Learning signs with NAO: humanoid robot as a tool for helping to learn Colombian Sign Language

Juan E. Mora-Zarate[1], Claudia L. Garzón-Castro[1]* and Jorge A. Castellanos Rivillas[2]

[1]Engineering Faculty, Research Group CAPSAB, Universidad de La Sabana, Chía, Colombia,
[2]Engineering Faculty, Universidad de La Sabana, Chía, Colombia

Sign languages are one of the main rehabilitation methods for dealing with hearing loss. Like any other language, the geographical location will influence on how signs are made. Particularly in Colombia, the hard of hearing population is lacking from education in the Colombian Sign Language, mainly due of the reduce number of interpreters in the educational sector. To help mitigate this problem, Machine Learning binded to data gloves or Computer Vision technologies have emerged to be the accessory of sign translation systems and educational tools, however, in Colombia the presence of this solutions is scarce. On the other hand, humanoid robots such as the NAO have shown significant results when used to support a learning process. This paper proposes a performance evaluation for the design of an activity to support the learning process of all the 11 color-based signs from the Colombian Sign Language. Which consists of an evaluation method with two modes activated through user interaction, the first mode will allow to choose the color sign to be evaluated, and the second will decide randomly the color sign. To achieve this, MediaPipe tool was used to extract torso and hand coordinates, which were the input for a Neural Network. The performance of the Neural Network was evaluated running continuously in two scenarios, first, video capture from the webcam of the computer which showed an overall F1 score of 91.6% and a prediction time of 85.2 m, second, wireless video streaming with NAO H25 V6 camera which had an F1 score of 93.8% and a prediction time of 2.29 s. In addition, we took advantage of the joint redundancy that NAO H25 V6 has, since with its 25 degrees of freedom we were able to use gestures that created nonverbal human-robot interactions, which may be useful in future works where we want to implement this activity with a deaf community.

KEYWORDS

Colombian Sign Language (CSL), neural networks, machine learning, social robots, education, human-robot interaction (HRI)

1 Introduction

People with disabilities constitute 16% of the global population, with 31% experiencing hearing problems, which significantly impacts their quality of life (World Health Organization, 2023; World Health Organization, 2024). In Colombia, where around 483000 people have hearing impairments, access to education is limited, with only
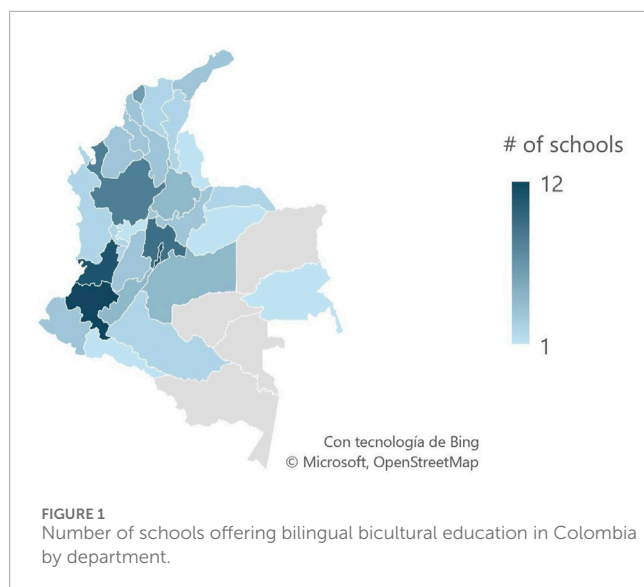
FIGURE 1
Number of schools offering bilingual bicultural education in Colombia by department.

TABLE 1 Recognition accuracies for sign language detection techniques.

Technique	Sign Language	Recognition accuracy	References
Computer Vision + Machine Learning	Japanese	96%–99%	Kondo et al. (2024)
Computer Vision + Machine Learning	Indonesian	94%–96%	Ilham et al. (2024)
Sensory Glove + Machine Learning	American	94%	Pal et al. (2024)
Sensory Glove + Machine Learning	Chinese	97%–98%	Ji et al. (2023)

one Colombian Sign Language (CSL) interpreter for every 1,152 deaf individuals and just 36% of interpreters working in basic education (Instituto Nacional Para Sordos, 2020). Moreover, out of more than 75,000 educational institutions in the country, only 109 offer bilingual and bicultural education, and some regions lack from CSL institutions as seen in Figure 1 (Instituto Nacional Para Sordos, 2024).

To address this issue, data gloves and computer vision (CV) technologies have been the most frequent developed for sign language recognition, achieving performance levels above 90% as the recent studies in Table 1 show. However, both have their limitations: data gloves cannot predict the position of the hands relative to the body (Liang et al., 2024), while CV approaches require high computational resources (González-Rodríguez et al., 2024). Sign language recognition technologies have also been integrated into sign language translation systems (Teran-Quezada et al., 2024; Jintanachaiwat et al., 2024; Shahin and Ismail, 2024) and
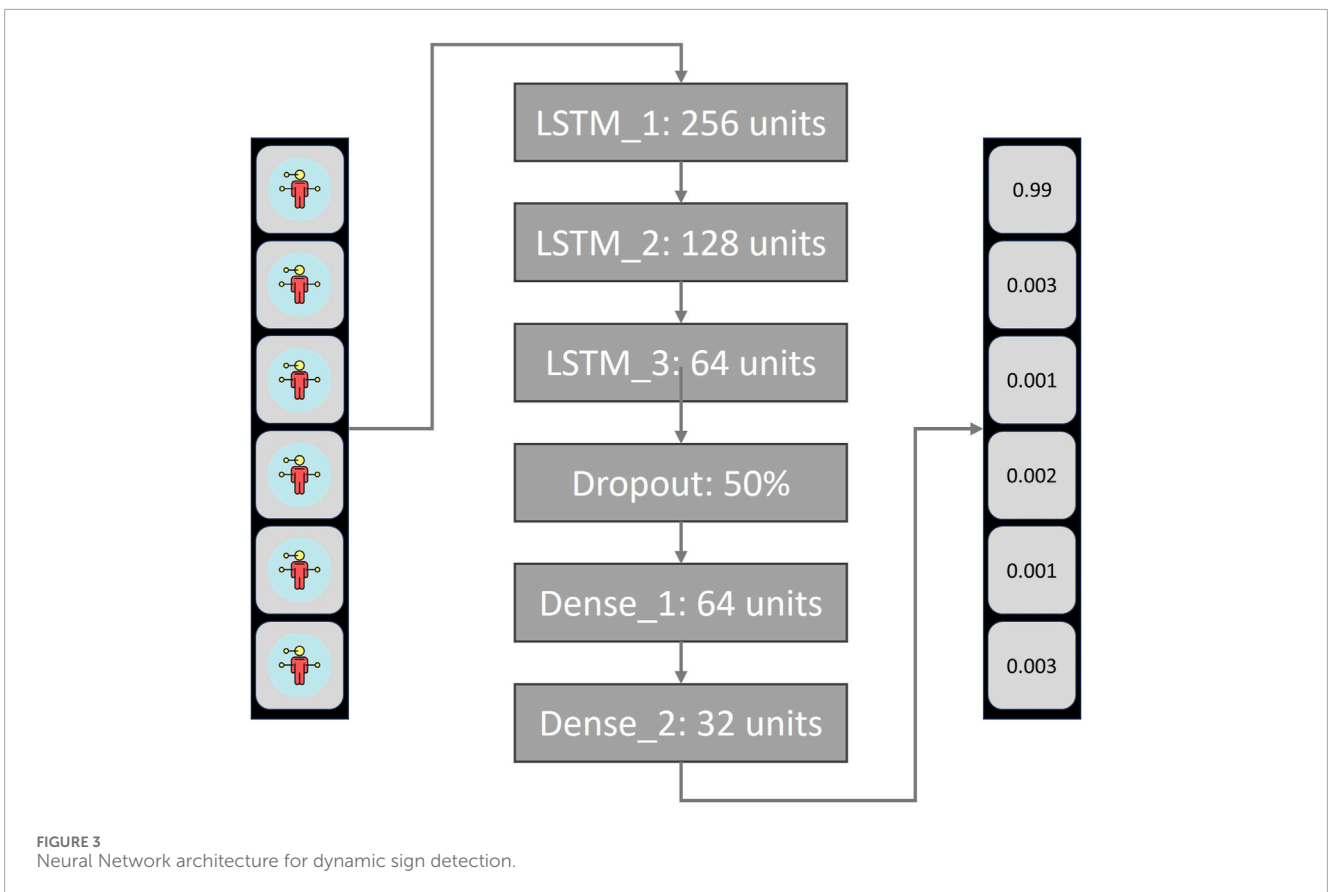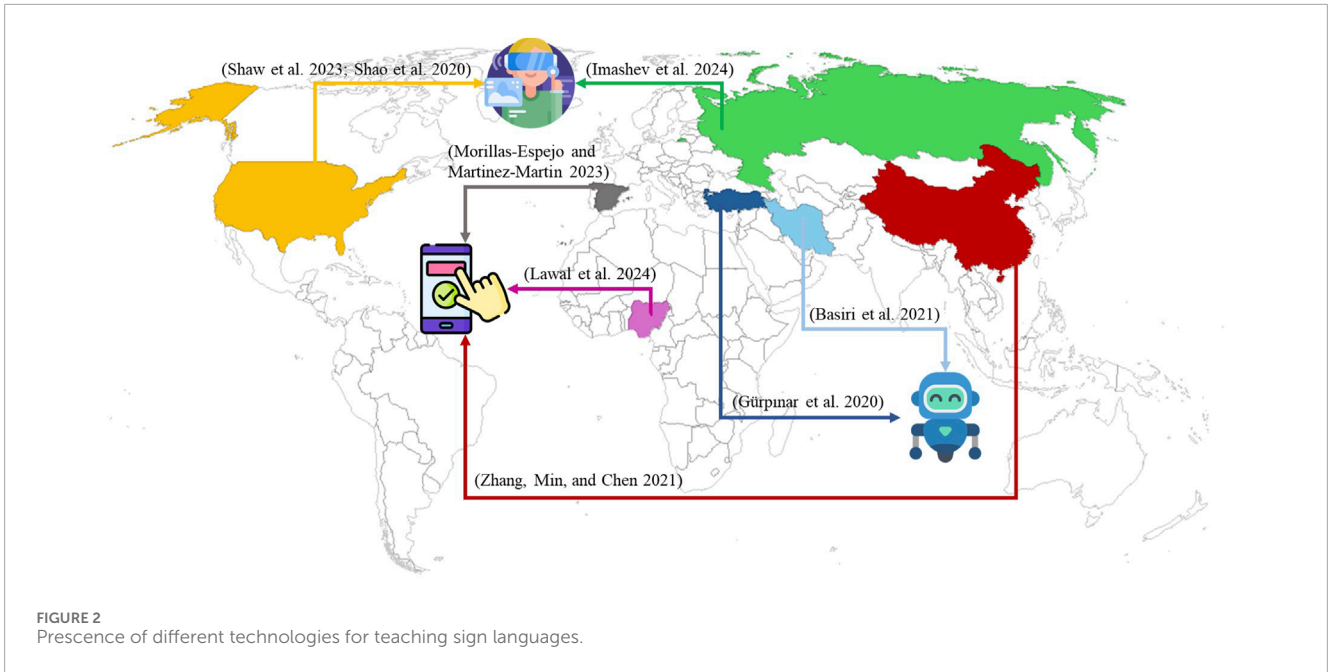
educational tools such as social robots (Gürpınar et al., 2020), mobile applications (Zhang et al., 2021), and Virtual Reality (Shao et al., 2020), all of which have proven effective in enhancing sign language learning and enabling these tools to be adopted by some other sing languages. However, as shown in Figure 2, the development and implementation of these technologies are concentrated in less than the 4% of the countries. Despite the promising results that may help to mitigate the gap in access to education for the deaf population, progress remains scarce in countries like Colombia. Works such as those of Pereira-Montiel et al. (2022); Gonzalez et al. (2024) are the most recent in the translation of CSL sentences and words, but there is still a gap on developing educational tool for this sign language.

In Colombia, early taught vocabulary in CSL includes topics like greetings, pronouns, numbers, and colors (Universidad de Bogota Jorge Tadeo Lozano, 2024; Universidad de Los Andes, 2024; Universidad ECCI, 2024). Therefore, this study proposes a performance evaluation for an educational tool capable of recognize all the 11 color-based signs in the CSL. To achieve this, we utilized the humanoid robot NAO H25 V6, which features 25 degrees of freedom and can engage in human-like interactions through the execution of gestures. Additionally, the education tool is enhanced with RGB eye LEDs and tactile buttons to provide both visual and physical feedback. This setup may help learners quickly assess their performance during the activity and maintain their engagement. Furthermore, NAO has shown effective as an educational tool that engages users through interactive experiences (Kurtz and Kohen-Vacs, 2024; Cruz-Ramírez et al., 2022; Mutawa et al., 2023). The proposed sign recognition system consists of the design and training of a Recurrent Neural Network (RNN), that uses as input hand and torso landmarks extracted with MedaPipe from a parallel video capture process. When validating the Neural Network (NN) performance, was found an F1-Score above the 90%, however, was also identified, that the latency needs to be improved when continuous sign recognition from the NAO H25 V6 camera. As future work, the range of sign topics recognized by the educational tool will be expanded to include a broader selection of early-taught vocabulary in CSL.

2 Materials and methods
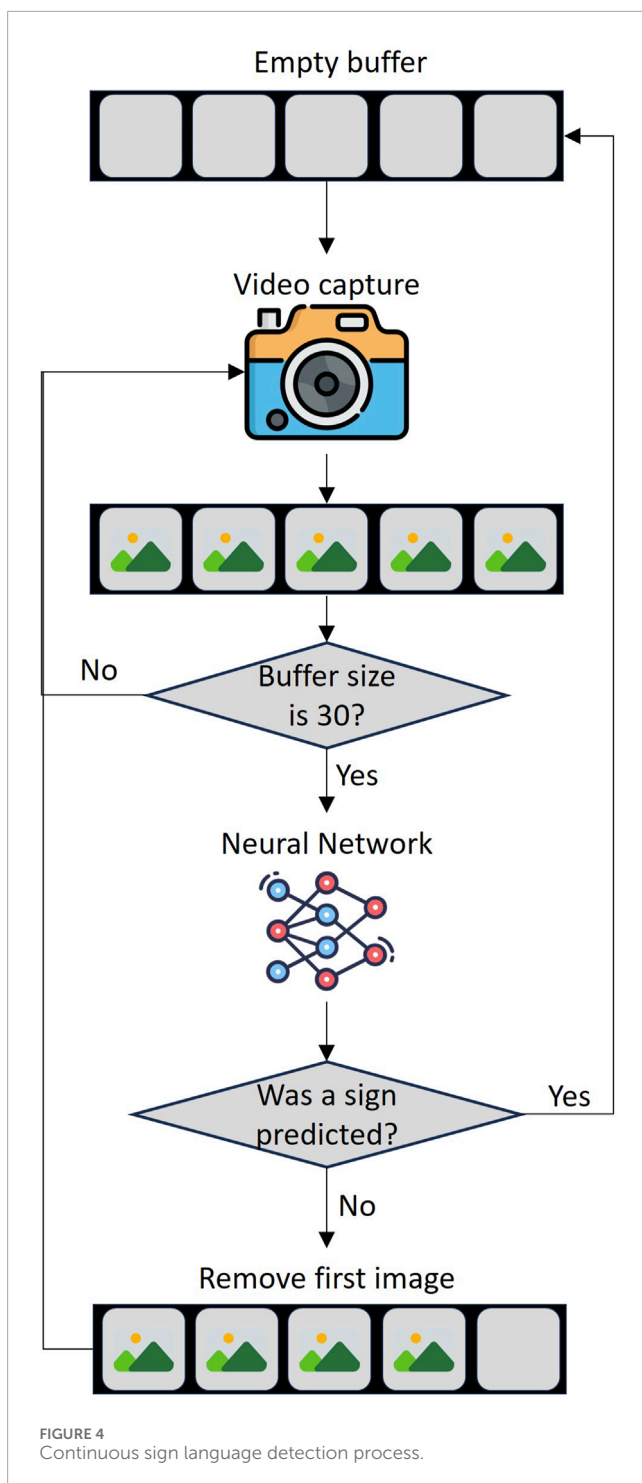
2.1 Neural network design

A RNN of the Long-Term Short-Term Memory (LSTM) type was built, because when working with time-dependent information, such as dynamic or moving signs, an architecture that preserves information from the past is needed to make the classifications (See Figure 3). The NN architecture consists of a sequential model with three stacked LSTM layers of 256, 128, and 64 units, respectively, using ReLU activation, followed by a dropout layer (0.5). It then has three dense layers with 64, 32, and 11 units, where the final layer uses SoftMax activation for multi-class classification, finally, when training, a stochastic gradient descent method was used for the optimization (Adam) and a loss function for multi-class classification models (categorical cross entropy) where used. The NN obtained an F1-Score of 92% and was evaluated under a training-test split in the ratio 80–20, with a dataset for the CSL colors (white, brown, gray, purple, orange, black, red, pink and green). The data was obtained from videos of 20 signers with a length

FIGURE 2
Prescence of different technologies for teaching sign languages.

FIGURE 3
Neural Network architecture for dynamic sign detection.

of 30 frames and were post-processed to extract with the MediaPipe tool the coordinates of the torso and right hand. It is important to clarify that the training and test set did not share people, to avoid biases when training and to guarantee that the NN was able to detect signs of new users.

2.2 Sign detection system

The sign detection system operates in a Python 3.12.3 script that runs on a laptop equipped a Core i7 8665U processor and 8 GB of RAM was made. The code has two simultaneous processes, the first

FIGURE 4
Continuous sign language detection process.

capture video by a designed activity (See Section 2.4) and send it to the Python 3.12.3 script via TCP/IP communication.

2.3 Neural network evaluation

When evaluating a classification algorithm that is going to run continuously, it is crucial to assess both its speed and accuracy. For speed, timestamps were used when the buffer fills and empties. To evaluate accuracy, three tests performing each sign 10 time with a rest between signals of 1–2 s was made. The choice of 3 experiments with a sample size of 10 each was based on the binomial proportion formula, considering an expected 5% performance drop from results on the testing data and a theorical performance of 99% due to the higher results found in literature. Those considerations led to a calculated sample size of 29, that was divided into three experiments to prevent user fatigue. Additionally, the test was performed ensuring that the user was framed as shown in the first column of Figure 5, since the NN was trained under these same conditions. Also, there was a cold Light Emitting Diode (LED) lightning in a room where no direct light from outside enters, to ensure that the MediaPipe tool would always detect the coordinates of the torso and hand.

2.4 Actions performed by NAO H25 V6

Unlike robotic platforms such as mobile or industrial ones, social robots such as the NAO H25 V6 robot (SoftBank Group Corp., Japan) manage to have more friendly interactions with humans, because they can give the appearance that they are expressing an emotion (Hellou et al., 2023). For this reason, the NAO H25 V6 was controlled by means of a Python 2.7.18 code, which with the import of the SDK 2.8.6 accessed the modules shown below:

- ALLeds: Control of LED's
- ALVideoRecorder: Video recording
- ALTouch: Reading capacitive sensors
- ALRobotPosture: Set predefined positions
- ALAnimationPlayer: Execute predefined movements
- ALMotion: Joint control

The modules allowed the development of an activity as described below (see Figure 6).

1. Starting position: The NAO H25 V6 is placed in a standing position and turns its head forward up to 4.5°. This is because the robot is located on a table of 1.1 m high and the user is at approximately 1.5 m.
2. Start of the activity: The user can choose between 2 evaluation modes. The first one will be activated by pressing the first touch button on the head of the NAO H25 V6. This will allow to select any of the 11 colors by means of a touch menu that will change selection each time the first button on the head of the robot is pressed. The second mode will be activated by pressing the second button on the head of the NAO H25 V6, and the color will be randomly selected. Once the color is selected, the robot greets with the gesture "animations/Stand/Gestures/Hey_1"and lights up its eyes in the color of the sign that should be made.

one is capturing a video signal from the integrated webcam of the laptop which it will take images every 4 frames and then by means of the MediaPipe tool create a list with the 54 coordinates of the torso. The second process fills a data buffer with the list created in the first process, when the buffer reaches a size of 30 lists, it enters the NN to make the prediction, if it predicts with a tolerance higher than 99%, the buffer is emptied because a word has already been detected (see Figure 4). On the other hand, when capturing video from the NAO H25 V6, the processes will not run simultaneously, it will first
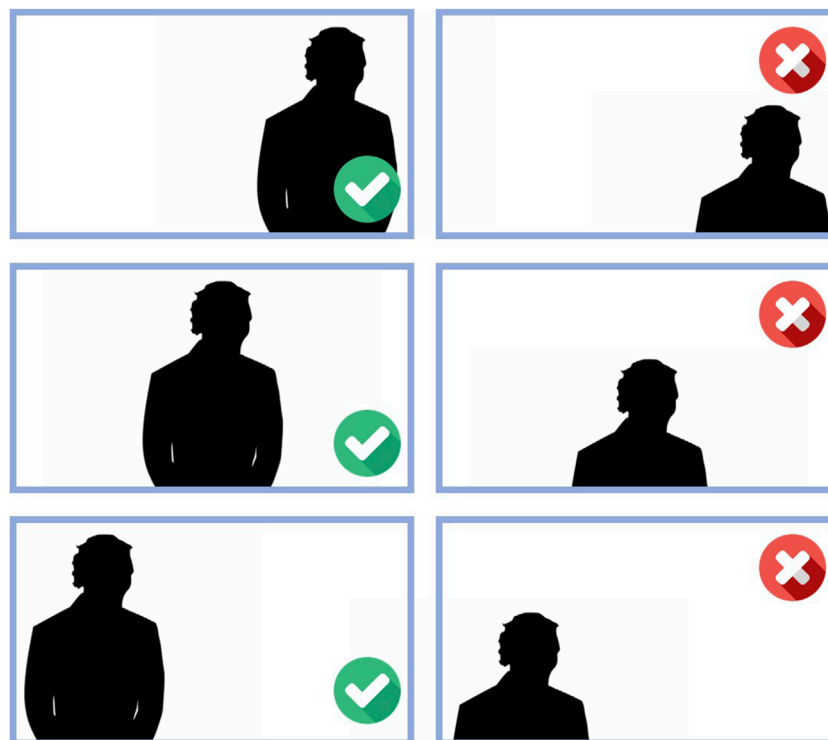
FIGURE 5
Description of how the user should be in frame for the sign detection process.

3. Recording of the signal: Recording is at a rate of 8 Frames Per Second (FPS) for 4 s and starts after the LEDs of the robot's eyes blink 3 times, once in the recording state the eight LEDs of each eye are turn on and will be turn off progressively as the 4 s of recording are over, thus giving feedback to the user of the remaining time.

4. Prediction of the signal: The video is stored in the NAO H25 V6 memory and is sent via TCP/IP to a second code in Python 3.12.3 that is responsible for making the prediction and respond with the color of the executed signal.

5. Result: With the received message, the robot lights up its eyes with the color that the user made. In addition, if the signer made the sign properly, the NAO H25 V6 affirms by making the gesture "animations/Stand/Gestures/Yes_ 2", or otherwise, denies by making the gesture "animations/Stand/Gestures/No_9".

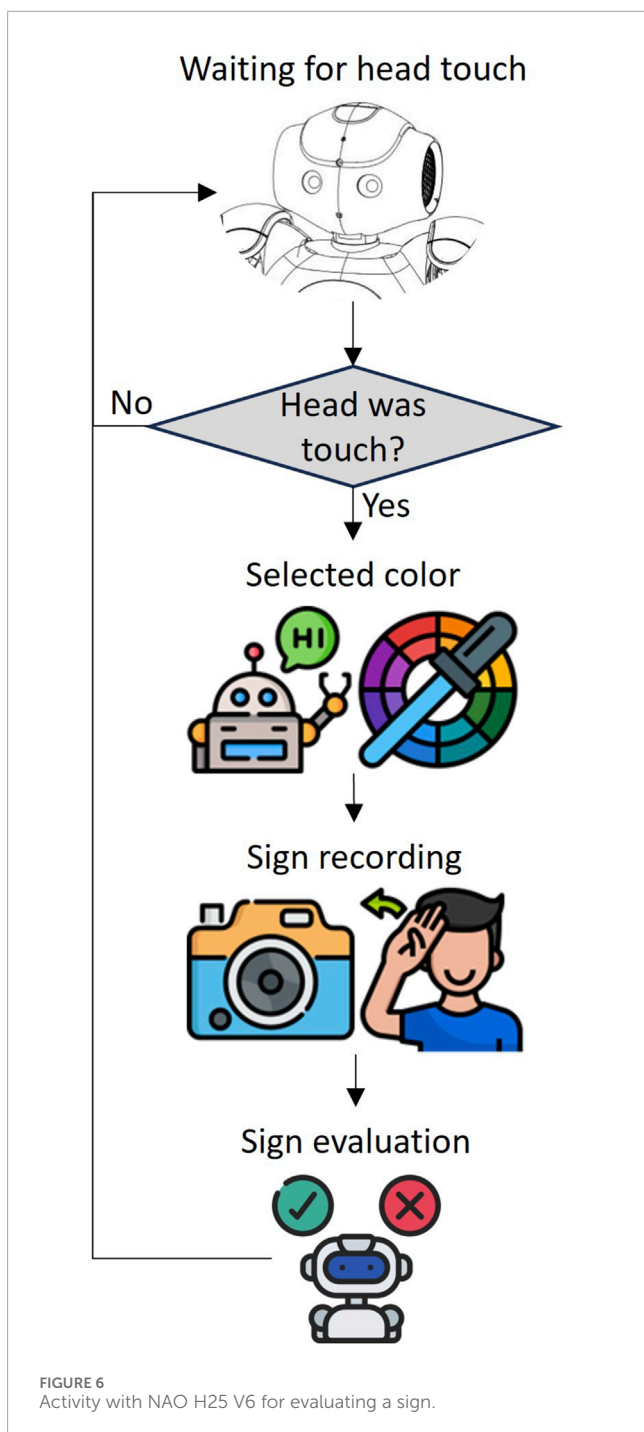# 3 Results

3.1 Neural network evaluation

As mentioned in the methodology, each class is considered as a binomial problem in which three tests with a sample size of 10 were conducted on the NN to evaluate its performance. The results for accuracy are shown in Figure 7, where the confusion matrix illustrates that most predictions fall within the true positive and true negative zones. The lowest data point in this zone is 7, corresponding

to a categorical accuracy of 70% for that class. Also, Figure 8 displays the average metrics across the three tests, with the purple sign achieving a F1-Score of 100%, which means it was always recognized, and no sign was confused with it. The F1 scores for other signs range from 98.3% to 83.1%, resulting in overall metrics of 91.6% for the F1 score, 92.6% for recall, and 91.5% for accuracy. Additionally, the precision is more affected than recall when the F1-score decreases, suggesting that the NN has more difficulty with precision than recall.

On the other hand, Figure 9A shows the time required by the algorithm to process each image using MediaPipe and send it to the sign prediction process, with an average time of 139.8 m per image for 150 frames. This translates to a frame rate of 28.6 FPS, which represents only a 4.7% reduction compared to the camera's capture rate of 30 FPS. Additionally, Figure 9A displays the time it takes for the algorithm to send data from the first process and capture it in the second, averaging 0.2 m between data emission and reception, ensuring no synchronization issues between processes. Finally, Figure 9B shows the average prediction time of 85.2 m for 27 predictions, indicating that processing the 30-frame buffer is faster than the rate of incoming data.

3.2 Activity with the robot NAO H25 V6

Five trials were conducted for each color sign, with the only prediction error occurring for the pink sign, which, as shown in the confusion matrix in Figure 10, was confused twice with

FIGURE 6
Activity with NAO H25 V6 for evaluating a sign.

4 Discussion

This paper shows the use of the NAO robotic platform developing that could supports de CSL learning. For it, a NN that executes on a computer was design for running continuously in two different scenarios, video capture from the integrated computer webcam and wireless streaming from the NAO H25 V6 camera. The NN detection capability and latency were evaluated, finding promising results in sign detection accuracy for both scenarios, and low latency for the webcam scenario. However, latency remains a limitation when using the NAO H25 V6, as it is 2.2 s slower than the webcam scenario. This issue arises due to the MediaPipe preprocessing, which is not done in parallel with the video capture, since the sign detection script only processes images after recording is completed. To address this, future works, should aim to run all operations in the NAO H25 V6 computer and use parallel processing to improve latency. Additionally, it would be relevant to increase the number of signs that the NN is able to detect, to design activities that involve the construction of sentences and not only the handling of words.

On the other hand, Table 2 provides a comparison of related studies that involve the use of humanoid robots for sign language recognition. For instance, Li et al. (2023) achieved 100% accuracy with a NN containing fewer than 3,000 parameters; however, it was only trained with two signers and could recognize only four signs, which limits its generalization to new users. Similarly, Hosseini et al. (2024) utilized a custom humanoid robot capable of mimicking various signs, thanks to its five-fingered hands with independent motors. Their focus, however, was on recognizing and imitating signs using a data glove, while our work demonstrates the novelty of using humanoid robots as a tool to support sign language education. Additionally, we leveraged the robot's camera for sign detection, with only the data processing occurring externally.

The studies summarized in Table 2 also highlight the importance of humanoid robots' joint redundancy in mimicking and teaching signs. Flexibility in a robot's joints is crucial for accurately mimicking signs, as finger position, hand orientation, and hand position relative to the torso are critical in sign language (Meghdari et al., 2019). Hosseini et al. (2024) addressed this with their RASA robot, which has 29 degrees of freedom in its upper torso. Redundancy is also significant in educational activities, as physical interaction with the robot enhances user engagement (Timmerman and Ligthart, 2023). In this study, we aimed to foster user engagement through gestures and non-verbal interactions, alongside the use of buttons for tactile feedback.

The NN implemented in this study was based on an LSTM architecture, demonstrating its potential in sign detection systems without being constrained by processing limitations. This opens opportunities to work with deeper NNs in the future, capable of classifying a greater variety of signs. MediaPipe landmarks proved to be an effective tool for feature extraction that could outperformed CNN-based feature maps, which generate larger data per frame. For instance, the NN proposed Sincan and Yalim Keles (2020) outputs 131,071 features per frame, compared to just 162 features produced by MediaPpie.
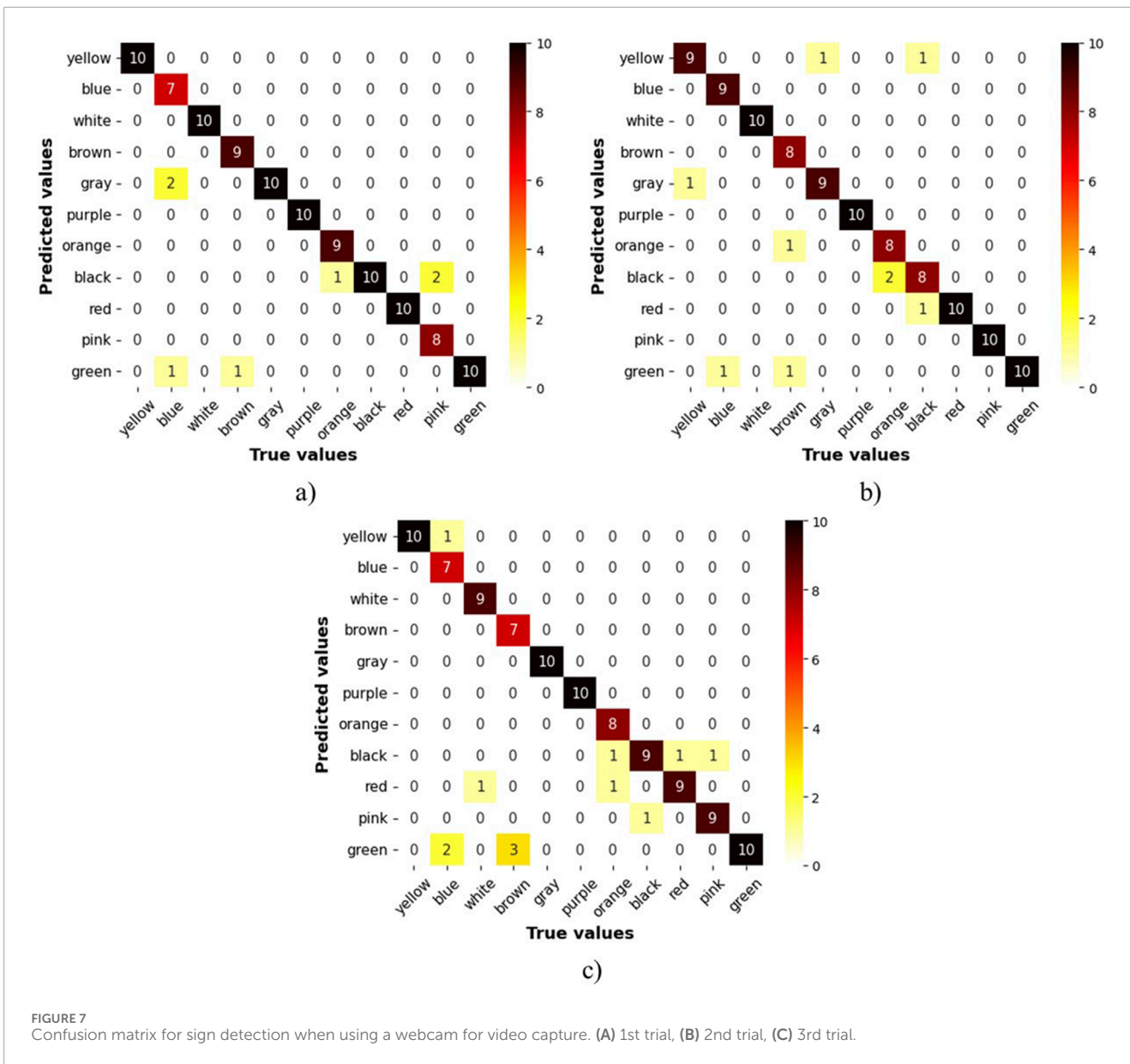
red and once with orange. Furthermore, as Figure 11 shows, 8 signs achieved an F-1 score of 100% and for the colors red, orange and pink, F-1 scores of 90.9%, 83.3% and 57.1% were obtained respectively. These results translate into an overall F-1 score of 93.8%.

On the other hand, the response time between the end of recording and when the NAO H25 V6 illuminates its eyes with the predicted sign color was also measured. As shown in Figure 12, the initial iterations took longer; however, after the model passed its warm-up stage, the response time decreased to an average of 2.29 s.
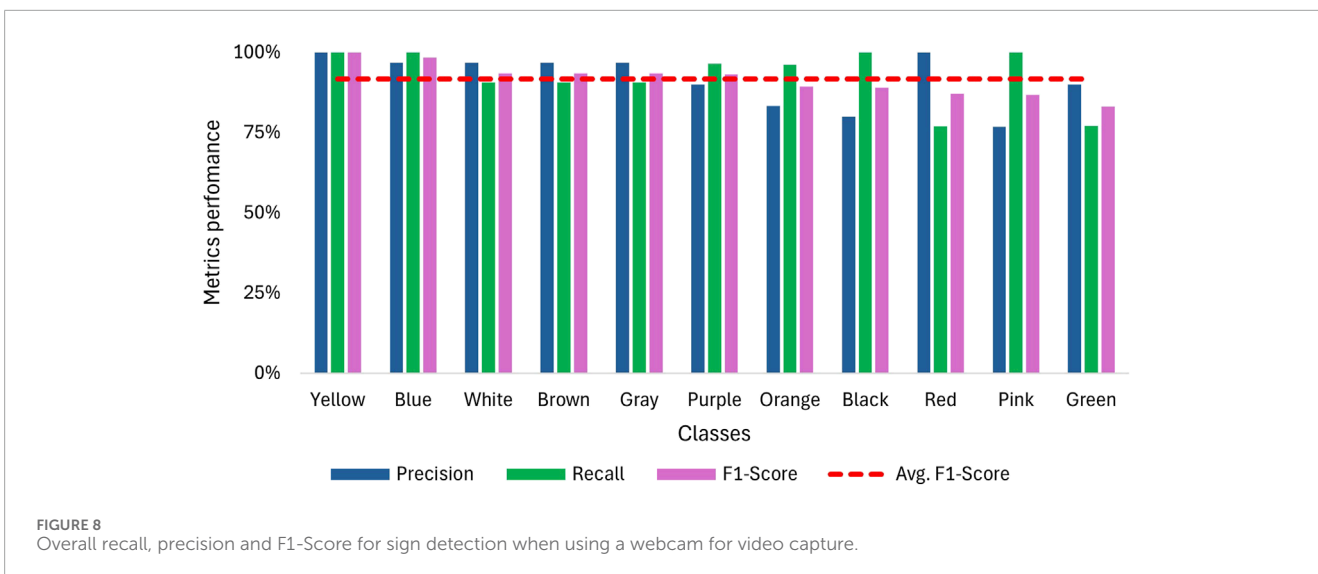
FIGURE 7
Confusion matrix for sign detection when using a webcam for video capture. **(A)** 1st trial, **(B)** 2nd trial, **(C)** 3rd trial.

FIGURE 8
Overall recall, precision and F1-Score for sign detection when using a webcam for video capture.
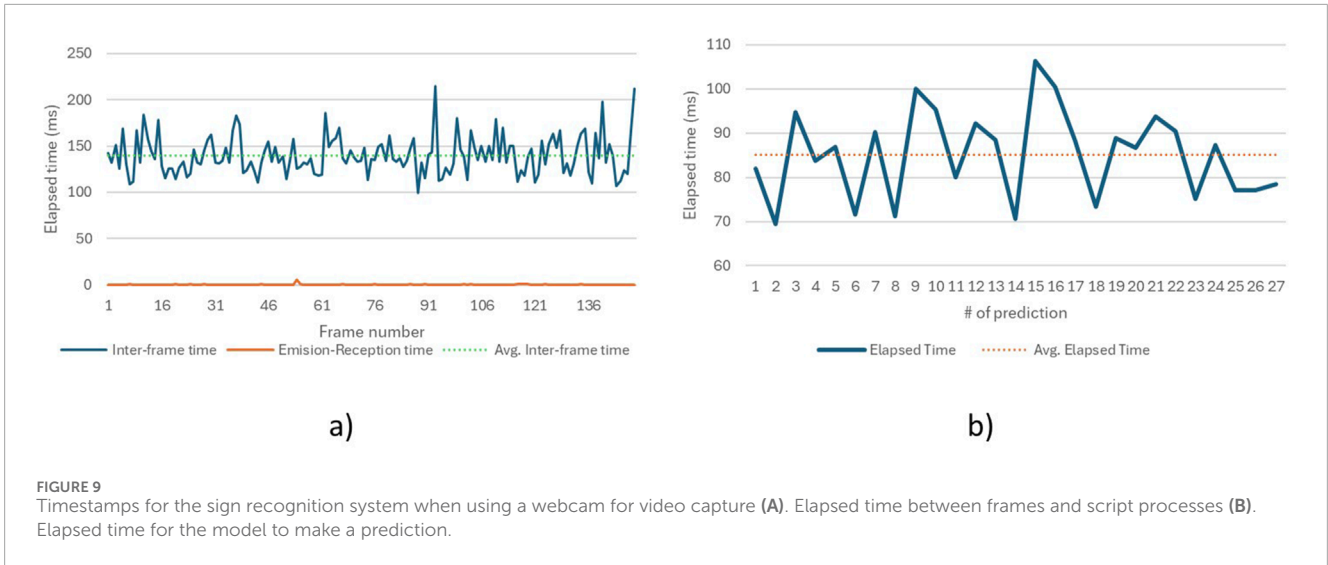
FIGURE 9
Timestamps for the sign recognition system when using a webcam for video capture **(A)**. Elapsed time between frames and script processes **(B)**. Elapsed time for the model to make a prediction.
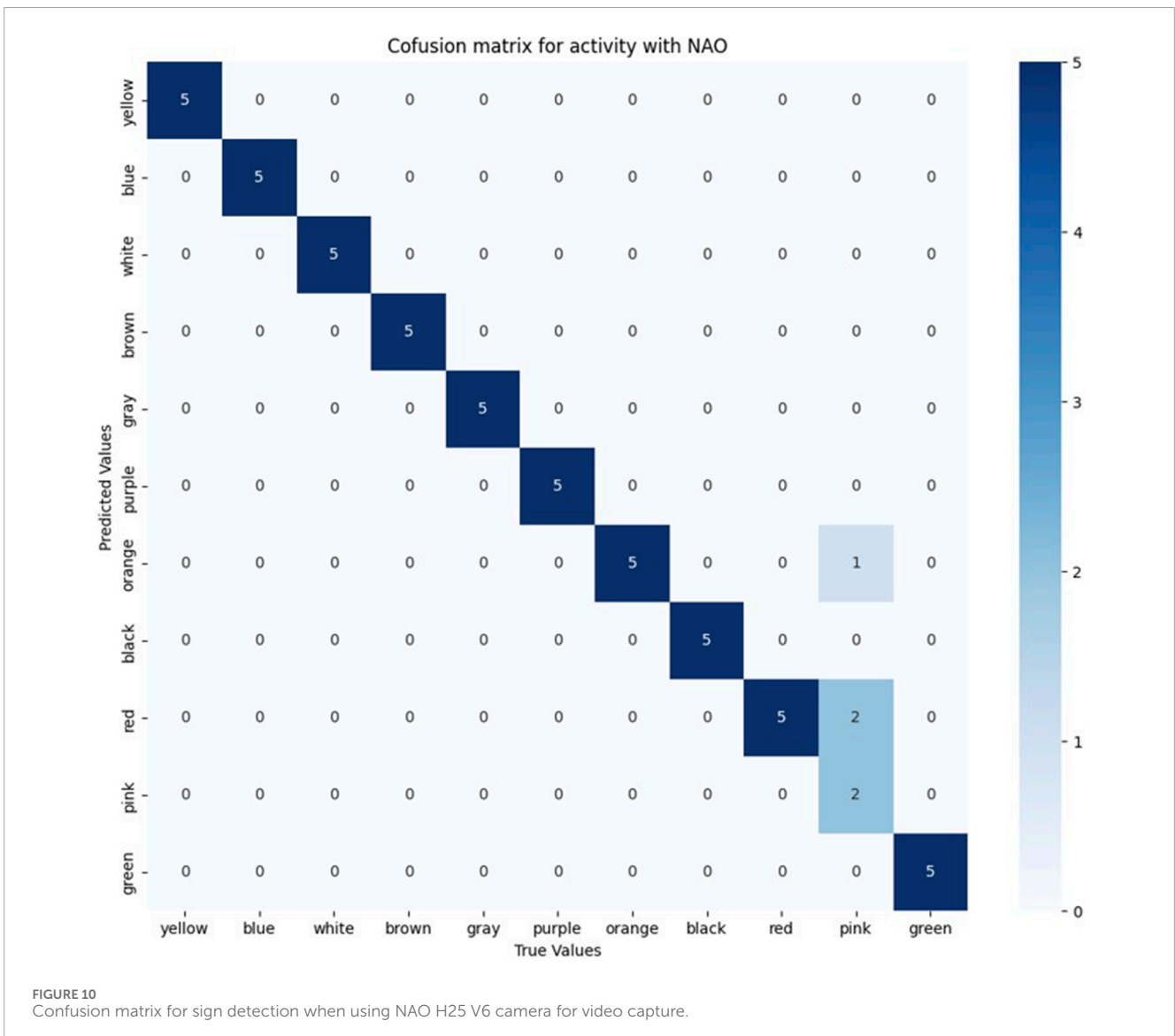

FIGURE 10
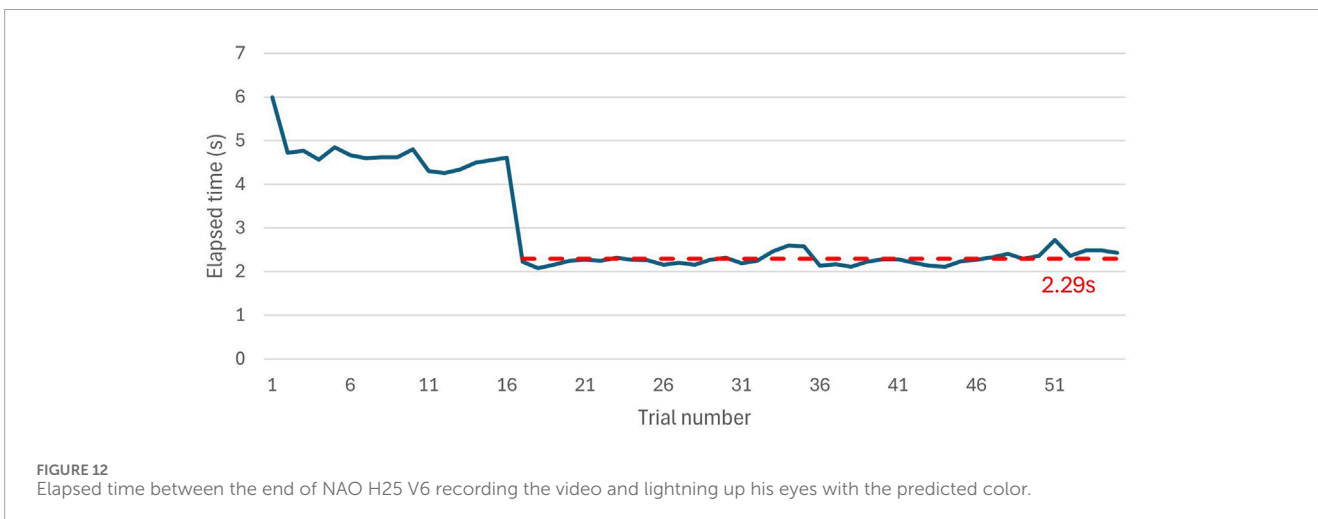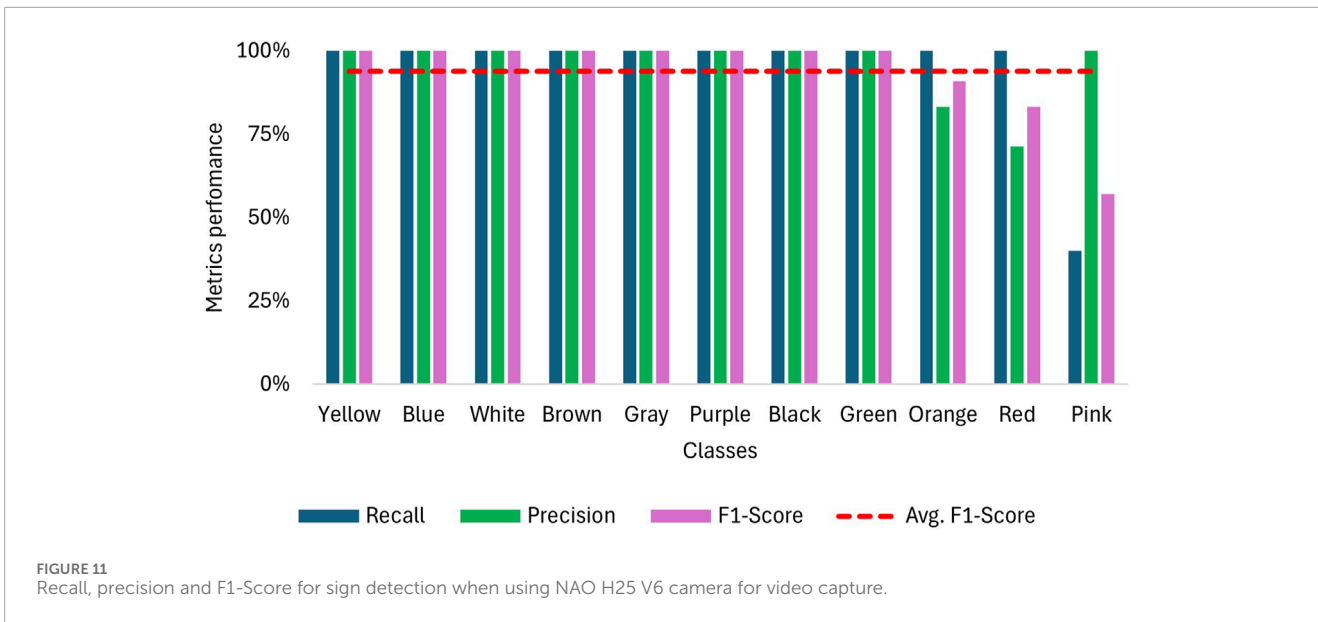Confusion matrix for sign detection when using NAO H25 V6 camera for video capture.

FIGURE 11
Recall, precision and F1-Score for sign detection when using NAO H25 V6 camera for video capture.

FIGURE 12
Elapsed time between the end of NAO H25 V6 recording the video and lightning up his eyes with the predicted color.

TABLE 2 Use of humanoid robots for sign language imitation and/or recognition.

References	Li et al. (2023)	Hosseini et al. (2024)	Ours
Robot	NAO	RASA	NAO H25 V6
# of signs	4	16	11
# of signers in dataset	2	12	20
Sensor	Kinect + Myo-armband	DataGlove	RGB camera
# of Neural Network parameters	2,764	2,462,912	682,187
Feature map	1,450	600	4,860
Accuracy	100%	70%	93.8%

TABLE 3 Performance for continuous sign recognition models.

References	Rahaman et al. (2024)	Sharma et al. (2024)	Joshi and Patel, (2024)	Ours
Accuracy	95.43%	99.7%	95.45%	91.6%
Computational cost (ms/frame)	56	500	-	35
Sign Language	Bangala	Indian	Indian	Colombian
# of signs	100	12	59	11

However, future work should consider enhancing this feature map by including joint angles and excluding irrelevant landmarks. As the confusion matrices on Figure 7 shows certain signs, such as blue and green, were confused due to similar movements but differing hand orientations. Including joint angles could provide more precise characteristics, reducing these types of errors.

Also, Table 3 compares studies that have explored continuous dynamic sign recognition across various sign languages. This study achieved an accuracy of 91.6%, which is comparable to other works that report accuracies above 90%. However, it stands out for its significantly lower computational cost, requiring only 35 m per frame, which is 93% less time than the study with the highest detection accuracy. This demonstrates the efficiency of the proposed system in terms of processing speed while maintaining competitive accuracy.

Additionally, it is important to highlight some of the technical limitations found in terms of detection capability and latency. Regarding detection, the system was prone to making incorrect predictions when the user was not performing any signs. As for the algorithm's latency, it increased by 2.2 s when making a wireless connection to the NAO H25 V6 camera from the computer. Although this latency is acceptable in terms of computational performance, further studies are needed to determine whether this delay might affect user interaction.

Since the objective of this work is to support CSL teaching for both hearing and non-hearing individuals who want to learn this language, the first step was to conduct a performance evaluation of the proposed educational tool. This evaluation demonstrated reliability in terms of sign detection accuracy and latency in a controlled environment, as presented in this paper's results. Based on these findings, the next step is to test the tool with real users to explore if features such as non-verbal interactions and visual feedback provided by the NAO H25 V6, are perceived as effective and engaging by the users, also, we aim to determine if the latency impacts the learning experience. To achieve this, the tool will be tested at both a bilingual bicultural school (secondary education) and at the Universidad de La Sabana (higher education), with the intention of reporting the results in a coming paper. Finally, in the future, we aim to expand the set of signs that the system can recognize by integrating additional vocabulary such as greetings, pronouns, and numbers, as experts suggest these are the primary topics taught in the early stages of CSL learning (Universidad de Bogota Jorge Tadeo Lozano, 2024; Universidad de Los Andes, 2024; Universidad ECCI, 2024).

5 Conclusion

An activity was designed to support the teaching of CSL using the NAO H25 V6 robotic platform. For sign detection we used the robot's upper camera that records the user executing a sign, we also took advantage of the redundancy in joints that the robot has, to use human-like gestures which allow non-verbal human-machine interactions and can be useful when working with deaf people. In addition, the integration of tactile buttons allows to have a physical interaction with the robot, which with gestures and LEDs integration can allow the user to feel more involvement of the robot in the activity. Likewise, a RNN of the type LSTM was built for the detection of 11 colors of the CSL, the architecture has three LSTM and dense layers with 682187 parameters. Where promising results were obtained for sign detection accuracy and computational cost, with an F1-Score of 91.6% and response time of 35 m (28.6 FPS) per frame for the test using a webcam, while the test using the NAO camera through a TCP/IP connection obtained results with an F1-Score of 93.8% and 2.29 s to predict a sign. Finally, as future work, we will run all the processes on the NAO computer to improve the response time per sign and we hope to take this activity to a bilingual bicultural school; to measure the impact, it would have on people who are learning basic CSL vocabulary.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

JM-Z: Data curation, Formal Analysis, Investigation, Methodology, Validation, Writing–original draft, Writing–review and editing. CG-C: Conceptualization, Funding acquisition, Methodology, Project administration, Supervision, Writing–review and editing. JC: Conceptualization, Formal Analysis, Methodology, Supervision, Writing–review and editing.

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The handling editor JLOA declared a past co-authorship with the author CG-C.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Cruz-Ramírez, S. R., García-Martínez, M., and Olais-Govea, J. M. (2022). NAO robots as context to teach numerical methods. *Int. J. Interact. Des. Manuf.* 16 (4), 1337–1356. doi:10.1007/s12008-022-01065-y

Instituto Nacional Para Sordos (2024). Oferta educativa. Available at: https://educativo.insor.gov.co/oferta/.

Gonzalez, H., Hernández, S., and Calderón, O. (2024). Design of a sign language-to-natural language translator using artificial intelligence. *Int. J. Online Biomed. Eng.* 20 (03), 89–98. doi:10.3991/IJOE.V20I03.46765

González-Rodríguez, J. R., Córdova-Esparza, D. M., Terven, J., and Romero-González, J. A. (2024). Towards a bidirectional mexican sign language–spanish translation system: a deep learning approach. *Technologies* 12 (1), 7. doi:10.3390/TECHNOLOGIES12010007

Gürpınar, C., Uluer, P., Akalın, N., and Köse, H. (2020). Sign recognition system for an assistive robot sign tutor for children. *Int. J. Soc. Robotics* 12 (2), 355–369. doi:10.1007/s12369-019-00609-9

Hellou, M., Gasteiger, N., Kweon, A., Lim, J. Y., MacDonald, B. A., Cangelosi, A., et al. (2023). "Development and validation of a motion dictionary to create emotional gestures for the NAO robot," in IEEE International Workshop on Robot and Human Communication, Busan, Korea, Republic of, 28-31 August 2023 (RO-MAN), 897–902.

Hosseini, S. R., Taheri, A., Alemi, M., and Ali, M. (2024). One-shot learning from demonstration approach toward a reciprocal sign language-based HRI. *Int. J. Soc. Robotics* 16 (4), 645–657. doi:10.1007/s12369-021-00818-1

Ilham, A. A., Nurtanio, I., Ridwang, S., Syafaruddin, (2024). Applying LSTM and GRU methods to recognize and interpret hand gestures, poses, and face-based sign language in real time. *J. Adv. Comput. Intell. Intelligent Inf.* 28 (2), 265–272. doi:10.20965/JACIII.2024.P0265

Instituto Nacional Para Sordos (2020). *Caracterización de ciudadanos*, Usuarios y grupos de interés.

Ji, A., Wang, Y., Miao, X., Fan, T., Ru, B., Liu, L., et al. (2023). Dataglove for sign language recognition of people with hearing and speech impairment via wearable inertial sensors. *Sensors* 23 (15), 6693. doi:10.3390/S23156693

Jintanachaiwat, W., Jongsathitphaibul, K., Pimsan, N., Sojiphan, M., Tayakee, A., Junthep, T., et al. (2024). Using LSTM to translate thai sign language to text in real time. *Discov. Artif. Intell.* 4 (1), 17–11. doi:10.1007/s44163-024-00113-8

Joshi, J. M., and Patel, D. U. (2024). GIDSL: Indian-Gujarati isolated dynamic sign language recognition using deep learning. *SN Comput. Sci.* 5 (5), 527–612. doi:10.1007/s42979-024-02776-7

Kondo, T., Narumi, S., He, Z., Shin, D., and Kang, Y. (2024). A performance comparison of japanese sign language recognition with ViT and CNN using angular features. *Appl. Sci.* 14 (8), 3228. doi:10.3390/APP14083228

Kurtz, G., and Kohen-Vacs, D. (2024). Humanoid robot as a tutor in a team-based training activity. *Interact. Learn. Environ.* 32 (1), 340–354. doi:10.1080/10494820.2022.2086577

Li, J., Zhong, J., and Wang, N. (2023). A multimodal human-robot sign language interaction framework applied in social robots. *Front. Neurosci.* 17 (April), 1168888. doi:10.3389/fnins.2023.1168888

Liang, Y., Jettanasen, C., and Chiradeja, P. (2024). Progression learning convolution neural model-based sign language recognition using wearable glove devices. *Computation* 12 (4), 72. doi:10.3390/COMPUTATION12040072

Meghdari, A., Alemi, M., Zakipour, M., and Kashanian, S. A. (2019). Design and realization of a sign language educational humanoid robot. *J. Intelligent Robotic Syst. Theory Appl.* 95 (1), 3–17. doi:10.1007/s10846-018-0860-2

Mutawa, A. M., Mudhahkah, H. M. A., Al-Huwais, A., Al-Khaldi, N., Al-Otaibi, R., and Al-Ansari, A. (2023). Augmenting mobile app with NAO robot for autism education. *Mach. 2023* 11 (8), 833. doi:10.3390/MACHINES11080833

Pal, D., Kumar, A., Kumar, V., Basangar, S., and Tomar, P. (2024). Development of an OTDR-based hand glove optical sensor for sign language prediction. *IEEE Sensors J.* 24 (3), 2807–2814. doi:10.1109/JSEN.2023.3339963

Pereira-Montiel, E., Pérez-Giraldo, E., Mazo, J., Orrego-Metaute, D., Delgado-Trejos, E., Cuesta-Frau, D., et al. (2022). Automatic sign language recognition based on accelerometry and surface electromyography signals: a study for Colombian Sign Language. *Biomed. Signal Process. Control* 71 (January), 103201. doi:10.1016/J.BSPC.2021.103201

Rahaman, M. A., Ali, M. H., and Hasanuzzaman, M. (2024). Real-time computer vision-based gestures recognition system for bangla sign language using multiple linguistic features analysis. *Multimedia Tools Appl.* 83 (8), 22261–22294. doi:10.1007/s11042-023-15583-8

Shahin, N., and Ismail, L. (2024). From rule-based models to deep learning transformers architectures for natural language processing and sign language translation systems: survey, taxonomy and performance evaluation. *Artif. Intell. Rev.* 57 (10), 271–351. doi:10.1007/S10462-024-10895-Z

Shao, Q., Sniffen, A., Blanchet, J., Hillis, M. E., Shi, X., Haris, T. K., et al. (2020). Teaching american sign language in mixed reality. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4 (4), 1–27. doi:10.1145/3432211

Sharma, V., Sharma, A., and Saini, S. (2024). Real-time attention-based embedded LSTM for dynamic sign language recognition on edge devices. *J. Real-Time Image Process.* 21 (2), 53–13. doi:10.1007/s11554-024-01435-7

Sincan, O. M., and Yalim Keles, H. (2020). Autsl: a large scale multi-modal Turkish sign language dataset and baseline methods. *IEEE Access* 8, 181340–181355. doi:10.1109/ACCESS.2020.3028072

Teran-Quezada, A. A., Lopez-Cabrera, V., Rangel, J. C., and Sanchez-Galan, J. E. (2024). Sign-to-Text translation from panamanian sign language to Spanish in continuous capture mode with deep neural networks. *Big Data Cognitive Comput. 2024* 8 (3), 25. doi:10.3390/BDCC8030025

Timmerman, E., and Ligthart, M. E. U. (2023). Let's roll together: children helping a robot play a dice game. *ACM/IEEE Int. Conf. Human-Robot Interact.* 20, 476–480. doi:10.1145/3568294.3580130

Universidad de Bogota Jorge Tadeo Lozano (2024). Curso básico lengua de señas colombiana. *Utadeo.* Available at: https://www.utadeo.edu.co/es/santa-

marta/continuada/educacion-continua/78091/curso-basico-lengua-de-senas-colombiana.

Universidad de Los Andes (2024). Introducción a la lengua de señas colombiana. *Uniandes*. Available at: https://educacioncontinua.uniandes.edu.co/es/programas/introduccion-la-lengua-de-senas-colombiana.

Universidad ECCI (2024). Lengua de señas colombiana – nivel a1. *Ecci*. Available at: https://www.ecci.edu.co/programas/lengua-de-senas-colombiana-nivel-a1/.

World Health Organization (2024). Deafness and hearing loss. Available at: https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss.

World Health Organization (2023). Disability. Available at: https://www.who.int/es/news-room/fact-sheets/detail/disability-and-health.

Zhang, Y., Min, Y., and Chen, X. (2021). Teaching chinese sign language with a smartphone. *Virtual Real. and Intelligent Hardw.* 3 (3), 248–260. doi:10.1016/J.VRIH.2021.05.004