Indonesian Journal of Electrical Engineering and Computer Science
Vol. 41, No. 1, January 2026, pp. 320~330
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v41.i1.pp320-330

Deep-fuzzy personalisation framework for robot-assisted learning for children with autism

Rose-Mary Owusuaa Mensah Gyening[1], James Ben Hayfron-Acquah[1], Michael Asante[1], Kate Takyi[1], Peter Appiahene[2]
[1]Department of Computer Science, Faculty of Physical and Computational Sciences,
Kwame Nkrumah University of Science and Technology (KNUST), Kumasi, Ghana
[2]Department of Information Technology and Decision Sciences,
University of Energy and Natural Resources, Sunyani, Ghana

Article Info

Article history:

Received Apr 16, 2024
Revised Oct 13, 2025
Accepted Dec 13, 2025

Keywords:

Africa
Autism
Deep learning
Fuzzy logic
Personalization
Robots

ABSTRACT

Research exploring the efficacy of robots in autism therapy has predominantly relied on the Wizard-of-Oz method, where robots execute predetermined behaviours. However, this approach is constrained by its heavy reliance on human intervention. To address this limitation, we introduce a novel deep-fuzzy personalization framework for social robots to enhance adaptability in interactions with autistic children. This framework incorporates a deep learning model called singleshot emotion detector (SED) with a mean average precision of 93% and a fuzzy-based engagement prediction engine, utilizing factors such as scores, IQ levels, and task complexity to estimate the engagement of autistic children during robot interactions. Implemented on the humanoid robot RoCA, our study assesses the impact of this personalization approach on learning outcomes in interactions with Ghanaian autistic children. Statistical analysis, specifically Mann Whitney tests ($U=3.0$, $P=0.012$), demonstrates the significant improvement in learning gains associated with RoCA's adoption of the deep-fuzzy approach.

Corresponding Author:

Rose-Mary Owusuaa Mensah Gyening
Department of Computer Science, Faculty of Physical and Computational Sciences
Kwame Nkrumah University of Science and Technology
Kumasi, Ghana
Email: rmo.mensah@knust.edu.gh

1. INTRODUCTION

Robots can perform repetitive tasks and can be programmed to behave in predictable ways under consistent conditions consistently. This makes them potentially suitable companions for autistic children who struggle with adapting to sudden changes in their surroundings [1]–[3]. In autism therapy, robots can assume various roles including but not limited to: serving as therapeutic aids, companions during play, facilitators of social interactions, and model social agent [4]. Research by [3], [5], and many other single-day and longitudinal research indicates that children diagnosed with autism spectrum disorder (ASD) exhibit a preference for socially assistive robots (SARs). While both short-term and longitudinal studies highlight the potential of robots as beneficial tools in autism therapy, there remains a gap in leveraging social robots within classroom settings to support the academic development of children with autism. Huijnen *et al.* [6] indicate that the use of social robots in autism education within classroom settings has not yet been extensively evaluated. Consequently, there is a pressing need for research to investigate how these robots can effectively facilitate the development of social, academic, imitation, and joint attention skills in children with autism.

Thill *et al.* [7] research in robot-assisted autism therapy should prioritize exploring methods to alleviate the workload on therapists. This can be achieved by enhancing social robots with supervised autonomy, enabling them to operate more independently while still under human oversight.

There is a demand for advanced robotic models that are not tied to specific platforms and have the ability to deduce the internal states of children in order to respond appropriately [7]. In light of recent technological advancements like machine learning (ML), there is a timely opportunity to explore how social robots can perceive data from their interaction environment, analyze it, and adjust their behavior to foster a supportive learning atmosphere that sustains children's interest. Lately, attention has shifted towards employing ML for diagnosing and intervening in autism, with most of the works concentrating on ML methodologies for autism diagnosis [8]. Crippa *et al.* [9] developed a supervised machine learning (SML) algorithm that categorizes children with low-functioning autism based on upper limb motion. Other proposed SML algorithms for predicting autism rely on cortical thickness in specific brain regions [10], multiparametric magnetic resonance imaging data [11], and abnormalities in face processing [12].

During robot-assisted therapy sessions for autistic children, researchers can utilize ML techniques to analyze and predict the emotional states and behaviors of children to guide social robots in providing tailored interventions to address the unique needs of each child. To address the demands for speed, responsiveness, and energy efficiency in implementing ML for human-robot interactions, there is a necessity to explore deep learning, a novel branch of ML that utilizes artificial neural networks, as traditional ML methods have limitations. Presently, enhancing autonomy in interactions between autistic children and robots poses a notable challenge. Given the unpredictability of autistic children, expecting social robots to possess precise knowledge of the interaction context is impractical. Fuzzy logic, renowned for its capability to manage imprecision and uncertainty, could offer a promising solution [13] and be appropriate for controlling the behaviour of social robots in autism therapy. ML and fuzzy computational models, leveraging multimodal data including emotional states, personality traits, social circumstances, and physiological condition, could be formulated to estimate the emotional state and levels of engagement of autistic children during live interactions with social robots.

This paper presents a novel deep fuzzy framework tailored for personalized robot-assisted learning in interactions involving autistic children and robots. Consisting of two core elements, this framework leverages state-of-the-art technology to enrich the quality of these interactions. The initial component, termed the singleshot emotion detector (SED), employs advanced deep learning methods to analyze the emotions expressed by autistic children via real-time video captured by a robot's camera. The second component is a fuzzy-based engagement estimation system, capable of integrating diverse input factors such as scores, IQ levels, and task complexities to evaluate the engagement levels of autistic children during their interactions with social robots. The framework's implementation was executed on the humanoid robot RoCA, followed by empirical investigations aimed at assessing the influence of the personalized approach facilitated by the deep fuzzy framework on learning outcomes within interactions between autistic children and robots.

Affect recognition in human-robot interaction (HRI): Affect refers to the physiological aspect of a person's emotions. Automatically detecting and categorizing human affect ensures a more natural and bi-directional communication, enhancing human-computer interaction. In the realm of special needs education, particularly for children with autism who may struggle with verbal communication, researchers can employ robots as therapy partners to perceive nonverbal cues from a child and anticipate their emotional state to adjust interventions as needed. Facial expressions are crucial in social interactions and can provide insights into a person's emotions or intentions. The automatic recognition of facial emotions (FER) in dynamic environments, such as human-robot interactions, presents a challenging task in computer vision. Detecting emotions automatically is complex due to computational demands, technological limitations, and the need for real-time processing of emotional states. Similar to human behavior, social robots must integrate multisensory data from their surroundings and interpret them to deduce emotional states and levels of engagement. They then adapt their behavior accordingly. Many FER systems proposed for robots in human-robot interaction have relied on conventional methods of feature extraction and ML.

In a study by Giorgan and Ploeger [14], gabor features were extracted from facial images, and two ML algorithms, support vector machines (SVM) and Adaboost, were used to create an emotion classification system for domestic robots. In Alonso-Martin *et al.* [15], openCV was used for facial detection, while two external software systems, SHORE and CERT, were employed for a facial emotion recognition (FER) system integrated into the robot Maggie. Cid *et al.* [16] used gabor filters for feature extraction, and a dynamic Bayesian network served as the classifier in enabling the robot Muecas to identify people's emotions. Leo *et al.* [17] developed a FER system based on the histogram of oriented gradients (HOG) descriptor, and SVM were implemented on the Robokind™ R25 robot. Liu *et al.* [18] designed a FER system using 2D gabor filters, local binary patterns (LBP), and an extreme learning classifier for implementation on the Nao robot.

However, these conventional approaches rely on shallow learning or manually crafted features like LBP, scale invariant feature transform (SIFT), HOG, and non-negative matrix factorization. These approaches require extensive computational resources, making them impractical for real-time human-robot interaction scenarios.Conventional ML algorithms like SVMs have effectively predicted emotions for images obtained in controlled laboratory settings but demonstrate limited applicability to spontaneous images captured in natural environments [19]. In complex environments where various noise sources are prevalent, emotion recognition is addressed through deep neural networks. Numerous FER systems based on deep learning adopt an image classification approach. For instance, in [20], faces were detected using OpenCV, and feature extraction was conducted utilizing a deep convolutional neural network (CNN) on the CK+ and JAFFE datasets for emotion classification. Yu and Zhang [21] observed facial expressions and categorized emotions through an ensemble of face detectors and classifiers. In human-robot interaction, robots can enhance decision-making by understanding the types of objects present in the environment and their precise positions. Thus, it is opportune to investigate ML models based on object detection for human-robot interaction. Object detection refers to a computer vision technique wherein objects (such as cats, dogs, human faces, books) are identified in digital images and video sequences, localized with bounding boxes, and subsequently classified [22].

Object recognition represents an extended form of image classification. Whereas image classification typically assigns an image to a single class, with or without localization, object detection algorithms identify multiple objects, localize them using bounding boxes, and assign each object to its predicted class. In human-robot interaction, object detection algorithms offer notable advantages as they can swiftly detect objects' presence and precise location. In scenarios such as robot-assisted learning for autistic children, an object detection algorithm enables a robot to recognize objects in the environment, facilitating appropriate actions if a child approaches a potentially hazardous or unsuitable object. By employing a fast object detection algorithm, the robot can predict the emotional states of autistic children and adapt its behavior accordingly. Although object detection has proven to be highly effective and efficient in domains like airplane and ship detection, environmental perception for robots, and object manipulation, its potential for real-time emotion recognition by social robots during interactions with autistic children has yet to be fully explored.

Personalization of autistic child–robot interactions: Each child with autism has unique strengths and challenges, so it is essential to create individualized intervention plans that cater to their specific needs. Robots should be able to perceive and analyze data, such as the child's emotional state and level of engagement, during interactive sessions to allow for real-time adjustments, leading to optimize learning outcomes. Despite the potential benefits, there has been limited research on using deep learning emotion recognition systems to enhance robot-mediated therapy for ASD. The existing emotion recognition systems mainly rely on traditional ML algorithms trained on datasets of typically developing individuals and audio cues [17], [23].

The scarcity of domain-specific datasets leads to suboptimal prediction accuracies of FER systems in human-robot interaction contexts. To tackle these issues, a has been proposed based on transfer learning on SSDLite, an object detection algorithm. The proposed SED constitutes a deep learning-driven system tailored for both face localization and classification of facial expressions. By combining face detection and classification processes into a unified pipeline, this model ensures high frames per second (FPS), minimal latency, and efficient power consumption, facilitating real-time emotion detection tasks.

To optimize learning outcomes, social robots must be capable of assessing children's engagement levels during real-time interactions between autistic children and robots (ACRI). By measuring a child's current emotional state and estimated level of engagement, the robot can tailor learning sessions to suit individual needs. However, engagement estimation in ACRI settings has not received extensive attention from researchers thus far. Existing engagement estimation models typically rely on single sensory modalities such as facial images, head movements, and body gestures. Since ASD is complex, a combination of multimodal data encompassing affective states, IQ levels, learning progress, and physiological indicators would offer a more accurate estimation of engagement levels in real-time. This paper also proposes and evaluates the use of fuzzy logic technology a method renowned for its efficacy in uncertain and imprecise environments in autistic child-robot interactions.

2. METHOD

In the field of autism management, social robots are used as interaction facilitators that can customize their behavior according to the individual profiles, emotional states, and learning progress of autistic children. These social robots can be equipped with some autonomy to perceive the emotional states and engagement levels of autistic children during real-time interactions and adapt the learning process accordingly. To achieve this goal, this paper proposes a personalized interactive framework that combines a novel SED with a fuzzy inference system (FIS) to foster sustained and meaningful long-term interactions between autistic children and robots. We implemented and evaluated this framework by integrating the deep learning-based SED model and the FIS system, modeled as a C# plugin, into the RoCA robot. We assessed the impact of personalization on learning outcomes in child-robot interactions involving autistic children through numerous experiments.

2.1. Deep learning-based singleshot emotion detector

Emotion recognition systems do not perform well in multicultural settings since emotions can vary across cultures [24] or may carry different meanings across different cultural contexts [25]. Existing FER systems, which are not trained on facial expressions specific to children with autism, will likely underperform in applications targeting the autism domain. This research aims to develop an FER system for the humanoid robot RoCA, intended for use in robot-mediated therapy for Ghanaian children on the autism spectrum. To achieve this, two datasets were collected: the GHANED dataset, comprising images of Ghanaians displaying six emotional classes (happy, sad, angry, fearful, neutral, and surprised), and the Autistic Children's Dataset (ACD), consisting of facial expressions captured during some preliminary experiments and longitudinal studies involving autistic children. Object detection systems such as faster region-based convolutional neural networks (Faster-RCNN) and You-Only-Look-Once (YOLO) are computationally demanding [26]. As a result, they might operate slowly for real-time interactions between autistic children and robots. In contrast, the Single Shot MultiBox Detector stands out for its rapid processing and precise results [26].

The proposed emotion prediction model utilizes transfer learning on SSDLite, a streamlined version of SSD designed for mobile applications [27]. Transfer learning was employed to accelerate the training process, leveraging pre-trained weights from the SSDLite model. These pre-trained weights encapsulate salient object features, thus reducing the time and data required to achieve the desired accuracy level. Given that the initial layers of most CNNs extract low-level features relevant to various tasks, freezing all layers except the final one of the SSDLite model was deemed appropriate. Subsequently, the last layer underwent retraining using custom datasets. This approach encompassed a two-stage fine-tuning process. Initially, the model was trained on the GHANED dataset, comprising 532 images depicting Ghanaians displaying both "acted" and "in the wild" emotions. This was followed by training on the ACD dataset, consisting of 287 facial expressions of autistic children gathered from previous preliminary and longitudinal experiments involving the children and a robot. The training images were annotated in pascal visual object classes (VOC) format. In the Pascal VOC, the annotation consists of the training folder name, file name, path to the file, bounding box coordinates and size of the image (width, height and depth) compiled in an XML file.The trained model achieved high accuracy by employing supervised fine-tuning on facial expression datasets of selected people before training on smaller autistic childrens' datasets. Evaluation of the proposed model was conducted using the mean average precision (mAP) metric, a standard measure for evaluating object detection algorithms. mAP score, derived from precision and recall, is calculated by averaging the interpolated average precisions across all classes. The SED model attained an impressive mAP of 93%. Following training, the trained SED was configured as a plugin using Visual Studio 2017 and seamlessly integrated into the custom software interface control of the RoCA robot. The process of training the deep learning-based emotion recognition model is illustrated in Figure 1. Table 1 presents an overview of the GHANED and ACD datatsets.

2.2. Fuzzy-based engagement estimation system

The second component of the proposed framework, the fuzzy logic reasoning model, is an artificial intelligence-powered algorithm that empowers the robot to reason and respond appropriately to the individual needs of each child. Fuzzy logic provides a robust methodology for handling uncertainties and deriving conclusions based on imprecise, ambiguous, or noisy data. This reasoning model consists of three crisp inputs, or linguistic variables: a child's score at a specific point in the learning session, the difficulty level of the task, and the child's IQ level. The output variable is the engagement level of the child. During the fuzzification stage, fuzzy sets were created for both the input and output variables. Figure 2 presents triangular membership functions were employed for fuzzification, as they are easily transferable to microcontrollers, despite the availability of other fuzzifiers like trapezoidal, singleton, and Gaussian.

The child's score was fuzzified into three membership functions using linguistic terms: high, average, and low. Task difficulty was fuzzified into easy, medium, and hard membership functions, while the child's IQ level was given membership functions of low, medium, and high. The engagement level of the child was fuzzified into three membership functions: highly engaged (HE), engaged (E), and not engaged (NE). Fuzzy rules were devised based on consultations with domain experts specializing in the care of children with autism.

The Mamdani inference approach was employed to map the three input variables (IQ, score, and task difficulty) to the output variable (engagement) using IF … THEN rules. For instance, if the score is low, the IQ is low, and the task difficulty is easy, THEN the child is deemed "not engaged." The output from the rule evaluation is always a fuzzy set. In the defuzzification step, the fuzzy output is converted into a crisp value for utilization by the robot. Defuzzification techniques such as the center of gravity, the center of sums, and the max criterion were considered, with the center of gravity chosen due to its computational efficiency and widespread usage compared to other approaches.

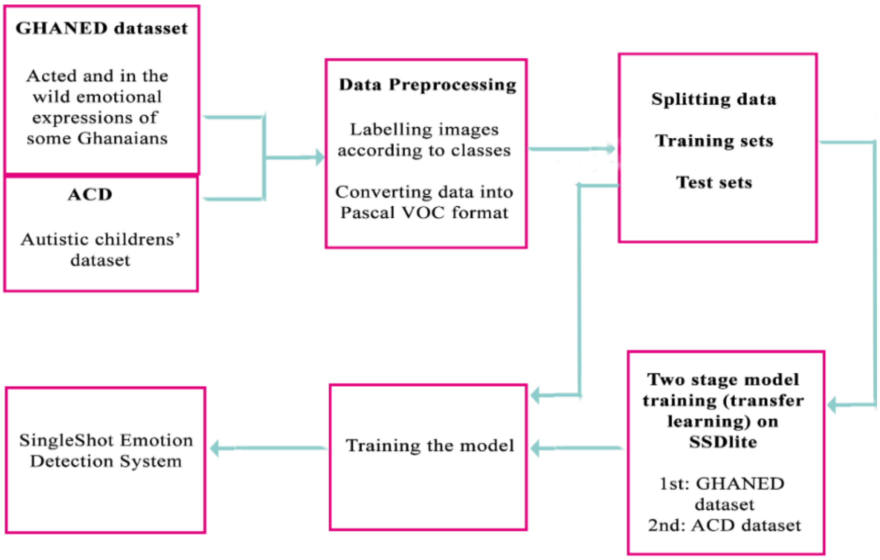

Figure 1. Development process of the proposed deep-learning based SED

Table 1. Overview of the GHANED and ACD datasets

	Number of labels						
Dataset	Happy	Sad	Fear	Surprise	Neutral	Anger	Total
GHANED	92	87	88	90	89	86	532
ACD	50	47	43	49	50	48	287
Total							819

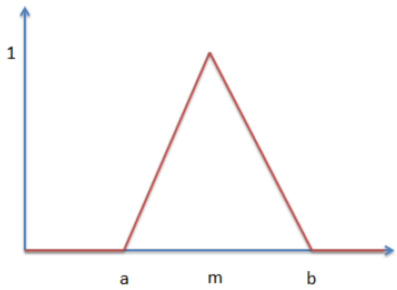

Figure 2. Triangular membership

2.3. The deep-fuzzy personalization framework

The proposed personalized interactive framework, as depicted in Figure 3, comprises the SED and the fuzzy-based engagement estimation system. This framework can be applied to social robots to facilitate human-robot interaction and promote sustained and meaningful long-term sessions between autistic children and robots. It has been implemented and tested on the humanoid robot RoCA, operating according to the procedure outlined in this section. The robot's onboard camera device is activated to stream live videos during interactions with autistic children. Additionally, an HTTP Server is initiated within Ez-Builder, an Integrated Development Environment utilized for programming RoCA. This HTTP Server enables external applications to access the robot's camera via a Uniform Resource Locator (URL), which is configured within the deep learning-based SED program written in Python. OpenCV captures live image streams from the robot's camera during the child-robot interaction sessions, which are then inputted into the SED model. The SED model conducts a singleshot object localization and classification of facial expressions in the received images, with the inferred emotional classes of the localized faces being logged into a file. The robot continues to capture live feeds during the interaction session, sends them to the SED model for predictions, and logs the results into a file.

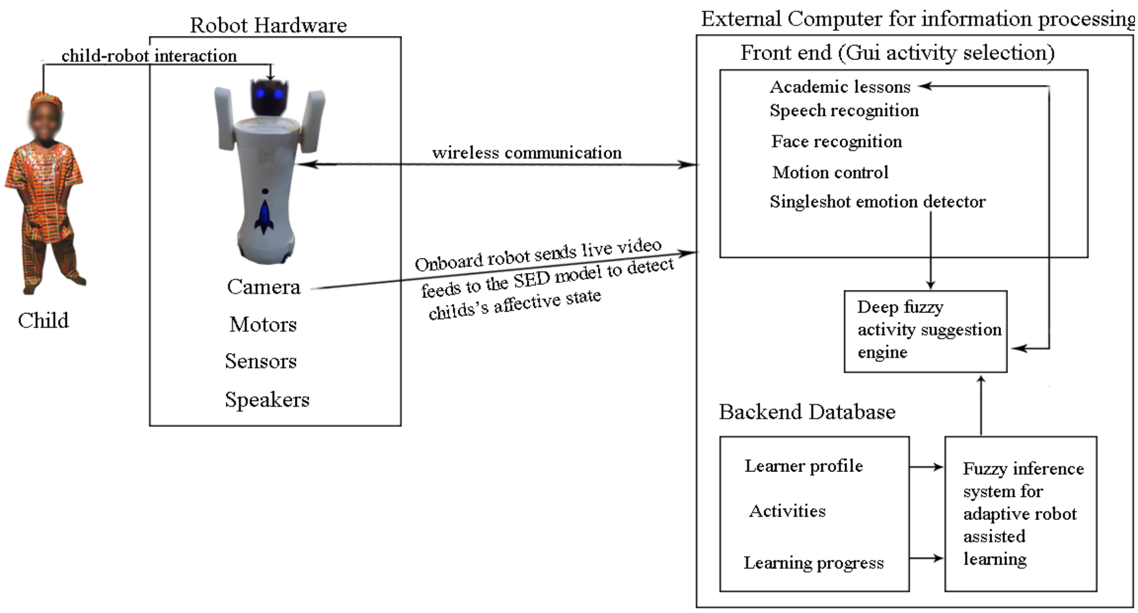

Figure 3. The proposed deep fuzzy personalization framework

At a predetermined time point, denoted as t (e.g., halfway through the lesson), the robot executes the fuzzy logic controller to estimate the child's engagement level. The fuzzy logic controller utilizes information such as the child's name, IQ level, task difficulty, and score to estimate the engagement level (HE, E, or NE). The inference engine of the fuzzy controller consists of 27 rules for estimating engagement, which were developed collaboratively with caregivers of children with autism. The output from the fuzzy inference system, along with the highest aggregated emotional class, is then processed by the deep fuzzy activity selection engine. This engine suggests to the teacher the appropriate action to take if the child's predicted engagement is low or average (for instance, changing the lesson or taking a break).

2.4. Evaluation of the deep fuzzy robot behaviour adaptation framework

An empirical investigation was conducted utilizing two distinct configurations of the RoCA robot: the RSEDFuzzy setup, wherein the robot, equipped with the SED model and fuzzy inference engine, operated semi-autonomously, and the RWOZ setup, where the robot was manually operated. In the RSEDFuzzy setup, the robot extensively employed the deep fuzzy personalization framework, leveraging the highest predicted emotional state, score, and task difficulty to deliver customized lessons to children. For instance, if the emotional states detected by the SED model indicate that a child's predominant emotional state during the learning session is sadness, with a low score but an easy task difficulty level, the robot intervenes by playing music or altering the lesson to re-engage the child. Conversely, when operated in RWOZ mode, the robot disregarded all factors and proceeded to deliver lessons until the sessions concluded. The empirical study aimed to evaluate the impact of RSEDFuzzy-based personalization on learning outcomes in autistic children compared to a randomly operated robot (RWOZ). The research question posed was: Would the task scores of children with ASD be higher in "fruit learning lessons" conducted with an RSEDFuzzy-based personalized robot compared to a Wizard of Oz-operated robot? The independent variables were RSEDFuzzy and RWOZ, while the dependent variable was the task scores.

Twelve participants (11 males and one female) between the ages of eleven and twenty-six, diagnosed with ASD and attending the garden city special school in Kumasi, Ghana, were recruited for the study. A randomized controlled research design was employed, with six participants randomly assigned to the RSEDFuzzy group and the remaining six to the RWOZ group. The mean age in the RSEDFuzzy group was 18.0, with a standard deviation of 5.4, while in the RWOZ group, the mean age was 16.7, with a standard deviation of 5.5. A Mann-Whitney U test was conducted to assess the significance of the age difference between the RSEDFuzzy and RWOZ groups. The results, U=16 and P=0.747, indicated no significant difference in the ages of children between the two groups.

2.4.1. Experimental setup

The experiments were conducted at the computer laboratory of the garden city special school in Ghana, utilizing the humanoid robot RoCA. Informed consent was obtained from the school. Each child participated in nine interaction sessions with the robot, totaling 108 sessions. The study protocol involved two pretest sessions (identical for both groups), followed by five training sessions and two immediate post-tests. Throughout the experiment, a teacher accompanied the children to the experimental room. The robot provided positive reinforcement by saying "good job" and offered verbal prompts as needed. A tablet was affixed to RoCA to display video recordings of selected fruits, while two tables were positioned beside RoCA, one with a banana and the other with an apple. The initial two sessions served as pretest lessons to assess the children's ability to identify two fruits, an apple and a banana. Before each session, the operator entered the child's name into a textbox on the robot control system, which was used to retrieve the child's information from a database, including IQ level and task difficulty. RoCA introduced itself to the child by saying, "Hello + “child's name”, my name is RoCA.

Today, we will study fruits. “Child's name”, please take the apple from the table." If the child selected the correct fruit, they were awarded a score, and the robot instructed them to give the fruit to their teacher. The same process was repeated for the other fruit, in this case, the banana. Following each question, the robot awaited confirmation by the operator (by clicking the "reward" button in the software interface) if the task was performed correctly. If the "good job" button was not clicked within ten seconds, the robot repeated the instruction and waited for confirmation or issued a final prompt after an additional fifteen seconds before proceeding to the next question. After the question and answer session, the robot concluded the session by playing music and dancing, signaling the end of the session to the child.

From sessions 3 to 7, RoCA employed video modeling techniques and interactive sessions to instruct the children on identifying and consuming apples and bananas. To prevent sensory overload, common in autistic children, the lessons focused solely on these two fruits. Each session commenced with RoCA addressing the child by name and announcing the lesson's topic, for example, "Hello + “child's name”, today we will learn about fruits." A picture of the designated fruit, such as an apple, appeared on the tablet attached to the robot. RoCA vocalized the name of the fruit, followed by a brief video displaying the fruit on the tablet. RoCA then gestured and turned its head toward the specific fruit on the nearby table while reiterating its name. The child was prompted to retrieve the designated fruit, with rewards or additional verbal cues provided as needed. This process was repeated for the other fruit. Finally, RoCA concluded the interaction session with a song. In the RWOZ group, the robot functioned in a non-adaptive mode, delivering standardized learning sessions without considering the child's emotional state, learning progress, or task difficulty level. Conversely, in the RSEDFuzzy group, RoCA acted as an adaptive agent, tailoring learning sessions based on inputs received by the deep-learning model and fuzzy-based engagement prediction engine. These inputs included the child's predominant emotional state, learning progress, IQ, and task difficulty. The last two sessions served as post-tests to evaluate the children's knowledge following the teaching sessions. The script delivered by RoCA in these sessions mirrored that of sessions 1 and 2. In both the pretest and post-test sessions, each correct task completion by the child was scored 5.

2.4.2. Data analysis

To assess the impact of personalization on learning outcomes, the children's baseline knowledge of fruits was measured during the pretest phase, and their knowledge acquisition following the learning sessions was gauged through post-tests. Each correct task completion during both pretest and post-test sessions earned a score of 5. A Mann-Whitney test was employed to determine any significant differences between the pretest scores of both groups, as well as to examine differences in post-test scores. The utilization of the Mann-Whitney test was due to its non-parametric nature and suitability for small sample sizes and test scores, ensuring robust analysis despite limited data.

3. RESULTS AND DISCUSSION

Table 2 illustrates the pretest and post-test scores of children in the two groups. The aggregated pretest scores for both groups differ by 20. A Mann-Whitney test conducted on the pretest scores indicates no notable distinction between the two groups, with U=13.0 and P=0.37. Conversely, the aggregated post-test scores differ by 55 (Table 1). A Mann-Whitney test performed on the post-test scores reveals a significant difference between the two groups, with U=3.0 and P=0.012. These findings indicate that children who received personalized lessons from RoCA, employing the RSEDFuzzy system, demonstrated superior performance in the post-tests, thus affirming the favorable impact of personalization through the deep-fuzzy personalization framework.

Table 2. Pretest and post-test scores of the children in the RSEDFuzzy and RWOZ groups

Condition	Subject	Pretest	Post-test
RSEDFuzzy	1	15	20
	2	0	10
	3	0	15
	4	10	20
	5	0	20
	6	15	10
Aggregated RSEDFuzzy group test scores		40	95
RWOZ	7	10	10
	8	0	10
	9	10	10
	10	0	5
	11	0	0
	12	0	5
Aggregated RWOZ test scores		20	40

3.1. Discussion

Robot-mediated therapy for children with autism has gained significant attention in recent years, with researchers exploring various approaches to enhance the effectiveness of interactions between robots and autistic children. The conventional Wizard-of-Oz approach, where robots execute pre-defined behaviors, has been the primary method used in many studies. However, this approach is inherently limited as it requires substantial human effort and lacks adaptability in real-time interactions. In this study, we introduced a novel deep-fuzzy personalization framework designed to address these limitations and enhance the efficacy of robot-mediated therapy for autistic children.

The lack of significant disparity in the pretest scores between the RSEDFuzzy and RWOZ groups suggests that all children commenced the experiment with comparable knowledge of apple and banana fruits. Following the intervention, both groups exhibited enhanced learning gains, as evidenced by their post-test scores. Notably, some children who initially struggled to identify fruits showed improvement, while others with low pretest scores demonstrated enhanced performance in the post-test. These findings align with previous research [28]–[30], indicating that autistic children may benefit from repeated lessons with robots over time. Examination of Table 1 reveals that while both groups experienced increased learning gains, a Mann-Whitney test revealed higher gains in the RSEDFuzzy group, which received personalized interactions from the robot. This underscores the significance of fostering rapport and sustaining engagement in interactions between autistic children and robots. The adoption of a deep-fuzzy personalization framework by the robot facilitated cooperation, prolonged bi-directional communication, and heightened attention in the RSEDFuzzy group.

Many studies in literature utilized the Nao robot, and the concentration areas for experiments were joint attention, eye contact exercise, imitation, and communication skills. The levels of participation among the children also varied depending on their mental and cognitive abilities. Deep learning in human-robot interaction is gradually proving beneficial due to its ability to provide personalization. Albeit the supposed benefits, the high cost associated with the design and development of robots and the sophisticated nature of training deep learning algorithms makes it difficult for low-income countries to benefit. The findings from this research are consistent with [31], who opines that supervised autonomy could help autistic child-robot interactions. There is a need to ascertain whether the learning gains would transcend beyond the classroom and be lifelong.

The deep-fuzzy personalization framework leverages advancements in deep learning techniques and fuzzy logic to imbue social robots with adaptability and responsiveness during interactions with autistic children. The deep learning-based emotion detection system, with a high mean average precision of 93%, enables the robot to perceive and interpret the children's emotional states accurately. The fuzzy engagement estimator enables real-time adjustments in interactions. The personalized and adaptive interventions are tailored to the unique needs and preferences of each autistic child. The robot can facilitate more meaningful and effective interactions by adjusting its behavior based on real-time assessments of emotional states and engagement levels, resulting in improved therapeutic outcomes. The proposed approach reduces the need for human intervention, making it more scalable and cost-effective in autism therapy.

In a developing country like Ghana, where technological resources are limited, the proposed framework, which is based on lightweight ML and fuzzy models, would be beneficial in complementing the efforts of skilled personnel. Albeit the supposed benefits, there should be conscious efforts to address privacy and ethical concerns associated with SARs, which could mar the advantages when left unattended. Multi-purpose activities need to be researched and designed for children based on their levels on the autism spectrum [32]. Future research should investigate whether autistic children could show aggressive behaviors toward robots long-term [33]. This study corroborates research by [34] and [35], which indicate that ML applications can be utilized in the design of interventions for autistic children.

4. CONCLUSION

This paper introduces a framework for personalizing robot behavior intended for use in interactions between autistic children and robots. This framework integrates a novel SED and a fuzzy-based engagement estimation system, implemented in Python and C#, respectively. Evaluation of the framework was carried out on the humanoid robot RoCA through multiple interaction sessions with autistic children, revealing that personalized lessons led to accelerated learning. The data generated from this research can facilitate further exploration of suitable robot-mediated interventions for autistic children in environments with limited resources.

While our study demonstrates promising results, several limitations and areas for future research should be acknowledged. Firstly, the sample size of our empirical studies was relatively small, necessitating larger-scale trials to further validate the efficacy of the deep-fuzzy personalization framework. Additionally, long-term follow-up studies are needed to assess the sustainability of the observed improvements in learning outcomes and to explore the potential for generalization across different settings and populations of autistic childrenAdditionally, considering the unpredictable behavior commonly observed in children on the autism spectrum, it is suggested that future research explore the potential of ensemble ML models incorporating facial expressions, audio cues, and body posture to predict the emotional states of these children. Moreover, it is imperative for researchers to dedicate efforts to gathering datasets specific to autistic children, enabling the development of ML applications tailored to meet the needs of both the children and their caregivers.

FUNDING INFORMATION

The authors state no funding is involved.

CONFLICT OF INTEREST STATEMENT

The authors state no conflict of interest.

DATA AVAILABILITY

Data availability is not applicable to this paper as no new data were created or analyzed in this study.

REFERENCES

[1] J. J. Cabibihan, H. Javed, M. Ang, and S. M. Aljunied, "Why Robots? A survey on the roles and benefits of social robots in the therapy of children with autism," *International Journal of Social Robotics*, vol. 5, no. 4, pp. 593–618, Nov. 2013, doi: 10.1007/s12369-013-0202-2.

[2] B. Scassellati, H. Admoni, and M. Matarić, "Robots for use in autism research," *Annual Review of Biomedical Engineering*, vol. 14, no. 1, pp. 275–294, Aug. 2012, doi: 10.1146/annurev-bioeng-071811-150036.

[3] B. Robins, "A humanoid robot as assistive technology for encouraging social interaction skills in children with autism," p. 173, 2005, [Online]. Available: https://uhra.herts.ac.uk/handle/2299/14273

[4] K. Dautenhahn, "Roles and functions of robots in human society: Implications from research in autism therapy," *Robotica*, vol. 21, no. 4, pp. 443–452, Aug. 2003, doi: 10.1017/S0263574703004922.

[5] K. Dautenhahn and A. Billard, "Games children with autism can play with robota, a humanoid robotic doll," in *Universal Access and Assistive Technology*, London: Springer London, 2002, pp. 179–190, doi: 10.1007/978-1-4471-3719-1_18.

[6] C. A. Huijnen, M. AS Lexis, and L. P. de Witte, "Robots as new tools in therapy and education for children with autism," *International Journal of Neurorehabilitation*, vol. 04, no. 04, 2017, doi: 10.4172/2376-0281.1000278.

[7] S. Thill, C. A. Pop, T. Belpaeme, T. Ziemke, and B. Vanderborght, "Robot-assisted therapy for autism spectrum disorders with (partially) autonomous control: challenges and outlook," *Paladyn*, vol. 3, no. 4, pp. 209–217, Jan. 2012, doi: 10.2478/s13230-013-0107-7.

[8] F. Thabtah, "Machine learning in autistic spectrum disorder behavioral research: A review and ways forward," *Informatics for Health and Social Care*, vol. 44, no. 3, pp. 278–297, Jul. 2019, doi: 10.1080/17538157.2017.1399132.

[9] A. Crippa *et al.*, "Use of machine learning to identify children with autism and their motor abnormalities," *Journal of Autism and Developmental Disorders*, vol. 45, no. 7, pp. 2146–2156, Jul. 2015, doi: 10.1007/s10803-015-2379-8.

[10] Y. Jiao, R. Chen, X. Ke, K. Chu, Z. Lu, and E. H. Herskovits, "Predictive models of autism spectrum disorder based on brain regional cortical thickness," *NeuroImage*, vol. 50, no. 2, pp. 589–599, Apr. 2010, doi: 10.1016/j.neuroimage.2009.12.047.

[11] Y. Zhou, F. Yu, and T. Duong, "Multiparametric MRI characterization and prediction in autism spectrum disorder using graph theory and machine learning," *PLoS ONE*, vol. 9, no. 6, p. e90405, Jun. 2014, doi: 10.1371/journal.pone.0090405.

[12] W. Liu, M. Li, and L. Yi, "Identifying children with autism spectrum disorder based on their face processing abnormality: A machine learning framework," *Autism Research*, vol. 9, no. 8, pp. 888–898, Aug. 2016, doi: 10.1002/aur.1615.

[13] M. Nazemizadeh, M. Taheri, and S. Nazeri, "The application of fuzzy-logic method to control of robots: a review study," *International Journal of Mechanical Engineering and Robotics Research*, vol. 3, no. 2, p. 229, 2014.

[14] G. Giorgana and P. G. Ploeger, "Facial expression recognition for domestic service Robots," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 7416 LNCS, 2012, pp. 353–364, doi: 10.1007/978-3-642-32060-6_30.

[15] F. Alonso-Martín, M. Malfaz, J. Sequeira, J. F. Gorostiza, and M. A. Salichs, "A multimodal emotion detection system during human-robot interaction," *Sensors (Switzerland)*, vol. 13, no. 11, pp. 15549–15581, Nov. 2013, doi: 10.3390/s131115549.

[16] F. Cid, J. A. Prado, P. Bustos, and P. Nunez, "A real time and robust facial expression recognition and imitation approach for affective human-robot interaction using Gabor filtering," in *IEEE International Conference on Intelligent Robots and Systems*, IEEE, Nov. 2013, pp. 2188–2193, doi: 10.1109/IROS.2013.6696662.

[17] M. Leo *et al.*, "Automatic emotion recognition in robot-children interaction for ASD treatment," in *Proceedings of the IEEE International Conference on Computer Vision*, IEEE, Dec. 2015, pp. 145–153, doi: 10.1109/ICCVW.2015.76.

[18] Z. Liu *et al.*, "A facial expression emotion recognition based human-robot interaction system," *IEEE/CAA Journal of Automatica Sinica*, vol. 4, no. 4, pp. 668–676, 2017, doi: 10.1109/JAS.2017.7510622.

[19] A. Mollahosseini, D. Chan, and M. H. Mahoor, "Going deeper in facial expression recognition using deep neural networks," in *2016 IEEE Winter Conference on Applications of Computer Vision, WACV 2016*, IEEE, Mar. 2016, pp. 1–10, doi: 10.1109/WACV.2016.7477450.

[20] V. Mayya, R. M. Pai, and M. M. Manohara Pai, "Automatic facial expression recognition using DCNN," *Procedia Computer Science*, vol. 93, pp. 453–461, 2016, doi: 10.1016/j.procs.2016.07.233.

[21] Z. Yu and C. Zhang, "Image based static facial expression recognition with multiple deep network learning," in *ICMI 2015 - Proceedings of the 2015 ACM International Conference on Multimodal Interaction*, New York, NY, USA: ACM, Nov. 2015, pp. 435–442, doi: 10.1145/2818346.2830595.

[22] L. Wang *et al.*, "Multi-channel convolutional neural network based 3D object detection for indoor robot environmental perception," *Sensors (Switzerland)*, vol. 19, no. 4, p. 893, Feb. 2019, doi: 10.3390/s19040893.

[23] J. C. Kim, P. Azzi, M. Jeon, A. M. Howard, and C. H. Park, "Audio-based emotion estimation for interactive robotic therapy for children with autism spectrum disorder," in *2017 14th International Conference on Ubiquitous Robots and Ambient Intelligence, URAI 2017*, IEEE, Jun. 2017, pp. 39–44, doi: 10.1109/URAI.2017.7992881.

[24] M. A. Quiros-Ramirez and T. Onisawa, "Considering cross-cultural context in the automatic recognition of emotions," *International Journal of Machine Learning and Cybernetics*, vol. 6, no. 1, pp. 119–127, Feb. 2015, doi: 10.1007/s13042-013-0192-2.

[25] H. A. Elfenbein and N. Ambady, "On the universality and cultural specificity of emotion recognition: A meta-analysis," *Psychological Bulletin*, vol. 128, no. 2, pp. 203–235, Mar. 2002, doi: 10.1037/0033-2909.128.2.203.

[26] W. Liu *et al.*, "SSD: Single shot multibox detector," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9905 LNCS, 2016, pp. 21–37, doi: 10.1007/978-3-319-46448-0_2.

[27] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen, "MobileNetV2: inverted residuals and linear bottlenecks," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 4510–4520, 2018, doi: 10.1109/CVPR.2018.00474.

[28] A. Duquette, F. Michaud, and H. Mercier, "Exploring the use of a mobile robot as an imitation agent with children with low-functioning autism," *Autonomous Robots*, vol. 24, no. 2, pp. 147–157, Feb. 2008, doi: 10.1007/s10514-007-9056-5.

[29] J. Wainer, B. Robins, F. Amirabdollahian, and K. Dautenhahn, "Using the humanoid robot KASPAR to autonomously play triadic games and facilitate collaborative play among children with autism," *IEEE Transactions on Autonomous Mental Development*, vol. 6, no. 3, pp. 183–199, Sep. 2014, doi: 10.1109/TAMD.2014.2303116.

[30] V. Pennazio, "Social robotics to help children with autism in their interactions through imitation," *Research on Education and Media*, vol. 9, no. 1, pp. 10–16, Jun. 2017, doi: 10.1515/rem-2017-0003.

[31] M. A. Islam, M. M. Hasan, and S. A. Deowan, "Robot-assisted training for children with autism spectrum disorder: a review," *Journal of Intelligent and Robotic Systems: Theory and Applications*, vol. 108, no. 3, p. 41, Jul. 2023, doi: 10.1007/s10846-023-01872-9.

[32] A. Zhanatkyzy, Z. Telisheva, A. Amirova, N. Rakhymbayeva, and A. Sandygulova, "Multi-purposeful activities for robot-assisted autism therapy: What works best for children's social outcomes?," in *ACM/IEEE International Conference on Human-Robot Interaction*, New York, NY, USA: ACM, Mar. 2023, pp. 34–43, doi: 10.1145/3568162.3576963.

[33] A. Langer, P. J. Marshall, and S. Levy-Tzedek, "Ethical considerations in child-robot interactions," *Neuroscience and Biobehavioral Reviews*, vol. 151, p. 105230, Aug. 2023, doi: 10.1016/j.neubiorev.2023.105230.

[34] N. Kaur and G. Gupta, "Refurbished and improvised model using convolution network for autism disorder detection in facial images," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 29, no. 2, pp. 883–889, 2023, doi: 10.11591/ijeecs.v29.i2.pp883-889.

[35] M. Ikermane and A. El Mouatasim, "Web-based autism screening using facial images and convolutional neural network," *Indonesian* Journal *of Electrical Engineering and Computer Science*, vol. 29, no. 2, pp. 1140–1147, 2023, doi: 10.11591/ijeecs.v29.i2.pp1140-1147.

BIOGRAPHIES OF AUTHORS

Rose-Mary Owusuaa Mensah Gyening is a lecturer at the department of computer science. She holds Ph.D. Computer Science from Kwame Nkrumah University of Science and Technology (KNUST). She has keen interest in machine learning, robotics, IoT, human-computer interaction, and cybersecurity. She can be contacted at email: rmo.mensah@knust.edu.gh.

James Ben Hayfron-Acquah is an associate professor and holds PhD Degree from Southampton University in the UK. His research interests are in automatic gait recognition, human identification at a distance, pattern recognition, mobile technology, databases and cloud computing. He can be contacted at email: jbha@yahoo.com.

Michael Asante is a professor and holds PhD in Systems Engineering from the University of Reading. His areas of specialization include data communication, computer security, distributed systems, computer networking, and infrastructure security. He can be contacted at email: mickasst@yahoo.com.

Kate Takyi holds PhD Degree at Lovely Professional University, Punjab – India in Computer Applications. She is currently a lecturer in the Computer Science, Department of KNUST. Her research areas of interest include machine learning, data science wireless networks, internet of things, and network communications. She can be contacted at email: takyikate@knust.edu.gh.

Peter Appiahene is a senior lecturer with over 10 years working experience in the Computer Science Industry and Academic Institutions both Domestic and International. He has a Ph.D. in computer science with specialization in reinforcement learning, and machine learning. He can be contacted at email: peter.appiahene@uenr.edu.gh.