SIGNIFY: Leveraging Machine Learning and Gesture Recognition for Sign Language Teaching Through a Serious Game

future internet

MDPI

Article

SIGNIFY: Leveraging Machine Learning and Gesture Recognition for Sign Language Teaching Through a Serious Game

Luca Ulrich [1,*], Giulio Carmassi [2], Paolo Garelli [2], Gianluca Lo Presti [2], Gioele Ramondetti [3], Giorgia Marullo [1], Chiara Innocente [1] and Enrico Vezzetti [1]

[1] Management and Production Engineering, Politecnico di Torino, C.so Duca degli Abruzzi, 24, 10129 Torino, Italy; giorgia.marullo@polito.it (G.M.); chiara.innocente@polito.it (C.I.); enrico.vezzetti@polito.it (E.V.)
[2] Biomedical Engineering, Politecnico di Torino, C.so Duca degli Abruzzi, 24, 10129 Torino, Italy; s319796@studenti.polito.it (G.C.); s312316@studenti.polito.it (P.G.); s315812@studenti.polito.it (G.L.P.)
[3] Computer Engineering, Politecnico di Torino, C.so Duca degli Abruzzi, 24, 10129 Torino, Italy; gioele.ramondetti@studenti.polito.it
* Correspondence: luca.ulrich@polito.it

Abstract: Italian Sign Language (LIS) is the primary form of communication for many members of the Italian deaf community. Despite being recognized as a fully fledged language with its own grammar and syntax, LIS still faces challenges in gaining widespread recognition and integration into public services, education, and media. In recent years, advancements in technology, including artificial intelligence and machine learning, have opened up new opportunities to bridge communication gaps between the deaf and hearing communities. This paper presents a novel educational tool designed to teach LIS through SIGNIFY, a Machine Learning-based interactive serious game. The game incorporates a tutorial section, guiding users to learn the sign alphabet, and a classic hangman game that reinforces learning through practice. The developed system employs advanced hand gesture recognition techniques for learning and perfecting sign language gestures. The proposed solution detects and overlays 21 hand landmarks and a bounding box on live camera feeds, making use of an open-source framework to provide real-time visual feedback. Moreover, the study compares the effectiveness of two camera systems: the Azure Kinect, which provides RGB-D information, and a standard RGB laptop camera. Results highlight both systems' feasibility and educational potential, showcasing their respective advantages and limitations. Evaluations with primary school children demonstrate the tool's ability to make sign language education more accessible and engaging. This article emphasizes the work's contribution to inclusive education, highlighting the integration of technology to enhance learning experiences for deaf and hard-of-hearing individuals.

Keywords: hand sign alphabet; social inclusion; machine learning; gesture recognition; gamification; serious game

Citation: Ulrich, L; Carmassi, G.; Garelli, P.; Lo Presti, G.; Ramondetti, G.; Marullo, G.; Innocente, C.; Vezzetti, E. SIGNIFY: Leveraging Machine Learning and Gesture Recognition for Sign Language Teaching Through a Serious Game. *Future Internet* **2024**, *16*, 447. https://doi.org/10.3390/fi16120447

Academic Editors: Marco Romano and Teresa Onorati

Received: 14 October 2024
Revised: 14 November 2024
Accepted: 24 November 2024
Published: 1 December 2024

1. Introduction

Sign language is a vital means of communication for the deaf and hard-of-hearing communities. Sign language (and sign alphabet) is considered a non-verbal language developed for deaf people community to communicate with each other and with normal people [1]. Early exposure to and learning of sign language and the sign alphabet are crucial for cognitive development, social interaction, and academic achievement in children who are deaf or hard of hearing. The main reasons for hearing loss include aging, genetics, high volume noise exposure, a variety of infections and, in some cases, certain toxins or medications [2]. For deaf children with hearing parents, or those who become deaf due to illnesses, the opportunity to learn and study sign language at a young age is particularly

pivotal. In such contexts, the people surrounding these children may not be experts in sign language, potentially slowing the learning rate for both the children and their families if not supported with adequate learning processes.

Deaf children face significant communication barriers in school, impacting their academic and social development. Limited teacher training and awareness further hinder their educational experience and social isolation and bullying due to communication difficulties can even lead to emotional illness [3]. Learning sign language provides many benefits not just for deaf and hard-of-hearing individuals, but even for hearing persons. Indeed, hearing individuals can greatly benefit by learning sign language, like, first of all, being able to communicate with one's deaf friends, relatives, and coworkers for better inclusion and understanding. Another critical benefit of learning sign language includes better cognitive performance. In fact, some studies show that learning a second language, particularly sign language, improves memory, creativity, and problem-solving skills. This is partly because sign languages stimulate other brain centers than spoken languages, which optimizes the flexibility of the mind [4]. As a consequence, sign language can be a way to improve general communication skills. Through sign language education, one may enhance awareness and sensitivity for facial expressions and gestures since it is pretty essential in the process of a non-verbal form of communication. The use of signs can improve interpersonal communication in everyday contexts, making an individual more sensitive and receptive to non-verbal clues. From this perspective, also teaching LIS to hearing children plays a vital role in promoting inclusivity and fostering a more accessible society. By learning LIS, children develop an early understanding of diversity, empathy, and communication beyond spoken language. It empowers them to interact with peers who are deaf or hard of hearing, breaking down social barriers and encouraging mutual respect. Moreover, exposure to sign language enhances cognitive abilities, including improved spatial awareness, problem-solving skills, and non-verbal communication. This approach not only benefits those who rely on sign language, but also nurtures a generation that values diversity and inclusiveness. The slogan for the new educational pathway is "learning for all" [5], emphasizing the need for inclusive education. Introducing sign language as an innovative tool in primary schools or kindergartens could be a key factor in fostering inclusivity and awareness already from a young age.

In recent years, the integration of technology into educational environments has revolutionized traditional teaching methods, providing innovative and engaging ways to improve learning experiences. One such innovation is the development of educational video games, which combine entertainment with learning goals to create an interactive and stimulating environment for students [6]. Currently, students can be considered digital natives, and their interest is usually aligned with immersive and interactive environments. In this sense, using serious games to teach sign language can facilitate that process by making the learning much more engaging and interactive [7].

In this work, the serious game SIGNIFY is introduced. SIGNIFY is an application specifically designed to teach Italian Sign Language (LIS) to primary school children. To deal with young learners, the necessary aim is to offer a tailored, engaging, and interactive learning environment. For this purpose, Machine Learning-based hand gesture recognition techniques were applied to assist in learning and easily mastering sign language gestures. By detecting and overlaying 21 hand landmarks and a bounding box on live camera feeds, the solution leverages an open-source framework to offer real-time visual feedback. Furthermore, different data sources were compared to evaluate the effectiveness of RGB versus RGB-D (depth) data for sign language recognition for the specific purpose of a children-friendly application. A new RGB-D database that records LIS gestures was needed to develop a tailored solution for the characteristic application, which serves as an essential resource for training and testing AI models. Finally, a usability analysis based on feedback collected from the children was performed, offering insights into their learning experience and interaction with the game. This analysis demonstrated to be remarkably beneficial for highlighting margins of improvement and ensuring the platform is engaging and accessible

for young users. These contributions advance the technological and educational aspects of sign language learning through AI.

The paper is structured as follows: Firstly, in the following section, related works are presented to better frame the project. Subsequently, in Section 3, all the technical aspects of the project are analyzed, ranging from the creation of the dataset to the machine learning method used to perform the classification. Furthermore, a brief overview of the developed serious game structure and functionalities is provided. The final two sections are dedicated, on the one hand, to the presentation of the results, in terms of the classifier's performances and ease of use of the game assessed in a case study with a group of children, and the future developments and perspectives on the project on the other hand.

2. Related Works

In the context of gesture recognition, there are two main sign language recognition approaches: image- and sensor-based. In particular, sensor-based approaches are built to obtain joints orientation, hands position, and hand velocity [8], and they include technologies such as microcontrollers and specific sensors, such as data gloves [9,10], power gloves [11], digital cameras [12], accelerometers [13,14], and leap motion controllers [15]. However, research is advancing primarily in the image-based field due to its advantage of not requiring users to wear devices on them [16].

Since LIS requires facing from multiple perspectives, being at the intersection of different disciplines, this section has been divided into three parts to better frame the previous works. The first subsection focuses on the impact of serious games in learning; the second and the third aim to better contextualize the work from a technological perspective, showing which gesture acquisition and classification methodologies have been explored in recent years, respectively.

2.1. Effectiveness of Serious Games in Learning

Serious games are intelligent tools of learning that combine game-like features and pedagogical content in an integrative manner to enhance motivation and involvement on the part of the user [17]. Implementation of serious games in education provides noticeable learning outcomes, and makes the process much more interactive and engaging. One of the main advantages of serious games is their ability to stimulate students. Playful elements such as challenges, rewards, and visible progress stimulates student's interest and commitment; thus, learning becomes an enjoyable and rewarding process. For example, the study by Kye et al. [18] mentioned that the learning game "Magic Touch Math" helped children master the basic operations in math. Besides that, serious games offer adaptability to varying learning styles and present personal experiences that can meet every student's particular needs. This approach is primarily practical with deep and durable learning. Serious games further enhance the students' transversal skills, such as critical thinking, problem-solving, and collaboration. Games tend to propose complex situations where students need to use these skills to move forward. Significantly, integrating this into the simulation of educational games will lead to an improvement in the abilities of students in real-life contexts [19]. Unavoidably, serious games can also be found to be particularly effective in the teaching of sign language. The visual and interactive elements of the serious game allow students to memorize and practice signs more effectively than traditional methods. For example, Lang et al. [20] created an edutainment game for teaching sign language using Kinect technology. They explained how the naturalness and real-time interaction of such applications could enhance learning by signs. Practical applicability was also shown through the serious games' effectiveness in learning gesture recognition in Portuguese sign language in an application developed by Soares et al. [21]. In further support, Pontes et al. [7] described a platform for learning numbers in Brazilian Sign Language through a serious game, MatLIBRAS Racing. Many students, including those waiting for their turn, attempted to replicate the hand configurations shown on the screen. A follow-up quiz conducted one week after the experiment assessed the retention of learned

signs. The results showed a positive distribution of correct answers, indicating that the majority of participants could identify the signs accurately. In recent years, several other languages have been included in serious games, such as American [22] and Pakistani [23], showing promising results in terms of learning, but also some drawbacks linked to the complexity of the proposed solutions, thus limited accessibility and language-specific issues. Nonetheless, serious games propose a different and exciting learning approach, combining amusement with education for improved educational outcomes. Their capability to motivate learners, adapt to the diversity of their learning styles, and enhance their transversal skills renders them valuable tools in modern education [20].

2.2. Image Acquisition Technologies

Image acquisition is the most critical part of any gesture recognition system, because it requires to find a trade-off between high data quality and overall system performance. In recent years, the ever evolving 3D acquisition technologies made it possible to obtain RGB and Depth information at the same time, in light of providing more robust and consistent solutions [24]. In 2D technologies, a single RGB camera is used for capturing two-dimensional images of the hands or body. This approach is widely adopted due to its cost-effectiveness and availability, since these kind of cameras are the most common ones to the general users. However, it presents some drawbacks, such as sensitivity to lighting conditions and difficulties in distinguishing between similar movements on different planes. Despite these limitations, 2D technologies can be quite effective when supported by robust image processing algorithms. For instance, Bora et al. [25] demonstrated the effectiveness of using MediaPipe and deep learning in real-time sign language recognition, showing that the 2D approach can effectively manage complex gestures.

Conversely, 3D technologies rely on depth sensors to capture three-dimensional information about the movement and position of hands and body. In contrast to 2D systems, these provide higher accurateness and robustness in the capturing process, since they can detect the change in depth and movement in space independently from lighting conditions. For example, Lang et al. [20] proposed a methodology using Kinect that may substantially increase the preciseness of gesture recognition, thereby making it especially effective for complex applications in the process of sign language teaching. One of the most important advantages of 3D technologies is their ability to capture complex and detailed movements, such as rotations and translations of the hands, demanding tracing using 2D technologies. This hardware is, as a rule, more expensive and less convenient for everyday use than 2D. In addition, Soares et al. [21] developed Kinect technology for the recognition of the whole body, expanding the functionality of sign language teaching with the help of recognizing gestures that would activate other parts of the body in communication, not just hand movements. Nowadays, 3D information is not only used to increase hand gesture algorithms recognition rate, but also to distinguish sign language linguistic from gestural expressions and increase the overall robustness, as proved by Stamp et al. [26].

Bora et al. [25] further compared the two technologies and noted several key differences. They found that while 2D systems are accessible and practical for many applications due to their lower cost and ease of use, they suffer from limitations in accuracy under varying lighting conditions and can struggle with depth perception. On the other hand, 3D systems, such as those utilizing Kinect, offer superior accuracy in detecting complex gestures and depth information, but at a higher cost and with greater hardware requirements. The authors concluded that the choice between 2D and 3D technologies should be based on the specific needs and resources of the application and in some cases, a combination of both may provide the optimal solution for accurate and reliable gesture recognition. Finally, both the technologies, 2D and 3D, come with their set of advantages and disadvantages. Their application would depend on the need and available resources.

2.3. Image Processing and Gesture Classification Methods

Image processing and gesture classification are crucial components of sign language recognition systems. These tasks can be divided into two primary parts: hand recognition and gesture classification. Various machine-learning techniques have been applied to each part, offering specific advantages in terms of accuracy, processing speed, and computational complexity.

Hand recognition involves detecting and tracking the hand in an image or video stream. Convolutional Neural Networks (CNNs) are widely used for this purpose due to their ability to extract meaningful features from images and videos. CNNs are particularly robust against variations in hand position, scale, and rotation as shown in various works, such as Nimisha et al. [27] and Uboweja et al. [28]. More recently, with the already mentioned widespread of hand sign recognition techniques to embrace different languages, Amirgaliyev et al. [29] developed a CNN-based methodology for the Kazakh sign language, obtaining 95.7% of recognition rate.

A recent application of this technology can be found in Google's MediaPipe library, used also in the work of Bora et al. [25] for gesture recognition. MediaPipe is Google's real-time hand landmark detection in which a hybrid of six models was combined in tracking and predicting a sequence of 21 3D points used in detecting/inferring hand gestures. This further facilitates sensitive and accurate tracking in this approach without the need for costly add-on hardware and, hence, it is practical for general use [30].

Once the hand is recognized, the next step is gesture classification, where the detected hand movements are interpreted as specific gestures. Several machine learning models have been used for gesture classification, including Random Forests, Support Vector Machines (SVMs), and Artificial Neural Networks (ANNs).

ANNs are powerful tools for gesture classification due to their ability to learn through data training. They can model complex non-linear decision boundaries, improving classification accuracy in real-time scenarios. Nimisha et al. proposed using ANN for the classification of sign language, demonstrating its effectiveness in capturing the intricacies of sign gestures [27]. Bajaj et al. also reported the great results of an ANN for gesture classification [31]. Although these are remarkable results, ANNs can sometimes be prone to overfitting, which can affect their performance.

Support Vector Machines (SVMs) are another popular machine learning tool for supervised learning, capable of handling both classification and regression tasks. This methodology classifies data based on probabilities, making them suitable for binary and multi-class classification problems. SVMs can perform both linear and non-linear classification by using different kernel functions. Nimisha et al. [27] used an SVM classifier to categorize signs into seven classes of the Pakistan Urdu language, showcasing its utility in sign language recognition. Kavana and Suma [30] used SVM for classification, reporting that this system outperformed other machine learning algorithms.

Random Forest models are robust for gesture classification due to their ability to handle non-linear data and aggregate multiple decision trees to improve overall accuracy. These models are effective in educational and gaming contexts, identifying complex gestures in variable environments. For instance, in the study of Ren Ewe et al. [32], a hybrid model combining VGG16 for hand recognition and Random Forest for gesture classification showed significant improvement in gesture classification accuracy, because the ensemble nature of Random Forest reduces variance and overfitting, enhancing robustness and generalization.

3. Materials and Methods

This section describes the methodologies employed to perform the automatic hand gesture recognition, from the databases creation to the hand detection and hand gesture recognition. Moreover, the design and development of the serious game presented to primary school children is detailed.

3.1. Database Acquisition

Since the aim was to compare different data acquisition systems (RGB and RGB-D), the first step of the work involved the creation of two datasets tailored for the Italian Sign Language (LIS) recognition. The premise is that the signs for letters *g*, *j*, *s*, and *z* were not considered, since they use a dynamic gesture that cannot be captured and identified in a single frame, as required by our real-time application.

Regarding the RGB database, a freely accessible online database was used [33]. This dataset is composed of around 250 pictures from 11 different people for each of the 22 letters considered. The images are in JPG format with dimensions 622 × 415 pixels, and granted a sufficient diversity in terms of hand shapes and possible orientations, as they were taken from 3 different angles (Figure 1).

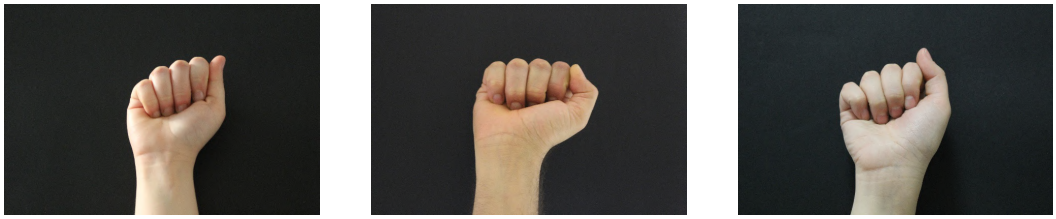
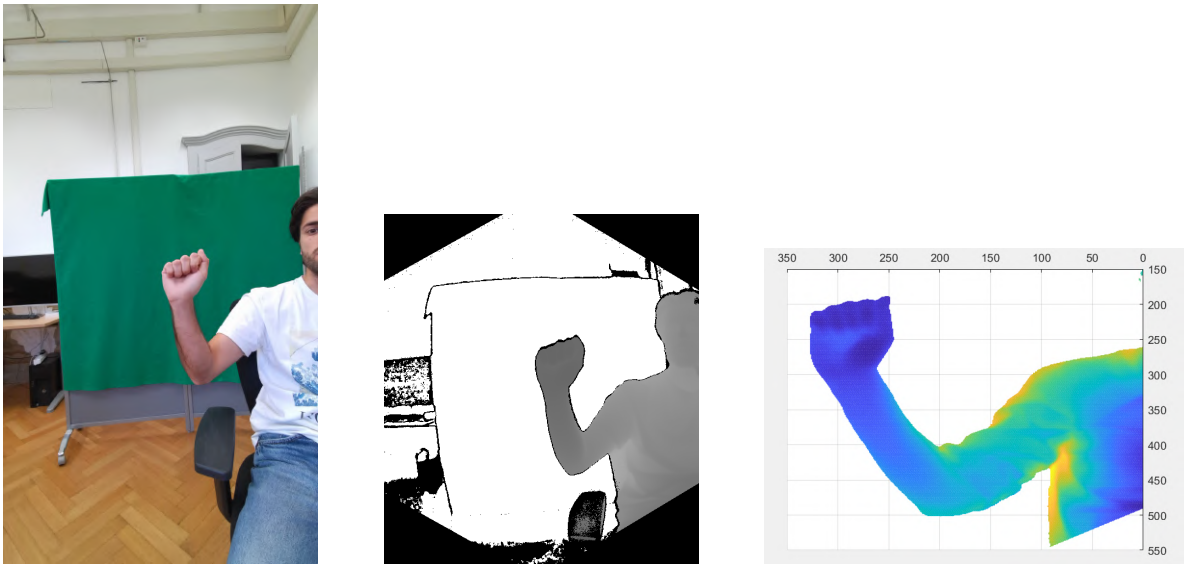


Figure 1. Three examples of images taken from the LIS-dataset for the letter A in three different orientations.

The dataset for RGB-D images was created from scratch. The existing RGB-D datasets concerning LIS, such as A3LIS-147 [34] and Montalbano [35], contain the subject's hands and body, offering a comprehensive picture of the motions necessary to represent LIS signs. Both include hand movement, body posture, and gestures. In contrast, the current application concentrates on hand movement to make learning as simple as possible for children. As a result, the information offered by these databases has insufficient resolution because the hand is not completely visible in the foreground. To solve this issue, a tailored database acquired using a state-of-the-art consumer-grade camera, the Kinect Azure, was created. The choice of a depth camera with time-of-flight (ToF) technology for depth acquisition has been performed to deal with the current depth camera market evolution. In fact, the most recent smartphones are equipped with ToF sensors because of their remarkable performance and small size. Moreover, the 4K camera allows for excellent RGB images to be associated with the corresponding depth map (Figure 2). Azure Kinect is a highly efficient device also used in other studies to obtain excellent quality 3D video tracking of different anatomical segments [36]. A 10 s video was recorded by seven people for each sign. The recorded videos from the Azure Kinect are in .mkv format, so a script to extract from the different channels the single frames and depth maps to create the dataset was developed. Since the videos were recorded at 30 fps, a total of 1500 images in PNG format for each hand sign were obtained. During the recording slight rotation and flexion of the wrist were performed to make the classifier more adaptable to different camera orientations and guarantee data generalization. Specifically, to train the algorithm for gesture recognition with the cleanest possible data, especially from the z-coordinate point of view, the videos were acquired by placing a green canvas behind each subject, so the hand had a uniform background. Nonetheless, as will be shown in Section 4, since the classifier is based solely on the landmarks coordinates, the algorithm performance is not negatively affected by a change in the background nor different values of distances between the subject and the camera (in accordance with the Kinect operating range). The acquisition device uses two different sensors to capture the RGB image and the Depth image, as shown in Figure 2a,b. Therefore, the dimension of each RGB image in this dataset is 1080 × 1920, while the dimension of the depth maps is 576 × 640. Consequently, depth map processing was needed to superimpose the two images and properly associate each x and y coordinate with the RGB color and depth value. To better visualize the 3D nature of the depth map, Figure 2c illustrates a cropped version that considers only the volume of interest, namely

the hand. This cropped version was obtained by thresholding the image on the z-axis and excluding everything behind the threshold. During the acquisition process, the camera was rotated 90 degrees to take advantage of an easier-to-read vertical image and return to the field-of-view conditions of a smartphone, which is the device that would be most suitable for exploiting the proposed application.



(**a**) RGB image (**b**) Depth map (**c**) Depth map processing (Matlab)

Figure 2. "A" hand sign captured by Azure Kinect.

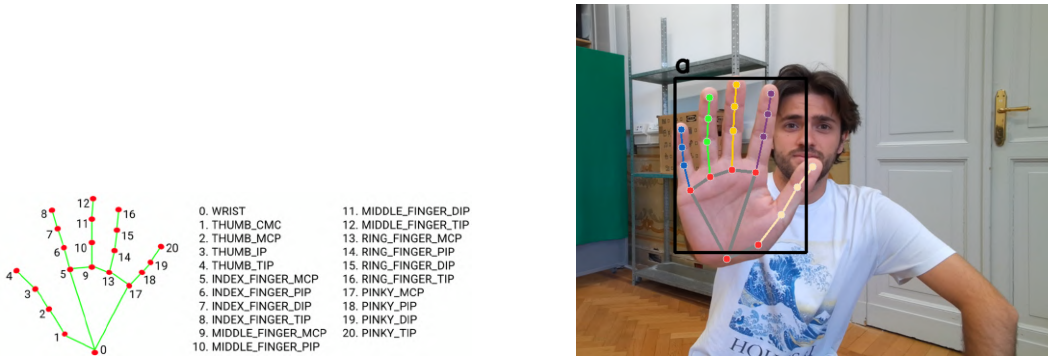
3.2. Hand Landmarks Identification

The choice of adopting a landmark-based instead of an image-based approach has been made for several reasons. Landmarks allows to focus on key points, such as finger joints and tips, which are crucial for accurately capturing the specific gestures used in sign language. In contrast, whole images include unnecessary background data and surrounding objects, introducing noise that can distract the AI model from focusing on the hand movements that matter most. Moreover, hand landmarks also reduce data complexity, allowing the model to process a small set of key points instead of every pixel in an image. This makes the AI system more efficient and faster, especially in real-time applications. By concentrating on the hand's structure and movement, landmark-based approaches can more precisely capture subtle differences between similar gestures. Full-image approaches, on the other hand, may obscure these details due to distractions or occlusions. Furthermore, focusing on hand landmarks helps the model generalize better across different users by identifying patterns based on the essential hand structure rather than relying on the entire visual scene. This approach improves recognition accuracy and consistency allowing also for the employment of a smaller dataset.

The Hand Landmarker of the Google package Mediapipe [37] has been used to automatically obtain the x and y coordinates of 21 landmarks in each hand picture in our datasets using the RGB image. Mediapipe Hands uses a two-stage tracking pipeline, including (1) a palm detector, which provides a bounding box of a hand inside the frame, and (2) a hand landmark model, which predicts the hand landmarks (Figure 3a). Specifically, the model can predict 21 hand landmarks triplets comprising x, y, and relative depth. However, since Mediapipe estimated the depth from the RGB image, a different approach was chosen in the current investigation to obtain a more precise value for the third coordinate. Indeed, the z coordinate was computed considering the depth map provided by the Kinect camera. Specifically, the z coordinate for each landmark was obtained using the normalized x and y coordinates multiplied by the Kinect depth map shape.

All of the coordinates obtained for each landmark from each frame were then normalized to make the classifier less dependent on the position of the hand or the distance from the camera. Figure 3b shows an example of hand landmark acquisition through Mediapipe during execution in real-time of the application.

This framework was chosen as it demonstrated real-time inference speed on mobile GPUs with high prediction quality. In particular, the Mediapipe Hand model was tested on a custom database composed of in-the-wild, in-house, and synthetic images and reached an average precision of 94% in the best configuration of the palm detector and an MSE of about 10 for the hand landmark model, with a latency on the full pipeline of about 17 ms on a CPU and 12 ms on a GPU.



(**a**) MediaPipe set of landmarks (**b**) Acquired landmarks

Figure 3. Hand landmarks.

3.3. Hand Gesture Classification

The data collected using Mediapipe was split randomly between a training set, used to train the machine learning algorithm, and a test set, for performance evaluation, with a standard 80:20 division of the images between the two sets. In particular, a Random Forest classifier was used to identify the corresponding hand sign from the coordinates of the Mediapipe landmarks. A Random Forest is a machine learning algorithm used for classification and regression tasks. It operates by creating multiple decision trees during training and then combines their outputs to make a final prediction. Each tree is built using a random subset of the data and features, which helps to reduce overfitting and increase the model's accuracy. By averaging the predictions (in regression) or taking a majority vote (in classification), the Random Forest provides more robust and accurate predictions compared to individual decision trees. Specifically, a Random Forest with 100 trees and default parameters was involved for the current investigation. The Random Forest technique was chosen above other machine learning algorithms for its excellent performance in multi-class classification tasks [38], and its distinct advantages in terms of interpretability, training efficiency, and robustness [39]. According to their structure, Random Forest classifiers enable us to clearly distinguish which hand positions most influence specific gestures, allowing us to fine-tune and increase detection accuracy [32,40]. Other algorithms, such as Artificial Neural Networks (ANNs), particularly those with multiple layers, are more difficult to interpret due to their black-box nature [41]. Random Forests are also more computationally efficient than deep learning models, which, while powerful, need significant computational resources and specific hardware to train and deploy [42]. This efficiency enables Random Forests to conduct real-time gesture recognition on limited hardware, which is critical for accessible responsive applications. Compared to alternative tree-based models, such as XGBoost [43], Random Forests have the advantage of easier hyperparameter adjustment and lower computing overhead, making them a more practical choice [31,44]. Furthermore, Random Forests' ensemble structure provides robustness against overfitting, a common problem with short gesture datasets, can handle large datasets, and naturally manages the noise that is frequently present in gesture data [42]. As a result, Random Forests provide a balanced approach that ensures accuracy, speed, and

interpretability, making them ideal for recognizing sign language gestures, particularly on mobile devices [45].

Depending on whether the input data were an RGB or RGB-D image, the database structure to be provided as input to the classifier is slightly different. In particular, a set of 21 coordinates is provided in both the cases. When using a 2D image, each coordinate consists of a doublet referring to the value of x and y. In the case of an RGB-D image, the z coordinate referring to depth is also added. Therefore, each coordinate will be characterized by a triplet of x, y, and z. Consequently, 21 × 3 coordinates and 21 × 2 were, respectively, obtained for each 3D or 2D frame. As a result, two different classifier models were trained to operate with the corresponding data sources.

3.4. SIGNIFY Design and Development

The serious game has been conceived to accomplish a "learning by doing" approach, and has a twofold goal: on the one hand, it is an educational tool allowing for learning LIS, even from scratch; on the other hand, it allows for perfecting the LIS knowledge through a game. In this perspective, the game has two main scenes: the tutorial and the game itself, both accessible from an initial menu. In other words, the objective of the developed application is to teach LIS and, subsequently, apply the acquired knowledge by reproducing the learned gestures to guess the hidden word in a game of Hangman [46].

To provide users with an immersive and more dynamic experience that includes 3D information, Unity 3D has been chosen as the development platform. Unity 3D is a comprehensive cross-platform game engine developed by Unity Technologies, widely used for creating three-dimensional (3D) games and interactive experiences [22]. One of its features is the Mechanism Animation System, which facilitates the integration of complex animations to enhance visualization and effectiveness. Additionally, Unity 3D benefits from a large and active community of developers, offering tutorials, forums, and documentation. Unity 3D was chosen to incorporate animations into objects, allowing the use of animations created in Blender to demonstrate how to perform gestures from Italian Sign Language to users.

Unity 3D (front end) has been integrated integrated with a Python script (back end), where the developed Machine Learning-based algorithm classifies the recognized hand gestures through a .txt file, according to Figure 4. The executable created by Unity 3D opens and reads the file where the prediction has been written to use the information for the game, both for the tutorial and the hangman game, as will be detailed in the following subsections.

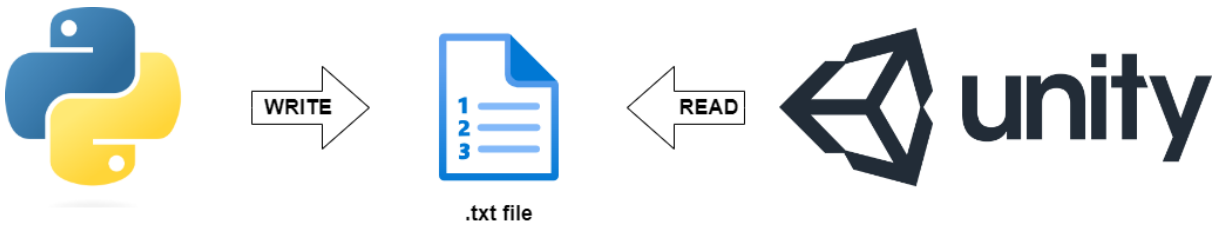


Figure 4. Communication diagram between Python script and Unity 3D.

3.4.1. The Tutorial Scene

The tutorial scene has been conceived to learn LIS [47]. The associated diagram is reported in Figure 5.

As represented, the Tutorial scene allows for selecting the menu button in order to switch to the Hangman Game scene once the user has enough confidence with the LIS. When the Tutorial starts, a message with all the needed information is provided; moreover, the camera is activated, and constantly updates the frames to feed the Machine Learning-based algorithm for the hand sign classification. If the performed sign is wrong, the user is suggested to retry, otherwise a tailored animation is displayed and a new letter is shown

for further training. It must be noticed that the same letter will not be displayed anymore in the current session. To motivate the user to learn and correctly reproduce the signs, a trophy system has been implemented, which can be earned only after a correct series of signs.

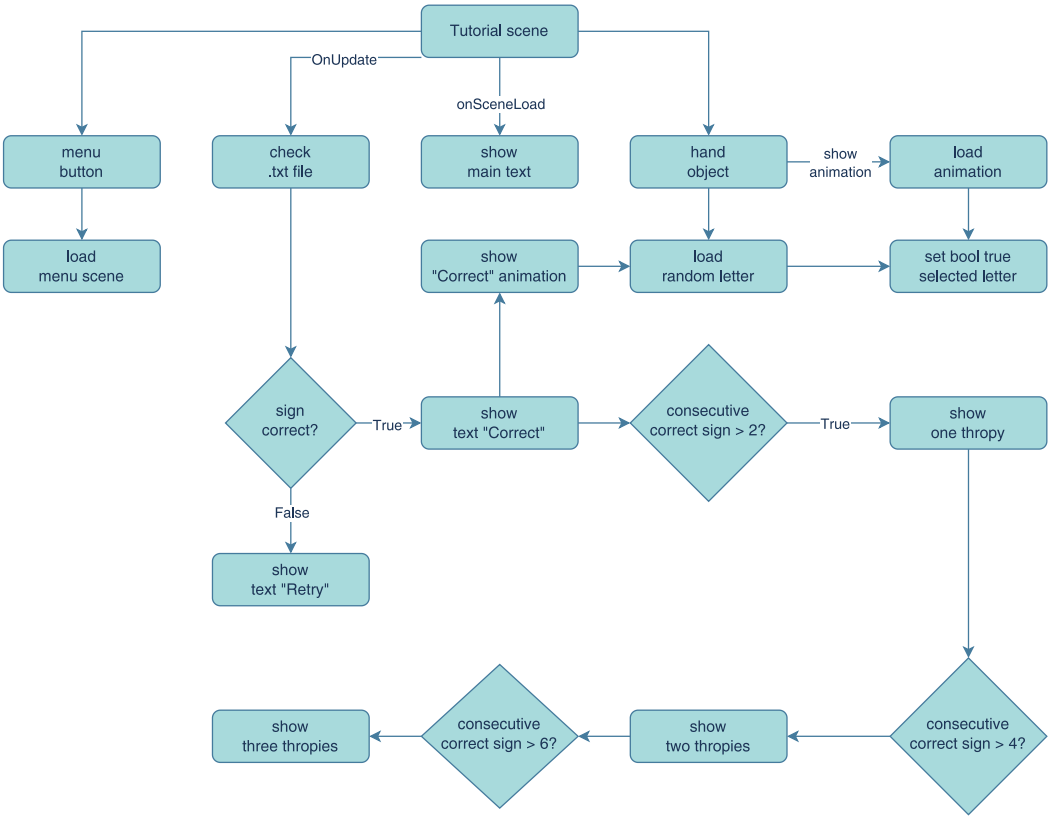
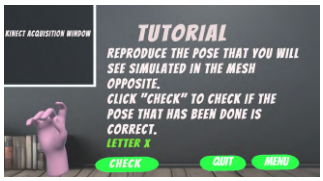
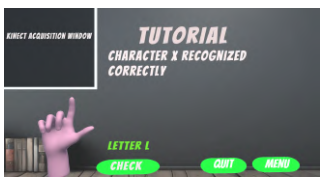


Figure 5. Diagram of the Tutorial scene.

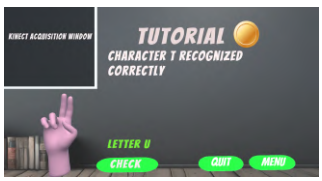
The User Interface (UI) consists of an upper left section, where the Python window is positioned. This window shows the camera image, i.e., the webcam or the Kinect Azure RGB-D video streams, with the overlaid prediction of the reproduced sign. Furthermore, UI is composed of a text box explaining to the user what needs to be performed to proceed, a button to close the application, and a button to proceed and check if the displayed sign is correct (Figure 6a). In the lower-left section, there is a simplified hand model that cyclically reproduces the required sign so that the user can understand, learn, and replicate the movement at her/his own speed. After understanding and reproducing the sign, the user can save it and, using the control button, and can verify through the updated text box if the reproduced gesture is correct (Figure 6b). Upon obtaining all of the trophies, a congratulatory message will be displayed, inviting the user to test the acquired skills in the Hangman game (Figure 6c).



(**a**) Starting image



(**b**) Correct hand sign



(**c**) Coin reward

Figure 6. Tutorial scene screenshots.

3.4.2. The Hangman Game Scene

The game scene is designed to test the user's knowledge of sign language by presenting a word to guess within a limited number of attempts, represented by the parts of the hangman's body. The diagram reporting the logic behind the adaptation of the Hangman Game is shown in Figure 7.

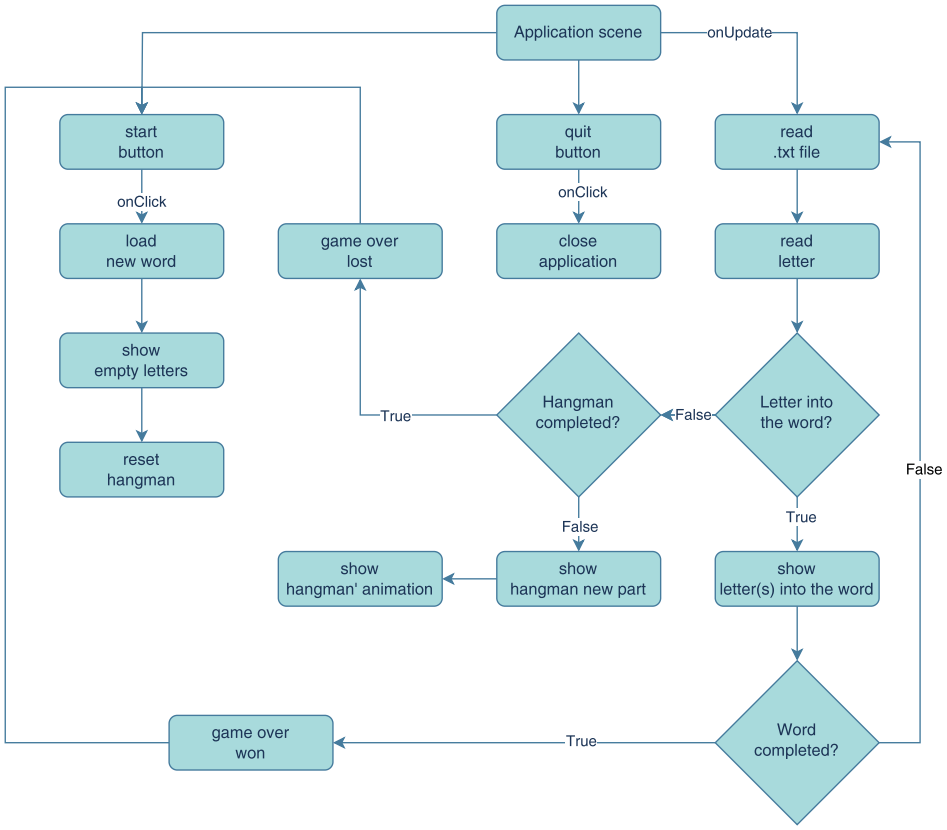
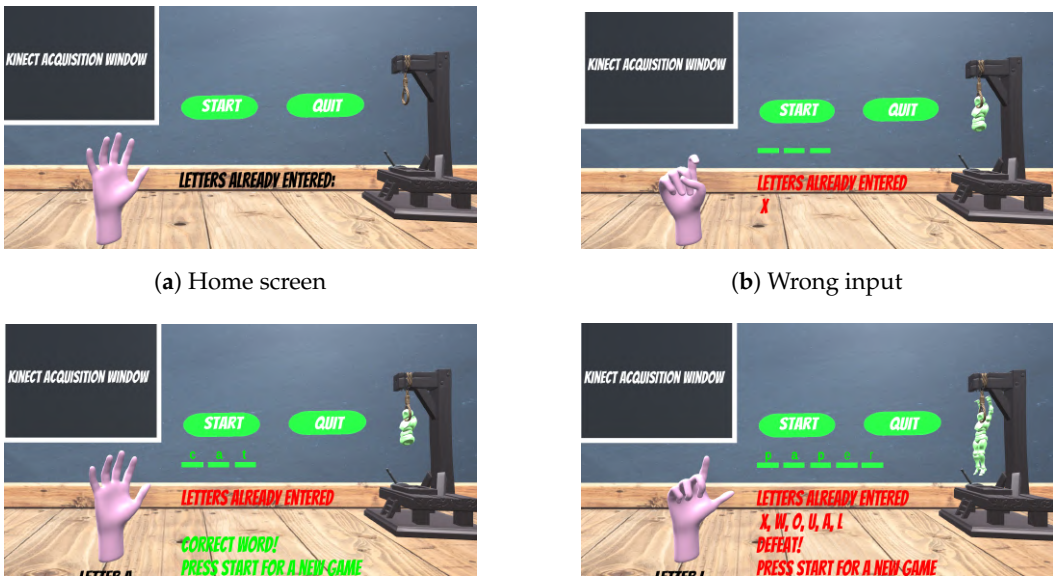


Figure 7. Diagram of the Hangman Game scene.

The Hangman Game scene allows to start the game or quit the application to roll back to the tutorial scene. Even in this scene, once it is deployed, the camera remains active to track the hand gestures and feed the Machine Learning-based algorithm with the acquired frames. The application automatically detects if the acquisition device is RGB-only or allows for the Depth stream. As will be shown in the next section, both of the options are suitable since gesture recognition rate is remarkable and real-time is granted even with RGB-D approach. If the performed gesture represents a letter within the word to be guessed, then the letter is displayed in the correct location and the list of the letters already inserted is updated. If the word is completed, then the game is won and another game can be started. Otherwise, if the word is not completed, then the user is allowed for performing another gesture. On the other hand, if the performed gesture is wrong, a check on the hangman status must be performed. If the hangman is not completed, another part of the body is shown; if the hangman is completed, the game is lost.

The UI includes an upper left section where the video stream is displayed, showing the image with the overlaid prediction of the reproduced sign to the user (Figure 8a). Next to it, there is a structure where the parts of the hangman will appear with the corresponding animation, enhancing the immersive experience and showing the user his/her hangman status. In the lower part of the screen, the letters already used are displayed, and in the lower left section, the animation of the last shown sign will be reproduced. This is a sort of double-check that the user can exploit for self-assessment. Once a sign is reproduced, the sign is be saved for providing the possibility of retracing the gaming experience. The

corresponding letter will be automatically inserted into the word or added to the list of incorrect characters, resulting in the appearance of a new part of the hangman (Figure 8b). Upon correctly completing the word, a caption will appear, prompting the user to start a new game (Figure 8c). In case of defeat, a different message with the same intent will be displayed (Figure 8d).



(**a**) Home screen (**b**) Wrong input

(**c**) Game won (**d**) Game lost

Figure 8. Hangman Game screenshots.

4. Results and Discussion

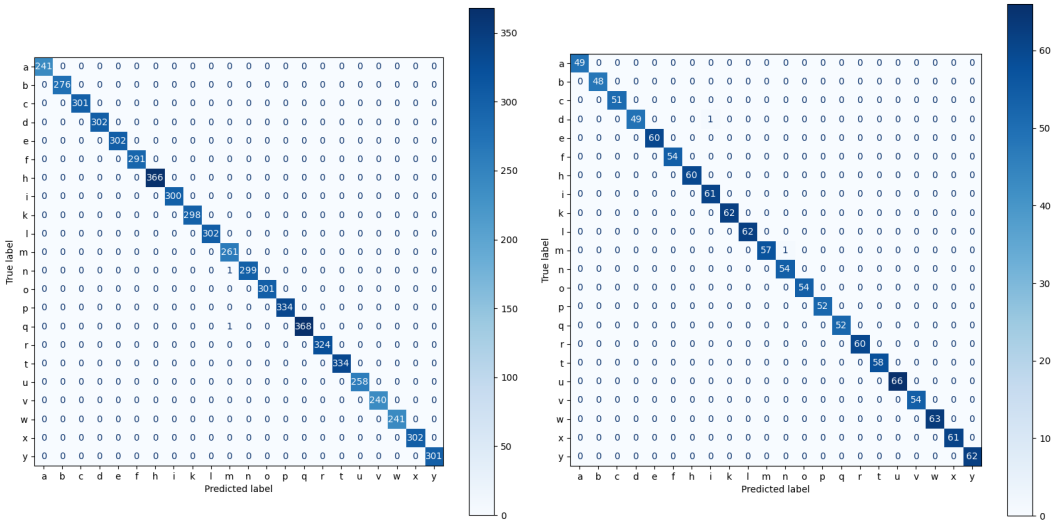
This Section is split into three Subsections to highlight the results obtained by the Machine Learning-based algorithm, the results obtained through the case-study making use of SIGNIFY in terms of usability and effectiveness, and the limitations that could inspire future improvements.

4.1. Automatic Hand Gesture Recognition

The results obtained from our analysis are truly remarkable, as illustrated by the confusion matrices in Figure 9, for both RGB-only and RGB-D acquisition processes.

The algorithm was tested on Windows 11 on an Intel(R) Core(TM) i5-9300H CPU @ 2.40 GHz and an NVIDIA GeForce GTX 1660 Ti. Both the RGB and RGB-D devices have demonstrated exceptional performance, showcasing their effectiveness in accurately classifying the data. The matrix highlights the high levels of precision and recall achieved, indicating a strong capability in distinguishing between classes. The capability of both the models to classify different hand signs with great accuracy is also proved by computing significant metrics, such as balanced accuracy and f1 m score [48]. Balanced accuracy takes into account both sensitivity (true positive rate) and specificity (true negative rate), providing a more comprehensive view of a model's performance across different classes. It mitigates the bias that can arise when one class dominates the dataset, ensuring that the model is assessed fairly, regardless of class distribution. The F1 m score, which is a variant of the F1 score that can accommodate multiple classes, offers a harmonic mean of precision and recall. This balance is crucial when dealing with imbalanced datasets, as it emphasizes both the ability to correctly identify positive instances and the minimization of false positives. According to the achieved results, the algorithm working with RGB images achieved 99.83% of balanced accuracy and 99.84% regarding F1 m score. On the other hand, RGB-D images allow for 99.98% and 99.98% concerning balanced accuracy and F1 m score, respectively. This achievement underscores the robustness of our approach, that further

overcome the already remarkable results obtained in the same field [20], and the potential for both the categories of devices.



(**a**) RGB-D model (**b**) RGB model

Figure 9. Confusion matrices.

The excellent performance achieved on the dataset has been similarly observed in the real-time case study described in the following paragraph. While some minor delays were noted during the application execution, it is important to emphasize that these delays are primarily attributable to the hardware used, such as the reliance on a USB 2.0 port instead of a USB 3.0 connection for linking the external camera to the PC. Despite these small setbacks, the overall results reaffirm the effectiveness of our approach, demonstrating that the high accuracy and reliability seen in the dataset translate well into real-world applications.

Although the application is designed for usage in a moderately controlled setting favorable for learning, further qualitative tests were performed to check the model's generalizability and adaptability under uncontrolled situations. Specifically, the application was tested indoors and outdoors under boundary circumstances such as fuzzy backdrops, suboptimal illumination, and varying hand orientation and distance from the camera. Figure 10 illustrates some of the obtained results, which allowed us to demonstrate the system's proper functioning, even under less controlled conditions, and its potential range of usage in various scenarios.

Figure 10. Qualitative tests in real-world uncontrolled settings. Different colors refer to different subgroups of hand landmarks.

The application operated in real-time on a qualitative level, but a more accurate analysis has been carried out to quantitatively assess the latency added for automatic gesture recognition. A human being has a reaction time that is around 150 ms [49]. We assumed that latency would be calculated by adding the processing times of the MediaPipe hand landmark library and the classification algorithm. On a GPU, the former can achieve

a latency of approximately 12 ms [37]. The latter was tested and was determined to have a maximum latency of 12 ms on our GPU. Considering the entire latency time is significantly less than the human reaction time, the algorithm's latency can be deemed negligible, as it does not interfere with the application's correct functioning.

When comparing RGB and RGB-D solutions, it is noteworthy that their performance levels are strikingly similar, with only subtle distinctions between the two approaches. However, the RGB-D solution exhibits a slight advantage, particularly in scenarios characterized by light variations. This added robustness allows the RGB-D approach to effectively address challenges that may cause performance drops in purely RGB systems. Furthermore, in previous works utilizing RGB-D, we observed that this approach reduces, though does not completely eliminate, the need to generalize the dataset [50,51]. This characteristic is crucial for applications requiring adaptability to diverse environmental conditions, reinforcing the value of incorporating depth information in enhancing classification accuracy and reliability.

The RGB-only solution offers the significant advantage of utilizing common devices like smartphone and tablet cameras, making it accessible and convenient for a wide range of users. In contrast, the RGB-D approach necessitates a more specialized 3D acquisition system, which can limit its applicability in everyday scenarios, and has been evidenced by the present study also by the need of creating an ad hoc database. However, advancements in technology, particularly with time-of-flight cameras, are rapidly changing this landscape. These innovations have led to improvements in both camera performance and compactness, making it increasingly feasible to integrate RGB-D capabilities into personal devices. As a result, the distinction in accessibility between RGB and RGB-D solutions is becoming less pronounced, potentially leveling the playing field for both approaches in these applications, and allowing it to be available to a wider range of users having devices with different hardware capabilities. This is in line also with previous findings related to the perceived user experience by learners using 2D and 3D gamified Virtual Reality on American Sign Language [22].

4.2. SIGNIFY Qualitative Assessment

As previously stated, the main objective of this work is the development of a serious game that can facilitate LIS learning for young children. To have a more robust evaluation of the application and to verify its usability and simplicity, a comprehensive assessment was required. In the previous section, the performance assessment of the automatic hand gesture recognition algorithm has been reported. In this section, the application has been evaluated in terms of usability, with a focus on its effectiveness, visual immediacy, and practical use in an educational context [52]. For this reason, the application was tested by a heterogeneous group of 11 children aged between 10 and 12 to evaluate the effectiveness of the game and its potential as a learning tool, under the supervision of three teachers. The most widely used evaluation methodologies for usability and engagement, System Usability Scale (SUS) [53] and User Engagement Scale (UES) [54], have been reworked to make them suitable for primary school children.

As shown in Figure 11, all users appreciated the tutorial section, since none of them were familiar with sign language, and only one of them expressed the need to revise gestures several times due to unfamiliarity with technology. Given the engagement and the rapid learning curve shown by the children, the teachers reported the gamification technique very useful, expressing the desire to use it for other purposes as well.

The graphical user interface was considered adequate by the majority of subjects, with 73% appreciating the UI and finding it effective, 18% finding it fairly effective, and the remaining 9% thinking it needs improvements. Eventually, 36% of the subjects considered the application easy to use, 28% encountered slight difficulties in use, and the remaining 36% found the application fairly easy to use (Figure 11).

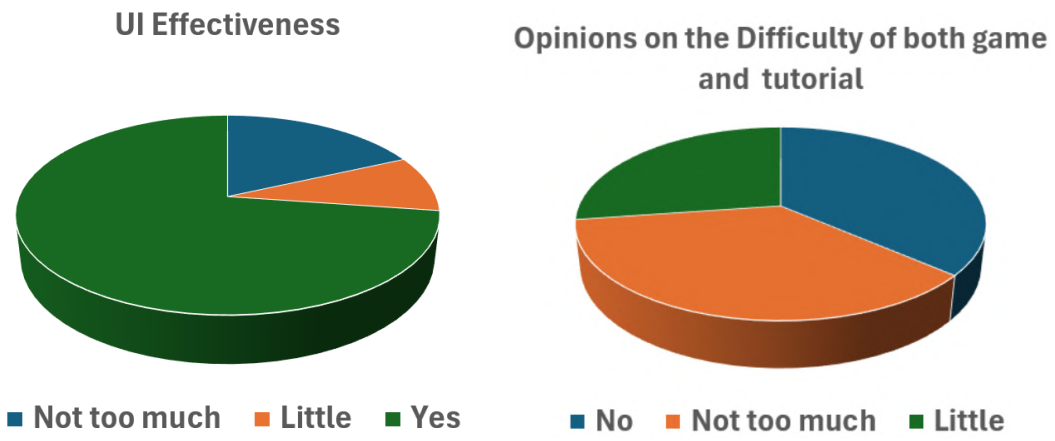


Figure 11. Questionnaire results. Allowed answers were: no, not too much, little, quite a lot, yes.

Therefore, unlike existing serious games in the literature, such as Magic Touch Math [18], the system for Portuguese sign language learning, and MatLIBRAS Racing [7], the proposed application has the great advantage of having been already validated from the actual end-users, children, and teachers. Indeed, the results have proven to be crucial in assessing the effectiveness of the proposed system, testing it directly on what would be the main target to highlight the feasibility of this approach. This difference is substantial in a context where automatic recognition has already proven to be particularly effective over the years [20] and, consequently, requires more attention from the perspective of integrating artificial intelligence with extended reality. In particular, it is this latter element that has already proven to be decisive in other contexts for facilitating the learning process through more engaging and immersive activities [55,56].

4.3. Limitations and Future Work

Although SIGNIFY resulted to be an effective tool both in terms of automatic hand gesture performance and application usability, there is still room for improvements. The hand sign dataset should be expanded by including images with different lighting conditions, varied and non-uniform backgrounds, and hand signs performed at different distances from the camera. This should be particularly significant, especially for the RGB-only database, since the RGB-D frames contains the 3D depth information useful to make the recognition more robust. Furthermore, the models are trained solely with signs made using the right hand. So, training the models with signs made using both hands will improve accessibility by ensuring the efficacy of recognition regardless of which hand is used. Another significant development will be the inclusion of dynamic letters [57] in the sign alphabet, such as *g*, *j*, *s*, and *z*, and also other signs that can be recognized only considering the dynamic nature of these gestures. This will likely require the inference to be performed on a brief video sequence instead of single frames.

Regarding the communication between Python and Unity 3D [58], the current application uses a text file to read and utilizes the information derived from the prediction of the model used in Python. However, this method of communication between the two channels is not optimal from a security point of view, and is the cause of minor delays if the available hardware is underperforming. A client–server architecture could mitigate this issue.

In addition, based on the results obtained from the questionnaires administered to potential users of the application, some criticalities emerged regarding the graphical component of the animations of the signs to be reproduced, which were deemed unrealistic. A possible development would be to use a more realistic hand model based on the 3D capture to make the hand movements more easily understandable to the end user in the tutorial section.

Finally, to reach a broader audience, developing a mobile version of the application will be a key goal. As already mentioned, both RGB and RGB-D methodologies could be

adopted, thanks to the integration of 3D depth acquisition cameras in the cutting-edge personal devices.

5. Conclusions

This study demonstrated the potential usefulness of serious games in educational contexts when combined with advanced machine learning technologies. Specifically, the applicability of cutting-edge gesture recognition technologies in children's Italian Sign Language (LIS) learning experience was explored. The comparison between the Azure Kinect and a standard RGB laptop camera reveals that both systems are capable of supporting gesture recognition for educational purposes, each with its own set of strengths and limitations. The inclusion of a tutorial section and a classic hangman game within the tool allows users to learn and practice the LIS alphabet interactively, making the learning process both educational and enjoyable. The positive feedback and engagement from primary school children during evaluations highlight the tool's effectiveness and potential impact in making sign language more accessible and engaging. In conclusion, this work underscores the significant contributions that technology can make in the field of inclusive education. By integrating gesture recognition with interactive gaming, this tool will not only support learning LIS, but also will promote a more inclusive and engaging learning environment.

Author Contributions: Conceptualization, G.C., P.G., G.L.P. and G.R.; methodology, G.C., P.G., G.L.P. and G.R.; software, G.C., P.G., G.L.P. and G.R.; validation, L.U., G.M., C.I. and E.V.; formal analysis, G.C., P.G., G.L.P. and G.R.; investigation, G.C., P.G., G.L.P. and G.R.; resources, E.V.; data curation, L.U., G.M. and C.I.; writing—original draft preparation, G.C., P.G., G.L.P. and G.R.; writing—review and editing, L.U., G.M., C.I. and E.V.; visualization, L.U., G.C., P.G., G.L.P. and G.R.; supervision, L.U., G.M., C.I. and E.V.; project administration, E.V.; funding acquisition, E.V. All authors have read and agreed to the published version of the manuscript.

Funding: This study was carried out within the Ministerial Decree no. 1062/2021 and received funding from the FSE REACT-EU—PON Ricerca e Innovazione 2014–2020. This manuscript reflects only the authors' views and opinions, neither the European Union nor the European Commission can be considered responsible for them.

Data Availability Statement: The data can be shared upon request. The data are not publicly available due to subjects' specific requests reported in the informed consent.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
ANN	Artificial Neural Network
CNN	Convolutional Neural Network
LIS	Italian Sign Language
ML	Machine Learning
RGB	Red Green Blue
RGB-D	Red Green Blue-Depth
SUS	System usability scale
SVM	Support Vector Machine
UES	User Engagement Scale
UI	User Interface

References

1. Mindess, A. *Reading Between the Signs Workbook: A Cultural Guide for Sign Language Students and Interpreters*; Hachette: London, UK, 2004.
2. World Health Organization. Deafness and Hearing Loss. In *Fact Sheet N 300*; World Health Organization: Geneva, Switzerland, 2015.

3. Edmondson, S.; Howe, J. Exploring the social inclusion of deaf young people in mainstream schools, using their lived experience. *Educ. Psychol. Pract.* **2019**, *35*, 216–228. [CrossRef]
4. Spence, C. How Learning a New Language Changes Your Brain | Cambridge English. 2022. Available online: https://www.cambridge.org/elt/blog/2022/04/29/learning-language-changes-your-brain/ (accessed on 25 June 2024).
5. Levesque, E.; Duncan, J. Inclusive education for deaf students: Pass or fail. *Deaf. Educ. Int.* **2024**, *26*, 125–126. [CrossRef]
6. Li, Y.; Chen, D.; Deng, X. The impact of digital educational games on student's motivation for learning: The mediating effect of learning engagement and the moderating effect of the digital environment. *PLoS ONE* **2024**, *19*, e0294350. [CrossRef]
7. Pontes, H.P.; Duarte, J.B.F.; Pinheiro, P.R. An educational game to teach numbers in Brazilian Sign Language while having fun. *Comput. Hum. Behav.* **2020**, *107*, 105825. [CrossRef]
8. Chouhan, T.; Panse, A.; Voona, A.K.; Sameer, S. Smart glove with gesture recognition ability for the hearing and speech impaired. In Proceedings of the 2014 IEEE Global Humanitarian Technology Conference-South Asia Satellite (GHTC-SAS), Trivandrum, India, 26–27 September 2014; pp. 105–110.
9. Assaleh, K.; Shanableh, T.; Zourob, M. Low complexity classification system for glove-based arabic sign language recognition. In Proceedings of the Neural Information Processing: 19th International Conference, ICONIP 2012, Doha, Qatar, 12–15 November 2012; Proceedings, Part III 19; Springer: Berlin/Heidelberg, Germany, 2012; pp. 262–268.
10. Shukor, A.Z.; Miskon, M.; Jamaluddin, M.; Ali@Ibrahim, F.; Asyraf, M.; Bahar, M. A New Data Glove Approach for Malaysian Sign Language Detection. *Procedia Comput. Sci.* **2015**, *76*, 60–67. [CrossRef]
11. Mohandes, M.; A-Buraiky, S.; Halawani, T.; Al-Baiyat, S. Automation of the Arabic sign language recognition. In Proceedings of the 2004 IEEE International Conference on Information and Communication Technologies: From Theory to Applications, Damascus, Syria, 19–23 April 2004; pp. 479–480.
12. Hongo, H.; Ohya, M.; Yasumoto, M.; Niwa, Y.; Yamamoto, K. Focus of attention for face and hand gesture recognition using multiple cameras. In Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580), Grenoble, France, 26–30 March 2000; pp. 156–161.
13. Zhang, H.; Wang, Y.; Deng, C. Application of gesture recognition based on simulated annealing BP neural network. In Proceedings of the IEEE International Conference on Electronic and Mechanical Engineering and Information Technology, EMEIT 2011, Harbin, China, 12–14 August 2011; pp. 178–181. [CrossRef]
14. Zhang, X.; Xiang, C.; Li, Y.; Lantz, V.; Wang, K.; Yang, J. A Framework for Hand Gesture Recognition Based on Accelerometer and EMG Sensors. *IEEE Trans. Syst. Man Cybern.-Part A Syst. Hum.* **2011**, *41*, 1064–1076. [CrossRef]
15. Chuan, C.H.; Regina, E.; Guardino, C. American sign language recognition using leap motion sensor. In Proceedings of the 2014 13th IEEE International Conference on Machine Learning and Applications, Detroit, MI, USA, 3–6 December 2014; pp. 541–544.
16. Qi, J.; Ma, L.; Cui, Z.; Yu, Y. Computer vision-based hand gesture recognition for human-robot interaction: A review. *Complex Intell. Syst.* **2024**, *10*, 1581–1606. [CrossRef]
17. Tolks, D.; Schmidt, J.J.; Kuhn, S. The role of AI in serious games and gamification for health: Scoping review. *JMIR Serious Games* **2024**, *12*, e48258. [CrossRef]
18. Kye, N.; Mustapha, A.; Samsudin, N. Gesture Recognition for Educational Games: Magic Touch Math. *IOP Conf. Ser. Mater. Sci. Eng.* **2017**, *226*, 012078. [CrossRef]
19. Zhan, Z.; Tong, Y.; Lan, X.; Zhong, B. A systematic literature review of game-based learning in Artificial Intelligence education. *Interact. Learn. Environ.* **2024**, *32*, 1137–1158. [CrossRef]
20. Lang, S.; Block, M.; Rojas, R. Sign Language Recognition Using Kinect. In *Artificial Intelligence and Soft Computing*; Rutkowski, L., Korytkowski, M., Scherer, R., Tadeusiewicz, R., Zadeh, L.A., Zurada, J.M., Eds.; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2012; Volume 7267, pp. 394–402. [CrossRef]
21. Soares, F.; Esteves, J.S.; Carvalho, V.; Lopes, G.; Barbosa, F.; Ribeiro, P. Development of a serious game for Portuguese Sign Language. In Proceedings of the 2015 7th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT), Ghent, Belgium, 30 October–1 November 2015; pp. 226–230. [CrossRef]
22. Wang, J.; Ivrissimtzis, I.; Li, Z.; Shi, L. The Impact of 2D and 3D Gamified VR on Learning American Sign Language. *arXiv* **2024**, arXiv:2405.08908.
23. Arooj, S.; Altaf, S.; Ahmad, S.; Mahmoud, H.; Mohamed, A.S.N. Enhancing sign language recognition using CNN and SIFT: A case study on Pakistan sign language. *J. King Saud Univ.-Comput. Inf. Sci.* **2024**, *36*, 101934. [CrossRef]
24. Ulrich, L.; Vezzetti, E.; Moos, S.; Marcolin, F. Analysis of RGB-D camera technologies for supporting different facial usage scenarios. *Multimed. Tools Appl.* **2020**, *79*, 29375–29398. [CrossRef]
25. Bora, J.; Dehingia, S.; Boruah, A.; Chetia, A.A.; Gogoi, D. Real-time Assamese Sign Language Recognition using MediaPipe and Deep Learning. *Procedia Comput. Sci.* **2023**, *218*, 1384–1393. [CrossRef]
26. Stamp, R.; Cohn, D.; Hel-Or, H.; Sandler, W. Kinect-ing the dots: Using motion-capture technology to distinguish sign language linguistic from gestural expressions. *Lang. Speech* **2024**, *67*, 255–276. [CrossRef]
27. Nimisha, K.P.; Jacob, A. A Brief Review of the Recent Trends in Sign Language Recognition. In Proceedings of the International Conference on Communication and Signal Processing, Melmaruvathur, India, 28–30 July 2020.
28. Uboweja, E.; Tian, D.; Wang, Q.; Kuo, Y.C.; Zou, J.; Wang, L.; Sung, G.; Grundmann, M. On-device Real-time Custom Hand Gesture Recognition. In Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), Paris, France, 4–6 October 2023; pp. 4275–4279. [CrossRef]

29. Amirgaliyev, Y.; Ataniyazova, A.; Buribayev, Z.; Zhassuzak, M.; Urmashev, B.; Cherikbayeva, L. Application of neural networks ensemble method for the Kazakh sign language recognition. *Bull. Electr. Eng. Inform.* **2024**, *13*, 3275–3287. [CrossRef]
30. Kavana, K.M.; Suma, N.R. Recognization of Hand Gestures Using MediaPipe Hands. *Int. Res. J. Mod. Eng. Technol. Sci.* **2022**, *4*, 4149–4153.
31. Bajaj, Y.; Malhotra, P. American sign language identification using hand trackpoint analysis. In *Proceedings of the International Conference on Innovative Computing and Communications: Proceedings of ICICC 2021*; Springer: Berlin/Heidelberg, Germany, 2022; Volume 1, pp. 159–171.
32. Ren Ewe, E.L.; Lee, C.P.; Lim, K.M.; Kwek, L.C.; Alqahtani, A. LAVRF: Sign language recognition via Lightweight Attentive VGG16 with Random Forest. *PLoS ONE* **2024**, *19*, e0298699. [CrossRef]
33. M. Donnici, G.M. Italian Sign Language Fingerspelling Recognition. 2018. Available online: https://github.com/maghid/italian_fingerspelling_recognition (accessed on 18 May 2024).
34. Fagiani, M.; Principi, E.; Squartini, S.; Piazza, F. A new Italian sign language database. In Proceedings of the Advances in Brain Inspired Cognitive Systems: 5th International Conference, BICS 2012, Shenyang, China, 11–14 July 2012; Proceedings 5; Springer: Berlin/Heidelberg, Germany, 2012; pp. 164–173.
35. Escalera, S.; Baró, X.; Gonzalez, J.; Bautista, M.A.; Madadi, M.; Reyes, M.; Ponce-López, V.; Escalante, H.J.; Shotton, J.; Guyon, I. Chalearn looking at people challenge 2014: Dataset and results. In Proceedings of the Computer Vision-ECCV 2014 Workshops, Zurich, Switzerland, 6–7, 12 September 2014; Proceedings, Part I 13; Springer: Berlin/Heidelberg, Germany, 2015; pp. 459–473.
36. Romeo, L.; Marani, R.; Malosio, M.; Perri, A.G.; D'Orazio, T. Performance analysis of body tracking with the microsoft azure kinect. In Proceedings of the 2021 29th IEEE Mediterranean Conference on Control and Automation (MED), Puglia, Italy, 22–25 June 2021; pp. 572–577.
37. Zhang, F.; Bazarevsky, V.; Vakunov, A.; Tkachenka, A.; Sung, G.; Chang, C.L.; Grundmann, M. Mediapipe hands: On-device real-time hand tracking. *arXiv* **2020**, arXiv:2006.10214.
38. Talla, S.; Venigalla, P.; Shaik, A.; Vuyyuru, M. Multiclass Classification Using Random Forest Classifier. *Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol.* **2019**, *5*, 493–496. [CrossRef]
39. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
40. Su, R.; Chen, X.; Cao, S.; Zhang, X. Random forest-based recognition of isolated sign language subwords using data from accelerometers and surface electromyographic sensors. *Sensors* **2016**, *16*, 100. [CrossRef] [PubMed]
41. Tang, D.; Taylor, J.; Kohli, P.; Keskin, C.; Kim, T.K.; Shotton, J. Opening the black box: Hierarchical sampling optimization for estimating human hand pose. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 3325–3333.
42. Goenawan, A.D.; Hartati, S. The Comparison of K-Nearest Neighbors and Random Forest Algorithm to Recognize Indonesian Sign Language in a Real-Time. *Sci. J. Inform.* **2024**, *11*, 237–244. [CrossRef]
43. Chen, T.; Guestrin, C. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM Sigkdd International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
44. Adeyanju, I.A.; Bello, O.O.; Adegboye, M.A. Machine learning methods for sign language recognition: A critical review and analysis. *Intell. Syst. Appl.* **2021**, *12*, 200056. [CrossRef]
45. Joshi, H.; Golhar, V.; Gundawar, J.; Gangurde, A.; Yenkikar, A.; Sable, N.P. Real-Time Sign Language Recognition and Sentence Generation. Available at SSRN: Joshi, Harita and Golhar, Vaibhav and Gundawar, Janhavi and Gangurde, Akash and Yenkikar, Anuradha and Sable, Nilesh P, Real-Time Sign Language Recognition and Sentence Generation. Available online: http://dx.doi.org/10.2139/ssrn.4992818 (accessed on 13 October 2024).
46. Logothetis, I.; Papadourakis, G.; Katsaris, I.; Katsios, K.; Vidakis, N. Transforming classic learning games with the use of AR: The case of the word hangman game. In Proceedings of the International Conference on Human-Computer Interaction, Bari, Italy, 30 August–3 September 2021; Springer: Berlin/Heidelberg, Germany, 2021; pp. 47–64.
47. Caserman, P.; Hoffmann, K.; Müller, P.; Schaub, M.; Straßburg, K.; Wiemeyer, J.; Bruder, R.; Göbel, S. Quality Criteria for Serious Games: Serious Part, Game Part, and Balance. *JMIR Serious Games* **2020**, *8*, e19037. [CrossRef]
48. Grandini, M.; Bagli, E.; Visani, G. Metrics for multi-class classification: An overview. *arXiv* **2020**, arXiv:2008.05756.
49. Thorpe, S.; Fize, D.; Marlot, C. Speed of processing in the human visual system. *Nature* **1996**, *381*, 520–522. [CrossRef]
50. Violante, M.G.; Marcolin, F.; Vezzetti, E.; Ulrich, L.; Billia, G.; Di Grazia, L. 3D facial expression recognition for defining users' inner requirements—An emotional design case study. *Appl. Sci.* **2019**, *9*, 2218. [CrossRef]
51. Ulrich, L.; Dugelay, J.L.; Vezzetti, E.; Moos, S.; Marcolin, F. Perspective morphometric criteria for facial beauty and proportion assessment. *Appl. Sci.* **2019**, *10*, 8. [CrossRef]
52. Malvasi, V.; Gil-Quintana, J.; Bocciolesi, E. The Projection of Gamification and Serious Games in the Learning of Mathematics Multi-Case Study of Secondary Schools in Italy. *Mathematics* **2022**, *10*, 336. [CrossRef]
53. Vlachogianni, P.; Tselios, N. Perceived usability evaluation of educational technology using the System Usability Scale (SUS): A systematic review. *J. Res. Technol. Educ.* **2021**, *54*, 392–409. [CrossRef]
54. O'Brien, H.; Cairns, P. An empirical evaluation of the User Engagement Scale (UES) in online news environments. *Inf. Process. Manag.* **2015**, *51*, 413–427. [CrossRef]
55. Damaševičius, R.; Maskeliūnas, R.; Blažauskas, T. Serious games and gamification in healthcare: A meta-review. *Information* **2023**, *14*, 105. [CrossRef]

56. Freire, M.; Serrano-Laguna, Á.; Manero Iglesias, B.; Martínez-Ortiz, I.; Moreno-Ger, P.; Fernández-Manjón, B. Game learning analytics: Learning analytics for serious games. In *Learning, Design, and Technology: An International Compendium of Theory, Research, Practice, and Policy*; Springer: Berlin/Heidelberg, Germany, 2023; pp. 3475–3502.
57. Wadhawan, A.; Kumar, P. Sign Language Recognition Systems: A Decade Systematic Literature Review. *Arch. Comput. Methods Eng.* **2021**, *28*, 785–813. [CrossRef]
58. Bustamante, A.; Belmonte, L.M.; Morales, R.; Pereira, A.; Fernández-Caballero, A. Video Processing from a Virtual Unmanned Aerial Vehicle: Comparing Two Approaches to Using OpenCV in Unity. *Appl. Sci.* **2022**, *12*, 5958. [CrossRef]