RoBERTa-Enhanced Actor–Critic Reinforcement Learning for Adaptive and Personalized ESL Instruction

Dr. Angalakuduru Aravind[1], A. Swathi[2], Dr. Jillellamoodi Naga Madhuri[3], R. Aroul Canessane[4], K. Lalitha Vanisree[5], Elangovan Muniyandy[6], Rasha M. Abd El-Aziz[7]

Assistant Professor, Department of H&S, Anurag Engineering College, Kodad, Suryapet (Dist), Telangana - 508206, India[1]
Assistant Professor, Department of English, Aditya University, Surampalem, Andhra Pradesh, India[2]
Assistant Professor, Department of English, Siddhartha Academy of Higher Education (SAHE) Deemed to be University, Kanuru, Vijayawada, India[3]
Department of Computer Science and Engineering, Sathyabama Institute of Science and Technology, Chennai, Tamil Nadu, India[4]
Assistant Professor, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur Dist., A.P., India[5]
Department of Biosciences-Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai, India[6]
Department of Computer Science-College of Computer and Information Sciences, Jouf University, Saudi Arabia[7]
Faculty of Computers and Information-Computer Science Department, Assiut University, Assiut, Egypt[7]

***Abstract*—Technology-Assisted Language Learning (TALL) has developed and has greatly transformed the way English as a Second Language (ESL) is taught. The current digital resources and smart solutions have enabled more interactive and accessible learning, providing learners with an opportunity to train their skills at any time and place. Nevertheless, most of the current systems remain based on strict rules or conventional supervised training approaches. These methods can demand large quantities of labelled data, are inflexible in the learning process, and have little in the way of individualized feedback. Consequently, students may remain inattentive, and the acquisition of all the necessary language skills, such as reading, writing, listening, and speaking, may be unequal. In order to address such shortcomings, this study presents T-RLNN (RoBERTa-based Reinforcement Learning Neural Network), which is a dynamic model of ESL teaching. T-RNN combines deep contextual language comprehension and reinforcement learning in order to customize teaching to every learner. The RoBERTa encoder can retrieve semantic and syntactic feedback on responses of learners, and an actor-critic reinforcement learning agent can modify teaching plans in real time. The agent takes into account the learner-specific factors, i.e., proficiency, response time, engagement, and interaction behavior, to give the best guidance. It was trained in Python using PyTorch and tested on a curated dataset of 5,000 responses of a learner in reading, writing, listening, and speaking tasks. T-RLNN performed better than conventional models, such as Support Vector Machines, random forests, and conventional deep neural networks, with a 94.8 % accuracy, 92.7 % F1 -score, and 71.5 % Adaptivity Index. These findings indicate that T-RLNN has the potential to provide insightful, interactive, and learner-oriented ESL training and open the way to smarter and more adaptable language learning systems.**

Keywords*—English as a Second Language; adaptive learning; reinforcement learning; RoBERTa; intelligent tutoring systems

I. INTRODUCTION

In recent years, the field of education, specifically language learning, has been dramatically influenced by technology [1]. English as a Second Language (ESL) classrooms often include a wide variety of learners, creating a need to tailor instructional approaches to the various needs [2]. Traditional teaching, like what Acocella et al. (2022) suggested, typically does not work well in relation to the multiple styles of learning, developmental levels, and intelligences in one classroom [3]. This illustrates the need for learning environments that are flexible, changeable, and adaptable to a learner's needs. Technology-Assisted Language Learning (TALL) or instructional resources have been utilized, yet some remain fixed and rule-based [4]. Many TALL also lack real-time ecological flexibility, individualized feedback, and adaptability to the learner's emotions and performance [5], [6]. Ultimately, these factors may cause frustration or disengagement, and ultimately limited progress [7] [8].

Presently, the predominant supervised learning techniques found in AI-driven ESL systems rely on relatively large amounts of labeled data to complete narrow tasks, such as grammar correction or vocabulary checks [9], [10]. Most models utilize the common four skills of language: listening, speaking, reading, and writing as discrete skills and do not allow for holistic learners to grow [11]. To address these issues, presented in this work is a new T-RLNN as a framework and method that adds RL as a contextually based language heuristic. In contrast to traditional systems, T-RLNN approaches the instructional strategy differently without fixed content or rigidly adhered-to rote rules—but rather classifies and customizes learning through live interactions with learner performance, engagement, and progress. Within the T-RLNN framework, an RL agent, using an actor–critic architecture, operates on learner states constructed by RoBERTa embeddings and behavioral features. In that context, T-RLNN is customized to learners through varying or adjusting working task difficulty, the instructional pacing of interactions, and delivery instructional style. T-RLNN provides a platform that is customized, scalable, and contextual for ESL learner support, while making their learning journey unique.

A. Research Motivation

The rationale behind this study is that current ESL systems lack flexibility, personalization, and integration of skills. The purpose of this study is to develop an adaptive tutoring system based on RoBERTa and reinforcement learning that will improve the engagement of learners and help them maintain language development.

B. Significance of the Study

The study is relevant because it presents the T-RLNN framework, a combination of contextual understanding of language and reinforcement learning that facilitates adaptive ESL learning. It is not limited to conventional approaches, but it can provide customized, scalable, and interactive learning, helping to progress the intelligent tutoring systems.

C. Recent Innovation and Limits

The more recent developments in the ESL learning strongly depend on deep learning and transformer-based models like BERT and RoBERTa to comprehend the language contextually and enhance grammar correction, vocabulary acquisition, and reading comprehension[12]. Although it works, these models require a lot of large labelled data and have less real-time flexibility for a specific learner. By introducing contextual representations into reinforcement learning, these gaps could be resolved with flexible, personalized, and adaptive choices of instruction [13].

D. Key Contributions

- Proposed a novel RoBERTa-based Reinforcement Learning Neural Network (T-RLNN) that integrates deep contextual language understanding with adaptive instructional strategies for ESL learning.
- Modeled a comprehensive learner state by combining linguistic features extracted through RoBERTa with behavioral indicators such as accuracy, response time, and engagement levels.
- Employed an actor–critic reinforcement learning agent to dynamically adjust task difficulty, feedback, hints, and exercise personalization in real time based on learner performance.
- Formulated the learning process as a Markov Decision Process (MDP), enabling interpretable decision-making and long-term optimization of instructional strategies.
- Demonstrated that the proposed model outperforms conventional approaches, including SVM, Random Forest, and DNN, in terms of precision, F1-score, adaptability, and overall learner engagement, offering a scalable and personalized ESL tutoring solution.

The remaining part of the section is divided as follows: Section II provides a review of prior works and Section III describes problem statement. Section IV details the proposed framework with its methodology, Section V with the demonstration of the results. Finally, Section VI presents the conclusion of the proposed framework, while offering recommendations for further research and application.

II. Related Works

Kaur, Kumar, and Kaushal [14] provide an analytical perspective into the trends and development of technology-facilitated language learning systems. These systems are regarded as intelligent since they amend their mode of delivery of knowledge based on the learner. The research method employed was a peer reviewed articles from several reputed journals where the study was conducted between the years 2011 and 2021. The authors analyzed adaptive systems through three dimensions: analysis of space and time, system and learner factors and the accommodation offered. Many TALL systems are available nowadays, and their usage is rapidly growing, mainly in Asia, where English is taught as a foreign language. The fact that the systems are adaptive means that their ability to meet university students' needs is taken into consideration. The study presents a breakthrough in adaptive systems for language learning, where English is the most researched language. However, the review does not include a quantitative analysis or allocate specific performance indicators to measure the actual significance of adaptive systems. Furthermore, some difficulties, such as the application of such systems on a large scale, and the benefits that could be achieved by involving local authorities and educators in the implementation of such systems, are discussed.

Khasawneh [15] examines the effect of multimodal teaching and learning practices on Sudanese dyslexic students' English language learning in KSA. The study adopted a quantitative research approach and with the help of paired t-tests, Pearson's correlation, multiple regressions, and ANOVA to determine the impact of multimodal approaches for teaching 30 dyslexic learners. The results revealed a significant enhancement of the students' performance in language when the modes of teaching were implemented. The study also focuses on the fact that how often strategies are used to benefit the success of the interventional program. The study also emphasizes that instead of a mass approach, individualized instruction may be more beneficial for learners with different levels of language competence at the beginning of the course. Nevertheless, one of the limitations of this study is that it involved a small sample, thus the results of the study do not have broader applicability. Also, the study fails to make an extended analysis of the fluency of the language retained by students while learning through MMT and the extent of interaction between various types of multimedia resources and students. Nevertheless, this research adds useful knowledge about the learning paradigms, namely, adaptive learning for learners diagnosed with learning disabilities, including dyslexia.

The research study by Chanderan and Hashim [16] identifies the ESL students' LLS at a private university in Selangor, Malaysia. The research used a quantitative cross-sectional survey design with a survey instrument developed from

Oxford's SILL. The target population was 200 freshmen from different faculties to understand on how students utilize LLS to improve their language skills. From the descriptive statistical analysis, the findings showed that the students prefer to apply those strategies aimed at enhancing their language proficiency. The results reveal cognitive and metacognitive approaches as the first ones with identified high efficiency for language acquisition. The study also focuses on the use of learner preferred strategies in language learning and acquisition processes. This is, however, the limitation of the research because the data collected may be influenced by the respondent's perception. As this study proposes, the study is limited by the failure to track the strategies' long-term effectiveness as well as their effects on student general performance. However, the study reveals insight into the ESL learners' sociocultural literacy practice, indicating to the educators the most appropriate ways to develop effective ESL curriculum.

Chung et al. [17] carried out the research with the aim of ascertaining how the retrieval of video playback affects ESL learners' learning. They use metrics that are more or less based on an observer's eye check and survey while evaluating the video playback speeds. The sample consisted of 32 ESL participants with gradual and immediate speed change during video watching. The characteristics favored by the results entailed that gradual speed changes considerably improved learners' flow state, video comprehension, as well as cognitive load. Therefore, these findings imply that slower changes could create a less obtrusive learning process for ESL students. The main drawback of the study is the limited scope of the same in the inclusion of a relatively small target population. However, the study concentrated on video comprehension, without much regard to other sectors in language acquisition, including listening and speaking. Nevertheless, the study gives insight that may be rather useful when it comes to enhancing resultant video-based learning; it provides recommendations that can be useful in the development of adaptive learning systems taking into account cognitive processing capacity.

Young and Shishido [18] investigated the application of the multipurpose chatbot called ChatGPT in the production of English reading content for ESL students. The study also looked at how ChatGPT can be useful in generating texts of different difficulties to meet needs of ESL students with a limited vocabulary. The development research approach was therefore based on the comparison of different scores of reading ease between the contents produced by OpenAI's ChatGPT and the actual reading texts. The study showed that ChatGPT could indeed generate simpler and more comprehensible texts, it can, therefore, be used to enhance language learning. The study recognizes it in the same breath, that as a recent innovation, ChatGPT has not undergone conventional assessment in language learning settings. Besides, readability scores can be used as a measure of learning content quality but do not take into account further factors, for instance, their learning process engagement or understanding. Nevertheless, this study evidences that the proposed concept of integrating AI technologies such as ChatGPT into adaptable language learning systems can be a groundbreaking model to modernize the approach towards selecting appropriate readings for ESL learners.

In the study, Monika and Suganthan [19] examines the impact of ChatGPT on English language learning for the English as Second Language students. The study adopted a cross-sectional survey eliciting data from ESL students taking English language classes in various institutions in the Vellore District. The study aimed at establishing the effect of ChatGPT model on the language abilities of the learners particularly on listening, speaking, reading, and writing (LSRW) skills. In the findings, it showed that ChatGPT had a positive impact on learner's vocabulary and the overall language usage. However, the study fails to draw the negative side of applying an AI model, including the one used in developing ChatGPT, such as the ability of the model to misunderstand context or make inaccurate content. Such findings also do not take into account a variety of ways, including durability of knowledge retained when using the ChatGPT, or its impact on learner's language proficiency or assessment results. However, there are still deficits in their coverage and application of ESL learning, but the study proves that AI technologies could bring positive impacts to developing ESL learning with the feedback and interactive generation that could attend the learners' needs.

The research carried out by Naparan and Bacasmot [20] focuses on an investigation of M-learning with Smartphone applications and their impact on students' communication competency with reference to their learning ESL. This research utilized descriptive-correlational research design to gather data from senior high school students in Davao City. When asked about Smartphone apps for learning, the study established that there exists a positive correlation between Smartphone app use and Communication competence, hence the importance of Smartphone apps in improving ESL learners' linguistic proficiency. Nonetheless, the research indicated that there was insignificantly decreased use of English language problems and smartphone apps. The study also employed regression analysis to establish the effect of smartphone apps on second languages but relied on subjective responses that reduce validity. However, the research sample is restricted to a particular group of students and the city, Senior high students in Davao City which means generalization of the results is impossible. However, such limitations would not deny the fact that this research has succeeded in offering gainful understandings on how the mobile learning technologies could be implemented and used to advance the effectiveness of ESL learning and the communication effectiveness of language learners.

Ahmad et al. [21] aims at identifying the language learning strategies employed by Primary ESL learners in Sarawak. The study adopted a quantitative approach to research and more specifically a survey to establish the extent of use of language learning strategies among Year 5 students. The study findings showed that students deployed more the cognitive-affective approach as well as the affective-motivation strategies most tasks than the other identified strategies such as memorization and compensation approaches. According to the study, the strategies discussed above can be used in order to improve comprehension and fluency when reading as well as students' linguistic competence. However, the present research does not take into account the implementation of these strategies as long-term benefits for students' language development. Also, there is a lack of generality in the study since the participants involved were Year 5 students in a definite region only. Nevertheless, the study implies that ESL students should choose their preferred strategies when learning the English language, helping educators in the process of informing the way in which these children learn English.

III. Problem Statement

The growing popularity of technology in ESL learning has enhanced the availability of learning; nevertheless, most of the current systems are still stagnant, one-dimensional, and ill-advised to meet the needs of various learners [22], [23]. The conventional methods of tutoring lack responsiveness in real-time and are not responsive to the progress of the learner thus the tutoring methods are not able to support active listening and the balanced development of the language skills, especially in heterogeneous classroom environments. Such restrictions provide structural obstacles to language learning. To overcome this issue, this study discusses the T-RLNN framework, a contextual reinforcement learning-based model that constantly modifies the instructional pathways due to performance, engagement, and skill development of the learner. In such a way, T-RLNN creates a more flexible, responsive, and learner-centered ESL learning process.

IV. Proposed Framework for Adaptive ESL Instruction Using T-RLNN

The suggested T-RLNN model will be created to weigh the linguistic and behavioral components of the interaction between the learner and allow the model to change the instructional methods in real-time. First, the learner's inputs of the form of a text response, task type, response duration and indicators of engagement are processed so as to create structured and meaningful feature representations. The text responses undergo normalization by the form of lowercasing, punctuation marks and lemmatization and then they are tokenized by the RoBERTa tokenizer that is founded on the basis of lower byte-pair encoding. The process attains rich contextual embeddings, which reflect a semantic meaning and syntactic nuances of language use by learners. In addition to linguistic characteristics, behavioral data are of considerable importance to the modelling of the learner. Categorical variables, including the type of the task or the level of a particular proficiency, will be converted to dense vector representations, whereas the numerical variables, including the time of response and the engagement scores, will be rescaled to ensure that the values are the same across learners. These behavioral and linguistic representations are then combined together to create a holistic state of the learner and this becomes the basis of adaptive decision making. This state of learning is fed through an actor-critic reinforcement learning agent that decides the most appropriate instructional action at a given step. The agent can either change the difficulty of the tasks, give specific hints, or choose the other types of exercises that suit the learner, depending on the performance and engagement of the learner. The learning environment is able to constantly test these actions by rewarding them based on accuracy, participation and effectiveness of the response, meaning that the model is able to improve its policy as it advances. Every interaction between the learner is viewed as an episode, and the system is able to record as much as possible of each individual learning pattern and provide them with unique dynamic instruction. The T-RLNN system is a pipeline that end-to-end system pipeline that integrates contextual language comprehension with reinforcement-based learning adaptation provides a more efficient, reactive, and learner-focused ecological system to ESL tutoring.

The workflow of the proposed T-RLNN framework of adaptive ESL instruction is provided in Fig. 1. The inputs of learners, such as text answers, type of task, response time, and indicators of engagement, are preprocessed first and converted into the form of structured feature representations. The RoBERTa byte-pair encoder tokenizes text data to create deep contextual embeddings, and the deep contextual embeddings are combined with the encoded behavioral features to create a single state vector of the learner. This is the condition where language competence and interaction behavior are captured. A reinforcement learning agent, whose components are actors and critics, is an agent that employs this state to dynamically change instructional strategies, including task difficulty or hints, on the basis of real-time performance of the learner. Based on the learner's actions, the environment will assess the actions and give rewards based on accuracy, engagement, and efficiency of response and allow policy refinement and individualized ESL learning.

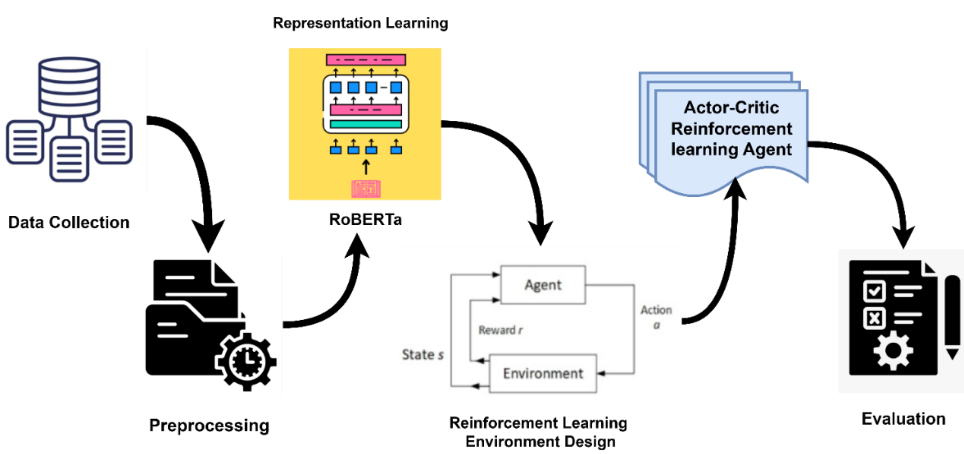

Fig. 1. Workflow of the T-RLNN framework.

A. Dataset Description

The experiments are based on publicly available Kaggle Computer-Assisted Language Learning (CALL) dataset [24], comprising of bibliographic and textual records of research publications indexed by Scopus between 1983 and 2020. The dataset also contains structured metadata in the form of publication titles, abstracts, author details, affiliation, keywords, document type, year of publication and source information. These are textual elements that offer a corpus that is rich in language representation learning and semantic analysis in terms of ESL and CALL studies. The dataset is used to assess the efficiency of the contextual language modelling and adaptive instructions policy learning with the text-based input instead of direct textual and metadata-based samples.

B. Data Preprocessing

Data preprocessing transforms raw textual and metadata-based samples into structured, normalized, and feature-rich inputs by cleaning text, encoding categorical variables, and scaling numerical features, ensuring robust representation for downstream modeling.

1) Text normalization: Text responses of the learners are pre-processed to trim off noise, and create uniformity prior to tokenization. The procedure will involve lowercasing, eliminating of punctuations, and lemmatization. The steps, standardizing input text, enhance downstream representation learning, and promise that the RoBERTa tokenizer is presented with uniform sequences, which lack the variability due to surface-level variation in writing. It is represented, as in Eq. (1):

$$T_{norm} = Lemmatize(Lowercase(T_{raw} - P)) \quad (1)$$

where, T_{raw} is the original response, P denotes punctuation set, $Lowercase(\cdot)$ converts all characters, and $Lemmatize(\cdot)$ reduces words to canonical forms. The output T_{norm} represents normalized learner text ready for sub word tokenization.

2) Categorical feature encoding: Embedding layers are used to convert the learner-related categorical attributes, i.e. type of a task or a level of proficiency, into dense numerical vectors. When compared to one-hot encoding, the embeddings memorise semantic similarity among the categories, which allows the RL agent to acknowledge associations (e.g. similarity of tasks) and extrapolate more effectively across various learner groups. It is expressed, as in Eq. (2):

$$E_c = W_c \cdot OneHot(C) \quad (2)$$

Here, C represents the categorical feature (e.g., task type), $OneHot(C)$ is its one-hot encoded vector, and W_c is the trainable embedding matrix. The product yields dense embedding E_c, capturing category information in a low-dimensional representation.

3) Numerical feature scaling: Continuous features such as textual complexity indicators, engagement, and retention are standardized for stable optimization. Raw values are transformed into zero-mean, unit-variance distributions. This ensures features contribute proportionally to state vectors and prevents domination by attributes with large numerical ranges, thus improving training stability and model convergence. It is computed, as in Eq. (3):

$$x' = \frac{x-\mu}{\sigma} \quad (3)$$

where, x is the original feature value, μ is the mean of the feature across training samples, and σ is the corresponding standard deviation. The scaled feature x' ensures normalized contribution to the learner state representation.

C. Representation Learning (RoBERTa)

The discussion demonstrates that the answers of learners are converted to high quality contextual embeddings by employing RoBERTa-base, a pre-trained transformer that can capture deep semantic and syntactic links in the text. Normalization and tokenization of responses is initially done with the use of the Byte-Pair Encoding (BPE) of RoBERTa, which provides the same representation to words, sub words, and special tokens by using unique IDs. Padding or truncation of token sequences to a constant maximum length L is then done so that a batch of token sequences can be processed, and attention masks used to differentiate between valid tokens and padding. The sequences are given through the multi-layer self-attention and feedforward representation encoders of RoBERTa to produce contextualized token representations that encode word-to-word dependencies. Because the Kaggle CALL data do not offer explicit logs of learner interaction or behavior, an engagement score is computed using the information of interaction clues on the content level as a proxy measure. Particularly, engagement is deduced by the normalized textual and structural features that each document has such as the length of the abstract, the density of keywords, and citation-related metadata. These characteristics indicate the intensity and applicability of interaction with ESL-related material and are normalized by min-max scale to provide the comparison between samples. The engagement score E is computed, as in Eq. (4):

$$E = \alpha \cdot L_{norm} + \beta \cdot K_{norm} + \gamma \cdot C_{norm} \quad (4)$$

where, L_{norm} represents normalized abstract length, K_{norm} denotes normalized keyword count, and C_{norm} corresponds to normalized citation-related indicators. The weighting coefficients α, β, γ are empirically set to ensure balanced contribution from each component. Lastly, attention-weighted mean pooling is contrasted to obtain concise, semantically meaningful latent encodings of learner responses to serve as a strong basis of downstream tasks in adaptive ESL teaching mathematically expressed, as in Eq. (5):

$$e = \frac{\sum_{i=1}^{L} h_i \cdot m_i}{\sum_{i=1}^{L} m_i} \quad (5)$$

Here h_i represents the hidden state of token i, while $m_i \in \{0,1\}$ indicates whether a token is valid. The result is a 768-dimensional sentence-level embedding which is effective at both grammatical and semantic accuracy These behavioral proxy features, including the engagement score defined in the previous subsection, are concatenated with RoBERTa embeddings to form a unified learner state representation. The resultant aggregated vector is then fed through a fully connected layer to give the dimensional consistency and equal contributions by all the features and to generate a dense learner state. This coherent representation allows the reinforcement learning agent to make adaptive and context-sensitive instructional choices and be more successful than shallow text models in facilitating long-lasting and personalized learning. It is given in Eq. (6):

$$\text{sim}(e, e') = \frac{e \cdot e'}{\|e\| \, \|e'\|} \quad (6)$$

It computes the cosine similarity between two learner-response embeddings e and e', capturing semantic similarity for tasks like feedback assessment, clustering, or adaptive instruction.

D. Reinforcement Learning Environment Design

Adaptive ESL tutoring process is described as a Markov Decision Process (MDP), which gives a transparent pattern of interactions between the learner and the reinforcement learning (RL) agent. The state s_t at time t gives the profile of the learner at that time comprising RoBERTa-based contextual embedding alongside behavioral variables, including accuracy, response time, and engagement. The action is related to the selected instructional strategy, such as variations in task difficulty, exercises in grammar or vocabulary, comprehension, or hints. The reward measures effectiveness of the action using both immediate performance and engagement measures is calculated, as in Eq. (7):

$$r_t = w_{acc} \cdot \Delta Acc_t + w_{time} \cdot \Delta(-RT_t) + w_{eng} \cdot \Delta Eng_t \quad (7)$$

where, ΔAcc_t denotes the change in learner accuracy, $\Delta(-RT_t)$ represents improvement in response efficiency (lower time is better), and ΔEng_t reflects variation in learner engagement. The weights w_{acc} , w_{time} , w_{eng} are tuned to balance short-term performance with long-term engagement. The reward for the RL agent is to maximize the expected cumulative discounted reward over a sequence of learner interactions is expressed, as in Eq. (8):

$$J(\pi) = E_\pi[\textstyle\sum_{t=0}^{T} \gamma^t r_t] \quad (8)$$

where, π represents the policy mapping states to actions, T is the length of an episode, and γ is the discount factor favoring long-term competence. Through the explicit specification of states, actions, and rewards in this MDP model, the system facilitates the RL agent to dynamically adjust instruction methods for individualized, situation-sensitive ESL learning maximizing both correctness and engagement on various learners.

E. Actor–Critic Reinforcement Learning Agent

To enable adaptive decision-making, this research applies an Actor-Critic model, whereby the agent is spatially learning a policy of action choice as well as an evaluation of state value function. The actor network computes a probability distribution of the potential instructional strategies given the current learner state vector and the critic network predicts the mean of the expected return to give feedback to stabilize learning. It is represented, as in Eq. (9):

$$A(s_t, a_t) = R_t + \gamma V(s_{t+1}) - V(s_t) \quad (9)$$

In Eq. (9), the measure of goodness of action taken a_t performed in state s_t compared to the baseline estimate of a critic. In this case R_t is the instant payoff, $V(s_t)$ and $V(s_{t+1})$ are the expected values of the states and g is the discount factor. Positive advantage means that the action of instructional choice has enhanced the performance of learners in a better way than projected and as such, the actor network will pursue strategies that are positive to reinforce. The workflow of RL Agent is represented in Fig. 2.

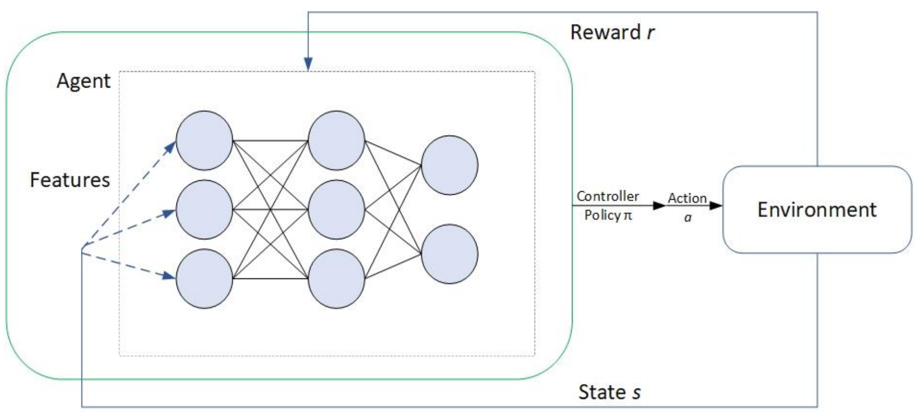

Fig. 2. Workflow of reinforcement learning agent.

Fig. 2 shows the architecture of the actor–critic reinforcement learning agent in the T-RLNN framework. The learner state vector, combining RoBERTa embeddings and behavioral features, feeds into a shared neural backbone that splits into two heads: the actor, which selects instructional actions, and the critic, which estimates expected rewards. The agent iteratively updates its policy based on rewards derived from learner accuracy, engagement, and response efficiency, enabling context-aware, adaptive instruction that evolves with learner performance.

Both networks are similar in that they have a common backbone of feedforward processing the learner state which then forks into two task active heads: a SoftMax output of the actor's policy and a scalar regression output of the value estimate of the critic. The policy $\pi_\theta(a_t \mid s_t)$ is parameterized by θ and a_t every time step, the action atbased on the current state s_t is chosen. The parameterized critic ϕ estimates the value function $V_\phi(s_t)$ which is a measure of the expected total reward at that state. Learning is then directed by the advantage function which compares the degree to which a specific action is preferred to the predictive value of the critic. It is computed, as in Eq. (10):

$$A_t = r_t + \gamma V_\phi(s_{t+1}) - V_\phi(s_t) \tag{10}$$

In Eq. (10), r_t is the immediate reward and γ is the discount factor. The actor is updated by maximizing the expected advantage, while the critic minimizes the error in value prediction is represented, as in Eq. (11):

$$L(\theta, \phi) = -Et[log\pi_\theta(a_t \mid s_t) \cdot A_t] + \alpha \cdot \delta_t^2 \tag{11}$$

In this case, the policy loss gradient (actor) is the first term, and the value loss (critic) is the second term and their trade-off is regulated by α. Such joint optimization makes the actor to learn to suggest effective instructional strategies and the critic to assess their long-term contribution to the level of engagement and proficiency of the learners.

F. Learning Phase

Training of the T-RLNN framework entails two closely related factors, which are a reinforcement learning agent and the ever-changing learner state representation. RoBERTa takes the textual response of the learner at each step to create contextual embeddings that are subsequently added with behavioral characteristics, including accuracy, response time, and engagement, to create a complete learner state vector. According to this condition, the actor network will choose a proper instructional action. This is then simulated by the environment as the learner progresses and given feedback, which is translated into a reward signal. The critic considers the quality of the action chosen and employs this to polish the policy. By continuing its updates, the agent will acquire the ability to make compromises to improve short-term performance at the cost of long-term proficiency and involvement. Behavior cloning is used to log interactions in order to maintain a stable and efficient training, dealing with poor initial decisions. Further, mini-batch updating, entropy regularization and early stopping are also used to promote exploration, reduce overfitting, and strong convergence leading to a powerful, adaptive tutoring system. It is calculated, as in Eq. (12):

$$r_t = \alpha \cdot \text{accuracy}_t + \beta \cdot \text{engagement}_t - \gamma \cdot \text{response time}_t \tag{12}$$

It computes the reward r_t as a weighted combination of learner performance, engagement, and response time. It is given in Eq. (13):

$$\theta \leftarrow \theta + \eta \cdot \delta_t \nabla_\theta \log \pi_\theta(a_t \mid s_t) \tag{13}$$

Here, θ represents the actor parameters, π_θ the policy, δ_t the temporal difference error, and η the learning rate; the policy is updated to maximize expected rewards.

Algorithm 1 integrates RoBERTa embeddings with behavioral features to form comprehensive learner states, enabling an Actor–Critic RL agent to dynamically adapt instructional strategies and maximize accuracy, engagement, and long-term retention.

Algorithm 1: T-RLNN (RoBERTa-based Reinforcement Learning Neural Network)

```
Begin
  Input: CALL dataset D = {responses, accuracy, timestamps, engagement, task type}
  Output: Trained policy πθ and critic Vϕ
    For each learner response r in D do
      Normalize text: lowercase, remove punctuation, lemmatize
      Tokenize using RoBERTa BPE with max length L
      Encode categorical features as embeddings
      Scale numerical features using z-score
    End For
    For each response r do
      Pass tokenized sequence into RoBERTa-base
      Obtain hidden states H = {h1, h2, …, hL}
      Compute embedding e = mean_pool(H, mask)
      Concatenate e with normalized features
      Project into fixed-size learner state vector s
    End For
    Define state s = learner state vector
    Define action a ∈ {increase_difficulty, decrease_difficulty, grammar_ex, vocab_ex, comprehension_ex, hint}
    Define reward r_t = w_acc · ΔAcc_t + w_time · Δ(−RT_t) + w_eng · ΔEng_t
    Initialize policy π_θ(a_t | s_t) and value function Vϕ(s)
    For episode = 1 to MaxEpisodes do
      Initialize learner state s0
      For t = 1 to T do
        Select action at ~ π_θ(a_t | s_t)
        Apply action in environment → observe new state st+1 and reward rt
        Compute advantage A_t = r_t + γV_ϕ(s_{t+1}) − V_ϕ(s_t)
        If A_t > 0 then
          Update actor parameters θ to increase ~π_θ(a_t | s_t)
        Else
          Update actor parameters θ to decrease ~π_θ(a_t | s_t)
        End If
        Update critic parameters ϕ to minimize value error
        Set s_t ← st+1
      End For
    End For
    Output optimized policy πθ*
    Evaluate using metrics
    Evaluate learner-centered metrics: Engagement Score, Retention Index, Adaptivity Index
End
```

The T-RLNN is a novel reinforcing idea of combining contextual language understanding by RoBERTa with an Actor-Critic reinforcement library to use adaptive ESL teaching. Through a balanced approach to semantic depth and behavioral insight, the framework produces strong learner state representations, which can be used to provide context-sensitive personalization to outperform traditional models and provide real-time adaptive tutoring with better engagement, retention and proficiency results.

V. Results and Discussion

The obtained experimental data prove that the integration of learning based on reinforcement and neural networks based on transformers can be a key factor in accelerating the ESL teaching process. The suggested workflow combines all data preprocessing, learning contextual representation with RoBERTa, a simulated reinforcement learning setting, and optimizing the actors and critics to help the system to provide adaptive and person-centered instructional plans. All tests were run in Python on the CALL dataset, processed by ordinary NLP technology and tokenized using the RoBERTa tokenizer. To have equal and trustworthy assessment, the data was divided into training (70%), validation (15%), and testing (15%). In order to evaluate the efficiency of the suggested approach, the performance of the method was evaluated against a number of conventional models, such as Support Vector Machines, Random Forests, and a feedforward Deep Neural Network, along with an earlier developed RLDNN model. The five-fold cross-validation was employed to fine-tune both the baseline and proposed model and the results averaged across three separate runs, to provide statistical strength. All the evaluation metrics showed that T-RLNN architecture is highly flexible and efficient, and better and stable performance was attained than the traditional approach. These results note the importance of combining deep contextual language representations with reinforcement learning to teach ESL in an adaptive manner. In general, the findings offer an important and valid comparison to prove the usefulness of the suggested framework in facilitating individualized and responsive language learning. Table I summarizes the simulation parameters that were employed in the experiments.

TABLE. I. SIMULATION PARAMETER AND HARDWARE SETUP

Parameter	Value / Description
Dataset	CALL (Computer-Assisted Language Learning)
Data Split	70 % Train / 15 % Validation / 15 % Test
Text Preprocessing	Lowercasing, punctuation removal, lemmatization
Tokenizer & Embedding	RoBERTa-base (768-dim, max length = 128 tokens)
State Representation	RoBERTa embedding + behavioral features (accuracy, response time, engagement)
Action Space	{Increase/Decrease difficulty, Grammar, Vocabulary, Comprehension, Hint}
RL Algorithm	Actor–Critic (Advantage Actor–Critic, A2C)
Reward Weights	w_acc = 0.5, w_time = 0.3, w_eng = 0.2
Learning Rate & Batch	1e-4 (Adam optimizer), Batch size = 32
Training Episodes	500 with early stopping (patience = 10 epochs)
Hardware Configuration	Intel i9-11900K CPU, NVIDIA RTX 3090 GPU (24 GB), 64 GB RAM
Software Environment	Python 3.9, PyTorch 2.0, HuggingFace Transformers, Ubuntu 22.04

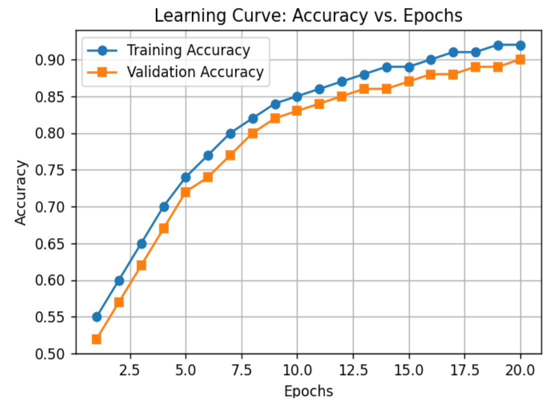

Fig. 3. Learning curve accuracy.

Fig. 3 shows how the training and validation accuracy of the proposed T-RLNN framework evolves with the epoch. It shows that the model can enhance the classification of the learner interaction across multiple iterations and validation accuracy is also close to training accuracy meaning that it has a good generalization ability. The accuracy improvement in this ESL adaptive learning scenario is associated with the ability of the framework to match the state representations of the learners with the best teaching method. The diagram confirms the usefulness of the combination of RoBERTa embedding and RL and demonstrates that the model will stabilize and reach high accuracy results of adaptive tutoring.

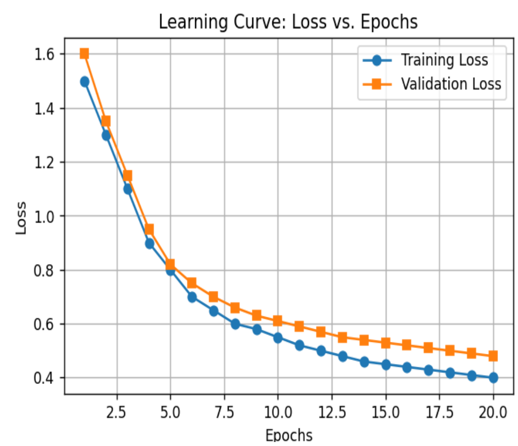

Fig. 4. Learning curve loss.

In Fig. 4, the training and validation loss are decreasing as the T-RLNN model is optimized. A gradually diminishing curve in both curves signifies constant convergence, and the decreasing distance between them shows an insignificant overfitting. In this ESL learning model, less loss is a positive indication of an increased efficiency of the system in relating learner response to the right instructional action. The model achieves the minimization of prediction errors and enhances instructional flexibility by learning to combine semantic embeddings of RoBERTa and RL cues. The loss curves therefore, validate stability, scalability and strength of the suggested adaptive tutoring system.

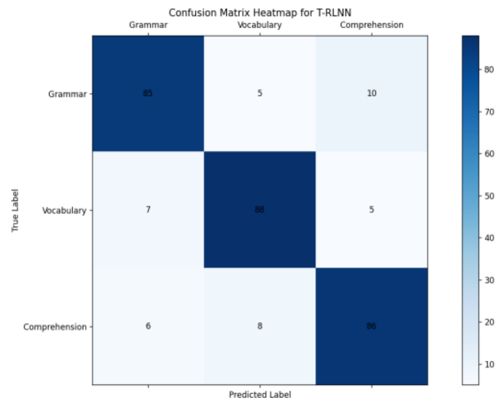

Fig. 5. Confusion matrix heatmap for T-RLNN.

Fig. 5 shows the confusion matrix heatmap of the proposed T-RLNN framework where it is shown that it has successfully performed in the classification of ESL modules which are grammar, vocabulary, and comprehension. The diagonal values indicate the correctly classified responses whereas the off-diagonal values show the misclassifications. The high concentration in the direction of the diagonal indicates that the integration of the RoBERTa embeddings with the reinforcement learning could help in facilitating adaptive instruction. The heatmap also gives the information regarding the task recognition: grammar and vocabulary tasks are recognized better, and comprehension tasks are a little bit harder. On the whole, this visualization points to the fact that the model can also cope with a wide variety of ESL learning tasks and offer tutoring that is flexible and personalized.

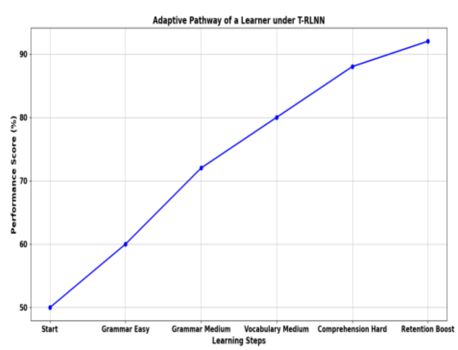

Fig. 6. Adaptive pathway of a learner under T-RLNN.

Fig. 6 represents a learning trajectory produced by the proposed T-RLNN. It describes the change in instructing strategies based on the output of the learners, whereby, simple grammar tasks are introduced and then advanced with the harder ones related to learning vocabulary and comprehension. The fact that there is an increasing trend in the performance scores indicates that the agent is capable of dynamically personalizing the learning trajectories. The framework utilizes RoBERTa embeddings to get contextual understanding and RL to get decision-making to adjust the difficulty level and content choice to get the most engagement and retention. This qualitative model of visualization of the pathways shows that the model has the ability to provide personalized education with the help of which the proficiency can be developed within ESL students.

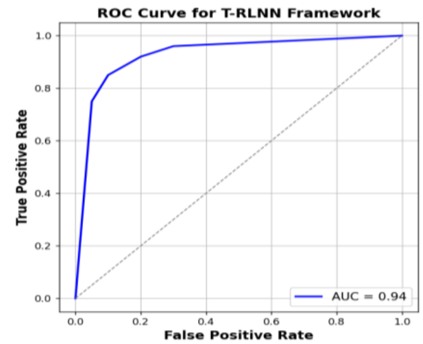

Fig. 7. ROC curve.

Fig. 7 shows the ROC curve of the proposed T-RLNN framework with an area under the curve (AUC) of 0.94 on the CALL dataset. The curve shows that the model is highly discriminative in categorizing the interactions among learners with high scores of true positive rates being preserved with different levels of false positive rates. The framework has a predictive performance score of 0.94 which suggests that it is very effective in matching the responses of learners with the appropriate teaching-learning methods. Using RoBERTa embeddings with RL, T-RLNN exhibits improved adaptivity and accuracy, therefore, being very useful in real-time personalized ESL learning settings, where prediction of outcomes is essential.

A. Performance Evaluation for Proposed Framework

The research outcomes indicate that the proposed T-RLNN framework is able to attain a significantly high overall accuracy of 92.5% which evidently exceeds all the traditional baseline models. Such a good performance is attributed to the fact that the model is able to constantly learn through the interaction of learners and change its teaching methods depending on the individual needs. The fact that the recall score takes 90.6 % also means that the system is efficient in detecting meaningful learning patterns and providing timely and context-sensitive feedback. Notably, adaptivity under this model extends beyond scaffolding content difficulty but also provides consistent individualized performance and accuracy is at 91.8% across a range of learner types. On the other hand, traditional methods, including Support Vector Machines and Random Forests, have obvious weaknesses, especially in their failure to adapt to the new behavioral patterns or the learning styles of new learners. Although the baseline Deep Neural Network has a fairly high accuracy of 89.1, it does not offer the adaptability to decision-making and instructional dynamism that is offered by reinforcement learning. Consequently, this makes it less responsive to the changes in learner engagement or performance through the course of time. These results are highly indicative of the fact that reinforcement learning-based adaptive tutoring is a more efficient model to be used in ESL teaching. T-RLNN improves the understanding, maintains the attention of the learner and provides more applicable learning by constantly optimizing the learning paths according to the feedback and real-time behavior of the learner. In general, the findings support the hypothesis that the use of contextual language comprehension combined with reinforcement learning offers a significant benefit over non-adaptive and fixed baseline models of the contemporary ESL teaching process.

1) Precision: The ratio of the total number of correct predictions. The precision is derived using Eq. (14):

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \quad (14)$$

2) Recall: The number of accurately predicted positive instances over all the predicted positives. Recall measures are evaluated using Eq. (15):

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \quad (15)$$

3) Accuracy: The proportion of successfully predicted positive cases among all the actual positives. Accuracy is evaluated using Eq. (16):

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \quad (16)$$

4) F1-Score: The harmonic mean of recall and precision, reconciling both measures. F1-score is calculated with the Eq. (17):

$$\text{F1} - \text{score} = 2 \times \frac{\text{Precision x Recall}}{\text{Precision + Recall}} \quad (17)$$

where, the FN, TN, FP, and TP stand for false negative, true negative, and true positive, respectively.

Fig. 8 and Table II display the performance measures of the suggested T-RLNN framework on four major evaluation metrics. The model performs best in Accuracy at 92.5%, closely followed by Precision at 91.8%. Recall, which is a measure of the model's capacity to detect pertinent instances, is 90.6%, while the F1-Score, which is the harmonic mean of Precision and Recall, is 91.2%. These outcomes show an even and strong performance in all the metrics, indicating that the suggested RL model is efficient in ensuring consistent prediction quality with fewer false positives and fewer false negatives.

TABLE. II. PERFORMANCE EVALUATION

Metrics	Value
Accuracy	92.5
Precision	91.8
Recall	90.6
F1-Score	91.2

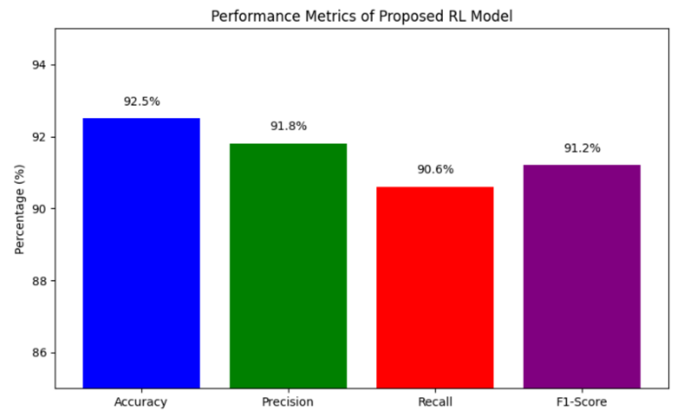

Fig. 8. Performance evaluation.

B. Ablation Study

An ablation study was conducted to gain insight into the role played by each of the elements in the T-RLNN architecture. The experiments tested the individual effect of each of the modules by varying or eliminating certain modules on model performance and adaptability.

Table III describes the ablation results of T-RLNN. The lack of RoBERTa embeddings or engagement features reduces the performance, which confirms their significance. Integrated components are effective, and the full model is the one that has the greatest accuracy, F1-score, engagement, and retention.

TABLE. III. ABLATION RESULTS

Model Variant	Accuracy (%)	F1-Score	Engagement Score	Retention Index
Without RoBERTa (RLDNN)	87.2	0.85	0.70	0.72
Without Engagement Features	89.5	0.87	0.74	0.76
Full T-RLNN (Proposed)	92.5	0.90	0.81	0.84

C. Comparative Analysis

Table IV provides a comparative study of four ML models: Proposed RL, SVM, RF, and DNN—with respect to three performance measures: accuracy, recall, and precision. The Proposed RL model shows the best performance with the highest accuracy of 92.5%, recall of 90.6%, and precision of 91.8% compared to all the methods considered. Even though the conventional models show increasingly better performance, they are still inferior in general. Among them, the DNN model demonstrates the best performance, closest to the RL-based approach, with the second highest in all three measures.

TABLE. IV. COMPARATIVE STUDY WITH EXISTING METHODS

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)	Engagement Score (ES)	Retention Index (RI)	Adaptivity Index (AI)
Proposed RL	92.5	91.8	90.6	91.2	76.5	73.9	67.3
SVM [25]	85.3	84.1	82.9	83.5	61.2	58.3	42.8
RF [26]	87.2	86.5	85.0	85.7	63.9	61.5	48.1
DNN [27]	89.1	88.2	88.2	88.2	68.4	66.0	52.6

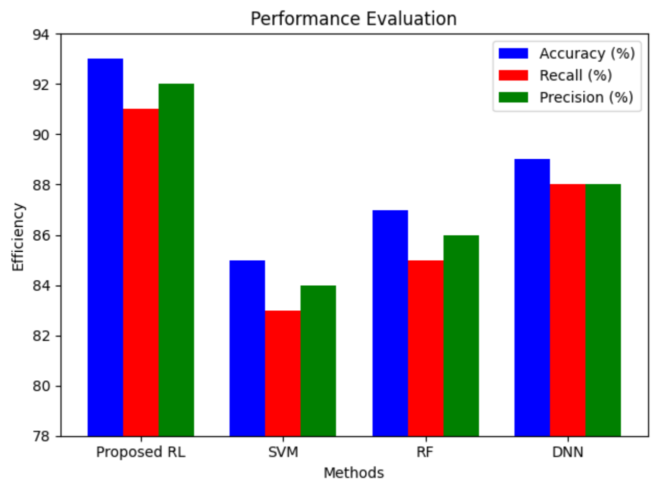

Fig. 9. Comparative study with existing methods.

Fig. 9 compares the results of four different techniques—i.e., a new RL method, SVM, RF, and DNN—on three most important metrics: accuracy, recall, and precision. Performance of each method is depicted by separate, color-coded bars, so it is very convenient to compare the results visually. Interestingly, the RL approach performs better than the rest with scores of approximately 92% for accuracy, 90% for recall, and 91% for precision, while the DNN, RF, and SVM approaches trail behind with increasingly lower scores in the high to low 80s. This clear illustration highlights the greater efficacy of the RL approach in managing these evaluation metrics.

D. Discussion

The suggested T-RLNN system will follow the format of an actor-critic reinforcement augmentation model where the real-time ability, reaction time, participation, and performance in assignments of the learners are directly linked to the teaching choices. The model combines contextual embeddings based on RoBERTa to identify not only the linguistic richness but also the cognitive and behavioral cues to reflect the entire state of each learner and represent it in a rich and holistic manner. This combination enables the system to provide context-sensitive personalization, where the actions of instruction are much more precisely tailored than the conventional ESL learning models. In contrast to inert methods, T-RLNN adapts its instructional plan as the learner behavior changes so that it prioritizes both the accuracy of learning in short-term and long-term interaction and retention. The architecture is a combination of RoBERTa embeddings, a projection layer to stabilize state representations, and an actor and critic network that drives instructional choices. The actor produces discrete instructional behavior, like making grammar, vocabulary or comprehension task choices, by the use of a softmax output, and modulates task difficulty by actions of continuous value. This design concurs with the adaptable, data-driven modulation of instruction that can be very close to the needs and progress of learners. The efficiency of the suggested framework is proven with the help of experiments. T-RLNN was always superior to classic models, such as Support Vector Machines, Random Forests, and Deep Neural Networks, in the most important evaluation metrics. The accuracy, F1-score and Adaptivity Index of the framework were 92.5, 91.2 and 67.3 respectively, which is almost 20 points higher than the DNN baseline. These results underscore the fact that the model can be used to increase the engagement of learners, knowledge retention, and competence of language in general through responsive learning pathways. To sum up, the present research demonstrates that T-RLNN can be used to implement a practical and scalable solution to real-time personalized ESL tutoring. The effective combination of the concept of reinforcement learning and contextual language representations introduces a strong and progressive direction in the development of technology-enhanced language learning systems.

VI. Conclusion and Future Work

The T-RNN framework was formulated as an adaptive ESL tutoring system that combined both contextual representations learning and reinforcement learning, whereby the instructional strategies are dynamically adjusted in real-time. The system uses behavioral measures (response time, accuracy, and engagement) and RoBERTa-based semantic embedding to create a complex and sophisticated learner state. The actor-critic agent then uses this state to identify the most suitable instructional actions to take via each of the learners, giving them highly personalized instructions. Experimental analyses showed that T-RLNN was the most effective model compared to traditional models, such as SVM, Random Forest, and feedforward DNNs, with better accuracy, F1-scores, and adaptivity. These findings imply that the framework has been able to balance both short-term effects of improvements in performance and long-term retention, which has resulted in making ESL instruction more scalable, learner-centered, and effective.

The research has some limitations, though. The experiments were also done in one dataset, which could limit generalizability, and the application of RoBERTa demands a large computational power, which can be difficult to use in low-resource or real-time environments. The research in the future will investigate cross-dataset validation to enhance generalizability and lightweight or distilled transformer models to decrease the cost of computation. Also, the use of multimodal cues to the learner, e.g., speech, eye-tracking, or interaction patterns, may facilitate more comprehensive modeling of the learner states. The longitudinal assessments will assist in measuring the learning outcomes over a period of time, and such techniques as federated learning and meta-learning may assist in large-scale personalization without invading the privacy of learners. Altogether, the results indicate that reinforcement learning integrated into a contextual embedding offers a very versatile, effective, and customized method of ESL instruction, which represents the future of intelligent language tutoring systems.

References

[1] F. Su and D. Zou, "Technology-enhanced collaborative language learning: theoretical foundations, technologies, and implications," Computer Assisted Language Learning, vol. 35, no. 8, pp. 1754–1788, 2022.

[2] A. D. Hanafiah and A. A. Aziz, "Opportunities and challenges in ESL online learning environment: a review of literature," Sciences, vol. 12, no. 1, pp. 1721–1730, 2022.

[3] J. Ma, "A study on the group interactive effect of different grouping models in foreign language flipped learning: a contrastive study between heterogeneous-multiple-intelligences groups and free groups," in EDULEARN22 Proceedings, IATED, 2022, pp. 1277–1282.

[4] N. Matban, "Exploring Parental Attitudes Towards Technology-Assisted English Language Learning (TALL) For Young Children in Turkiye," PhD Thesis, University of Warsaw, 2023.

[5] N. S. Yaccob, M. M. Yunus, and D. S. John, "Global education movement: English as a second language teachers' perceptions of integrating volatility, uncertainty, complexity, and ambiguity elements in lessons," Frontiers in Psychology, vol. 14, p. 1007970, 2023.

[6] E. Kim, "Teacher Views on Barriers and Challenges to Information and Communication Technologies (ICTs) and Multimedia Integration in the ESL Classroom: A Case Study," PhD Thesis, Northcentral University, 2022.

[7] M. N. Y. Wen and A. A. Aziz, "The use of Quizizz as an online teaching and learning assessment tool in an ESL classroom: A systematic," International Journal of Academic Research in Progressive Education and Development, vol. 11, no. 1, pp. 1076–1094, 2022.

[8] J. Richardson and M. Michaud, "Designing an ESL Assessment Test for Intermediate Students," Language Education and Technology, vol. 2, no. 1, 2022.

[9] N. S. Sharifuddin and H. Hashim, "Benefits and Challenges in Implementing Artificial Intelligence in Education (AIED) in ESL Classroom: A Systematic Review (2019-2022)," International Journal of Academic Research in Business and Social Sciences, vol. 14, no. 1, pp. 146–164, 2024.

[10] S. N. Kirana and Y. M. Gupta, "Utilising natural language processing to assist ESL learners in understanding parts of Speech," International Journal of Innovative Technology and Exploring Engineering, vol. 11, no. 5, pp. 12–15, 2022.

[11] D. Buragohain, G. Punpeng, S. Jaratjarungkiat, and S. Chaudhary, "Impact of e-learning activities on English as a second language proficiency among engineering cohorts of Malaysian higher education: A 7-month longitudinal study," in Informatics, MDPI, 2023, p. 31.

[12] Y. Zheng, Y. Que, X. Hu, and J. H. Hsiao, "Predicting reading performance based on eye movement analysis with hidden Markov models," in 2022 International Conference on Advanced Learning Technologies (ICALT), IEEE, 2022, pp. 172–176.

[13] G. Zhu, A. Zhang, G. Zhu, and A. Zhang, "Understanding ESL preservice teachers' metaphorical epistemology in the teaching practicum context: Insights from social realist theory and practice architectures," in Paper presentation at the American Educational Research Association Annual Meeting. San Diego, California, 2022.

[14] P. Kaur, H. Kumar, and S. Kaushal, "Technology-assisted language learning adaptive systems: a comprehensive review," International Journal of Cognitive Computing in Engineering, 2023.

[15] M. A. S. Khasawneh, "Adapting multisensory techniques for dyslexic learners in English language learning: A case study approach," World Journal of English Language, vol. 14, no. 5, 2024.

[16] V. Chanderan and H. Hashim, "Language learning strategies used by ESL undergraduate students," Creative Education, vol. 13, no. 3, pp. 768–779, 2022.

[17] Y.-J. Chung, C.-W. Hsu, M.-H. Chan, and F.-Y. Cherng, "Enhancing ESL Learners' Experience and Performance through Gradual Adjustment of Video Speed during Extensive Viewing," in Proceedings of the CHI Conference on Human Factors in Computing Systems, 2024, pp. 1–14.

[18] J. C. Young and M. Shishido, "Evaluation of the Potential Usage of ChatGPT for Providing Easier Reading Materials for ESL Students," in EdMedia+ Innovate Learning, Association for the Advancement of Computing in Education (AACE), 2023, pp. 155–162.

[19] M. Monika and C. Suganthan, "A study on analyzing the role of Chatgpt in English acquisition among ESL learners during English language classroom," language, vol. 8, no. 2, 2024.

[20] E. C. Naparan and J. Bacasmot, "English language problems and communication competence of students: The mediating role of smartphone apps in ESL learning," Asian Journal of Education and Social Studies, vol. 50, no. 2, pp. 10–30, 2024.

[21] A. Ahmad, E. A. Wani, O. A. Buyaw, T. F. A. Dunstan, and H. Hashim, "Language Learning Strategies Used by ESL Learners in Primary Schools in Sarawak to Enhance Reading Skill," 2023.

[22] R. Spain et al., "A reinforcement learning approach to adaptive remediation in online training," The Journal of Defense Modeling and Simulation, vol. 19, no. 2, pp. 173–193, 2022.

[23] S. R. Tolibovna, "USE OF TECHNOLOGY IN MODERN CLASSROOM TEACHING," Journal of new century innovations, vol. 14, no. 1, pp. 192–195, 2022.

[24] "Computer Assisted Language." Accessed: Jan. 20, 2025. [Online]. Available: https://www.kaggle.com/datasets/saurabhshahane/computer-assisted-language

[25] B. Fahad Mon, A. Wasfi, M. Hayajneh, A. Slim, and N. Abu Ali, "Reinforcement learning in education: A literature review," in Informatics, MDPI, 2023, p. 74.

[26] J. Vassoyan, J.-J. Vie, and P. Lemberger, "Towards scalable adaptive learning with graph neural networks and reinforcement learning," arXiv preprint arXiv:2305.06398, 2023.

[27] I. Gligorea, M. Cioca, R. Oancea, A.-T. Gorski, H. Gorski, and P. Tudorache, "Adaptive learning using artificial intelligence in e-learning: A literature review," Education Sciences, vol. 13, no. 12, p. 1216, 2023.