Adaptive Teaching of the Iranian Sign Language Based on Continual Learning Algorithms

Morteza Memari[1], Alireza Taheri[1]

[1] Social and Cognitive Robotics Lab., Mechanical Engineering Department, Sharif University of Technology, Tehran, Iran

Corresponding author: Alireza Taheri (e-mail: artaheri@sharif.edu).

“Author Alireza Taheri has received a research grant from the “Iranian National Science Foundation (INSF) (http://en.insf.org/) (Grant No. 4031030).”

ABSTRACT Research has demonstrated that intelligent systems significantly enhance the learning process. This study aims to design and implement an interactive computer-based platform for adaptive teaching of Iranian Sign Language (ISL). Unlike most sign languages that rely solely on hand movements, ISL also requires lip movements, which are crucial for distinguishing many words. This dual requirement presents a unique challenge in detecting ISL words from video frames. To address this, we created a dataset named ISLR101, containing videos of 101 ISL signs. We designed a neural network with three transformer encoder modules for different input features and one for generating output. This architecture enables the system to accurately learn and recognize ISL signs. To ensure the neural network can continuously learn new ISL words without forgetting previously learned ones, we employed the Elastic Weight Consolidation (EWC) method. This approach helps maintain an average accuracy of 82.92% across six training tasks, each comprising approximately 17 classes. Following the training process, we developed an interactive teaching system based on fuzzy logic. This system adapts to users’ needs and performance, enhancing their learning experience. The system’s effectiveness was evaluated using the UTAUT questionnaire in a preliminary exploratory study involving 20 hearing individuals (10 males and 10 females). The results indicated that the adaptive teaching architecture interacted effectively with different users (Cohen’s d = 0.57) and adapted to them (Cohen’s d = 0.63) over four training sessions. Additionally, users showed increased motivation to interact with the intelligent educational system (Cohen’s d = 0.92).

INDEX TERMS Adaptive Teaching, Continual Learning, Fuzzy Logic, Human-Machine Interaction, Iranian Sign Language.

I. INTRODUCTION

Humans are social beings and form societies to advance their life goals. Since language serves as the primary means of communication within a society, there is a strong need for individuals in a human society to learn a common language. A portion of the deaf community uses a language known as (local) sign language to communicate and interact with other humans. Deaf individuals living in Iran have chosen Iranian Sign Language (ISL) as their tool of communication with fellow society members. Therefore, the significance of teaching Iranian Sign Language to all members of society, both hearing and deaf, is evident. One of the main features of ISL is the simultaneous use of hand and lip movements. On the other hand, with the remarkable advancements in technology and the widespread utilization of intelligent systems such as social robots [1], computer-based platforms [2], and artificial intelligence in various industrial, welfare, and other sectors research in the emerging field of intelligent systems and AI has also surged [3]. The positive outcomes of research on the use of social robots and AI have encouraged researchers in this domain to further explore areas like education, therapy, and entertainment [4]. One notable example of such research involves employing robots to teach humans [5], especially those with special needs. The teaching of sign language to hearing-impaired people and/or hearing people interested in learning sign language using social robots has been one of the focal points in this field [6, 7].

Sign language is not only the primary means of communication for deaf children with those around them, but also plays a significant role in their intellectual and cognitive development. Additionally, incorporating social robots and/or intelligent apps into the teaching process can enhance their educational efficiency and contribute to their intellectual growth. Therefore, enhancing the utilization of intelligent systems to teach Sign Language to people is considered as an important research line in the field of social robotics.

Based on the articles [8] and [9] and following the suggestions made by [10] and [11], previous studies on adaptive teaching by robots/intelligent systems can be categorized into three groups according to the method of user profiling and the robot's adaptability mechanism:

- Adaptive systems without user modeling: These systems typically exhibit reactive behavior that aligns with the user's real-time feedback. However, the mentioned systems do not have the capability to store user information in a separate memory.
- Adaptive systems with static user models: In these systems, user-specific characteristics are initially defined, and this predefined information is subsequently used to tailor the system's behavior to each user.
- Adaptive systems with dynamic user models: Similar to the previous category, these systems maintain and utilize distinct user’s information. The key distinction lies in the fact that in this type of systems, user’s information is updated in real-time to align with the robot's current task during operation.

In 2014, Aly and Tapus [12] conducted a study focusing on the interplay between human and robot personalities and how they influence each other. Utilizing Nao and Alice robots in their research, they employed a novel fuzzy methodology to detect the user's emotions in real time. The primary objectives of the mentioned study were to adapt the robot's reciprocal behavior based on human personality and emotions, ultimately striving for improved long-term interaction. Fuzzy logic is widely recognized as a powerful tool for addressing complex or ill-defined problems. Consequently, it is proved to be valuable and efficient in discerning nuanced human emotions that lack precise and well-defined boundaries. When confronted with novel human emotional states, the robot/system classifies them into existing fuzzy categories or generates new categories to guide its verbal and non-verbal responses. Subsequently, it tailors its speech content, body movements, and facial expressions based on these generated categories. However, two notable weaknesses to consider are the limitations in the range of emotions available for the robot to express and the absence of diverse scenarios for the robot to select interaction content, which may impact its suitability as a teacher.

In 2015, Westlund, Gordon, and their colleagues [13] introduced a mobile robot designed to assist children (whose first language was English) in learning Spanish as a second language. The study involved thirty-four preschool children who participated in an interactive game. This game was played on a tablet, featuring a toucan bird displayed on the screen. During this interactive game, two factors were personalized for each individual: 1) the game content (i.e., selection of words to teach), and 2) the robot's emotional response to the child's emotional behavior and reactions. Subsequently, the child's motivation and educational progress were assessed. However, due to the extensive time required to examine the subjects' responses and performance, the final outcomes of the study were deferred to consequent reports.

Gordon and Breazeal [14] introduced an intelligent robotic educational system, aiming to assist children in learning and enhancing their reading skills. This system records and stores the user's reading skill level, periodically evaluating and updating the data through an active learning technique. The recorded information is then utilized to tailor the interactive game between the child and the robot, enhancing the teaching process. In this study, the robot adjusted its motivational strategy, incorporating verbal and non-verbal activities based on assessments conducted by an automated facial expression analysis system. These activities were chosen to align with the child's motivational level. The data analysis revealed that while the child's emotional valence changed in response to the robot's non-verbal activities, it did not significantly affect the child's overall motivational level. Nonetheless, personalizing instruction led to long-term increases in positive emotional valence. Another noteworthy result of this study was the demonstration of the system's versatility in interacting with individuals of various age groups. Additionally, it found that children who interacted with the adaptive system achieved more effective learning outcomes. Despite the significance of these findings, it is important to acknowledge certain limitations in the mentioned study. These include the relatively short duration of children's interactions and the exclusive reliance on individuals' emotional valence and motivational levels for personalizing education. Consequently, these results, while valuable, may not offer a comprehensive guideline for adaptive teaching.

In [15], the authors developed an empathetic robot to assist users in learning geography. The system records and stores individual skill levels, such as compass use and map symbol recognition, and adapt its responses accordingly. The authors measured significant improvements in perceived enjoyment, mutual understanding, and user trust as the key metrics. One of the main challenges highlighted in this article is the integration of individuals' emotional and social signals with their learning performance to generate appropriate teaching scenarios for the robot. Another notable strength of this work is the modular approach to create robotic perception and adaptability processes, organized within a literature-based architecture known as imitation. However, there are two notable limitations to be considered. The Nao robot, used in this study, lacks the ability to replicate emotions through facial expressions. Additionally, its size limits its authority as a geography teacher, which could be seen as weaknesses in its suitability for this role.

In another study [16], an adaptable cognitive architecture named R-Castle was implemented on the Nao social robot to facilitate interactive educational activities (such as teaching geometric shapes to children). This innovative architecture introduced a new goal for the robot. Within this framework, the system estimates the user's motivational level, and the robot is designed to adapt in order to maintain the user's attention throughout the interaction. During the robot's engagement with the child in a game-based scenario, the level of user attention becomes a crucial factor that influences the robot's behavior. It determines whether the robot makes mistakes or not, and whether it adapts a decisive or flexible teaching approach, transitioning between being a teacher or a learner. An additional innovation in this research is the utilization of a search engine to inform the selection of future interactions, enhancing the overall effectiveness of the educational experience.

Tapus et al. [17] introduced a robotic system designed to assist post-stroke patients with rehabilitation exercises. This system employs adaptability by adjusting its behavior and personality traits, specifically in terms of introversion or extroversion, through three parameters: interaction distance, movement speed, and verbal communication (i.e., voice intensity and talking speed). The Policy Gradient Reinforcement Learning (PGRL) algorithm played a pivotal role in this system's operation. The primary objectives of this adaptability are to help patients achieve higher functional levels by optimizing the aforementioned three parameters and tailoring them to each patient's personality. In this study, practical user feedback was utilized for adaptability, with a focus on user’s information such as personality traits that influence the robot's decision-making process. However, this approach prioritized user’s personality over the learning process. Notable disadvantages of this system include the absence of a reciprocal relationship between the robot and the human, as well as the robot's lack of awareness of the user's internal states.

The surge in smartphone technology has inspired researchers and developers to create new sign language apps. These apps leverage cutting-edge technology to bridge the communication gap between deaf and hearing individuals. Dahanayaka et al. designed an Android app that bridges this gap [18]. The app uses advanced AI (by using Convolutional Neural Networks) to translate sign language into spoken words for hearing people and spoken language into text for deaf and mute people. It focuses on using a phone's camera (vision) for both sign language recognition and converting spoken words to text. To test the app's effectiveness, the researchers trained the sign language recognition part using some letters on Google Colab, while a separate system trained speech recognition on machines. Once trained, these models were converted for use on mobile devices using TensorFlow Lite. The results were promising, with the letters being recognized correctly 95% of the time.

A 2018 study by Yousaf et al. introduced a mobile app called "Vocalizer to Mute" [19]. This app helps bridge the communication gap by using automatic speech recognition to understand speech from deaf or mute individuals and convert it into clear speech for others. The app also features a 3D avatar that can display sign language, acting as a communication bridge for those unfamiliar with sign language. To improve accuracy, the researchers analyzed specific speech features of deaf and mute individuals. In a small test with 15 deaf or mute children (aged 7-13) at a social welfare facility, the app achieved an impressive 97.9% accuracy. The study found that this mobile technology not only improved communication accuracy, but also enhanced social interaction for deaf and mute people.

A study by Liu et al. (2020) explored the potential of an app called FinGTrAC [20]. This app uses a wearable sensor (like a ring and smartwatch) to track complex finger movements with high accuracy. Previous technology could only recognize a few dozen hand gestures. FinGTrAC can recognize hundreds of gestures using minimal sensors. The study focused on how this technology could be applied to translate American Sign Language (ASL). Ten participants (7 men and 3 women) were involved in the evaluation. They learned 50 common ASL phrases (3-11 words each). While wearing the sensor platform, they signed these phrases ten times each. Analysis showed the system could identify the top 100 most common ASL gestures with 94.2% accuracy, even when used in different contexts.

Social robots and intelligent systems can play a valuable role in the teaching of sign language to individuals, provided that they can accurately and rapidly recognize the signs and movements of sign language performed by the user. To ensure the effectiveness of this recognition system across diverse environmental and practical conditions and in interactions with various users, it must possess sufficient robustness. Various tools have been employed to capture signs and patterns in sign language, including RGB cameras [21, 22], Kinect sensors [23], TOF cameras [24], Leap Motion sensors [25], data gloves [26], and/or combinations of these tools [27].

In the quest for automatic detection of sign language motion patterns, various methods have been proposed in the literature by utilizing different data collection mechanisms. Recognition algorithms commonly employed in research include various types of neural networks [22], classic classification methods such as Nearest Neighbor [28] and

Support Vector Machine [29], as well as Hidden Markov Models [30], and combinations of these approaches [31]. Among neural network types, there are standard neural networks, RBF neural networks, recurrent neural networks, deep neural networks, long-short term memories, etc.

In a study conducted in 2019 [30], deep neural networks were used to develop a pattern for continuous detection of sign language signs. In this study, sign language sentences were taken as input from a video containing RGB frames and optical flow images, and a sequence of uttered word labels was provided as output. This approach employed deep neural networks with temporal fusion layers as the feature extraction module and recurrent neural networks as the sequential learning module. Notably, this was a departure from the common use of Hidden Markov Models, which have limited capacity in capturing transient information. The primary aim of the mentioned study was to explore methods for achieving high recognition accuracy in expressing sign language signs with a small amount of learning data. This challenge arose from the constraints of having insufficient data to fully train a complex deep neural network.

In 2018, Hosseini [32] developed a real-time unit for detecting Iranian Sign Language signs using a data glove and Hidden Markov Model. Among 36 evaluated models, the best-performing model correctly identified 14 sign language signs with an accuracy of approximately 85%. While this method is less accurate compared to neural networks, the author's choice was driven by the fact that Hidden Markov Models not only detect motion but also have the ability to reproduce it.

In the latest research conducted in 2023 by Basiri [33], deep neural networks and the "state-image" method for data preprocessing were used to empower the RASA robot for adaptive teaching of Iranian Sign Language. This study tailored the teaching approach by adjusting selected words for teaching, robot speed, robot repetition time, and its emotional reaction to each user. The adaptive system achieved an accuracy of over 90% in correctly identifying 15 signs from Iranian Sign Language and demonstrated a significant positive impact on user performance.

The present study builds upon Basiri's [34] work, introducing a key innovation in the form of the continual learning algorithm. The main goal of this study is to design an intelligent computer-based app with the ability of adaptive teaching of the Iranian Sign Language based on continual learning algorithms to users. To this end, a dataset called ISLR101 was collected by our group encompassing 101 words from the Iranian Sign Language lexicon and performed by 11 individuals (resulting in a total of 4040 videos). We utilized YoloV5m and MediaPipe to extract the key points for the hands and lips, as well as coordinates for the wrists and elbows. To improve sign recognition accuracy, we incorporated additional features, including the Euclidean distance between hand positions and the angle created by the line connecting the hands to the horizontal axis. A transformer-based neural network, capable of continual learning, was employed to detect ISL words. This network was trained on a Core (TM) i7-7500U CPU with 16GB of RAM. After training the neural network, we created an interactive fuzzy logic-based architecture to teach Iranian Sign Language to users. To effectively implement this adaptive teaching system, we developed a straightforward software application using the PySimpleGUI library. In an initial exploratory acceptance study of the designed app, we conducted four training sessions with 20 hearing individuals (10 men and 10 women) and had them complete the UTAUT questionnaire. Figure 1 illustrates an overview of this study.

The first significant innovation of this research involves gathering/presenting a dataset called ISLR101 to encompass a wide variety number of ISL motions (in comparison to the previous/available ISL datasets). This expansion includes signs involving simultaneous hand and lip movements. Therefore, instead of relying on a data glove, an RGB camera is utilized for motion/sign detection. This not only enables the teaching of a broader range of signs, but also eliminates the need for a data glove, promoting the more general use of the intelligent system in this area.

Unlike most sign languages that rely solely on hand movements for communication, Iranian Sign Language also requires lip movements. These lip movements are essential because they distinguish many words in ISL. Therefore, a key challenge in ISL is the need to use both hand and lip movements, simultaneously. Hence, the second innovation of this study is considering the motions of both hands and lips in recognizing the performed ISL signs by users based on deep neural networks.

To enable users to operate our system with the ability to improve its vocabulary size over time (and without having technical knowledge), it is necessary for the system to gradually learn Iranian Sign Language words over time; and not forget previously learned words when learning new ones. Therefore, continual learning techniques are used to train the system. This algorithm allows the system to train itself while operating and enrich its training dataset continuously. In essence, the teaching system can learn during the teaching phase, mirroring the continual learning process observed in humans.

Furthermore, this research aims to enhance the system's user modeling speed. When encountering a new user, instead of building a dynamic model of the user from scratch, the system selects the closest existing model based on the similarity of the user's behavior to previous users and customizes it. Additionally, this study incorporates both general adaptability (i.e., improving the system's logic concerning the overall educational program) and specific adaptability (customizing the educational program for individual users) concurrently within the teaching system's learning process. Performance and effectiveness of this research have been evaluated through a preliminary exploratory study.

All in all, the main contributions of this study are as follows:

- Gathering/Providing the ISLR101 dataset (a unique dataset including the videos of 101 ISL signs) by our research group
- Simultaneous detection/process of hands and lip motions
- Using continual learning algorithm for ISL recognition
- Improving our recent designed computer-based app (Ali Ghadami, 2023) with the new ability of adaptive teaching of ISL signs to users
- Conducting a preliminary exploratory field study to investigate the designed app's performance/acceptance for 20 users

To the best of our knowledge, the first four mentioned items of our contributions have not been presented in previous ISL studies so far.

II. Methodology

To achieve our research objectives, first, we gathered an appropriate dataset. Then, we initiated the process by training a neural network to recognize the words in Iranian Sign Language (ISL). We employed our recorded videos demonstrating the execution of each word/sign in the ISL for training. An essential feature of this neural network is its capability for continual learning, enabling it to be trained for new words/signs while maintaining high accuracy in recognizing all previously learned words/signs. This eliminates the need to train the system on all words simultaneously. The continual learning algorithm ensures that the system retains its learning capacity, enabling it to recognize more words over time. Following the completion of the neural network training, we designed an interactive fuzzy logic-based architecture for teaching Iranian Sign Language to users. This teaching architecture offers adaptability at both general and specific levels. This design approach makes the training system emulate human teaching methods and enhances its effectiveness in facilitating high-quality learning experiences. Finally, we evaluated the performance of the teaching system within the context of sign language teaching software through field studies involving diverse users. We developed a simple software to implement the interactive training architecture designed in the previous phase and used it to teach the Iranian Sign Language to users. We collected user feedback through validated questionnaires and drew conclusions regarding the performance of the adaptive teaching architecture based on this data.

A. Data Collection and our Dataset

The dataset utilized in this research, named Iranian Word-Level Sign Language Recognition Dataset (ISLR101), comprises videos demonstrating the execution of 101 words/signs in Iranian Sign Language. We have collected this dataset at the Islamic Azad University, Fereshtegan Branch (which is a unique university for individuals with hearing problems in Iran), with the invaluable assistance of esteemed sign language interpreters. Each video features one individual performing a single word, with diverse backgrounds representing the data collection environment. The only constraint was the presence of a sole person within each frame. All videos possess a resolution of 600 * 800 pixels and a frame rate of 25 frames per second. This dataset was collaboratively compiled in conjunction with other research projects at the Social and Cognitive Robotics Lab; encompassing 101 words from the Iranian Sign Language lexicon, performed by 11 individuals, and resulting in a total of 4040 videos (Figure 2). The length of the videos of our dataset ranges from 21 frames to 116 frames (with an average length of 57 frames).

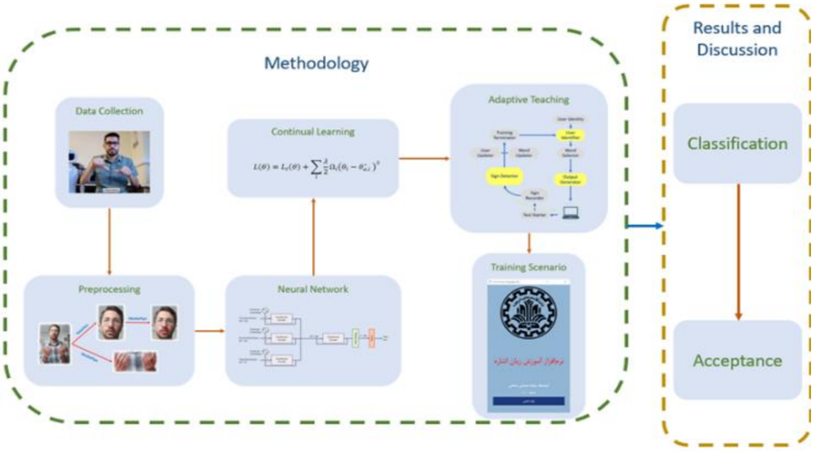


FIGURE 1. An overview of this study

This dataset has been divided into the training, validation, and test sets for use in training the neural network. The division of the data is such that the data from 9 (out of 11) individuals are used as the training data, the data from 1 individual is used as the validation data, and the data from 1 individual is used as the test data (in all of the tasks in this paper). The selection of individuals was done randomly. This type of division ensures that the performance of the neural network has less dependency on the specific characteristics of the individuals executing the signs/words.

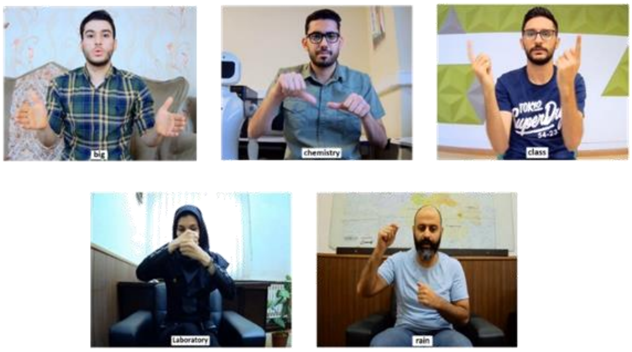


FIGURE 2. ***ISLR101: Iranian Word-Level Sign Language Recognition Dataset***

B. Preprocessing

The duration of the videos in the dataset was variable; therefore, the length of all videos was restricted to 40 frames for this reason in our study. While the emergence of deep neural networks has reduced the necessity for extensive data preprocessing, the advantages of utilizing clean and meaningful inputs remain undeniable, leading to benefits such as reduced training time and increased the network accuracy. In this research, instead of employing raw image data as input, we opted to extract and utilize meaningful features pertinent to sign language recognition. These features encompass hand key point coordinates, lip key point coordinates, and the length and angle of the line connecting the hands [35].

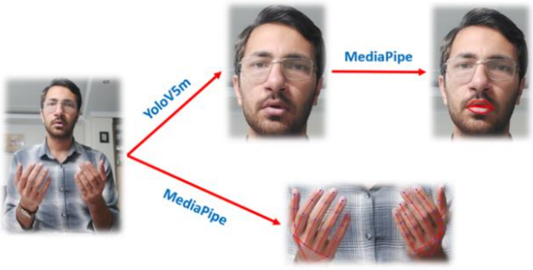


FIGURE 3. ***Using YoloV5m and MediaPipe to extract hands and lip keypoints***

For the detection of hand and face regions, we employed the pre-trained YoloV5m model due to its remarkable accuracy and speed [35]. The YoloV5m model was chosen for training due to its superior accuracy and speed. This model detects objects using rectangular bounding boxes, resulting in each training label containing a vector. This vector includes the identified object class, the normalized coordinates of the bounding box's center, and the bounding box's length and width. Our labels for hands and lips were manually prepared for each selected data point in [36] Training the network involved combining four different datasets: Roboflow hand data, Roboflow-FAST-model face data, several frames from Iranian deaf news, and some frames from our Iranian Sign Language word data. The dataset consists of 8543 images, which were split into training, validation, and test sets in an 8:1:1 ratio. Given the vital role of hands in sign language, we extracted the coordinates of 21 points from each hand using the MediaPipe [37] for each frame (Figure 3). Considering 3 components for each key point and 2 hands within the image, the length of this feature vector was totally 126. These extracted features primarily offer insights into the hand shapes. To capture motion information, we also extracted the coordinates of two points on each hand, specifically the wrist and elbow, utilizing the MediaPipe library. Consequently, with 3 components per point and 2 hands, the length of the hand motion feature vector was 12. Furthermore, considering the significance of lip movement in Iranian Sign Language, we employed the MediaPipe library to extract the lip keypoint coordinates. As shown in figure 3, initially, we used the YoloV5m network to obtain the lip region; and subsequently, we extracted the coordinates of 40 points on the lips for each frame. Taking into account the bounding box detected for each hand, we determined the hand's position based on the center of the enclosed rectangle. Additionally, we normalized the hand positions with respect to the image's width and height. To enhance our sign recognition, we also included features such as the Euclidean distance between the hand positions and the angle formed by the connection line between the hands and the horizontal axis. The Extracted features' structure is shown in figure 4. The idea of feature extraction in this study is as the same as one of our recent works for recognizing ISL without using continual learning algorithms [36].

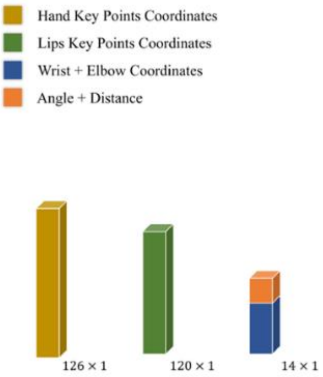


FIGURE 2. ***Extracted features' structure***

C. Neural Network's Architecture

In this research, input streams are processed independently within the neural network, with each stream undergoing separate treatment. As shown in figure 5, initially, each input

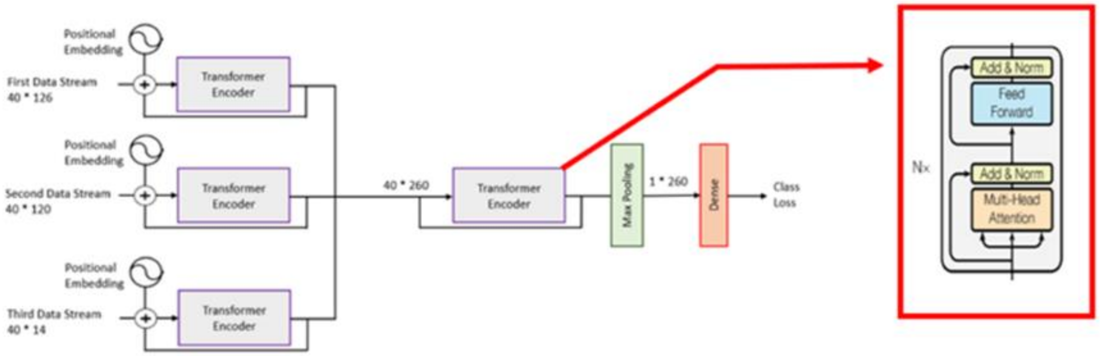


FIGURE 3. ***Our used neural network architecture***

stream proceeds through a transformer encoder module (followed by Gelu and linear activation functions). It is notable that before the first encoders, positional embeddings layers are added. The output of this module is a vector of the same dimension and number as the input vector, encompassing not only information from the current frame but also the details from the preceding and succeeding frames. Essentially, these encoder modules analyze the relationship of each input stream with itself, producing a context-aware output vector. Subsequently, the three resulting output vectors (each corresponding to one of the input streams) are concatenated to form a unified vector. This combined vector then traverses another transformer encoder module; and following the Max Pooling module, it passes through a dense layer with a neuron count equal to the number of words (i.e., 101); ultimately determining the neural network's output. The advantage of this architectural approach lies in its ability to separately convey the unique information and features of each input stream to the model. This information transfer significantly enhances the model's performance. Furthermore, this structure enables the effective amalgamation of the strengths of each stream, fostering diversity and combinability of information, which is instrumental in enhancing the model performance.

The model comprises four transformer encoders, each equipped with 12 heads. The encoder handling the hand keypoint input stream and the one responsible for the wrist and elbow input stream consist of 256 neurons in their respective Feed Forward layers (in the Transformer Encoder module). The lip keypoint encoder features 64 neurons in its Feed Forward dense layer, while the final encoder in the network contains 512 neurons. The model employs the Softmax activation and the Categorical Cross-entropy loss function. The model has totally 5106151 trainable parameters. Training is facilitated using the Adamax optimizer with a learning rate of 0.0012 and a decay rate of 0.0001. Our designed network was trained on a Core (TM) i7-7500U CPU with 16GB of RAM. We have used Tensorflow and Keras API for building/training our neural network model.

D. Continual Learning

When dealing with neural network training scenarios where not all training data is available simultaneously, challenges arise in the realm of deep neural networks [38]. One such challenge is the catastrophic forgetting; which occurs when the network forgets previously learned tasks, attempting to learn new ones. This phenomenon severely impairs network performance. To illustrate, in our research, if the introduced neural network is initially trained to recognize 10 words/signs, achieving an accuracy of around 90%; subsequent training to recognize 10 additional words causes its accuracy (on the previous ten learnt words) to plummet to approximately 10%. Continual learning algorithms have emerged as a solution to address issues like catastrophic forgetting. These algorithms aim to imbue the neural network with the capability to maintain high performance on previously learned tasks while acquiring proficiency in new ones.

In our research, we employed the Elastic Weight Consolidation (EWC) method [39] for the continual learning of the neural network. This method involves a multi-step process. Initially, the neural network undergoes the standard training for the first task. Subsequently, the importance of network weights is computed using the Fisher information matrix. Finally, the network loss function is adjusted using (1), ensuring that during the training for a new task, the weights deemed more crucial for mastering the previous task undergo minimal changes. This preservation of performance on prior tasks is a key objective.

$$L(\theta) = L_t\,(\theta) + \sum_i \frac{\lambda}{2}\, \Omega_i \left(\theta_i - \theta^*_{o,i}\right)^2 \tag{1}$$

In Eq. (1), L(θ) represents the network loss function for future tasks, comprising two components. The first term, denoted as $L_t(\theta)$, pertains to the loss associated with the new task. The second term accounts for the changes in network weights. Ω_i corresponds to the Fisher information matrix for task I; θ_i signifies the modified network weights, and $\theta^*_{o,i}$ represents the trained weights from the previous tasks. The coefficient λ (which is a network hyperparameter) determines the extent to which maintaining performance on prior tasks influences the learning of new tasks. In practical terms, setting this coefficient to zero (0) signifies that the network focuses solely on learning the new task, effectively bypassing the continual learning algorithm during the neural network training. In this research, the value of the lambda parameter is set to 2.

E. Adaptive Teaching Architecture

The primary objective in designing an adaptive teaching architecture for teaching Iranian Sign Language is to enable each user to interact with the system in accordance with their unique performance level, ensuring that the teaching process is fully customized to their capabilities. To achieve this level of personalization, it is essential to maintain separate information for each word/sign and user, allowing the system to adapt to individual needs. Information pertaining to each user is stored within their user profile, which records the performance metrics. Correctly executed words are added to the 'correct words' list, while incorrectly executed words are recorded in the 'incorrect words' list. Each word is assigned a 'word weight', influencing its likelihood of being chosen for the user's practice. Initially, all words are assigned a weight of 1. The 'repetition' metric indicates the average number of attempts required for the user to execute a word correctly. 'Speed' and 'accuracy' represent the average execution speed and accuracy of the user; while the 'user score' is determined on a scale from 0 to 100, considering repetition, speed, and accuracy. Similarly, information related to each word/sign is stored in its respective 'word profile'; with a 'word score' also calculated between 0 and 100, based on frequency, speed, and accuracy.

The overall structure of the adaptive teaching architecture for teaching Iranian Sign Language in this study is depicted in figure 6. The process begins by obtaining the user's name, allowing the 'user identifier' module to access the user-specific profile or create a new one if it is the user's first interaction with the system. Subsequently, the 'word selector' module chooses a word for instruction based on the user's profile. The 'output generator' module determines the training system parameters such as playback speed and repetition count, based on both user and word profiles. The system then presents the correct execution video of the chosen word to the user, adjusting the playback speed as per the 'output generator's' settings. Following this, the user attempts to replicate the sign, and the 'detector' module processes the execution video using the trained model from the neural network section to identify the performed word and assess the execution accuracy. The outcome of this detection process is presented to the user. These steps are repeated according to the repetition count defined by the 'output generator.' The 'user' and 'word updater' modules update the respective profiles based on the user's performance. Finally, the 'training termination' module determines whether to conclude the training process for the user or proceed with teaching the next word. This comprehensive approach outlines the design of the adaptive teaching architecture for our Iranian Sign Language teaching system.

The 'User Identifier' module serves the crucial role of user recognition. It employs the user's name as the criterion for identification. If the user is accessing the system for the first time, this module initiates the creation of a new user profile. The initial user score is set to 50 for newcomers. For returning users, the module locates their existing profile within the system and provides access to it. The 'Word Selector' module operates based on the user profile and is responsible for selecting a word/sign to teach the user. Word selection is carried out randomly from the system's set of words, but it is influenced by each word's weight in the user profile. This approach ensures that words with higher weights are more likely to be chosen. Consequently, the selected word varies for each user, catering to their specific capabilities. The 'Output Generator' module plays a pivotal role in tailoring the system to the user's needs. This module extracts user and word scores from their respective profiles and, using fuzzy logic, determines the word's playback speed and repetition count for that user. Initially, it employs fuzzy membership functions (illustrated in figure 7) to gauge the user's level and word's level.

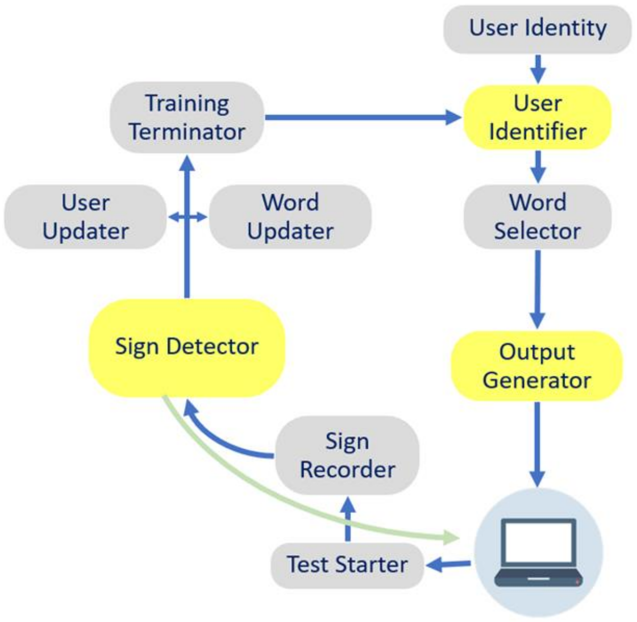
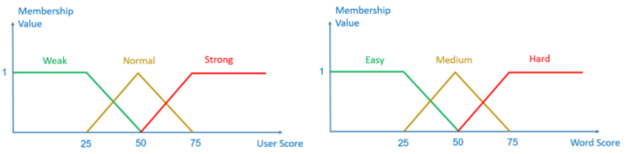


FIGURE 4. ***The designed adaptive teaching structure in this study***

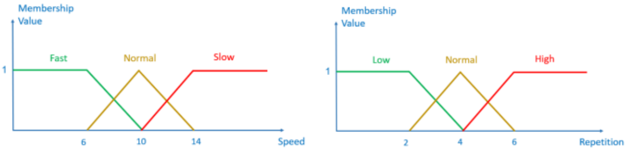


a) User Score

a) Word Score

FIGURE 5. ***Fuzzy membership functions for a) user score, and b) word score***

Once the user and word levels are determined, the module utilizes the fuzzy rule base (illustrated in table 1) for playback speed and repetition count, to make precise calculations.
Subsequently, the module quantifies the qualitative values of the system parameters (i.e., the playback speed and the repetition count) by employing fuzzy membership functions (as shown in figure 8) to map these values to quantitative equivalents.



b) Playback speed

b) Repetition

***FIGURE 6.** Fuzzy membership functions for the a) playback speed, and b) repetition*

The 'Detector' module processes the user's execution video, extracting hand and lip key points, as well as hand connection line length and angle (as explained in the preprocessing section). Using the trained model, this module identifies the performed word and evaluates the execution accuracy.
To improve user experience and enhance system detection accuracy, the module employs the windowing technique. Essentially, the recorded user video is divided into clips, each consisting of 40 frames with a 5-frame stride. The word detection process is conducted on each clip, and the system reports the word with the highest accuracy among all these clips as the final detection result.
Once the training session is completed for each word/sign, the 'User Updater' module updates the user profile based on their performance. It calculates and records the average repetition number, execution speed, and execution accuracy separately in the user profile. Subsequently, it computes the user's score based on these three factors. The score determination utilizes interpolation. A set of diverse executions was assessed by the sign language teaching experts, who assigned scores ranging 0 to 100 for each execution. These expert scores were logged alongside the execution features. An approximation of the experts' scoring method was achieved using a second-order polynomial function.
The module also updates word weights based on the user's correct and incorrect executions of each word/sign. For each correct execution, the word's weight decreases by 0.2, reducing its likelihood of being selected by the 'Word Selector' module. Conversely, for each incorrect execution, the word's weight increases by 0.5, boosting its chance of selection. The 'Word Updater' module operates in a manner analogous to the 'User Updater'. Following the training process for each word, it adjusts the word profile based on the user performance. Word scoring is similarly calculated using a second-order polynomial function, approximated through the interpolation.

***TABLE 1.** Fuzzy rule base for the system parameters*

Speed		Word		
		Easy	Medium	Hard
User	Weak	Normal	Slow	Slow
	Normal	Fast	Normal	Slow
	Strong	Fast	Fast	Normal

a) Playback speed

Repetition		Word		
		Easy	Easy	Easy
User	Weak	Normal	High	High
	Normal	Low	Normal	High
	Strong	Low	Low	Normal

b) Repetition

The 'Training Terminator' module is responsible for deciding whether the training should continue or conclude for the user. If it selects continuation, the 'Word Selector' module proceeds to choose the next word, and the training loop persists. Alternatively, if the 'Training Terminator' module decides on the training termination, the user's training process comes to an end, and the system becomes available for the other users. This module can operate either manually (with human supervision), or automatically after a specified number of word training sessions for the user.

F. Training Scenario

To implement the designed adaptive teaching architecture effectively, a suitable platform for user interaction was needed. Recently, we developed a simple software/app for teaching Iranian Sign Language using the PySimpleGUI library, which offers the capability to create a user-friendly graphical interface [36]. We have improved this app to be used for adaptive teaching of ISL (see figure 9).

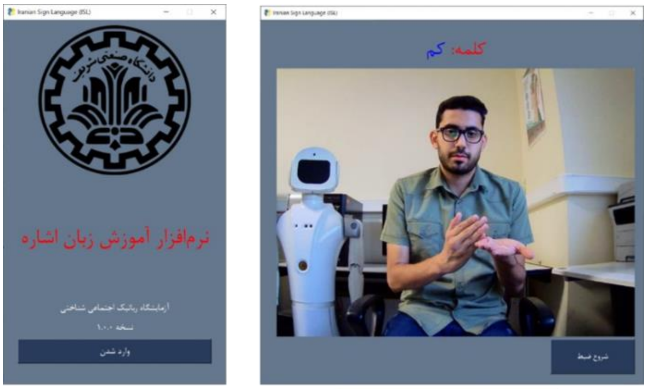

Iranian Sign Language (ISL)
دانشگاه صنعتی شریف
نرم‌افزار آموزش زبان اشاره
آزمایشگاه رباتیک اجتماعی شناختی
نسخه ۱.۰.۰
وارد شدن
Iranian Sign Language (ISL)
کلمه: کم
شروع ضبط

FIGURE 9. Sample snapshots of our improved recent developed software with the new ability of adaptive teaching of ISL

As a preliminary exploratory acceptance study, to assess the performance of the training system, as well as the quality of interaction and adaptation to the user, we designed a scenario for utilizing this system; allowing us to evaluate both the system's performance and the users' satisfaction. According to this scenario, each training session involved teaching the user 5 words/signs through the system. In this research, we trained a group of 20 hearing individuals, comprising 10 men and 10 women aged between 18 and 25 years old, none of whom had any prior familiarity with the Iranian Sign Language. Each user completed 4 training sessions, with the duration of each session varying between 10 to 30 minutes based on the repetition count determined by the system for each word/sign. The training process for individuals involved all participants completing their first training session before proceeding to the second session. At the end of the second and the fourth training sessions, each user completed a questionnaire regarding the system's performance. It is worth noting that all stages of the field study for this research were conducted at the Social and Cognitive Robotics Lab. within the Mechanical Engineering Department at Sharif University of Technology, Iran (see figure 10).

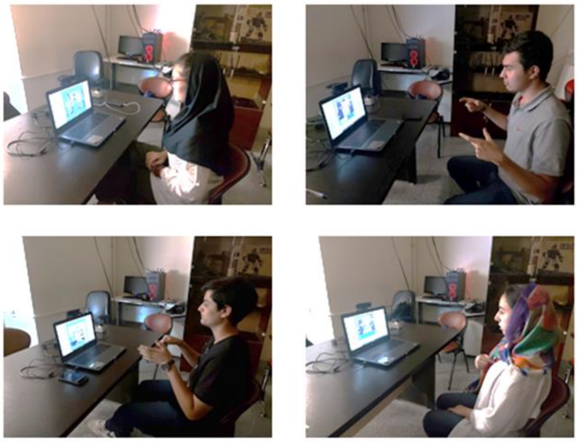


FIGURE 10. The field study environment during the acceptance sessions

G. Evaluation criteria

Of the common evaluation methods in technology-based educational studies is using standard questionnaires to assess the acceptance and adaptability of systems in social and cognitive interactions. Through statistical analysis of users' responses to the questionnaire items, we can not only observe the acceptance of non-human systems, but also evaluate their adaptability to the users. In this study, we designed a questionnaire based on the Unified Theory of Acceptance and Use of Technology (UTAUT) questionnaire with some minor modifications to align with our research objectives [40]. Our used questionnaire assesses seven key parameters: Anxiety, Attitude toward Technology, Facilitating Conditions, Intention to Use, Perceived Adaptiveness, Perceived Ease of Use, and Perceived Usefulness. Respondents answered questions using a 5-point Likert scale [41].

III. Results and Discussion

A. The performance of the designed neural network

The neural network explained earlier was used to learn 101 words/signs from the Iranian Sign Language lexicon involving simultaneous hand and lip motions. This network has the capability to recognize the trained words with an average accuracy of 86.71%. Considering the application intended for this research, the accuracy of the neural network in recognizing sign language words is assessed as good and sufficient.

To examine the catastrophic forgetting phenomenon, the 101 target words were divided into 6 groups. Each group (except the last one) contained 17 words and the last group contained 16 words. Thus, the 101-class classification problem was divided into 6 classification problems with 17 or 16 classes. Each of these new classification problems is called a task. First, the neural network is trained with the training data for task 1 and evaluated on the test data for the same task. The obtained accuracy is reported in the first column of the first row in table 2. Then, the same network is trained using the training data for task 2 and evaluated on the test data for tasks 1 and 2. The obtained accuracy is reported in the first and second columns of the second row in table 2, respectively. This process continues until task 6.

As shown in table 2, after learning an activity, the neural network shows a good performance in recognizing that activity; however, it forgets the activities learned in the past. In such problems, continual learning algorithms will be helpful. To prevent the occurrence of the catastrophic forgetting, the EWC method has been implemented in the neural network. This algorithm was used to learn 6 activities with 16 or 17 total words. The results of the neural network's performance using the continual learning algorithm are presented in table 3. As can be seen, the accuracy of the neural network in recognizing previous activities has had significant growth compared to table 2.

TABLE 2. ***Results of the classification without continual learning algorithms***

		Accuracy					
		1	2	3	4	5	6
Learned Task	1	96.02					
	2	17.92	95.89				
	3	13.56	16.68	90.12			
	4	8.20	13.47	10.48	93.26		
	5	7.63	11.09	8.60	16.71	91.93	
	6	5.11	6.34	4.03	12.40	13.62	90.50

TABLE 3. ***Results of the classification with EWC algorithms (i.e., using continual learning algorithm)***

		Accuracy					
		1	2	3	4	5	6
Learned Task	1	95.98					
	2	84.50	93.27				
	3	83.43	82.39	90.26			
	4	83.11	81.87	81.31	91.02		
	5	82.99	81.53	80.79	82.60	90.42	
	6	82.62	81.34	80.31	82.09	81.97	89.16

The neural network introduced in this study successfully learned 101 words of Iranian Sign Language, encompassing hands and lip movements, across six tasks. Crucially, during each task, the neural network did not have access to the training data from the previous tasks. Nevertheless, it achieved a fairly promising average accuracy of 82.92% in identifying the full set of words at the end of the study. Notably, this recognition accuracy is only 5% lower than when the network learns all 101 words at once. Comparing the obtained accuracy with results reported in the literature utilizing the EWC method for continual learning of the network [39], we indicated that our study may not surpass other continual learning research in terms of recognition accuracy; however, it is important to consider the unique aspects/applications of our research. The specific application, the diversity of words in the dataset, and the relatively small amount of the training and test data in comparison to other studies may make the results of this study unprecedented within its category. It should be noted that due to the lack of similar studies in ISL recognition using continual learning algorithms, (as a limitation,) we could not have a systematic comparison of the obtained results of this research with similar papers in the literature.

B. Results of the experimental study

The designed architecture, according to the explanations provided in the adaptive teaching architecture section, has been used for teaching the selected Iranian Sign Language signs through the developed software. To analyze the performance of the educational system and how it interacts and adapts to the user, the users' responses to the questionnaire were examined. Considering the scores that the users gave to each question, the mean and the standard deviation related to each of the seven parameters were calculated. It should be noted that the reported averages (in table 4) are the mean scores that the users gave to the questions and are reported in the range of 1 to 5. Also, calculating the Cohen's d statistical parameter, the effect size of the users' opinions in the second questionnaire (at the end of the fourth teaching session) compared to the first questionnaire (at the end of the second teaching session) was also examined.

TABLE 4. ***Results of the questionnaires in the experimental study for our 20 participants***

Parameter	Mean (1st questionnaire)	Mean (2nd questionnaire)	Deviation	Cohen's d
Anxiety	1.53	1.45	0.08 decrease	0.09
Attitude Toward Technology	4.04	4.39	0.35 increase	0.47
Facilitating Conditions	4.63	4.51	0.12 decrease	0.21
Intention to Use	3.50	4.08	0.58 increase	0.92
Perceived Adaptiveness	3.63	4.23	0.6 increase	0.63
Perceived Ease of Use	3.75	3.98	0.23 increase	0.27
Perceived Usefulness	4.06	4.39	0.33 increase	0.57

As expected, no noticeable change is observed in the users' anxiety level when learning through the system. Also, the users stated that they experienced very little anxiety when interacting with the system. Users have a very positive attitude towards this app, and this positive attitude has also increased due to the adaptability of the system. Such an attitude can be a good motivation for future activities. Users found the system very easy to use and believed that one could use this system without a prior knowledge. However, there is a slight decrease in the perceived ease of use of the system in the users' opinions, which may be due to the prolonged training time in each session and users' fatigue. Users did not have much motivation at the beginning of working with the system, but after completing the training process through the system, their motivation increased noticeably. The reason for this increase could be the appropriate intelligence level of the system and the attractiveness of using it over time. Analysis of the results conveys that the designed system has been able to attract different users and make the learning process enjoyable for them. The obtained information shows that users initially did not have a specific view on the adaptivity of the teaching algorithm, but over time they well understood this adaptability. These statistics indicate that the educational architecture designed in this research has succeeded in adapting itself to different users in a way that users also understand this adaptability. As can be seen in Table 4, after completing the learning process and more interaction with the system, users felt more comfortable with the system. This positive change indicates the effective communication of the system with the user, which is one of the strengths of the designed educational architecture. According to the users' viewpoints, the use of such intelligent systems for educational purposes and especially for teaching sign language is very useful. After four educational sessions, the users felt this

usefulness more and established more connections with the system. The strength of this feeling as well as its noticeable increase confirms that the designed system has had the necessary capability to effectively communicate with different users and positively influence the teaching process.

Fig. 11 shows the trend of changes in the score of 10 selected words according to the training sessions (considering all of the participants and conducted sessions). It should be noted that the mentioned words/signs were not necessarily performed in all the sessions for all the participants. The score of each word represents its difficulty, from 0 (the easiest) to 100 (the most difficult). Initially, the score of all words is set to 50. The density of the graphs during the first 20 times indicates that initially the score of each word fluctuates around the initial score and the system does not distinguish between words in terms of difficulty. As the number of training sessions increases, the scores of the words become more separated and the system gradually distinguishes between the easy, medium, and difficult words. This process of differentiation, which is called general adaptability, makes it possible for the system to more effectively select training words and session characteristics when faced with a new user. General adaptability can be considered equivalent to the experience of a human teacher in the teaching process. The more the intelligent system resembles a human teacher, the more effective the teaching will be. A comparison of the final results of the system's word scoring with the users' ranking of word difficulty shows that the system correctly identifies the difficulty of the words and adapts itself to their opinion.

Figure 12 shows the relative difficulty of those selected 10 words from the system's and users' perspectives. The results indicate that the order of words from the easiest to the most difficult from the system's perspective largely aligns with the average of users' opinions which indicates that the proposed fuzzy system performed appropriately. The only discrepancy lies in the order of the words "student," "lesson," and "see." The system assigned the same difficulty level to all three words, while users had varying opinions. It is anticipated that this discrepancy will be significantly reduced as the number of users and training sessions increases.

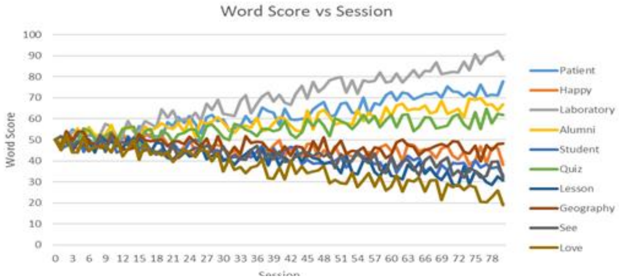
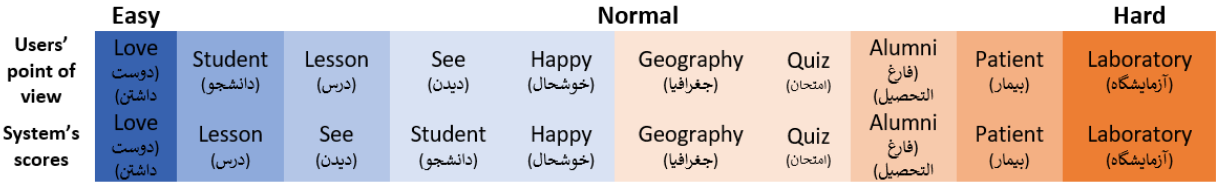


FIGURE 7. *Word score changes during training sessions*

After the end of the training sessions, the users' learning quality was evaluated by a human supervisor, and all users were able to perform well on 82% of the words they were taught through the system. The more the training sessions and the number of users, the better performance in the human assessment is expected.

The adaptive teaching architecture designed in this study succeeded in gaining high scores from users in the evaluated parameters. Most importantly, during our preliminary exploratory study, this architecture interacted well with different users and in just four training sessions, adapted itself to them in such a way that in addition to effective learning of sign language, users are more motivated to interact with the intelligent teaching system.

Compared to the most related research done in this field [34], the performance of the results obtained in the present study was obviously better in the Attitude Toward Technology (Cohen's d: 0.47 vs. 0.38), Intention to Use (Cohen's d: 0.92 vs. 0.27), and Perceived Usefulness (Cohen's d: 0.57 vs. 0.09) items; which makes this study a step forward in improving the quality of adaptive intelligent teaching systems.



	Easy				Normal					Hard
Users' point of view	Love (دوست داشتن)	Student (دانشجو)	Lesson (درس)	See (دیدن)	Happy (خوشحال)	Geography (جغرافیا)	Quiz (امتحان)	Alumni (فارغ التحصیل)	Patient (بیمار)	Laboratory (آزمایشگاه)
System's scores	Love (دوست داشتن)	Lesson (درس)	See (دیدن)	Student (دانشجو)	Happy (خوشحال)	Geography (جغرافیا)	Quiz (امتحان)	Alumni (فارغ التحصیل)	Patient (بیمار)	Laboratory (آزمایشگاه)

FIGURE 8. *The comparison of words difficulty ranking: the average of users' viewpoints vs. the system's scores. It should be noted that the equivalent Persian words are shows in the figure*

Regarding the Perceived Adaptiveness item, both of the studies reach the Cohen’s d effect size more than 0.6 (with the mean score of 4.23 vs. 3.73) which shows that the used fuzzy algorithm for adaptive teaching of ISL signs work appropriately (especially in this study) while the adaptiveness of teaching is perceived by the users.

IV. Limitations and Future Works

Despite continuous efforts to implement the adaptive teaching system on the Rasa robot, which is the robot available in the Social Robotics Laboratory with the capability of finger movement, unfortunately, due to some hardware challenges and limitations, this possibility has not been realized. It is anticipated that if this system is implemented on social robots, it will have a more significant impact on users and enhance the quality of their learning experience.

Since linear fuzzy membership functions are selected to ensure feasibility in real-time teaching scenarios, the app's initial assessments may be overly "sharp." This results in drastic changes in the system's teaching behavior during the initial sessions. However, this issue is resolved after a few sessions, as the system collects sufficient data to adjust its output parameters more logically.

The results of this study can be used to improve the performance of recognizing sign language words by using multiple continual learning methods and developing neural networks capable of gradual learning and simultaneous recognition of different sign languages. Also, the adaptive teaching architecture designed in this study can be used to teach other concepts and engage users' emotions during interaction with the system.

V. Conclusion

The primary goal of the present research was to teach Iranian Sign Language, focusing on the utilization of continual learning algorithms and the adaptability of the training system/app to users. Data collection for this study was conducted in collaboration with the Social Robotics Laboratory at Sharif University of Technology and the Islamic Azad University, Fereshtegan Branch, Iran. The implementation of the preprocessing methods (i.e., our feature extraction methods) has empowered the trained neural network in this study to independently recognize sign language words/signs, irrespective of people's faces or the surrounding environment. This aspect significantly enhances the system's robustness and performance. Furthermore, the neural network structure incorporates transformer modules in conjunction with the continual learning algorithms. This innovation allows the sign language teaching system to gradually expand its vocabulary over time while providing adaptive training to the users. The adaptive teaching architecture developed in this research not only promotes more effective interaction between intelligent systems and users, but also represents a significant stride toward enhancing human-machine interaction during a preliminary exploratory study. In conclusion, it is anticipated that this research will pave the way for the widespread adoption of intelligent systems and social robots in the realm of human-robot/computer interaction. Within this context, emerging and advanced technologies will be harnessed to enhance the overall quality of life for all individuals. It should be noted that, due to the absence of similar studies on ISL recognition using continual learning algorithms, we were unable to systematically compare our research results with those of other papers in the literature.

ACKNOWLEDGMENT

The funding for this study was provided by the “Iranian National Science Foundation (INSF)” (http://en.insf.org/) and “Dr. Ali Akbar Siassi Memorial Research Grant Award”. We would like to express our gratitude to our friends Mr. Ali Ghadami and Mr. Hossein Ranjbar for their cooperation and helps with the data collection, as well as all the participants. We also thank Dr. Mehran Memari for the English editing of the final manuscript. In order to reduce the ambiguity of our writing, we also used ChatGPT for paraphrasing some sentences of the manuscript.

Conflict of interest

Author Alireza Taheri has received a research grant from the “Iranian National Science Foundation (INSF)” (http://en.insf.org/) (Grant No. 4031030)”. The author Morteza Memari declares that he has no conflict of interest.

Availability of data and material (data transparency)

All data from this project (the collected dataset on ISL, photos and videos of the sessions, results of the questionnaires, etc.) are available in the Social & Cognitive Robotics Laboratory archive.

Code availability

All of the robots’ codes are available in the archive of the Social & Cognitive Robotics Laboratory.

Authors’ contributions

Both authors contributed to the study's conception and design. Material preparation, data collection, and analysis were performed by Morteza Memari. Alireza Taheri supervised this research. The first draft of the manuscript was written by Morteza Memari; and both authors commented on previous versions of the manuscript. Both authors read and approved the final manuscript.

Ethical Approval

Ethical approval for the protocol of this study was provided by the IPM - Institute for Research in Fundamental Sciences.

Consent to participate

Informed consent was obtained from all individual participants included in the study.

Consent for publication

The authors affirm that human research participants provided informed consent for publication of all images. All of the participants have consented to the submission of the results of this study to the journal.

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3492056

REFERENCES

1. Taheri, A., et al. *Investigating the Impact of Human-Robot Collaboration on Creativity and Team Efficiency: A Case Study on Brainstorming in Presence of Robots*. in *International Conference on Social Robotics*. 2023. Springer.
2. Amiri, O., et al. *Virtual Reality Serious Game with the TABAN Robot Avatar for Educational Rehabilitation of Dyslexic Children*. in *International Conference on Social Robotics*. 2023. Springer.
3. Jebellat, I., et al., *A reinforcement learning approach to find optimal propulsion strategy for microrobots swimming at low reynolds number.* Robotics and Autonomous Systems, 2024: p. 104659.
4. Chen, L., P. Chen, and Z. Lin, *Artificial intelligence in education: A review.* Ieee Access, 2020. **8**: p. 75264-75278.
5. Brusilovsky, P., *AI in Education, Learner Control, and Human-AI Collaboration.* International Journal of Artificial Intelligence in Education, 2023: p. 1-14.
6. Debnath, J. and P.J. IR. *Real-Time Gesture Based Sign Language Recognition System*. in *2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS)*. 2024. IEEE.
7. Sindhu, K.S., et al. *Sign Language Recognition and Translation Systems for Enhanced Communication for the Hearing Impaired*. in *2024 1st International Conference on Cognitive, Green and Ubiquitous Computing (IC-CGU)*. 2024. IEEE.
8. Martins, G.S., L. Santos, and J. Dias, *User-adaptive interaction in social robots: A survey focusing on non-physical interaction.* International Journal of Social Robotics, 2019. **11**: p. 185-205.
9. Rossi, S., F. Ferland, and A. Tapus, *User profiling and behavioral adaptation for HRI: A survey.* Pattern Recognition Letters, 2017. **99**: p. 3-12.
10. Norcio, A.F. and J. Stanley, *Adaptive human-computer interfaces: A literature survey and perspective.* IEEE Transactions on Systems, Man, and cybernetics, 1989. **19**(2): p. 399-408.
11. McTear, M.F., *User modelling for adaptive computer systems: a survey of recent developments.* Artificial intelligence review, 1993. **7**: p. 157-184.
12. Aly, A. and A. Tapus. *Towards enhancing human-robot relationship: customized robot's behavior to human's profile*. in *2014 AAAI Fall Symposium Series*. 2014.
13. Westlund, J.K., et al., *Learning a second language with a socially assistive robot.* Almere, The Netherlands, 2015.
14. Gordon, G. and C. Breazeal. *Bayesian active learning-based robot tutor for children's word-reading skills*. in *Proceedings of the AAAI Conference on Artificial Intelligence*. 2015.
15. Aylett, R., et al. *I know how that feels—An empathic robot tutor*. in *eChallenges e-2015 Conference*. 2015. IEEE.
16. Tozadore, D., et al., *Project r-castle: Robotic-cognitive adaptive system for teaching and learning.* IEEE Transactions on Cognitive and Developmental Systems, 2019. **11**(4): p. 581-589.
17. Tapus, A., C. Ţăpuş, and M.J. Matarić, *User—robot personality matching and assistive robot behavior adaptation for post-stroke rehabilitation therapy.* Intelligent Service Robotics, 2008. **1**: p. 169-183.
18. Dahanayaka, D., B. Madhusanka, and I. Atthanayake, *A multi-modular approach for sign language and speech recognition for deaf-mute people.* Engineer, 2021. **97**: p. 1.
19. Yousaf, K., et al., *A Novel Technique for Speech Recognition and Visualization Based Mobile Application to Support Two-Way Communication between Deaf-Mute and Normal Peoples.* Wireless Communications and Mobile Computing, 2018. **2018**(1): p. 1013234.
20. Liu, Y., F. Jiang, and M. Gowda, *Finger gesture tracking for interactive applications: A pilot study with sign languages.* Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2020. **4**(3): p. 1-21.
21. Starner, T., J. Weaver, and A. Pentland, *Real-time american sign language recognition using desk and wearable computer based video.* IEEE Transactions on pattern analysis and machine intelligence, 1998. **20**(12): p. 1371-1375.
22. Kishore, P., et al. *4-Camera model for sign language recognition using elliptical fourier descriptors and ANN*. in *2015 International Conference on Signal Processing and Communication Engineering Systems*. 2015. IEEE.
23. Agarwal, A. and M.K. Thakur. *Sign language recognition using Microsoft Kinect*. in *2013 sixth international conference on contemporary computing (IC3)*. 2013. IEEE.
24. Zahedi, M. and A.R. Manashty, *Robust sign language recognition system using ToF depth cameras.* arXiv preprint arXiv:1105.0699, 2011.
25. Potter, L.E., J. Araullo, and L. Carter. *The leap motion controller: a view on sign language*. in *Proceedings of the 25th Australian computer-human interaction conference: augmentation, application, innovation, collaboration*. 2013.
26. Oz, C. and M.C. Leu, *American sign language word recognition with a sensory glove using artificial neural networks.* Engineering Applications of Artificial Intelligence, 2011. **24**(7): p. 1204-1213.
27. Brashear, H., et al. *Using multiple sensors for mobile sign language recognition*. in *Seventh IEEE International Symposium on Wearable Computers, 2003. Proceedings.* 2003. Citeseer.
28. Izzah, A. and N. Suciati, *Translation of sign language using generic fourier descriptor and nearest neighbour.* International Journal on Cybernetics and Informatics, 2014. **3**(1): p. 31-41.
29. Ye, J., H. Yao, and F. Jiang. *Based on HMM and SVM multilayer architecture classifier for Chinese sign language recognition with large vocabulary*. in *Third International Conference on Image and Graphics (ICIG'04)*. 2004. IEEE.
30. Cui, R., H. Liu, and C. Zhang, *A deep neural framework for continuous sign language recognition by iterative training.* IEEE Transactions on Multimedia, 2019. **21**(7): p. 1880-1891.
31. Kumar, B.P. and M. Manjunatha, *A hybrid gesture recognition method for American sign language.* Indian Journal of Science and Technology, 2017. **10**(1): p. 1-12.
32. Hosseini, S.R., et al. *Teaching persian sign language to a social robot via the learning from demonstrations approach*. in *Social Robotics: 11th International Conference, ICSR 2019, Madrid, Spain, November 26–29, 2019, Proceedings 11*. 2019. Springer.
33. Basiri, S., et al., *Dynamic iranian sign language recognition using an optimized deep neural network: an implementation via a robotic-based architecture.* International Journal of Social Robotics, 2023. **15**(4): p. 599-619.
34. Basiri, S., et al., *Design and implementation of a robotic architecture for adaptive teaching: A case study on Iranian sign language.* Journal of Intelligent & Robotic Systems, 2021. **102**(2): p. 48.
35. Redmon, J., et al. *You only look once: Unified, real-time object detection*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
36. Ali Ghadami, A.T., Ali Meghdari, *Developing a vision-based system for continuous translation of Iranian Sign Language*, in *Mechanical Engineering*. 2023, Sharif University of Technology.
37. Lugaresi, C., et al., *Mediapipe: A framework for building perception pipelines.* arXiv preprint arXiv:1906.08172, 2019.
38. Soni, M. and M.A. Shnan, *Scalable Neural Network Algorithms for High Dimensional Data.* Mesopotamian Journal of Big Data, 2023. **2023**: p. 1-11.
39. Kirkpatrick, J., et al., *Overcoming catastrophic forgetting in neural networks.* Proceedings of the national academy of sciences, 2017. **114**(13): p. 3521-3526.
40. Mazhari, A., P. Esfandiari, and A. Taheri. *Teaching Iranian sign language via a virtual reality-based game*. in *2022 10th RSI International Conference on Robotics and Mechatronics (ICRoM)*. 2022. IEEE.
41. Marschark, M. and P.C. Hauser, *How deaf children learn: What parents and teachers need to know*. 2012: OUP USA.

Morteza Memari was received the B.S. and M.S. degrees in mechanical engineering from Sharif University of Technology, Tehran, in 2023. He is now a Ph.D. candidate in mechanical engineering in Sharif University of Technology, Iran since Fall 2023. From 2023, he was a Research Assistant with the Social and Cognitive Robotics Laboratory. Also, he was a teaching assistant in more than 5 undergraduate and graduate courses in the field of robotics, AI, and control. His research interests include social robotics, artificial intelligence, human-robot interaction, and control theory. Mr. Memari was a recipient of the best Mechanical Engineering B.S. Project Award in 2021, and he is a member of Center of Excellence in Design, Robotics, and Automation (CEDRA).

Alireza Taheri is an Associate Professor of Mechanical Engineering with an emphasis on Social and Cognitive Robotics at Sharif University of Technology, Tehran, Iran. He is the Head of the Social and Cognitive Robotics Lab. and the Measurement Systems Lab. at Sharif University of Technology. The line of his research focuses on designing/using Social and Cognitive Robotics, Virtual Reality Systems, and Human-Robot Interaction (HRI) platforms for education and rehabilitation of children with special needs (e.g. children with autism, children with hearing problems, children with cerebral palsy). His researches include robots' design and fabrication, serious games' design, artificial intelligence and control, conducting educational/clinical interventions for children, developing cognitive architectures for social robots, mathematical modeling of participants' behaviors during HRI, and empowering robots to analyze users' behaviors automatically and then react adaptively.