management and Education Department, College of Nursing, Princess Nourah Bint Abdulrahman University, Riyadh 11671, Saudi Arabia

[6]  Computer Science and Information Department, Applied College, Taibah University, Madinah 41461, Saudi Arabia

[7]  Faculty of Computer Science & Engineering, New Mansoura University, Gamasa 35712, Egypt

[8]  Nile Higher Institute for Engineering and Technology, Mansoura, Egypt