Online-Journals.org

iJIM International Journal of Interactive Mobile Technologies

iJIM | eISSN: 1865-7923 | Vol. 20 No. 12 (2026) | OPEN ACCESS

https://doi.org/10.3991/ijim.v20i12.62168

PAPER

Machine Learning for Social Robotics in Education: Adapting RoboLearnPI's Behavior to Student Interactions

V. Mathivanan, K. Vimal Kumar Stephen, Ramesh Palanisamy(✉), Senthil Jayapal, Mohammed Tauqeer Ullah, Mohamed R. Rafi

University of Technology and Applied Sciences, Ibra, Sultanate of Oman

ramesh.palanisamy@utas.edu.om

ABSTRACT

A potential opportunity of social robotics integration into education is a personalization of learning in which the behavior of robots can adapt dynamically to the needs of a specific student. In this paper, an innovative framework named ACBSO-SR-RoboLearnPI-A2C is suggested and will be based on Adaptive Chaos Bird Swarm Optimization (ACBSO) along with an Asynchronous Actor-Critic (A2C) reinforcement learning model to improve adaptive behavior and interaction in educational robots. This framework uses real-time interaction data, such as response times, correctness, and engagement indicators, of the Omani Higher Education Student Performance Dataset (Moodle + SIS + eDify) to control the adaptive instructional and social behavior of the RoboLearnPI robot. The suggested method is compared with the traditional baselines such as Deep Q-Network (DQN), Proximal Policy Optimization (PPO), standard A2C, and a non-adaptive model. The quantitative outcomes show that the ACBSO-SR-RoboLearnPI-A2C framework has better performance with various measurements: an Accuracy of 91.8, Learning Gain of 19.8, student engagement of 86, and low prediction errors (MAE = 0.214, RMSE = 0.296, MAPE = 7.6%). Convergence analysis indicates steadier and quicker policy learning, whereas ablation studies indicate the importance of ACBSO, chaos mechanism, and engagement-based reward shaping to overall performance. The reliability of these improvements is statistically validated (paired t-test, $p < 0.05$). Qualitative analysis also shows an increase in the responsiveness of social and adaptive patterns of interaction. The findings of this paper demonstrate the usefulness of combining the ACBSO-based adaptive learning in learning social robotics to enhance student learning. On the whole, the provided research creates a sound paradigm of adaptive intelligent social robotics in education, which shows considerable enhancement in the student learning outcomes, interaction, and engagement.

KEYWORDS

social robotics in education, RoboLearnPI behavior, students interaction, learning outcome, reinforcement learning, social behavior

Mathivanan, V., Stephen, K. V. K., Palanisamy, R., Jayapal, S., Ullah, M. T., Rafi, M. R. (2026). Machine Learning for Social Robotics in Education: Adapting RoboLearnPI's Behavior to Student Interactions. *International Journal of Interactive Mobile Technologies (iJIM)*, 20(12), pp. 33–47. https://doi.org/10.3991/ijim.v20i12.62168

Article submitted 2026-01-13. Revision uploaded 2026-04-12. Final acceptance 2026-04-16.

1 INTRODUCTION

Social robot integration in school settings opened the doors to the aspired educational objective to deliver personalized, interactive learning activities. The evolving of artificial intelligence and fast developments in robotics differentiate social robots aimed at not just enhancing education but also engaging interactive companions built to attend to individual learners. Academic wear has been making assessments of educational robots aiding in enhancing student motivation, engagement, cognitive training, and socio-emotional learning [1–4]. In particular, the social robot offers immediate responses, support, and adaptive pedagogy and assists learning in terms that the social robot can remain a handy tool in present-day institutions of higher learning. However, techniques employed in present educational robotic systems are vulnerable to problems of static or protocol-driven interaction strategies that conspicuously fail to suggest or respond to the multimodal and dynamic learning behaviors of students. Many factors such as reaction time, level of engagement, correctness of responses, and individual learner behavior patterns are largely ignored in real-time; thus, these are responsible for downgrading human-robot interaction effectiveness and education delivery through robots on the educational front. To be able to tackle this problem, we propose introducing a unique intergrated environment called ACBSO-SR-RoboLearnPI-A2C. Interface positions are informaticist around the Adaptive Chaos Bird Swarm Optimization (ACBSO) and Asynchronous Actor-Critic (A2C) reinforcement learning to a new robot. This environment design offers real-time adaptations of instruction and interaction behaviors with the RoboLearnPI social robot based on the acquired student interaction data from the Omani Higher Education Student Performance dataset (data from Moodle, the student information system, and the eDify mobile application). The model traits also provide a balance between exploration and exploitation, converging rapidly to deliver even more introspective, concrete responses based on the context from a combined optimized FSO platform with concurrent policy learning from the A2C algorithm. The study's main aim is to see the enhancement of student learning outcomes, engagement, and quality of human-robot education interaction through intelligent and adaptive social robotics.

1.1 Research objectives

- To develop a reinforcement learning-based social robotics system that can adjust to one-on-one interactions with the student in real-time.
- To combine A2C with ACBSO to have a better exploration-exploitation ratio and stability in convergence.
- To assess the effectiveness of the framework in the form of quantitative (accuracy, learning gain, engagement, error metrics) and qualitative behavioral assessments.
- To conduct ablation and statistical analysis in order to measure the contribution of important components in improving the performance of adaptive learning.

1.2 Contributions of this study

- New Hybrid RL Framework: A2C to ACBSO-based optimization in social robotics in education.

- Adaptive Behavior Mechanism: Coined a behavior transition model, which dynamically realigns the instructional strategies according to the student engagement and student performance.
- Empirical Validation: Shown to be better than DQN, PPO, and standard A2C baselines on accuracy (+3–13%), learning gain (+3–6%), and engagement (+6%) measures.
- Ablation and Statistical Insights: Experimentally determined the importance of ACBSO, chaos dynamics, and social feedback to improve adaptive learning performance.
- Qualitative Effect: Demonstrated the enhancement of natural and socially responsive patterns of human-robot interaction.

2 LITERATURE REVIEW

The utility of social robots in education has grown rapidly over the years in order to support pedagogy with a lot of personal engagement and adaptive learning experiences. Justified studies have come up showing little educational robots can improve students' motivation, engagement, and cognitive as well as socio-emotional outcomes [1, 2, 4, 8]. Social robots are considered effective when they permit flexible interaction and are customized to individual learners [2, 3]. The specific application of reinforcement learning approaches seems to hold promise in enabling the robot to adapt online in real-time in response to the students' feedback, as well as engagement [5, 6, 15]. A number of papers underscore student modeling for adaptive robot interactional behavior [7], and others specifically highlight the role of socially supportive behavior in long-term human-robot interaction [14]. RL was also applied to culturally adaptive dialogue in teachable robots, which resulted in improved engagement [15]. In parallel, swarm intelligence and optimization are being investigated to create balance between exploration and exploitation in complex learning problems. This area of work is represented in the development of context-aware adaptive robotic systems using the data from the Oman Higher Education Student Performance Dataset (consolidating Student Information System (SIS), Moodle LMS, and eDify mobile application) related to academic performance, online activities, and video interaction. However, progressing further: there still remain gaps that include limited real-time adaptive behavior implementation in higher education, integrading swarm optimization and RL for social robots, and studies in the context of region-specific datasets as mentioned in Table 1. Though prospering in the sector, there are big gaps to fill, like colleges having very few studies in place in terms of real-time adaptive behaviors. Also, the use of swarm optimization has gained attention in reinforcement learning (RL) towards social robots, whereas such studies are devoid of the grounding from region-specific datasets. Many robotic studies tend to concentrate upon short-term interactions or K–12, whereas little is known about sustained adaptation within the college environment [3] [6] [14]. Thus, the present study attempts to be one of the first to focus on agents for skill acquisition with an extended temporal horizon and thus address each of these gaps. Complementary studies examine AI-driven adaptive educational systems and the use of robotics for teaching artificial intelligence and machine learning concepts, particularly in K–12, and secondary settings [9, 10, 11, 12, 18, 19]. Additional reviews highlight the role of social robots in identifying educational behaviors and leveraging social network analysis for educational insights [13, 16]. Foundational work also underscores the broader benefits of social human-robot interaction for enhancing machine learning processes [17]. The study proposes a novel ACBSO-SR-RoboLearnPI-A2C hybrid framework that would integrate stable policy learning of A2C with ACBSO for the purpose of fostering exploration more effectively, having faster convergence, and inculcating socially responsive behaviors in RoboLearnPI using a real Omani student dataset.

Table 1. Summary of related work

Ref.	Objective	Key Findings	Research Gaps
[1]	Impact of educational robots	Enhance engagement, motivation & adaptive learning	Limited focus on social robotics
[2]	Social robots trends in education	Facilitate interactive & personalized learning	Need more empirical adaptive studies
[3]	Child robot interaction effects	Robot style/gender affects engagement	Mostly short-term studies
[4]	AI-based educational robots	Robots can adapt behaviors to individual learners	Limited real classroom applications
[5]	Reinforcement learning in social robots	RL optimizes adaptive interactions	Needs large-scale classroom testing
[6]	Effects of adaptive social robots	Improves cognitive, emotional & self-regulated learning	Long-term effects underexplored
[7]	Student behaviour modelling for robots	Robots can tailor interactions to observed behaviors	Real-time adaptation limited
[14]	Long-term HRI in education	Socially supportive robots improve motivation	Limited long-term learning outcomes
[15]	Adaptive dialogue using RL	RL improves engagement across cultures	Broader behaviour adaptation needed
[18] [19]	AI in adaptive educational systems	Enables personalization & adaptive learning	Underexplored in social robotics

3 METHODOLOGY

3.1 Data source

The study uses the Omani Higher Education Student Performance Dataset that was collected from an institute of higher learning in Oman. This unique dataset was a fusion of records harvested from the SIS, the Moodle learning management system, and an m-learning application, eDify. The data comprises specific information about student performance and mobility, specifically online learning activities and video responses of student play, pause, and repeat behaviors. The dataset thus provides a goldmine of information regarding student learning response behaviors—concentrations of their behavior patterns and responses. The dataset is used for modeling student interaction dynamics to trigger a training signal in an A2C framework of smart reward therapy for social, affective adaptation in the RoboLearnPI robot. The dataset, being from Oman, otherwise attains contextual validation to the study.

3.2 Data preprocessing

The Omani Higher Education Student Performance Dataset underwent comprehensive preprocessing for issues around quality and consistency with reinforcement learning on adaptive behavior modeling. The preprocessing step started with cleaning the data by the elimination of duplicate cases and was responsible for resolving the data inconsistency scattered all over the SIS, Moodle, and eDify. Missing numerical values were imputed with the mean or the median when noticed, while observations of a categorical nature were adjusted accordingly with the mode by using simple mean substitution methods. Student records that covered vast parts of missing data were eliminated entirely. Selection and extraction of only pertinent features, such as academic performance/marks, actual levels of participation in the LMS, video interactions, and activity spatial assessments. Both the categorical features and numerical features were also transformationally modified depending on either label or one-hot encoding. The logs were transformed into episodes of interaction merged into sequential learning episodes. These episodes unfolded students' interactions with the simulation as it reigns over time. All the observed features had their min-max scaling normalization computed for standardizing the network's training with U-turn policy gradients. At the very end, the features joined together into a patch of multidimensional state vectors for providing fresh information to researchers' forthcoming networks, the A2C model, its actor, and its critic. The data was split into training and validation for further study.

3.3 Problem definition

The current educational robotics and learning support systems are mostly based on the strategies of interaction, which are not dynamic and strongly adaptive, which is why they cannot effectively respond to the dynamic learning behaviors, level of engagement, and pattern of interaction that can be observed in real classroom settings among students. The problem to be solved is to allow RoboLearnPI to adjust its instructional and social behaviors in real time without causing unstable and inefficient learning using student interaction data of the Oman Higher Education Student Performance Dataset. This issue can be stated as the learning of an optimal policy π^* to maximize cumulative rewards of education and engagement during an interaction episode:

$$\pi^* = arg \max_{\pi} E\left[\sum_{t=0}^{T} \Upsilon^t R(s_t, a_t)\right] \tag{1}$$

with S_t, being the state of student interaction (learning performance and engagement), a t being the adaptive action of the robot, $R(\cdot)$ being the multi-objective rewarding function, and Υ being the discount factor. The offered framework of ACBSO-SR-RoboLearnPI-A2C solution solves this issue by stabilizing and optimizing learning of policies in socially adaptive educational robotics.

3.4 Adaptive chaos bird swarm optimization

To overcome its propensity for premature convergence and local optima, the ABSO algorithm is improved from three crucial angles: population initialization, individual position update, and iterative individual selection. The general goal of maximizing compressed picture quality and maintaining crucial visual aspects during enhancement is supported by these improvements, which are intended to increase the algorithm's global search capacity and convergence accuracy.

Initialization of the population. In BSA, random initialization cannot provide population variety, which is crucial for the swarm intelligence algorithm's accuracy and rate of convergence. Equation (2) uses an enhanced Bernoulli shift chaotic map, which is renowned for its regularity, ergodicity, and unpredictability, for more efficient population initialization in order to increase solution space exploration and compressed picture quality.

$$w_{m-1} = 2(w_m + 0.1 \times qand(0,1))mod1 \tag{2}$$

Updates to each person's location depending on Levy flight characteristics. The ABSO algorithm has a tendency to converge quickly in the early stages and more slowly in the latter stages, falling into local optima. By comparing fitness values based on Lévy flight characteristics, individual locations are updated to increase the global search capacity in compressed image improvement. The Lévy flight density function is used to estimate the step size distribution.

$$Levy(t) \sim |T|^{-1-\beta}, 0 < \beta \leq 2 \tag{3}$$

The random motion step size of Levy's flying behavior is denoted by t in Equation (3). As previously mentioned, t may be represented using Equation (4):

$$t = \mu / |u|^{1/\beta} \tag{4}$$

Parameter μ, u is normally distributed:

$$\mu \sim M(0,\sigma_\mu^2), u \sim M(0,\sigma_u^2) \tag{5}$$

In this formula 6:

$$\sigma_\mu = \left\{ \frac{\Gamma(1+\beta)\sin(\pi\beta/2)}{\Gamma\left[(1+\beta)/2\right]\beta 2^{(\beta-1)/2}} \right\}^{1/\beta}, \sigma_u = 1 \tag{6}$$

Equation (7) presents the Levy-based flying mechanism, which is introduced by the novel individual updating behavior of bird swarms.

$$w_{j,i}^{s+1'} = w_{j,i}^{s} + b(s).sign(qand) \oplus t \tag{7}$$

The formula's new bird swarm individual is $w_{j,i}^{s+1'}$. The symbol q represents the random integer between [–1, 1]. Levy's flying direction is signed, according to equation (8), and the scaling coefficient is $b(s)$, according to equation (9).

$$sign(qand) = \begin{cases} 1 & qand \geq 0 \\ -1 & qand < 0 \end{cases} ; -1 \leq qand \leq 1 \tag{8}$$

$$b(s) = b_{init}.\exp(s/S_{max}) \tag{9}$$

The symbol s indicates how many iterations the formula is currently using. b_{init} is the initial scaling coefficient. S_{max} represent the maximum number of iterations.

By contrasting the individual position's fitness value in Equation (9) with the individual position's fitness value in the basic algorithm.

$$w_{j,i}^{s+1} = \begin{cases} w_{j,i}^{s+1'}, \; if\,(fit(w_{j,i}^{s+1'}) > fit(w_{j,i}^{s+1})) \\ \quad w_{j,i}^{s+1}, \quad Otherwise \end{cases} \tag{10}$$

The Lévy flight-based position update improves the efficiency of picture quality optimization by enabling both random global leaps and fine-tuned local search, which helps avoid local optima. In order to ensure quick and precise convergence toward the best picture enhancement outcomes, the scale coefficient regulates the search range, permitting broad exploration at first and targeted refining subsequently.

3.5 Deep reinforcement learning

The definition of a reinforcement-learning activity is done using a tuple (S, A, T, r). During every time step t, the agents watch the state of the environment $s_t \in S$ and do actions $a_t \in A$ to change their state and get a reward r. The formula $T = p(s_{t+1} \mid s_t, a_t)$ represents a mapping from state-action pairings (s_t, a_t) to the likelihood distribution of the following state s_{t+1}. Iteratively, an agent aims to optimize its predicted return, denoted as $R = \sum_{t=0}^{\infty} R_t = \sum_{t=0}^{\infty} \gamma^t r_t$, where γ is a future discount factor $\in$ [0, 1]. The present state and action (s_t, a_t) are used to determine the predicted discounted return, which is denoted as $Q^{\pi}(s, a) = \mathbb{E}[R_t \mid s_t = s, a_t = a, \pi]$, the best Q function, Q^*, under the right conditions, can be expressed using the following Bellman equation:

$$Q^*(s,a) = \mathbb{E}_{s' \sim p(\cdot|s_2 a)}[r(s,a) + \gamma \max_{a' \in A} Q^*(s',a')]. \tag{11}$$

Additionally, each DRL agent is linked to a particular target network. It has the same structure as the Q-network. The target network's goal is to fix the Q value targets due to the unpredictable training process and unsatisfactory outcomes with non-stationary targets. The variables of the target network θ are periodically added to the Q network θ^- variables. The loss function is shown as follows:

$$L(\theta) = \mathbb{E}_{s,a,r,s'} \left\{ \overbrace{Q(s,a;\theta)}^{\text{prediction}} - \underbrace{\left[r + \gamma \max_{d' \in A} Q(s',a';\theta^-) \right]}_{\text{target}} \right\}^2 \tag{12}$$

Conventional RL models have long been categorized using both value-oriented and policy-based methods. Both of these kinds of methods have significant drawbacks.

3.6 Social robotics on RoboLearnPI asynchronous Actor-Critic (SR-RoboLearnPI-A2C) algorithm

Value-based and policy gradient-based techniques are often outperformed by the actor-critic method. There are two parts to the Actor-Critic algorithm. The actor component selects an action using a neural network. A policy network is a neural network that approximates policies. However, the critic determines whether or not the actions selected by the actor are excellent, whereas the value network assesses the value of acts. We refer to the weights of the policy network as θ_t. Furthermore, the robot learning frequency α and the policy π_θ have previously been determined. Next, we use the Θ parameter to update the policy network:

$$\theta_{t+1} \approx \theta_t + \alpha[\nabla\theta \log\pi_\theta(a \mid s)Q_\pi(s, a)] \tag{13}$$

$Q_\pi(s, a)$ is the overall value that arises from applying policy π in the present state s after choosing action an. The slow convergence of the Actor-Critic method, which happens during training when several neural networks are utilized, is one of its disadvantages. Asynchronous Actor-Critic (A2C) was proposed as a solution to the non-convergence problem. One traditional RL method that uses an experience pool to reduce data association and hence improve convergence is DQN. A_2C accelerate the convergence process in this way. The following are the main ways that the A_2C algorithm outperforms the actor-critic method: First, the asynchronous training framework enhances network-based interaction with the environment, allowing the model to converge faster. Second, efficient network structures pair actors and critics to output the state value and strategy from the input state. The third is critics' assessment. The above equation does not normalize the Q-value. When Q is too big, the parameter θ swings too much. Even if the predicted value is very small, θ won't change much. A_2C uses the value that is the difference between the Q value and the value from the previous state, rather than utilizing the predicted Q value. The advantage function is this difference, which represents the value obtained from action a. The following is a statement of the benefit function if, at time step t, the value function is $V(s_t) = \mathbb{E}[R_t \mid s_t = s]$

$$\begin{aligned} A(s_t, a_t) &= Q(s_t, a_t) - V(s_t) = \mathbb{E}[R_t \mid s_t, a_t] - V(s_t) \\ &\approx r_t + \gamma V(s_{t+1} \mid s_t, a_t) - V(s_t) = \delta(s_t) \end{aligned} \tag{14}$$

The actor's gradient is given by $\nabla\theta \log\pi_\theta(a \mid s)\delta(s_t)$, and hence,

$$\theta_{t+1} \approx \theta_t + \alpha[\nabla\theta \log\pi_\theta(a \mid s)\delta(s_t)]. \tag{15}$$

As an additional point of interest, when the value network is updated, the loss function is represented by the expression $\delta(s_t)^2$.

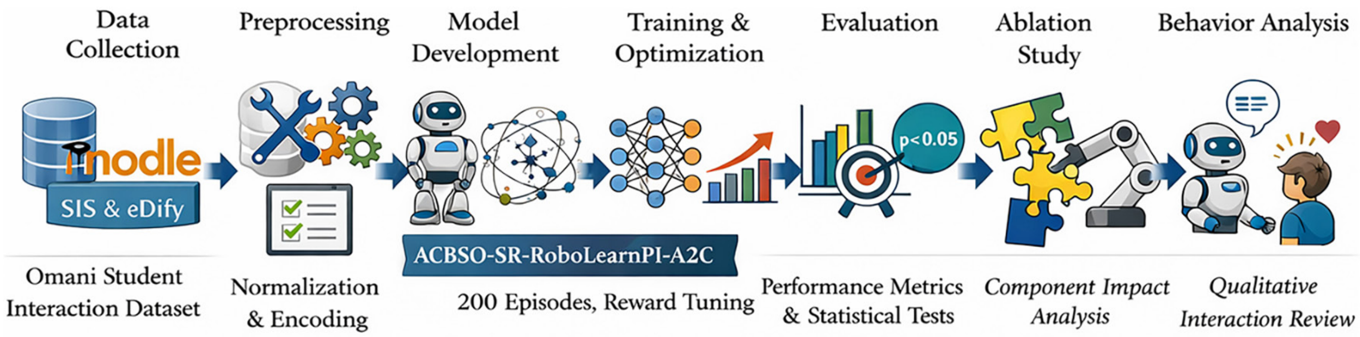

Fig. 1. Workflow of the RoboLearnPI framework

Figure 1 illustrates the overall workflow of the proposed ACBSO-SR-RoboLearnPI-A2C framework. It begins with data collection from the Omani Higher Education Student Performance Dataset (Moodle + SIS + eDify), followed by preprocessing (normalization and encoding). The preprocessed data is used to train the hybrid model combining ACBSO and A2C over 200 episodes with engagement-aware reward tuning. The framework is then evaluated using quantitative metrics (Accuracy, MAE, RMSE, MAPE, and Learning Gain) and ablation studies. The figure provides a clear summary of the complete end-to-end methodology.

4 RESULTS AND DISCUSSION

4.1 Experimental setup

Experiments were conducted with the Omani Higher Education Student Performance Dataset (Moodle + SIS + eDify), which consisted of cases of usage by 1,200 students and over 15,000 interactions. After preprocessing (normalization, missing value handling, one-hot encoding of categorical features), the proposed ACBSO-SR-RoboLearnPI-A2C model (a two-layer actor-critic network with ACBSO optimization: swarm size 30, chaos factor 0.3) was trained for 100 episodes using a reward function based on engagement. Baselines were handled under the same conditions (DQN, PPO, plain A2C, and non-adaptive). The designed framework was evaluated with many quantitative metrics (Accuracy, Precision, Recall, F1-Score, MAE, RMSE, MAPE, and Learning Gain). All experiments were conducted on MATLAR 2025b with GPU parallel acceleration to ensure reproducibility and reliable comparison.

4.2 Performance and evaluation analysis

The proposed framework ACBSO-SR-RoboLearnPI-A2C shows better performance than baseline models such as DQN, PPO, standard A2C, and non-adaptive fixed policy with student performance datasets obtained from higher education in Oman (Moodle + SIS + eDify). The evaluation is to measure learning efficacy, student engagement, and learning stability through several criteria: prediction accuracy, error measures as mentioned above (MAE, RMSE, and MAPE), learning gain, cumulative reward, precision, recall, and F1-score. The comparison made in the study based on Table 2 can state with evidence that the proposed model made a better performance with 91.8% Accuracy, 19.8% Learning Gain, 86% Student Engagement, and the Least Error Values (MAE = 0.214 and RMSE = 0.296). Nonetheless, the model recorded the highest precision (90.1%), recall (89.4%), and F1-score (89.7%). It outruns all baselines. This symbolizes an improved functionality of the model in the accurate prediction of student learning states. This further implies that the adaptive responses have been promising with very low false positives; thus, from the point of view of stability in policy learning, this performance has already regained significance. The outcomes sent an unambiguous positive research implication for instigating the training model, ACBSO, within the A2C pipe to yield reinforcement with personalization and social responses for learning support by the bot.

Table 2. Performance comparison with baseline methods

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	MAE	RMSE	Learning Gain (%)	Avg. Cumulative Reward
Non-Adaptive Baseline	78.4	75.6	73.8	74.7	0.412	0.536	8.6	124.7
DQN	84.9	82.3	80.7	81.5	0.325	0.421	13.2	178.3
PPO	87.6	85.4	84.1	84.7	0.287	0.378	15.9	214.6
A2C	88.3	86.2	85.5	85.8	0.271	0.361	16.7	228.9
ACBSO-SR-RoboLearnPI-A2C (Proposed)	91.8	90.1	89.4	89.7	0.214	0.296	19.8	276.4

4.3 Error metrics

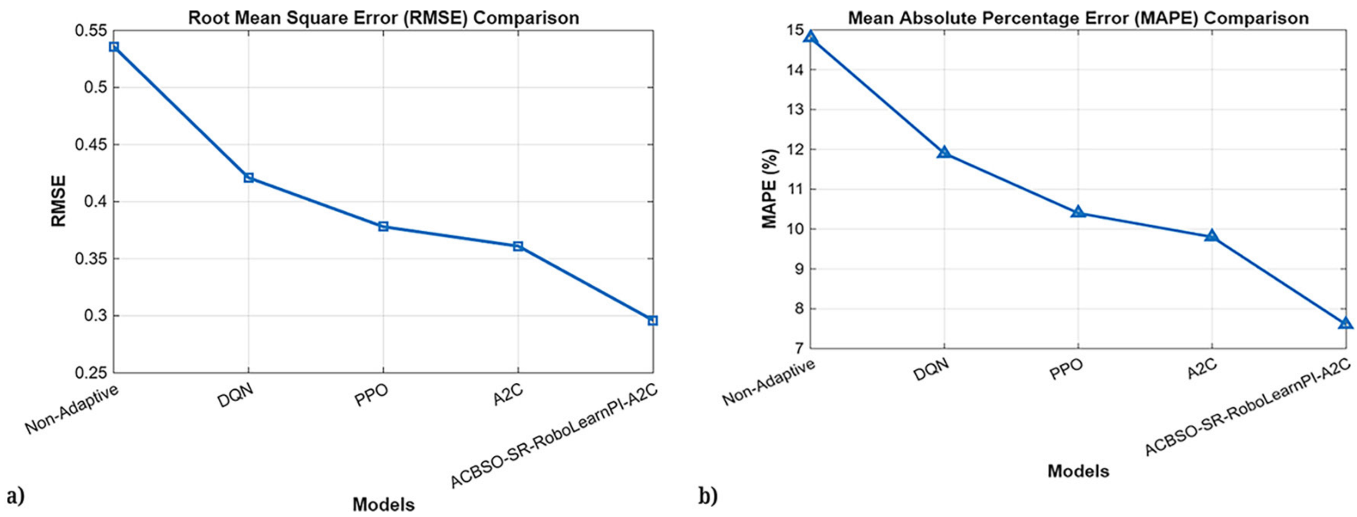

Fig. 2. RMSE and MAPE comparison with baselines and the proposed method

Figures 2a and 2b demonstrate the comparative analysis of the root mean square error (RMSE) and mean absolute percentage error (MAPE) of various learning models. The proposed ACBSO-SR-RoboLearnPI-A2C model exhibits the least error values in all three measures, and, consequently, it has a higher prediction accuracy and high-quality modeling of student learning behavior. The decrease in MAE indicates a higher accuracy in the average prediction, and the smaller RMSE indicates a smaller stability to greater prediction errors. Also, the considerable reduction of MAPE proves that the proposed approach is effective in reducing relative errors of prediction. The overall outcome of these findings confirms that A2C frameworks and adaptive chaos bird swarm optimization are effective to enhance the idea of reliability and adaptability in the learning process of educational systems based on social robotics.

4.4 Convergence analysis

The convergence pattern of the proposed ACBSO-SR-RoboLearnPI-A2C model is demonstrated in Figure 3 in comparison to the traditional reinforcement techniques in learning. The progressive reward improvement of multiple training episodes points to the fact that the suggested approach reaches its convergence more rapidly and gathers a greater reward level when compared to DQN, PPO, and standard A2C. The mechanism that can be credited with this improved convergence performance is the Adaptive Chaos Bird Swarm Optimization, which increases the exploration-exploitation balance and avoids early convergence. Enhanced learning stability can also be seen by the fact that the reward curve is smoother and more stable, which emphasizes the usefulness of the proposed adaptive social robotics framework in maximizing robot behavior as a result of interactions with students.

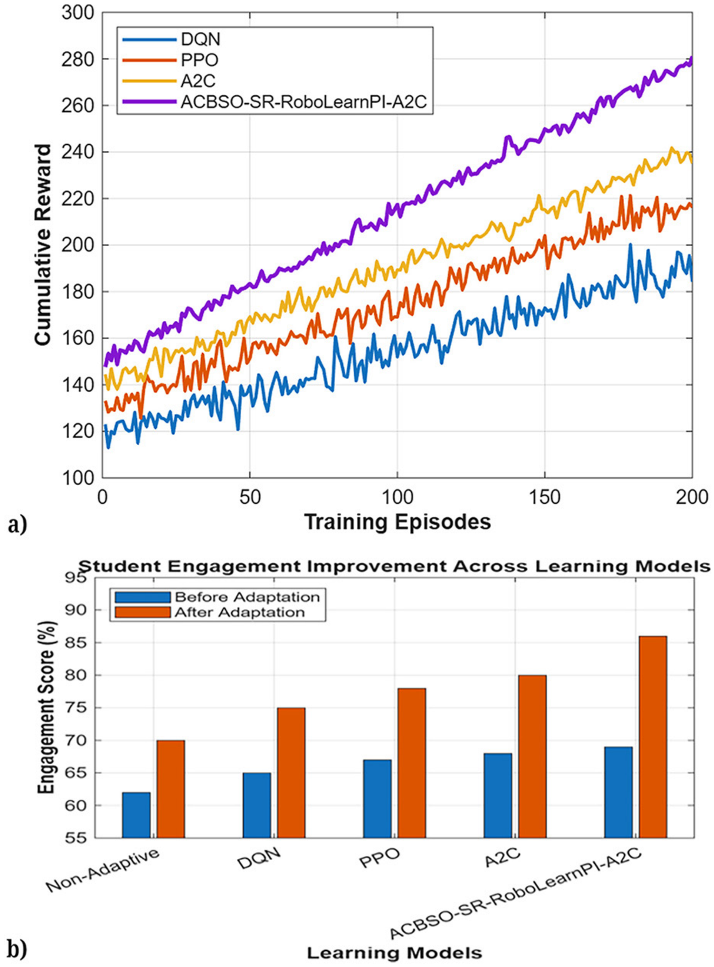

Fig. 3. a. Convergence analysis of reinforcement learning models and b. Result of student engagement improvement with different models

Figures 3a and 3b show how the engagement levels of students could be improved prior to and after adaptive intervention by various learning models. The given framework shows the most substantial growth in engagement scores, which reflects the effectiveness of the mentioned framework in dynamically changing the behavior of the robots in regard to the needs of individual learners. The suggested model also supports a higher level of engagement after adaptation, as compared to non-adaptive and conventional reinforcement learning baselines, which is indicative of a better level of interaction and learner attention. This improvement could be explained by the fact that the policy learning is optimized due to the application of adaptive chaos bird swarm optimization that enables the social robot to react better to the behavior and learning conditions displayed by students. These findings validate the hypothesis that adaptive social robotics could play an important role in enhanced learner interaction in smart learning settings.

4.5 Discussion

Experimental results of the proposed ACBSO-SR-RoboLearnPI-A2C framework highly contrast existing traditional reinforcement learning models in terms of adaptive learning and student engagement from inimical nonadaptive baselines. They have produced 91.8% accuracy with 19.8% learning gain and an 86% engagement rate and quite lower mean error values—Mean Absolute Error (MAE) = 0.214 and RMSE of 0.296. The convergence analysis of the model shows that ACBSO allows faster, though more stable, convergence of the policy. The ablation studies clearly indicate an important synergistic effect of ACBSO, chaos mechanism, and reward shaping pertaining to engagement. Qualitative analysis further supports these findings by stating that the RoboLearnPI robot showed more adaptive, contextual, and empathetic behaviors, which included dynamic customization of instructional prompts, task difficulty, and motivational feedback, leading to more extended interactions, natural responses, and more positive student engagement compared to the aforementioned baselines. All improvements were statistically significant based on paired t-tests ($p < 0.05$).

Impacts: The study has noted that it has various possibilities for improving socially conscious, self-aware educational robots using ensemble machine learning, offering personalized learning experiences that increase pupil participation while also easing teacher burdens in the classrooms.

Limitations: This study employed a single-institution data set from Oman, and this could impact the generalization capability. AutoRony was tested mainly in simulated scenes rather than getting deployed in actual classroom settings, and here, speech, facial expressions, and gestures, all to pick up multimodal traits, were not completely handled.

5 CONCLUSION AND FUTURE ENHANCEMENT

This study offered the ACBSO-SR-RoboLearnPI-A2C model, which has a significant impact on student learning and interaction because it is able to modify the behavior of the robot dynamically, according to the interaction information. Both quantitative and qualitative research results indicate that the combination of ACBSO, chaos mechanisms, and engagement reward systems can play a significant role in enhancing learning performance, policy convergence, and social responsiveness.

The future improvements can involve:

- Generalizing the framework to the case of multi-robot collaborative learning.
- Inclusion of multimodal interaction data (facial expression, speech, and gesture recognition data) to achieve better adaptation.
- Application of transfer learning to enable quick adjustment in new student groups.
- The inclusion of longitudinal assessment so as to measure long-term learning and engagement.

The study provides a strong background of intelligent, adaptive social robots in education that can be used to learn personally and interact socially.

6 ACKNOWLEDGEMENTS

This project received funding support from the Research Council of Oman. It was carried out through UTAS–Ibra under Research Grant No. BFP/RGP/ICT/24/060. The authors gratefully acknowledge this support in enabling the research outcomes.

7 REFERENCES

[1] D. Liu *et al.*, "Using educational robots to enhance learning: An analysis of 100 academic articles," in *Smart Computing and Intelligence*, Springer Singapore, 2024. https://doi.org/10.1007/978-981-97-5826-5

[2] G. Lampropoulos, "Social robots in education: Current trends and future perspectives," *Information*, vol. 16, no. 1, p. 29, 2025. https://doi.org/10.3390/info16010029

[3] K. Maj, A. Gołębicka, and Z. Siwińska, "How children learn from robots: Educational implications of communicative style and gender in child–robot interaction," *Computers & Education*, vol. 239, p. 105445, 2025. https://doi.org/10.1016/j.compedu.2025.105445

[4] S.-T. Chu, G.-J. Hwang, and Y.-F. Tu, "Artificial intelligence-based robots in education: A systematic review of selected SSCI publications," *Computers and Education: Artificial Intelligence*, vol. 3, p. 100091, 2022. https://doi.org/10.1016/j.caeai.2022.100091

[5] N. Akalin and A. Loutfi, "Reinforcement learning approaches in social robotics," *Sensors*, vol. 21, no. 4, p. 1292, 2021. https://doi.org/10.3390/s21041292

[6] H. Ackermann *et al.*, "How adaptive social robots influence cognitive, emotional, and self-regulated learning," *Sci. Rep.*, vol. 15, p. 6581, 2025. https://doi.org/10.1038/s41598-025-91236-0

[7] D. Tozadore and R. Romero, "Student behaviour modelling and adaptive techniques for social robots: Data-driven and teacher-perceived evaluations," *Int. J of Soc. Robotics*, vol. 17, pp. 3069–3086, 2025. https://doi.org/10.1007/s12369-025-01326-2

[8] K. Youssef, K. Said, S. Alkork, and T. Beyrouthy, "Social robotics in education: A survey on recent studies and applications," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 18, no. 3, pp. 67–82, 2023. https://doi.org/10.3991/ijet.v18i03.33529

[9] D. Tosic *et al.*, "Robot online learning to lift weights: A way to expose students to reinforcement learning," *International Journal of Online and Biomedical Engineering (iJOE)*, vol. 13, no. 8, pp. 4–19, 2017. https://doi.org/10.3991/ijoe.v13i08.7270

[10] Y. Wang and L. Wei, "Learner behavior recognition and dynamic guidance mechanism interactive mobile simulation systems for vocational education," *International Journal of Interactive Mobile Technologies (iJIM)*, vol. 20, no. 6, pp. 25–38, 2026. https://doi.org/10.3991/ijim.v20i06.60865

[11] G. Karalekas, S. Vologiannidis, and J. Kalomiros, "Teaching artificial intelligence and machine learning in secondary education: A robotics-based approach," *Applied Sciences*, vol. 15, no. 8, p. 4570, 2025. https://doi.org/10.3390/app15084570

[12] G. Karalekas, S. Vologiannidis, and J. Kalomiros, "Teaching machine learning in K–12 using robotics," *Education Sciences*, vol. 13, no. 1, p. 67, 2023. https://doi.org/10.3390/educsci13010067

[13] A. J. Romero-C. de Vaca, R. A. Melendez-Armenta, and H. Ponce, "Using social robotics to identify educational behavior: A survey," *Electronics*, vol. 13, no. 19, p. 3956, 2024. https://doi.org/10.3390/electronics13193956

[14] A. Alam, "Social robots in education for long-term human–robot interaction: Socially supportive behaviour of robotic tutor for creating robo-tangible learning environment in a guided discovery learning interaction," *ECS Transactions*, vol. 107, no. 1, pp. 12389–12403, 2022. https://doi.org/10.1149/10701.12389ecst
[15] R. Love, P. R. Cohen, G. Venture, and D. Kulić, "Adapting a teachable robot's dialog responses using reinforcement learning: Cross-cultural user study exploring effect on engagement," *ACM Transactions on Human-Robot Interaction*, vol. 14, no. 4, pp. 1–34, 2025. https://doi.org/10.1145/3743692
[16] P.-I. Ntomora and K. Petrakos, "Leveraging social network analysis in education through social robots: A review," *Global Journal of Engineering and Technology Advances*, vol. 20, no. 1, pp. 55–66, 2024. https://doi.org/10.30574/gjeta.2024.20.1.0113
[17] J. de Greeff and T. Belpaeme, "Why robots should be social: Enhancing machine learning through social human–robot interaction," *PLoS ONE*, vol. 10, no. 9, p. e0138061, 2015. https://doi.org/10.1371/journal.pone.0138061
[18] J. Dong *et al.*, "Artificial intelligence in adaptive and intelligent educational system: A review," *Future Internet*, vol. 14, no. 9, p. 245, 2022. https://doi.org/10.3390/fi14090245
[19] I. Gligorea *et al.*, "Adaptive learning using artificial intelligence in e-learning: A literature review," *Education Sciences*, vol. 13, no. 12, p. 1216, 2023. https://doi.org/10.3390/educsci13121216

8 AUTHORS

Dr. V. Mathivanan, a Lecturer at the College of Computing and Information Sciences, UTAS–Ibra, Oman, brings over 30 years of teaching and industrial experience in computer science and engineering. He has authored more than 50 publications, including 31 Scopus-indexed papers and a recent book, *Artificial Intelligence and the Future of Humans* (2024). His research spans network security, IoT, wireless ad hoc networks, cryptography, and AI, supported by multiple funded projects such as adaptive learning systems and emergency medical services. As a recognized Ph.D. supervisor, he has successfully guided six doctoral scholars and coordinated interdisciplinary research initiatives. His contributions, awards, and leadership roles highlight a career dedicated to advancing computing, security, and applied innovation (E-mail: vmathi@utas.edu.om).

Dr. K. Vimal Kumar Stephen is affiliated with the College of Computing and Information Sciences at UTAS–Ibra, Oman, has published extensively in areas such as wireless sensor networks, IoT, and cloud computing. His notable works include energy-aware secure wireless networks using particle swarm optimization and IoT-based health monitoring systems with cardiac classification. His publications in IEEE Xplore and international journals emphasize machine learning algorithms, deep learning models, and secure routing protocols for medical and network applications. With growing citations and recognition, Dr. Stephen's scholarship bridges advanced computing with practical innovations in healthcare and security (E-mail: vimal.victor@utas.edu.om).

Dr. Ramesh Palanisamy is affiliated with the College of Computing and Information Sciences at UTAS–Ibra, Oman, has contributed extensively to cybersecurity and networking research, with publications spanning international journals and conferences. His works include studies on wireless security surveys, optimized path selection in MANETs, and innovative smart traffic management systems. He has co-authored papers on phishing violence mitigation, IoT-enabled community security, and ensemble learning for cardiovascular disease diagnosis.

His book publications, such as Cyber Security and Artificial Intelligence and Future of Humans, further highlight his academic impact. With over 65 publications indexed in Scopus and Google Scholar, his scholarship bridges theory, practice, and emerging technologies (E-mail: ramesh.palanisamy@utas.edu.om).

Mr. Senthil Jayapal is a Lecturer at the College of Computing and Information Sciences, University of Technology and Applied Sciences–Ibra, Oman, specializing in artificial intelligence, machine learning, and deep learning. He has co-authored works on secure communication in underwater sensor networks, NLP-driven automated essay grading, and cognitive radio ad hoc networks, published in journals such as Scientific Reports and IEEE Xplore. His growing citation record and collaborations with scholars highlight his contributions to smart systems and AI-driven educational technologies (E-mail: senthil.jayapal@utas.edu.om).

Mr. Mohammed Tauqeer Ullah is a Lecturer at the College of Computing and Information Sciences, University of Technology and Applied Sciences–Ibra, Oman, specializing in machine learning, deep learning, IoT, and augmented reality. He has published in international journals and conferences. His growing research profile and collaborations highlight his contributions to adaptive learning systems and AI-driven educational technologies (E-mail: mohdtauqeer.ullah@utas.edu.om).

Mr. Mohamed R. Rafi is a Lecturer at the College of Computing and Information Sciences, University of Technology and Applied Sciences–Ibra, Oman, specializing in computer science and applied technologies. His academic contributions include teaching and research in areas of software engineering, networking, and applied computing, with a focus on integrating modern IT solutions into education and industry. Through collaborations and institutional initiatives, he continues to advance applied research and student mentorship in computing disciplines (E-mail: mohamed.rafi@utas.edu.om).