



Article

Design, Development, and a Pilot Study of a Low-Cost Robot for Child–Robot Interaction in Autism Interventions

Ilias A. Katsanis * , Vassilis C. Moulianitis and Diamantis T. Panagiotarakos

Department of Product and Systems Design Engineering, University of the Aegean, 84100 Syros, Greece; moulianitis@aegean.gr (V.C.M.); panagiotarakosd@gmail.com (D.T.P.)

* Correspondence: ikatsanis@aegean.gr; Tel.: +30-22810-97158

Abstract: Socially assistive robots are widely deployed in interventions with children on the autism spectrum, exploiting the benefits of this technology in social behavior intervention plans, while reducing their autistic behavior. Furthermore, innovations in modern technologies such as machine learning enhance these robots with great capabilities. Since the results of this implementation are promising, their total cost makes them unaffordable for some organizations while the needs are growing progressively. In this paper, a low-cost robot for autism interventions is proposed, benefiting from the advantages of machine learning and low-cost hardware. The mechanical design of the robot and the development of machine learning models are presented. The robot was evaluated by a small group of educators for children with ASD. The results of various model implementations, together with the design evaluation of the robot, are encouraging and indicate that this technology would be advantageous for deployment in child–robot interaction scenarios.

Keywords: social robots; socially assistive robots; autism spectrum disorder; child–robot interaction; robot-mediated intervention; robot design; machine learning



Citation: Katsanis, I.A.; Moulianitis, V.C.; Panagiotarakos, D.T. Design, Development, and a Pilot Study of a Low-Cost Robot for Child–Robot Interaction in Autism Interventions. *Multimodal Technol. Interact.* **2022**, *6*, 43. <https://doi.org/10.3390/mti6060043>

Academic Editor: Shruti Chandra

Received: 1 April 2022

Accepted: 31 May 2022

Published: 6 June 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The rapid progress of robotics offers great possibilities in the implementation of robots in human–robot interaction (HRI) settings and especially in populations relative with healthcare demands, elderly care, rehabilitation, physical impairments, etc. The use of socially assistive robots (SAR) aims at aiding these populations through social interaction, creating strong social bonds, and achieving measurable progress [1].

Recent studies have explored the potential use of SAR in special education and specifically for individuals with cognitive disorders such as autism. Autism spectrum disorder (ASD) is a developmental disorder (DD) that characterizes children with a set of eccentric patterns of behavior that appears in infant or toddler years [2], while the deficits fall into three categories: social interaction, communication, and stereotyped behavior [3]. Furthermore, due to the absence of permanent cure, specialized intervention plans are taking place to reduce the autistic behavior of children, while producing significant results [4].

The integration of SAR in autism interventions, which is referred to as robot-assisted therapy (RAT) or robot-mediated interventions (RMI), has been reviewed by several studies. The key elements that make SAR an effective interaction tool for autism interventions were explored in [5]. Robots usually are used as attractors, mediators, or as measurement tools in ASD interventions and diagnosis [6]. Huijnen et al. found that state-of-the-art robots could address many ASD interaction objectives, which are identified by ASD professionals [7]. In addition, it was found that children with autism perceived a robot as a social peer with mental states, interacting with it fairly [8]. Breazeal et al. state that children with autism retained what the robots told them and sought information for them during interaction tasks [9].

Furthermore, novel innovations in machine learning allowed robots to be equipped with advanced capabilities regarding the diagnostic purposes of autism such as identification and classification of symptoms [10–12].

During the review of the state-of-the-art robots that are used in autism intervention, it was realized that current commercially available robots are technically sophisticated, using advanced hardware equipment, while requiring a complex design to combine their behaviors with this hardware. This also means that the integration of software and algorithms is also complicated and it is achieved through the use of various expensive components. As such, these robots are quite expensive and require high financial resources that some organizations do not have.

To investigate this issue, this paper proposes the design of a low-cost, portable robot that could be implemented in autism interventions. The robot is equipped with inexpensive but powerful hardware that allows the implementation of advanced capabilities and modalities to the robot, which will enrich the interaction with ASD children. Exploiting the benefits of technological novelties, the robot is equipped with machine learning abilities that are being deployed straight to the hardware, minimizing both computational and power consumption demands. In this way, the cost of the robot is significantly reduced since the hardware and the software are forming an all-in-one solution with no separate equipment to be needed. Regarding the robot's machine learning abilities, two neural networks were developed to perform speech recognition and motion classification, giving the robot the basic capabilities for interactions, contributing to the adaption of its behavior accordingly. In addition, machine vision algorithms were developed for face detection and identification, which will provide the robot with better skills during tasks that require visual stimuli. This approach will benefit, on one hand, therapists, who can use a cheap assistive tool in their practices, and, on the other hand, children, who can communicate and interact with the robot more efficiently and naturally. Finally, a team of special educators evaluated the robot's prototype, together with its software prototype, giving valuable feedback and recommendations about its design and development.

The rest of the paper is organized as follows. Section 2 reviews the related work in the field of socially assistive robots for autism interventions. Section 3 introduces the design of the robot and the development of the technology and methods that equip the robot. Section 4 presents the evaluation results of the robot's and software's designs from a team of special education teachers who have daily experience with ASD children. Section 5 gives the discussion, and Section 6 gives the conclusion and future research work.

2. Related Work

2.1. Socially Assistive Robots and Autism Spectrum Disorder

Various designs of robots have been developed through the years, with different characteristics and expressions. Van Straten et al. stated that children with ASD are particularly attracted by technological artifacts, such as robots, due to the simpler way of interaction and the motivation that they establish [13]. Additionally, ASD children may possibly feel stress while interacting with other people due to unpredictable and complex human behavior. Thus, it has been reported that ASD individuals exhibit increased levels of communication abilities and social skills when interacting with robots rather than with humans [14,15]. Furthermore, Dautenhahn et al. found that children with ASD interact better with robots simply because they perceive them as predictable, controllable, and acceptable partners [15]. Usually, robots that are adopted in HRI studies for ASD interventions have anthropomorphic characteristics with realistic features, including integration of moving bodies, electronics, and software [16–18]. Other robots have cartoon-like characteristics with various interaction modalities and channels, capable of interacting without constraints [19,20]. In addition, robots with animal-like characteristics have been developed, exploiting the benefits of animal caring in social HRI tasks [21–23]. Finally, robot designs that resemble toys have been explored, forming different HRI settings and goals [24]. A collection of the various robot shapes and designs can be seen in Figure 1. The

presented robots have different appearances and morphological characteristics that distinguish each other, allowing ASD children to identify the potential social cues and generalize the learned skills towards human–human interactions [25], either when interacting with humanoid robots or with robotic animals. However, a balance between the appearance and interaction cues should be achieved, due to sensory overload problems that these children may face [26].



Figure 1. A selection of robots applied in autism interventions. (a) Robot Kaspar, (b) NAO robot, (c) Flobi robot, (d) TEO robot, (e) Probolino, (f) Keepon robot, (g) Aibo robot dog, (h) Roball robotic toy, (i) dinosaur robot Pleo.

3. Materials and Methods

In this section, the design and development of the robot is presented. First, the mechanical design process of the robot is described, including the design specifications and requirements, and the final detail design is also described. Second, the technological development of the robot is introduced, presenting the robot's hardware and the complete process of programming, in terms of software and algorithms.

3.1. Robot's Design

The design process of the robot is based on the functional requirements, characteristics, and constraints that were set and emerged from the observation of other solutions in HRI. Furthermore, a team of eight educators, who work daily with ASD and other DD individuals in a public school in Greece, was involved in the design process, reviewing the prototype of the robot. It should be noted that they did not have any prior experience in the use of robots or in robotics; thus, they did not introduce any bias into the procedure.

3.1.1. Design Requirements, Characteristics, and Restrictions

The design requirements, characteristics, and constraints are presented in Table 1, which presents the information under which the robot will be designed.

Table 1. Requirements, characteristics, and restrictions of robot development.

Requirements	Characteristics	Constraints
Convey feelings	Removable parts	Low cost, under 250 euros
Facial expressions	Portable	Designed with curves
Tactile interaction	Lightweight	Use of soft materials
Perform movements	Robust and solid	Speed limitations
Produce voice and sounds	Durable	Use of mild colors
Speech recognition	Low space footprint	Not too high sounds
Face recognition	Environmentally friendly	Not exceeding 250 mm height
		Not exceeding 1.2 kg weight

3.1.2. Mechanical Design of the Robot

Figure 2 presents the mechanical drawing of the robot's shape with its dimensions. It is a small robot, with a form capable of triggering natural interaction, expressing states that enable emotional interaction; its small size allows users to handle it nicely. Moreover, it allows personalization due to its modular characteristics, offering great opportunities in intervention plans. The robot has three DOF in total, one per arm and one in the torso.

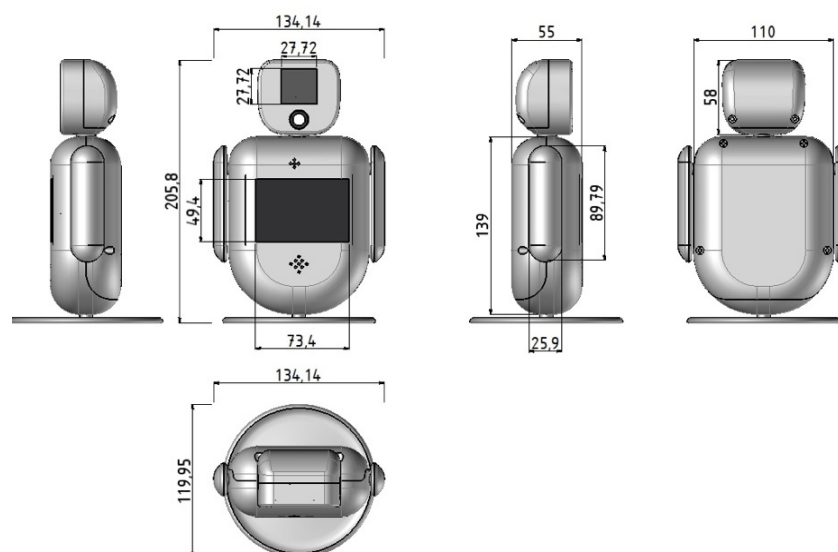


Figure 2. Robot's form, size, and the relevant dimensions.

In Figure 3, the placement of the robot's parts is presented. On the front side, the microphone and the speaker are located behind two spots, aiming at a more accurate acquisition of speech and sound. A vibration motor touches the front frame, so the transmission of any oscillation can be easily understood. Responses to violent touches, such as hits or falls, are recognized by the MCU's accelerometer, through the use of a trained neural network. Figure 4 presents, in more detail, the location of the electronic parts inside the body of the robot. Figure 5 shows the 3D-printed prototype of the robot.

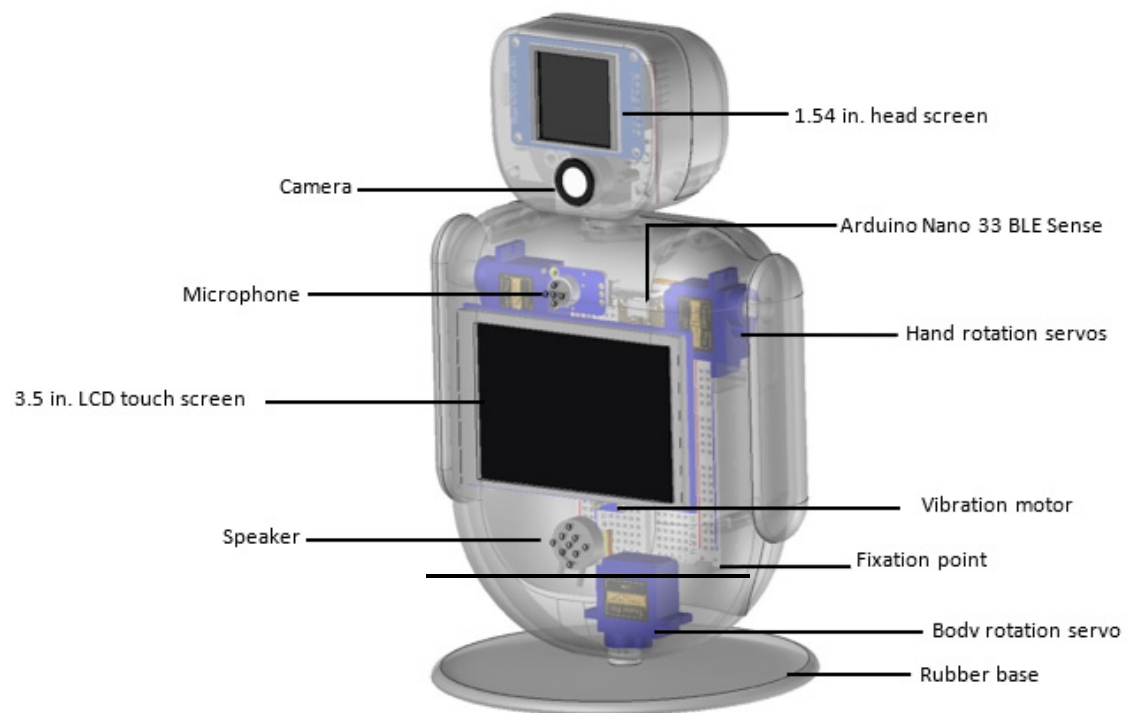


Figure 3. Robot's anatomy with the electronic and hardware parts.

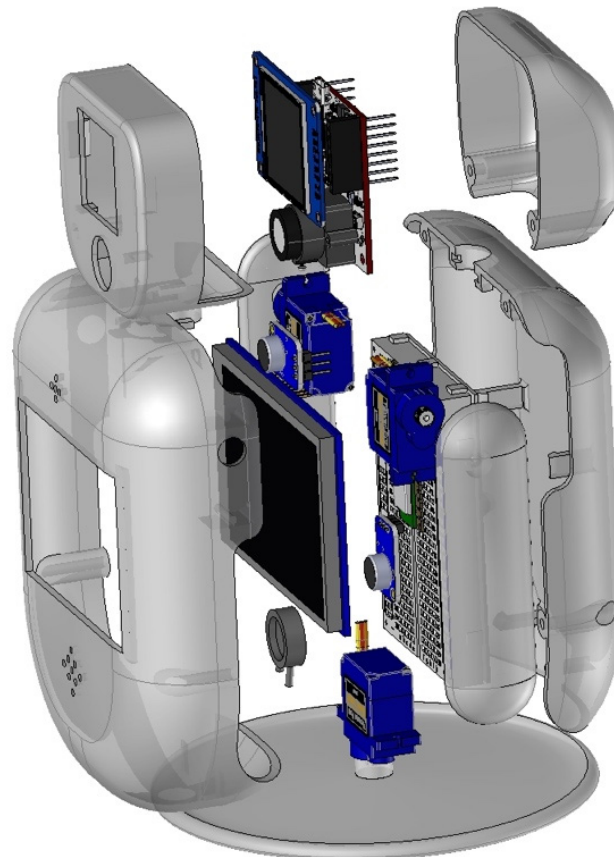


Figure 4. Exploded view of robot, with the electronic components and the fixations of the removable parts.

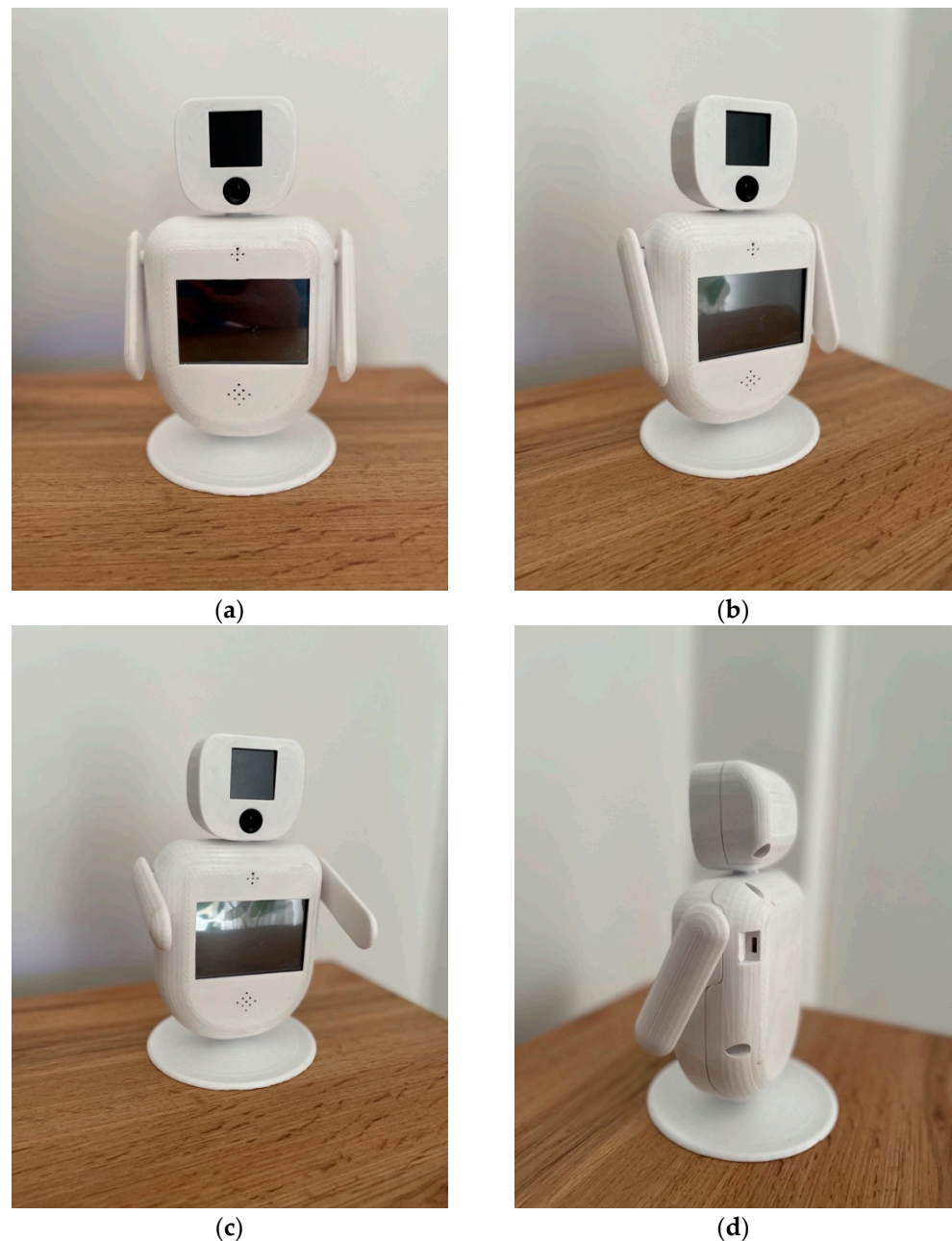


Figure 5. The 3D-printed robot prototype front (a), hand movement (b,c), and side view with MCU port and connection wholes (d).

3.2. Robot's Hardware and Electronics

Low-cost and efficient hardware components equip the robot. A powerful microcontroller unit (MCU) is used for controlling the robot and implementing the machine learning models. The list of all electronic parts that are used for the robot is presented below:

- **MCU:** An Arduino Nano 33 BLE Sense is used. The unit contains several sensors, with the most important for the proposed robot to be the nine-axis Inertial Measurement Unit (IMU). Moreover, it can be programmed with the Python programming language.
- **Camera:** A state-of-the-art machine vision camera with a 115° field of view is used. The camera is capable of performing vision tasks such as face, object, and eye tracking and recognition, which are essentials for the interaction tasks.

- **Screens:** Two screens are integrated on the robot. The first one is mounted on its head, displaying basic emotions such as happiness, sadness, surprise, fear, and anger. The second one is mounted on the center of the robot and allows tactile interaction.
- **Sensors:** The robot uses one vibration motor, offering the sense of shiver when the robot is being hit or shake, and a gyroscope and accelerometer from the MCU, detecting hit and fall.
- **Servos:** Three MG90S 2.8 kg micro servos with metal gears, due to their small size and their elastic response to external forces, are used. These motors are used to move the hands and torso.
- **Audio:** One loudspeaker is implemented, expressing the voice of the robot, and a microphone with a 60x mic preamplifier is used to capture and amplify sounds near the robot.

Table 2 summarizes the characteristics and the equipment that are used for the development of the robot.

Table 2. Summary of the equipment that is used for the development of the robot.

Size	205 × 134.1 mm	Servos	3 × MG90S 2.8 kg
Weight	1.2 kg	Gyroscope and Accelerometer	3-axis LSM9DS1 IMU sensor
Port	USB 2.0 Micro-B	Speaker	1 × 3 W loudspeaker
Camera	640 × 480 32-bit	Microphone	Electret 60× min preamplifier
Head Screen	1.54" 240 × 240	Central screen	3.5" touch LCD screen 420 × 320

3.3. Instructional Technology

In this section, the robot's control software prototype, the detailed development of the ML models, computer vision algorithms, and the deployment on the MCU are presented.

3.3.1. Software Prototype

The prototype of the robot's software was developed in order for specialists to assess and evaluate its design and overall user experience. The software prototype was designed to provide the necessary information about the intervention and children's interaction with the robot. Additionally, educators are capable of controlling the robot with a set of predefined functions, acquiring data about the progress of the children, the functionality of the robot, and the related team. Figure 6 presents the main page of the software.

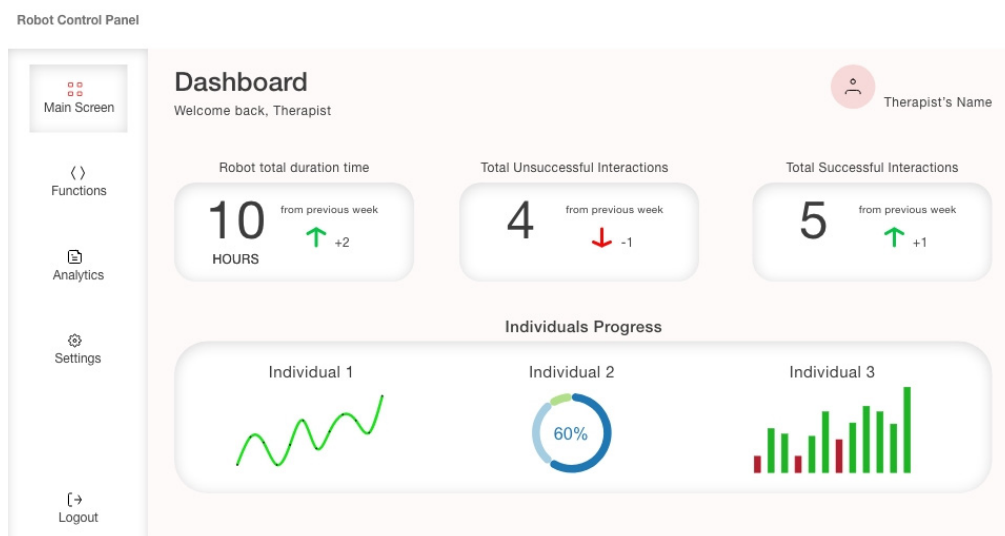


Figure 6. The main screen of robot's control software.

For the development of the software prototype, the necessary information is displayed with the use of simple diagrams that are easy to use and interpret by the users. Furthermore, a few restrictions were set on the users so that changes can be performed under consideration. The software's assessment results by the team of specialists are presented in the Section 4 of the paper.

3.3.2. Robot's Main Capabilities

The main capabilities of the robot include speech recognition and the classification of specific motions. Regarding speech recognition, the robot is able to understand specified commands that will be captured from its microphone. This capability allows the robot to participate in social initiation tasks, in which the child will be engaged in speech development assignments that will be determined by the specialist. For this purpose, a neural network was trained with a set of keywords that correspond to various functions of the robot. The dataset consisted of four main keywords, forming a minimum interaction with the robot, and included the *"hey_robot"* keyword that activates the robot, the *"dance"* keyword that enables the robot to dance, making specific movements with its body and hands, the *"photo"* keyword that makes the robot take a photo and save it in a specific folder, and, finally, the *"goodbye"* keyword, which is a farewell to the robot. Additionally, labels including *"unknown"* words and *"noise"* were used to train the NN against faulty inputs that may be captured while increasing its accuracy. In this paper, the keywords used for the development and training of the NN were in English, but, in the future, they will be captured in Greek. For more details about models' development process, please see the provided supplementary materials.

Concerning the classification of specific motions, a neural network was trained with a dataset of three classes including the acquired data from the microcontroller's IMU. The three classes in the dataset were *"stable"*, where the MCU is in idle position, resembling that the robot stands still, *"hit"*, where the captured motions resemble when the robot is being hit, and *"fall"*, where the captured movements resemble a possible fall of the robot. With the recognition of these motions, the robot is able to express its feelings about its current situation.

Further to the development of the previous capabilities, the robot is able to perform face tracking and identification using its camera. This ability allows the robot to participate in eye contact tasks, helping children to develop social communication abilities through conversations or other short tasks that would be defined by a therapist. Firstly, the robot detects a person's face and then tries to identify with whom it may interact. A set of known and unknown faces is created, training a machine learning recognizer, which, through the use of specific algorithms, is able to identify the person. This capability will help the robot to adapt its behavior accordingly and adjust its movements and motions to the current state of the child. The detailed development of neural networks and machine vision algorithms is described in the next section.

3.3.3. Machine Learning Models' Development

This section presents the complete process of ML models, starting from data acquisition and ending with the development and deployment of the neural networks that are capable of detecting specific words and classifying motions in real time. The process of the models' development was achieved through the use of Edge Impulse Studio [27]. For each neural network, the following procedure was followed: (1) data preparation, (2) data preprocessing, (3) feature extraction, and (4) neural network classifier training.

Concerning speech recognition, the data of this model were acquired through a mobile phone (iPhone 12), due to its ability for capturing audio spatially, and were uploaded to the Edge Impulse tool. Two datasets were developed; the smaller dataset was developed to train the neural network faster while the bigger one contained more variety in the data. Once this process was completed, the data were divided into approximately 360 samples

for each class. For this dataset, a refinement process was followed, where deteriorated and duplicate samples were removed, improving the model's accuracy.

Once the dataset was prepared, the audio processing parameters were set. The window size of the data that were processed per classification and the frequency were set to 1000 ms and 16,000 Hz, respectively.

For the processing of sound, the Mel frequency cepstral coefficients (Audio MFCC) technique was selected; this extracts the coefficients of a raw signal [28]. In this step, the raw audio signal was taken and, using signal processing, the cepstral coefficients were generated. Figure 7 presents the cepstral coefficients for the keywords *hey_robot*, *dance*, *photo*, and *goodbye*, as presented in the Edge Impulse tool.

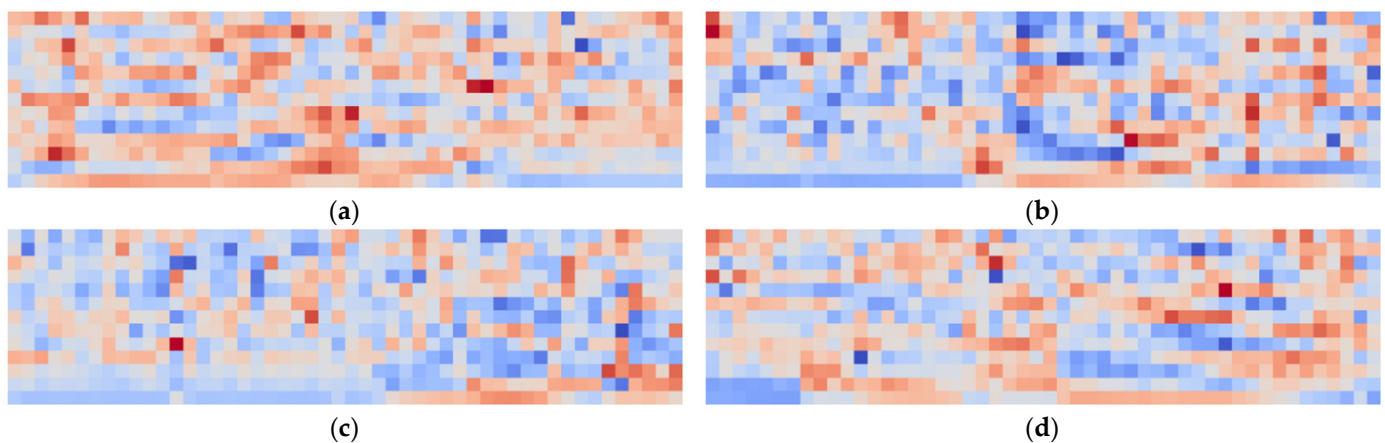


Figure 7. Cepstral coefficients for (a) *hey_robot*, (b) *dance*, (c) *photo*, and (d) *goodbye* labels. All coefficients were in a 1000 ms window.

Once the feature extractions were generated, the NN was developed and trained on these features using a TensorFlow framework [29]. The dataset was divided for training and testing; 80% of the speech samples were used for training, and the remaining 20% were used for testing the NN. The model's architecture is presented in Figure 8; its parameters were the following:

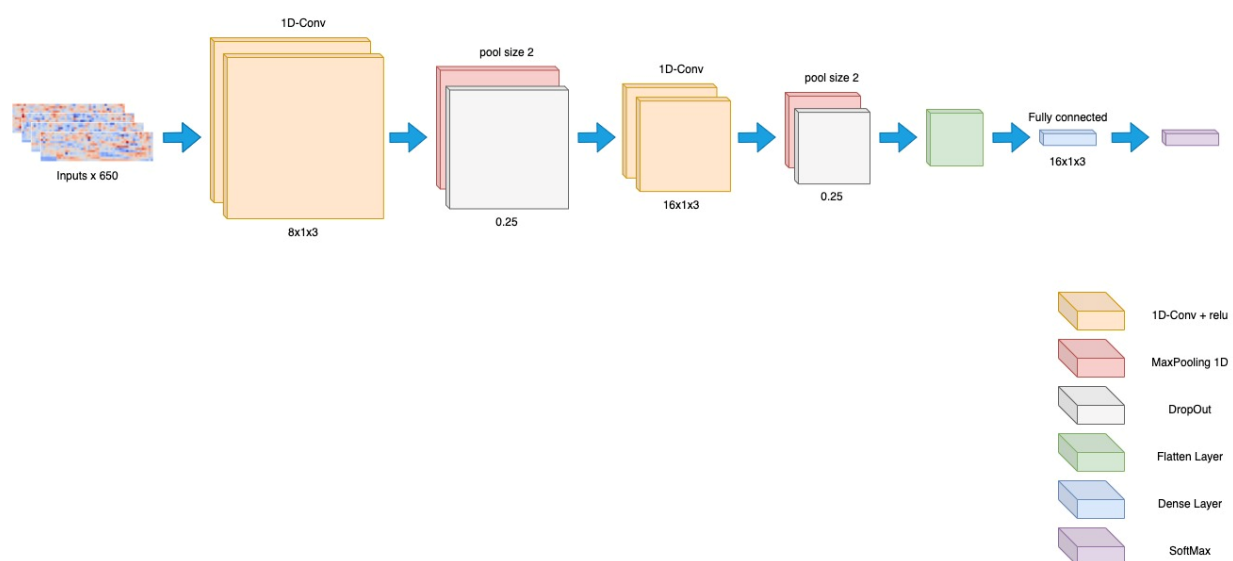


Figure 8. The architecture of the speech detection neural network.

Architecture:

- The input layer had 650 inputs that came from the feature extraction, providing the MFCCs of the raw data.
- Layer 1: a 1D convolutional layer was created with eight neurons and three kernel sizes, which was the length of the 1D convolutional window, one convolutional layer, and the Relu activation function.
- Layer 2: a MaxPooling 1D layer was set with a pool size of 2; the maximum element from each region of incoming data covered by the filter was selected.
- Layer 3: a dropout layer was created with a value of 0.25; 25% of input units will drop at each epoch during training, preventing the NN from overfitting.
- Layer 4: a 1D convolutional layer was set with 16 neurons, with the same kernel size and activation function as previously stated.
- Layer 5: a max pool 1D layer was set, as in Layer 2.
- Layer 6: a dropout layer was set, as in Layer 3.
- Layer 7: a flatten layer was created, converting the data into a one-dimensional array for inputting it to the next layer. The output of the convolutional layers was flattened and connected to the final classification layer, which was a fully connected layer.
- Layer 8: a dense layer was created to densely connect the NN layers, where each neuron in this layer received input from all the neurons of the previous layers.

Parameters:

- Parameter 1: the Adam optimizer was selected due to its little memory requirement and its computationally efficiency [30]; the learning rate, which means how fast the NN learns, was equal to 0.005. The batch size was defined to 32.
- Parameter 2: for training the neural network, the Categorical Crossentropy loss function was selected, accuracy was defined as metric in order to calculate how often the predictions equal the labels, and epochs were set to 500.
- Parameter 3: 20% of the samples were set for testing.

Regarding the motion classification model, the data were collected from microcontroller's three-axis accelerometer and consisted of three main labels. The collected data were 12 min of training data and 3.30 min of test data.

For the processing of data, the window size was set to 2000 ms and the spectral analysis was used, analyzing the repetitive motion of the accelerometer data while extracting the frequency and power of the signal over time [31]. Figure 9 presents an example of the frequency and power of a sample in the “fall” label, as generated in the Edge Impulse tool.

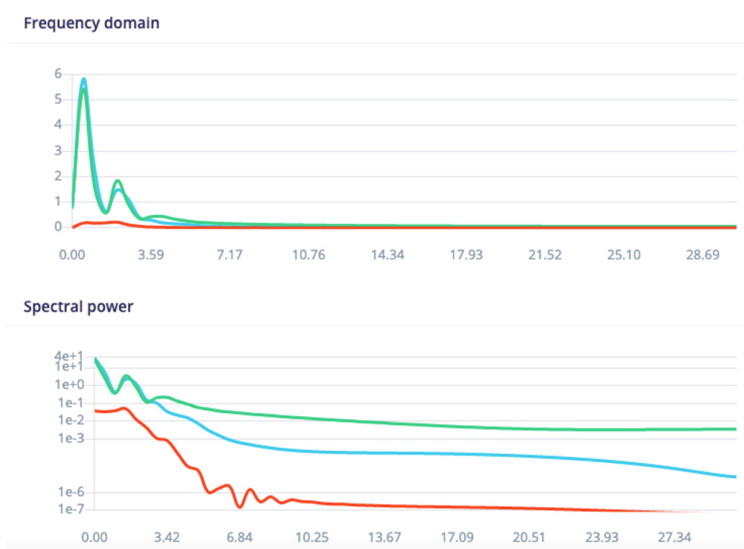


Figure 9. Frequency and power of a “fall” sample in a 2000 msec window, after the processing of the signal.

The proposed model that was developed for the motion classification was smaller than the previous one. Figure 10 shows the architecture with the following parameters:

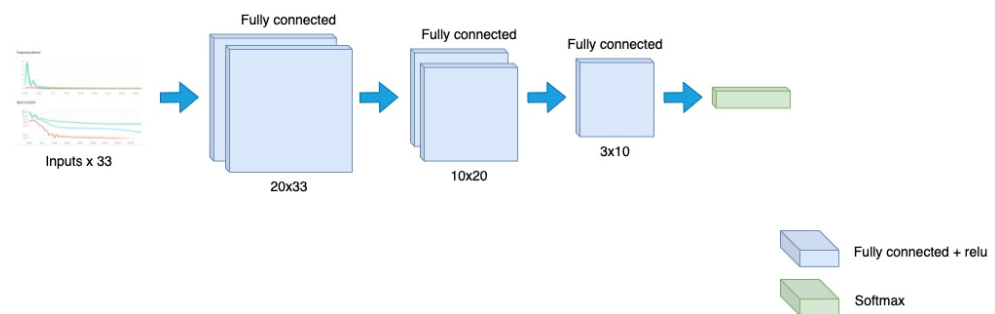


Figure 10. The architecture of the motion classification neural network.

Architecture:

- Layer 1: an input layer with 33 features that came from the spectral analysis, providing the frequency and spectral power of a motion.
- Layer 2: a dense layer with 20 neurons was set together with the Relu activation function.
- Layer 3: another dense layer with 10 neurons was set together with the Relu function.
- Layer 4: A final dense layer was used to connect the previous layers, with the SoftMax activation function.

Parameters:

- The Adam optimizer was selected with a learning rate of 0.0005, and the batch size was set to 32.
- The Categorical Crossentropy loss function was selected, accuracy was defined as metric, and the epochs were equal to 100.
- The data for testing were defined at 20% of the dataset (the remaining 80% of the dataset was used for training).

After the development of the models, they were deployed on the MCU. Since the development took place on TensorFlow, they had to be converted to TensorFlow Lite (TF Lite) models to fit onto the MCU using the TinyML technique [32,33].

Regarding the speech detection model, a two-buffer mechanism was set for implementing continuous audio sampling. In this method, audio is sampled in parallel with inference and output, and, while inference is still running, sampling continues in the background. Thus, no sample is missed, and new audio data are captured and analyzed. Using continuous inferencing, small buffers that contain audio samples are passed to the inferencing process, with the oldest to be removed and the new to be inserted at the beginning. The continuous inference process is presented in Figure 11. Concerning the motion classification model, it was implemented with the same two-buffer technique.

A confidence threshold was set to 80% for both models, meaning that if a feature was detected by the model with at least 80% confidence, this feature matched the actual label of the class.

The deployment of the model on an Arduino microcontroller followed an implementation of different scripts based on the Python language and OpenCV library. These scripts contained functions for face detection and recognition and the integration with the keyword detection model.

For the face detection task, the Haar cascade classifier was used based on the work made in [34]. Using OpenCV library and a pretrained classifier, the detection of the face was implemented by a detection method.

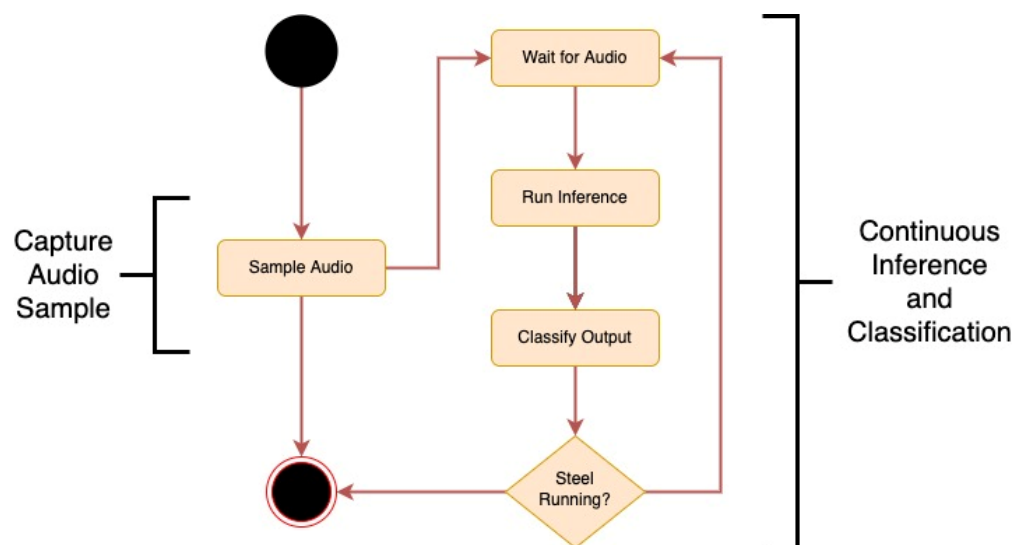


Figure 11. Activity diagram of inferencing, using the two-buffer mechanism.

Regarding the face identification task, a dataset of faces, with the name of each person as an ID, was created and was used to train an ML recognizer based on the histogram of oriented gradients (HOG) [35]. Firstly, the image was converted to grayscale and every single pixel on it was processed. Then, the gradients to the direction in which the image was darkening were determined, and the part of the image that looked similar to the created HOG was created. Next, the face landmark algorithm [36] was used to deal with the various face orientations, extracting 68 specific landmarks of the face, where eyes and lips were always in the same position, as shown in Figure 12.



Figure 12. Facial landmarks of (a) frontal face, (b) side face.

Once the landmarks were acquired, a neural network was trained to generate measurements, comparing known faces against unknowns. Pre-trained networks can be found at [37], which can be used to produce encoded values of these measurements. Finally, a trained classifier, which takes measurements from a new face, determined which known face matched the closest, with a confidence threshold of 85%. The result of the classifier is a person's name, and an example is presented in Figure 13.

After developing the models and deploying them to the microcontroller, the results of the models' performances, inferencing, and the educators' review are presented in the next section.

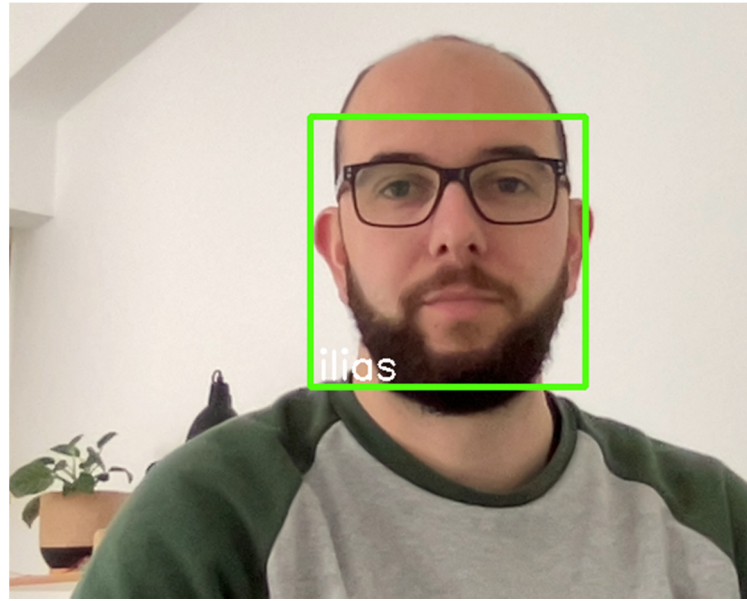


Figure 13. The resulting image from the trained classifier with the name of the person.

3.3.4. Evaluation Metrics

The speech and motion classification models' performance were evaluated using total accuracy ($TAcc$), accuracy (ACC), sensitivity or recall (TPR), specificity (TNR), precision (PPV), and F1-score (F_1).

Total accuracy is the model's overall accuracy and was determined by the fraction of the model's correct predictions for each class to all samples of the model:

$$TAcc = \frac{\sum_{i=1}^m p_i^c}{\sum_{i=1}^n s_i} \quad (1)$$

where p_i^c is the i -th correct prediction and s_i is the i -th sample.

Accuracy was calculated for every model class through the false/true positive/negative result of the classifier. It is defined by:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

where TP , TN , FP , and FN are the sums of the true positive, true negative, false positive, and false negative results of the classifier for every class, respectively.

Recall is the metric that evaluates the true positive rate of every class:

$$TPR = \frac{TP}{TP + FN} \quad (3)$$

Specificity is the metric that evaluates the true negative rate of every class:

$$TNR = \frac{TN}{TN + FP} \quad (4)$$

Precision indicates the positive predictive values of every class:

$$PPV = \frac{TP}{TP + FP} \quad (5)$$

The F1-score or harmonic mean indicates an average rate of the model's accuracy for each class:

$$F_1 = 2 \frac{Precision * Recall}{Precision + Recall} \quad (6)$$

4. Results

4.1. Speech Detection Model

Regarding the speech detection model, the total samples for the *dance*, *goodbye*, *hey_robot*, *noise*, *photo*, and *unknown* labels were 57, 54, 91, 67, 75, and 53, respectively. Table 3 presents the confusion matrix of the model. Based on the values of the matrix, $TAcc = 97.48\%$. According to Equation (1), $\sum_{i=1}^m p_i^c$ is the sum of the main diagonal values of the matrix, while $\sum_{i=1}^n s_i$ is the sum of the model's total samples.

Table 3. The confusion matrix of the model.

	Predicted Labels						
	DANCE	GOODBYE	HEY_ROBOT	NOISE	PHOTO	UNKNOWN	UNCERTAIN
Actual Labels	Dance	56	0	0	0	0	1
	Goodbye	0	53	0	0	0	1
	Hey_robot	0	0	91	0	0	0
	Noise	0	0	0	64	1	2
	Photo	0	0	0	74	0	1
	Unknown	0	0	0	2	49	2

In Table 4, the metrics defined in Section 3.3.4 were calculated for each class, when the model was being tested for the same data.

Table 4. Metrics' results for each class in model testing.

Class	ACC	TPR	TNR	PPV	F_1
<i>dance</i>	0.997	0.982	1.00	1.00	0.991
<i>goodbye</i>	0.997	0.981	1.00	1.00	0.991
<i>hey_robot</i>	1.00	1.00	1.00	1.00	1.00
<i>noise</i>	0.992	0.955	0.994	0.970	0.962
<i>photo</i>	0.997	0.987	1.00	1.00	0.993
<i>unknown</i>	0.990	0.925	0.997	0.980	0.951

It can be seen from Table 4 that the *hey_robot* class had the highest percentage for every metric. The *Noise* and *unknown* classes had the lowest rates in *TPR* metric, which indicates that there were fewer true positives; consequently, the F_1 for these classes was lower, too.

4.2. Motion Classification Model

Figure 14 presents the *TAcc* and the *TPR* for each class based on Equations (1) and (3), respectively, of the unoptimized and optimized model and which were extracted from the Edge Impulse tool. It can be noticed that there was a significant drop of accuracy between the unoptimized and optimized versions of the motion classification model due to quantization in order to fit onto the MCU. The *stable* class had the highest rate in both versions, due to the simplicity of the samples. On the contrary, the *fall* and *hit* classes had the lower rates in *TPR* due to the confused movements.

4.3. Models' Comparison

For increasing the models' accuracy, various implementations were deployed including the fine tuning of epochs and learning rate, giving different results. The development of the proposed model for speech detection was divided into two stages. In the small dataset, a neural network with MFCC feature extraction was developed and tested in four implementations compared with the pretrained MobileNet-v2 model, with Mel-filterbank energy (MFE) features' extraction involving two implementations. For this dataset, the

MFCC model was tuned with 100 and 200 epochs and the learning rate was from 0.001 to 0.005. Regarding the MFE model, the epochs were 100 and the learning rates were 0.01 and 0.05, respectively. The results indicated that $TAcc$ for the pretrained, MobileNet-v2 model, was almost 93% with 100 epochs and 0.05 learning rate and the F_1 score was 95%, which was the best implementation for this dataset. The results are shown in Figure 15.



Figure 14. $TAcc$ of unoptimized (a) and $TAcc$ of optimized (b) models for the motion classification model.

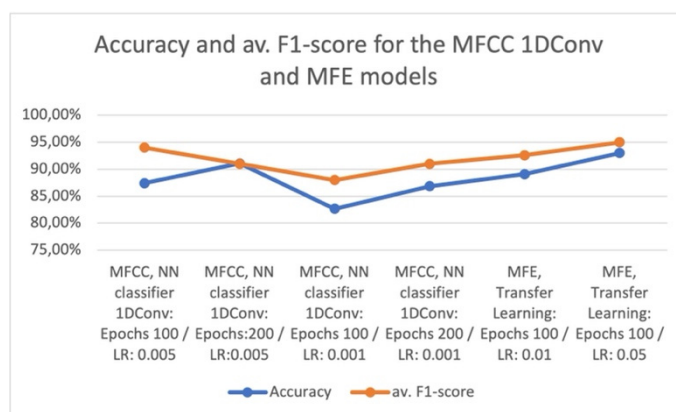


Figure 15. Accuracy and average F_1 -score of the MFCC and MFE models in different implementations.

The second stage of development, based on the big dataset, included the deployment of three implementations. The first two models were based on the MFCC feature extraction and were deployed with the same values in epochs and learning rate; however, there was a difference in their architecture, where the first was a 2D convolutional NN and the second was the model that was proposed in Section 3.3.3. The last model was the MFE model, which was implemented in the small dataset with 50 epochs. The results of this implementation showed that the proposed 1D conv neural network performed better than the other two, with $TAcc = 97.48\%$. Additionally, the MFE model performed poorly in this dataset, with $TAcc = 90\%$. Additionally, in the MFE models, changes to epochs seemed to have no effect on $TAcc$. Figure 16 presents the comparison results.

In Figure 18, it is evident that there was a reduction in Ram, Flash, and latency between the two models. Also, the TAcc of the optimized model remained almost the same, compared with the unoptimized.

On the other side, regarding the motion classification model, the comparison between the optimized and unoptimized model, extracted from the Edge Impulse Studio, can be seen in Figure 19.

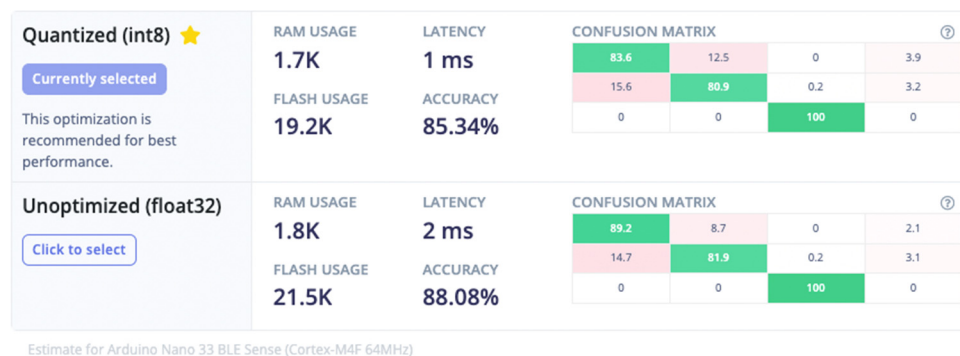


Figure 19. Comparison between optimized and unoptimized deployments for the motion classification model.

From this comparison, it was shown that on-device accuracy dropped significantly in the optimized model (3.15%), with the rate for uncertain values in the *fall* label being higher (3.9%). This low performance could be explained by the limited dataset of the captured data. The results of the classification and testing on the MCU, for the speech detection and motion classification models, are presented in Tables 5 and 6, accordingly.

Table 5. On-device model testing and classification for speech detection model.

	Dance	Goodbye	Hey_Robot	Noise	Photo	Unknown
Dance	0.99609	0.00000	0.00000	0.00000	0.00000	0.00000
Goodbye	0.00000	0.99609	0.00000	0.00000	0.00000	0.00000
Hey_robot	0.00000	0.00000	0.99609	0.00000	0.01172	0.00000
Noise	0.00000	0.00000	0.00000	0.98047	0.02734	0.00391
Photo	0.00000	0.00000	0.00000	0.00000	0.93359	0.09766
Unknown	0.00000	0.00000	0.00000	0.01953	0.02734	0.89844

Table 6. On-device model testing and classification for motion classification model.

Timestamp	Fall	Hit	Stable
80	0	0	1.00
560	0	1.00	0
80	1.00	0	0

In Table 5, the on-device model testing and classification results of the speech model's labels are presented. As can be seen, the results explain the capability of the model to perform fast, continuous audio capturing with impressive accuracy in each label.

In Table 6, the model was tested to classify motions for the three classes for 5 s sampling. For the *fall* and *stable* motions, the samples were correctly predicted, with ~100% accuracy; however, for the *hit* sample, the classification started ~500 ms after the initialization. Thus, it was clear that there were a few false positive samples in the *hit* label, which implies the need for a better dataset.

4.5. Evaluation Using Specialists

Eight educators of a public school in Athens for the specialized education for children with ASD and other developmental disorders ($n = 4$, autism specific; $n = 2$, other DDs; $n = 2$ general education, working across various special educational needs) participated in small focus group sessions (see Appendix A). Specialists varied widely in their levels of experience, ranging from <2 to 28 years in their current education setting ($M = 12.7$ years, $SD = 9.19$). The aim of this evaluation was to identify educators' ideas and perceptions about the robot and software prototype, taking their recommendations into consideration for future work while developing the robot for ASD interventions. It should also be noted that specialists had no previous experience in robotics.

4.5.1. Procedure Description

Seven educators completed one small focus group regarding the software prototype assessment in their working environment. The focus group included a brief introduction on the Human–Robot Interaction field and the context of RMI in ASD. After, the robot's software prototype was presented, and they had 5–10 min to navigate through its screens. When they finished, they were given a questionnaire to answer about their experience, indicating necessary corrections (see Appendix B). When this session ended, the second focus group took place with eight educators participating regarding robot assessment. Participants viewed the robots that were presented in the paper's Section 2.1, and they were not given any information about their capabilities or technical feasibility. Then, the robot's 3D-printed prototype was presented. Later, they assessed the prototype, and they were encouraged to consider its potential uses in autism interventions (see Appendix C).

The total duration of the first focus group was 100 min, with a 90 min time (12.85 min mean time) for answering. The total duration of the second focus group was 90 min, with an 80 min time (11.42 min mean time) for answering the specific questionnaire.

4.5.2. Results of Software Prototype Session

The first two questions were answered prior to the demonstration of the prototype and showed that 71.42% of specialists want to have the control of the robot, parameterizing its functions according to children's needs, and 85.71% of the specialists expect that the software will have buttons and images. All the specialists found it useful, user-friendly, and comprehensible. Of the educators, 71.42% faced difficulties in returning to a previous screen, while 42.85% indicated that the analytics should be displayed differently. Additionally, 71.42% of the educators indicated that the sound in button pressing, the commands in the main screen, and the limited available options were missing. Among the most important findings were a help button, a more colorful interface, and the option to note important reasons about the intervention. Of the specialists, 28.57% stated that the apply button was something unnecessary, while 71.42% found the analytics page the most important element. The answers for possible changes varied between aesthetic and usability opinions. In the last question, the specialists recognized equipment and robotic knowledge as the main prevention reasons.

When the general questions were finished, nine Likert Scale questions were answered. Concerning Q14, 57.14% of specialists rated the ease-of-use level sufficient. Regarding Q16, 71.42% of the specialists rated the overall experience as excellent, and, in Q17, three out of seven agreed that the process of use was clear. In Q18, 57.14% of the specialists agreed that the navigation was easy, and, in Q19, five out of seven stated that such a software would be extremely important. In Q20, five of seven educators had no previous experience of using similar software, and, in Q21, 85.71% had completely no experience in the use of robots. Finally, in Q22, 42.85% of the specialists understood completely the software.

4.5.3. Results of Robot Prototype

Concerning the first and second questions, the majority of the specialists described the robot as cute and safe, having the appropriate size for children, being minimally

expressive and child-friendly, with an accessible shape. In the third question, the emotions that the robot caused the specialists ranged from curiosity, to tenderness, to excitement, and pleasure. Additionally, the specialists found the central screen and face tracking and speech detection capabilities as the main advantages, which also gave value to the robot. In question 6, two out of eight educators said that they did not like that the head rotates with the body, and two out of eight noticed the lack of human characteristics. In general, attached parts, positive feedback, head rotation, and the guidelines to children seemed to be what specialists would add. In questions 10 and 11, we found that the attached accessory, human face screen, head rotation, feedback functionality, and a more stable base should be improved. Furthermore, the specialists indicated emotion expression tasks and speech activities as the robot's potential deployment domain. Lastly, in question 14, the educators would use robots Kaspar, Nao, Flobi, Aibo, and Roball as an alternative.

5. Discussion

In this study, important findings were found for the development of ML models that will be deployed on a robot for ASD interventions, as well as for the design of the robot's and software's prototypes using educators' reviews.

Regarding machine learning models, some differences were found. First, for the speech detection model, the small dataset was very limited, where the MobileNet-v2 model was shown to be the best, with 95% accuracy. For the refined dataset, the proposed model reached almost 98% accuracy when deployed on the microcontroller, while its accuracy was slightly increased. As for the F1-score, it was higher than the accuracy, which indicated good balance between precision and recall. Concerning the model for classifying motions, the findings suggest that more work has to be completed in terms of the data. Despite that accuracy reached almost 86%, there were misclassifications of samples, which may happen due to the limited amount of data and the way that data were captured.

In terms of on-device performance regarding the speech detection model, it took 105 ms for detection and 5 ms for classification. This explains the vast capabilities of TinyML in demanding tasks, such as child-robot interaction, where continuous and reliable sensing is required. On the other side, the on-device performance of the motion classification model showed that classification started later in time regarding the *hit* class, which explains that the motions in this class were not distinct.

The evaluation of the specialists is very important, both for the robot and its software, while their engagement in the process is significant. Firstly, regarding the software prototype, the specialists want to use it for controlling the robot. They mentioned aesthetic elements that would make the experience richer and the solution of other minor usability issues. Moreover, they found the analytics and preview panes as the most useful elements. Additionally, they indicated a lack of equipment and knowledge as reasons that may prevent them from using the robot. As a general result, the overall experience was interesting and no major usability issues that hindered interaction were found. Lastly, an understanding of what the software can accomplish remains questionable and needs to be defined appropriately.

Regarding the assessment of the robot's prototype, the specialists anticipate working with it in interventions when it is ready, with emotions and speech development tasks as the most important implementations. Furthermore, the educators characterized the robot as simple, making it an appropriate tool for children. Additionally, the central screen of the robot, which allows tactile interaction and the ability to recognize speech and faces, was defined as the most important function. The specialists also stated that there are things that need to be improved such as autonomous head rotation, anthropomorphic characteristics, and a more stable base. Consequently, the robot seemed to be attractive, but further improvements, which entail future work, should be accomplished both in the design and the machine learning models.

Despite the possible benefits and useful prospects, it should be noted that some limitations of the study still exist.

First, the robot has not been tested in real interventions with ASD children; therefore, the impact on them was not found. Additionally, another limitation is that there were not clear objectives to measure based on this deployment; thus, further investigation is needed. Furthermore, despite the positive results of the machine learning models, these need to be tested in real situations and interventions in order for their applicability in interaction tasks to be certified, while their functionality with the proposed robot will be verified by end users. In addition, the robot's hardware should also be tested in real intervention scenarios that will assure that it fits in the particular robot.

Therefore, the important findings of this study rely on the development of machine learning models, their positive outcomes through various implementations and approaches, and the useful results of educators' evaluation about the design of the robot and the functionality of its software. In addition, the limitations of the study include the applicability of the robot in real scenarios, the determination of clear deployment objectives, and the applicability of ML models and robot's hardware in interventions. All the above need to be examined and studied further, entailing future work, which is presented in the next section.

6. Conclusions and Future Work

This paper focuses on the design of a low-cost, portable, and intelligent robot, capable of being implemented in ASD interventions. Additionally, the development of two machine learning models, which were deployed and tested on a microcontroller, gives the ability to the robot of sensing real-world cues, interacting with the child in a certain degree of multimodality. The results of the models are promising but there are certain additions that need to be accomplished. Additionally, the evaluations of the specialists who engaged in the process are useful for future consideration, developing a more robust and efficient robot for autism interventions.

Future work includes research in three domains: (1) the expansion of the models' datasets with more data from various individuals with and without ASD; (2) improvements in the models' architectures with the refinement of the algorithms and the use of novel techniques; and (3) the development and improvement of the robot according to specialists' reviews and feedback.

Supplementary Materials: The models, datasets, and the related code are available online at https://github.com/ilkatsanis/RobotDesign_TinyML (1 April 2022).

Author Contributions: Conceptualization, I.A.K., V.C.M. and D.T.P.; methodology, I.A.K. and V.C.M.; software, I.A.K.; data curation, I.A.K.; writing—original draft preparation, I.A.K.; writing—review and editing, V.C.M.; visualization, I.A.K.; supervision, V.C.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Acknowledgments: The authors would like to thank the educators of the first public kindergarten of Athens for their valuable contribution and support to the study.

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

Appendix A

This appendix presents the background of the participating educators who took part in the two focus groups. The first table shows the participants of the first focus group and the second one shows the participants of the second focus group. Notes: PGES = *public general education school*; PSES = *public special education school*; MSEN = *multiple special education needs*; ODD = *other developmental disorders*.

Table A1. Participants of the first focus group.

Focus Group 1 n = 7	Specialist ID	Gender	Years of Experience	Type of School	Needs of Children
	Specialist ID.1	Female	5 years	PSES	Autistic children
	Specialist ID.2	Female	16 years	PGES	ODD
	Specialist ID.3	Female	2 years	PSES	Autistic children
	Specialist ID.4	Female	19 years	PGES	ODD
	Specialist ID.5	Female	28 years	PGES	MSEN
	Specialist ID.6	Female	17 years	PGES	Autistic children
	Specialist ID.7	Female	2 years	PSES	Autistic children

Table A2. Participants of the second focus group.

Focus Group 2 n = 8	Specialist ID	Gender	Years of Experience	Type of School	Needs of Children
	Specialist ID.1	Female	5 years	PSES	Autistic children
	Specialist ID.2	Female	16 years	PGES	ODD
	Specialist ID.3	Female	2 years	PSES	Autistic children
	Specialist ID.4	Female	19 years	PGES	ODD
	Specialist ID.5	Female	28 years	PGES	MSEN
	Specialist ID.6	Female	17 years	PGES	Autistic children
	Specialist ID.7	Female	2 years	PSES	Autistic children
	Specialist ID.8	Female	16 years	PGES	MSEN

Appendix B

In this appendix, the questionnaire regarding the assessment of the software's prototype from the therapists is presented. The first two questions were answered before the demonstration of the prototype. Questions 3–13 were answered descriptively, while questions 14–22 were in a Likert Scale.

Table A3. Assessment questions for the robot's software prototype.

Prior to the demonstration of the prototype	Question 1. What Do You Expect to Do with the Software Prototype?
	Question 2. How Do You Expect It to Look Like?
After the demonstration of the prototype	Question 3. How did the software prototype look to you as a whole?
	Question 4. What made it more difficult for you to use it?
	Question 5. Was there anything that did not look like as expected? If so, which one?
	Question 6. Was there something you expected to exist but was missing?
	Question 7. Is there any item that is not needed?
	Question 8. Which feature(s) of the prototype are most important to you?
	Question 9. Which feature(s) of the prototype is less important to you?
	Question 10. What is the most important function that You think should exist?
	Question 11. If you had a magic wand, what would you add to the software prototype?
	Question 12. What would you change in the appearance of the prototype?
	Question 13. What would prevent you from using the software?

Table A3. *Cont.*

Likert Scale Questions	Question 14. What was the level of ease of use? (i) Excellent, (ii) Sufficient, (iii) Medium/Normal, (iv) No, and (v) Disappointing
	Question 15. How satisfied are you with its use? (i) Excellent, (ii) Sufficient, (iii) Medium/Normal, (iv) No, and (v) Disappointing
	Question 16. How would you rate the overall experience? (i) Excellent, (ii) Sufficient, (iii) Medium/Normal, (iv) No, and (v) Disappointing
	Question 17. The process of using it was clear. (i) Strongly Agree, (ii) Agree, (iii) Disagree, (iv) Strongly Disagree
	Question 18. Navigating through the prototype was easy. (i) Strongly Agree, (ii) Agree, (iii) Disagree, (iv) Strongly Disagree
	Question 19. How important would such software be for using a robot? (i) Excellent, (ii) Sufficient, (iii) Medium/Normal, (iv) No, and (v) Disappointing
	Question 20. Have you ever had previous experience using similar software? (i) Very much, (ii) Very, (iii) Medium/Normal, (iv) Almost no, (v) Not at all
	Question 21. Do you have previous experience in applying robot to the educational process? (i) Very much, (ii) Very, (iii) Medium/Normal, (iv) Almost no, (v) Not at all
	Question 22. To what extent did you understand what software does? (i) Very much, (ii) Very, (iii) Medium/Normal, (iv) Almost no, (v) Not at all

Appendix C

In Appendix C, the questionnaire regarding the assessment of the robot prototype from therapists is presented. The specialists were given the robots presented in the paper's Section 2.1 and a summary of its functions, gaining better knowledge for the proposed robot.

Table A4. Assessment questions for robot's prototype.

Question 1. How would you describe the presented robot?
Question 2. How would you evaluate its shape and form?
Question 3. What emotions does its overall appearance evoke to you?
Question 4. Based on its morphological elements and functions, what do you think is its main advantage?
Question 5. Based on the table with its functions, which do you think are the most useful and important?
Question 6. What do you dislike about its morphological features?
Question 7. Why do you think this robot is worth? If not, please explain.
Question 8. What would you add about its morphological features?
Question 9. What would you add about its functions?
Question 10. What could we improve about the existing appearance?
Question 11. What could we improve about its existing features?
Question 12. How does our robot compare to the ones you saw before?
Question 13. What potential uses would you discern in ASD interventions?
Question 14. If not ours, which of the other robots would you like to use and why?

References

- Feil-Seifer, D.; Mataric, M.J. Defining socially assistive robotics. In Proceedings of the 9th International Conference on Rehabilitation Robotics ICORR 2005, Chicago, IL, USA, 28 June–1 July 2005; pp. 465–468.
- Association, A.P. *Neurodevelopmental Disorders: DSM-5® Selections*; American Psychiatric Publishing: Washington, WA, USA, 2015.
- Lord, C.; Risi, S.; DiLavore, P.S.; Shulman, C.; Thurm, A.; Pickles, A. Autism from 2 to 9 years of age. *Arch. Gen. Psychiatr.* **2006**, *63*, 694–701. [\[CrossRef\]](#)
- Huijnen, C.A.; Lexis, M.A.; Jansens, R.; de Witte, L.P. Roles, strengths, and challenges of using robots in interventions for children with autism spectrum disorder (ASD). *J. Autism Dev. Disord.* **2019**, *49*, 11–21. [\[CrossRef\]](#)
- Begum, M.; Serna, R.W.; Yanco, H.A. Are robots ready to deliver autism interventions? A comprehensive review. *Int. J. Soc. Robot.* **2016**, *8*, 157–181. [\[CrossRef\]](#)
- Pennisi, P.; Tonacci, A.; Tartarisco, G.; Billeci, L.; Ruta, L.; Gangemi, S.; Pioggia, G. Autism and social robotics: A systematic review. *Autism Res.* **2016**, *9*, 165–183. [\[CrossRef\]](#)
- Huijnen, C.A.; Lexis, M.A.; Jansens, R.; de Witte, L.P. Mapping robots to therapy and educational objectives for children with autism spectrum disorder. *J. Autism Dev. Disord.* **2016**, *46*, 2100–2114. [\[CrossRef\]](#) [\[PubMed\]](#)
- Kahn, P.H., Jr.; Kanda, T.; Ishiguro, H.; Freier, N.G.; Severson, R.L.; Gill, B.T.; Shen, S. Robovie, you'll have to go into the closet now: Children's social and moral relationships with a humanoid robot. *Dev. Psychol.* **2012**, *48*, 303. [\[CrossRef\]](#) [\[PubMed\]](#)
- Breazeal, C.; Harris, P.L.; DeSteno, D.; Westlund, J.M.K.; Dickens, L.; Jeong, S. Young children treat robots as informants. *Top. Cogn. Sci.* **2016**, *8*, 481–491. [\[CrossRef\]](#)
- Bone, D.; Bishop, S.L.; Black, M.P.; Goodwin, M.S.; Lord, C.; Narayanan, S.S. Use of machine learning to improve autism screening and diagnostic instruments: Effectiveness, efficiency, and multi-instrument fusion. *J. Child Psychol. Psychiatr.* **2016**, *57*, 927–937. [\[CrossRef\]](#)
- Liu, W.; Li, M.; Yi, L. Identifying children with autism spectrum disorder based on their face processing abnormality: A machine learning framework. *Autism Res.* **2016**, *9*, 888–898. [\[CrossRef\]](#) [\[PubMed\]](#)
- Bussu, G.; Jones, E.J.; Charman, T.; Johnson, M.H.; Buitelaar, J.K. Prediction of autism at 3 years from behavioural and developmental measures in high-risk infants: A longitudinal cross-domain classifier analysis. *J. Autism Dev. Disord.* **2018**, *48*, 2418–2433. [\[CrossRef\]](#)
- van Straten, C.L.; Smeekens, I.; Barakova, E.; Glennon, J.; Buitelaar, J.; Chen, A. Effects of robots' intonation and bodily appearance on robot-mediated communicative treatment outcomes for children with autism spectrum disorder. *Pers. Ubiquitous Comput.* **2018**, *22*, 379–390. [\[CrossRef\]](#)
- Linstead, E.; Dixon, D.R.; French, R.; Granpeesheh, D.; Adams, H.; German, R.; Kornack, J. Intensity and learning outcomes in the treatment of children with autism spectrum disorder. *Behav. Modif.* **2017**, *41*, 229–252. [\[CrossRef\]](#) [\[PubMed\]](#)
- Dautenhahn, K.; Werry, I.; Salter, T.; Boekhorst, R. Towards adaptive autonomous robots in autism therapy: Varieties of interactions. In Proceedings of the 2003 IEEE International Symposium on Computational Intelligence in Robotics and Automation. Computational Intelligence in Robotics and Automation for the New Millennium (Cat. No. 03EX694), Kobe, Japan, 16–20 July 2003; Volume 2, pp. 577–582.
- Wood, L.J.; Zaraki, A.; Robins, B.; Dautenhahn, K. Developing kaspar: A humanoid robot for children with autism. *Int. J. Soc. Robot.* **2021**, *13*, 491–508. [\[CrossRef\]](#) [\[PubMed\]](#)
- Gouaillier, D.; Hugel, V.; Blazelevic, P.; Kilner, C.; Monceaux, J.; Lafourcade, P.; Maisonnier, B. Mechatronic design of NAO humanoid. In Proceedings of the 2009 IEEE international conference on robotics and automation, Kobe, Japan, 12–17 May 2009; pp. 769–774.
- Hegel, F.; Eyssel, F.; Wrede, B. The social robot 'flobi': Key concepts of industrial design. In Proceedings of the 19th International Symposium in Robot and Human Interactive Communication, Viareggio, Italy, 13–15 September 2010; pp. 107–112.
- Bonarini, A.; Garzotto, F.; Gelsomini, M.; Romero, M.; Clasadonte, F.; Yilmaz, A.N.Ç. A huggable, mobile robot for developmental disorder interventions in a multi-modal interaction space. In Proceedings of the 2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), New York, NY, USA, 26–31 August 2016; pp. 823–830.
- Duquette, A.; Michaud, F.; Mercier, H. Exploring the use of a mobile robot as an imitation agent with children with low-functioning autism. *Auton. Robot.* **2008**, *24*, 147–157. [\[CrossRef\]](#)
- Kozima, H.; Michalowski, M.P.; Nakagawa, C. Keepon. *Int. J. Soc. Robot.* **2009**, *1*, 3–18. [\[CrossRef\]](#)
- Cao, H.L.; Pop, C.; Simut, R.; Furnémont, R.; Beir, A.D.; Perre, G.V.D.; Vanderborght, B. Probolino: A portable low-cost social device for home-based autism therapy. In *International Conference on Social Robotics*; Springer: Cham, Switzerland, 2015; pp. 93–102.
- Stanton, C.M.; Kahn, P.H.; Severson, R.L.; Ruckert, J.H.; Gill, B.T. Robotic animals might aid in the social development of children with autism. In Proceedings of the 2008 3rd ACM/IEEE International Conference on Human-Robot Interaction (HRI), Amsterdam, The Netherlands, 12–15 March 2008; pp. 271–278.
- Michaud, F.; Caron, S. Roball, the rolling robot. *Auton. Robot.* **2002**, *12*, 211–222. [\[CrossRef\]](#)
- Scassellati, B.; Admoni, H.; Mataric, M. Robots for use in autism research. *Annu. Rev. Biomed. Eng.* **2012**, *14*, 275–294. [\[CrossRef\]](#) [\[PubMed\]](#)
- Lord, C.; Bishop, S.L. Autism Spectrum Disorders: Diagnosis, Prevalence, and Services for Children and Families. Social Policy Report. *Soc. Res. Child Dev.* **2010**, *24*, 2.

27. Available online: <https://www.edgeimpulse.com/product> (accessed on 18 March 2022).
28. Sahidullah, M.; Saha, G. Design, analysis and experimental evaluation of block based transformation in MFCC computation for speaker recognition. *Speech Commun.* **2012**, *54*, 543–565. [[CrossRef](#)]
29. Abadi, M.; Barham, P.; Chen, J.; Chen, Z.; Davis, A.; Dean, J.; Zheng, X. {TensorFlow}: A System for {Large-Scale} Machine Learning. In Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16), Savannah, GA, USA, 2–4 November 2016; pp. 265–283.
30. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
31. Koopmans, L.H. *The Spectral Analysis of Time Series*; Elsevier: Amsterdam, The Netherlands, 1995.
32. Banbury, C.R.; Reddi, V.J.; Lam, M.; Fu, W.; Fazel, A.; Holleman, J.; Yadav, P. Benchmarking TinyML systems: Challenges and direction. *arXiv* **2020**, arXiv:2003.04821.
33. Warden, P. Why the Future of Machine Learning Is Tiny. 2018. Available online: <https://petewarden.com/2018/06/11/why-the-future-of-machine-learning-is-tiny/> (accessed on 11 March 2022).
34. Available online: https://docs.opencv.org/3.4/db/d28/tutorial_cascade_classifier.html (accessed on 12 March 2022).
35. Dalal, N.; Triggs, B. Histograms of oriented gradients for human detection. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 20–25 June 2005; Volume 1, pp. 886–893.
36. Kazemi, V.; Sullivan, J. One millisecond face alignment with an ensemble of regression trees. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 1867–1874.
37. Available online: <https://github.com/cmusatyalab/openface/tree/master/models/openface> (accessed on 14 March 2022).