A Resource-Aware Multi-Agent Reinforcement Learning Framework For Personalized English Teaching

Linzhi Shao*

Engineering & Technical College of Chengdu University of Technology, Leshan 614007, China

* *Corresponding author. E-mail: Shaodorothy@126.com*

Received: Aug. 19, 2025; Accepted: Oct. 27, 2025

English language teaching (ELT) systems often lack personalization and adaptive feedback. Traditional English teaching methods lack personalization, real-time feedback and engagement. Integrating Multi-Agent Reinforcement Learning (MARL) offers adaptive, interactive learning environments. Agents include Teacher, Student, Content, Evaluation and Interaction modules, collaboratively learning optimal teaching strategies. Lack of reward-based personalization limits adaptive lesson selection, reduces engagement and weakens real-time feedback, hindering effective English learning. Research aims to develop a resource-aware MultiAgent Proximal tuned deep edge Q-learning (R-MAP-DEQL) framework for a personalized English teaching system. Datasets include English vocabulary, grammar exercises, reading passages and audio samples of pronunciation. Data preprocessing involves tokenization and normalization of text to standardize input and remove noise. Mel-Frequency Cepstral Coefficients (MFCC) are extracted from audio samples to capture pronunciation and speech patterns. DEQL with proximal tuning enables agents to optimize policies efficiently, balancing exploration and exploitation while accounting for computational constraints and providing real-time personalized teaching interventions. The framework is implemented in Python using RL and deep learning libraries. Experiments demonstrate improved learner performance, engagement and personalized lesson adaptation. Visualizations show progressive improvement across vocabulary, grammar, reading and pronunciation metrics, confirming system effectiveness. Experimental results demonstrate that R-MAP-DEQL achieves an accuracy of 97%, a precision of 95%, a recall of 93%, and an F1-score of 96%. The proposed MARL-based English teaching system effectively personalizes learning, adapts dynamically to student performance and enhances engagement. Resource-aware multi-agent (MA) strategies ensure optimized teaching decisions. Results highlight potential for AI-driven education, scalable real-time deployment and improved language skill acquisition.

Keywords: English language teaching (ELT), Multi-Agent Reinforcement Learning (MARL), Deep edge Q-learning (DEQL), Reinforcement Learning (RL)

http://dx.doi.org/10.6180/jase.202607_30.033

1. Introduction

English has become the common language of communication, education and career growth. However, there are certain disadvantages of traditional ways of teaching English. These conventional techniques tend to include grammar-translation technique, which emphasizes on memorizing words and grammar rules, the direct technique, which places students in the target language with no explicit grammar instruction and the audio-lingual technique, which involves repetitive exercises and patterns. The accessibility has been improved with technology-based platforms such as language learning applications, web portals and intelligent tutoring systems (ITS), however, real-time flexibility, deep personalization and optimal feedback remains a challenge. Using AI and MARL, this study offers a new solution to dynamic and real-time instruction of the English language, which directly fills the main gaps in the current educational technologies and contributes to the development of machine learning in the educational process. ELT plays a very important role in this globalized age because the language has become the global language of the highest international usage in communication, education, business and technology [1]. Successful ELT focuses on developing learners' reading, writing, listening and speaking skills to make them capable of using the language accurately and fluently in authentic situations. Conventionally, courses tend to focus on teaching grammar rules, vocabulary and reading for comprehension, whereas contemporary approaches support interactive learning, critical thinking and real-life application [2, 3]. English education also varies according to the learners' different needs to achieve academic or career success, or to communicate globally. The introduction of digital media, multimedia support and online teaching has taken English teaching outside the classroom where it offers flexibility and interesting methods to the learners. Finally, English teaching is not only aimed at developing language competencies but also building confidence, cross-cultural understanding and interaction in a globalized society [4, 5]. The models of teaching English are traditional classroom instructions and technology-assisted systems. Conventional approaches focus on grammar translation and rote memorization, which provide solid backgrounds but are not personalized and interactive [6]. The standardized instruction is typically provided to students with different levels of skills and learning rates, which leads to the lack of enthusiasm and slower learning. In addition, pronunciation and speaking practice are still not developed in most models because of the lack of integration of advanced speech analysis tools. Computer-Assisted Language Learning (CALL) with multimedia tools and ITS as adaptive learning and Learning Management System (LMS) platforms such as Moodle and Google Classroom facilitate structured learning but lack personalization. Modern mobile apps enhance accessibility and pronunciation practice, yet still lack real-time adaptability and individualized strategies for diverse learners [7].

Various methods have been devised to improve ELT with AI and adaptive learning. Rulebased ITS generally use single-agent RL techniques to adapt exercises to student performance. Yet, they are limited by static adaptation and low scalability [8]. Single-agent RL models maximize the learning of individual students but do not encourage group interaction or multi-student interaction. Collaborative learning platforms, including gamified and interactive tools, encourage group learning and engagement but generally lack dynamic personalization and agent-based adaptability. Nonetheless, existing AI and MARL-based solutions do not offer effective personalization because of the lack of reward-based adaptation and cannot adapt to real-time because they cannot dynamically modify lessons in response to current learner performance. In addition, most systems are characterized by high computational costs, which restrict their scalability and resourceconstrained performance. Although these models are more engaging and offer a form of adaptive feedback, they fail to capitalize on MA interactions and collaborative intelligence, which are critical to the development of a more responsive, personalized and effective English learning environment [9]. These gaps highlight the pressing need for a more adaptive, resource-aware and smarter ELT system that can offer customized interventions, promote learner engagement and facilitate efficient skill mastery in vocabulary, grammar, reading and pronunciation. Gudivaka introduced an AI-based, large-scale structure to be used in music education, which uses student performance and interaction data to dynamically adjust learning pathways. The proposed Englishteaching framework based on this method is modified through MARL in which agents rely on learner analytics to customize lessons, modify difficulty and resource allocation. Its advantages are more personalized instruction, adaptive learning that can be scaled and more effective utilization of teaching resources [10]. The proposed R-MAP-DEQL approach can solve these issues by incorporating multiple agents and providing real-time flexibility, low computational expenses, and profound personalization. The system is able to balance exploration and exploitation and has low computational costs by using proximal-tuned DEQL.

Research introduces the proposed R-MAP-DEQL method, which utilizes MA Student, Teacher, Evaluation, Content and Interaction to adaptively personalize lessons, monitor learner progress in actual time and support collaborative knowledge. The agents are configured to achieve a particular reward goal: Student rewards are aimed at learning outcomes, Teacher rewards are aimed at instructional effectiveness, Content rewards are aimed at resource congruence, Evaluation rewards are aimed at feedback accuracy and Interaction rewards are aimed at keeping all modules in sync. Collaborative intelligence is achieved algorithmically by updating shared state, agents share policy information and coordinate actions to maximize the overall learning performance. The aim of this research is to progress a scalable, resource-efficient and adaptive English tutoring system that exploits learner engagement, accommodates diverse learning styles and vocabulary, grammar, reading and pronunciation performance are improved.

- A MARL-based ELT system was developed to ensure personalization, real-time feedback, scalability and enhanced learner engagement.
- The dataset is collected from Kaggle sources, comprising vocabulary, grammar exercises, reading passages and pronunciation audio samples from diverse learners.
- Preprocessing techniques such as tokenization, normalization and noise removal are applied in this research to standardize inputs and maintain structured, clean data.
- The MFCCs are used during the extraction stage to locate pronunciation patterns, while textual features help the analysis of vocabulary and grammar.
- The proposed R-MAP-DEQL framework is designed, integrating MA with proximal-tuned DEQL for adaptive and efficient teaching.

◇ Experimental results demonstrated significant improvements in vocabulary, grammar, reading and pronunciation performance, alongside increased engagement and personalized learning adaptation.

An introduction is provided in Section 1. Section 2 reviews related work on MARL in education systems. Section 3 describes the methodology, including dataset preparation, preprocessing, feature extraction and the proposed R-MAP-DEQL framework. Experimental results and discussions on learner performance are presented in Section 4. Finally, Section 5 concludes the research, highlighting the effectiveness of the MARL-based personalized English teaching system.

2. Methods and materials

In the research, existing studies on personalized English learning, adaptive feedback systems, MARL, Artificial intelligence (AI)-driven teaching strategies and DL and ML-based performance prediction were critically reviewed, as discussed in this section.

AI was explored [11] as a means to optimize learning and teaching by integrating insights from learning analytics. Secondary research data and a case study provided the methodology, while comparative analysis guided the design of an AI-based multi-agent system. Results showed effective analysis of learning processes, though limitations included reliance on secondary data and prototype-level validation. An intelligent integration system of AI English teaching resources was developed [12] using MA collaboration. The method combined Ant Lion Optimizer (ALO) and Moth Flame Optimization (MFO) algorithms within the Device-to-Device Communication (D2D) power allocation framework, analyzing population size, dimensions and convergence through simulations. Findings showed better integration of resources and energy efficiency.

Nevertheless, there were such limitations as the use of simulation settings and limited validation in actual educational settings.

A big data and cloud computing-based online English independent learning platform based on an MA model was evaluated [13]. The aim was to improve individualization, teamwork and resource combination. The approaches included the analysis of the features of online learning and the development of MA modules. Findings indicated better learner autonomy and collaboration. The drawbacks were the dependence on the internet infrastructure and the inability to ensure consistent engagement. The purpose of English upgrading was to overcome the language barrier and improve communication on the global level [14]. An agent-based, distance system was adopted with the help of AI to establish adaptive and flexible, learner-centered environments. Findings indicated better personalization and learner flexibility in prototype testing. The shortcomings were the absence of complete teacher-like interaction and limited scope of evaluation, which still needs to be refined to be applicable to a wider range of academic English.

An MA physical education (PE) instructional management system was tested [15] to enhance the quality of teaching, individualization and student involvement. A MA framework was used, which allowed collaboration, interaction and real-time feedback among agents. The findings indicated greater operational efficiency than the traditional systems (75% max, 67% min). Nevertheless, it had such constraints as reliance on technological infrastructure and complexity of implementation.

User-centred adaptive Human-Computer Interaction (HCI) in e-learning was enhanced [16] by integrating social computing and usability metrics into intelligent interaction agents. These agents use social computing data, including peer interaction and collaboration patterns, along with usability measures, including task completion time and error rates, to personalize the learning interface and give personalized feedback. MARL framework addresses the shortcomings of the current systems by introducing multi-agent cooperation, where each agent can adjust not only to the needs of individual students but also to optimize the content of the lesson, its difficulty and feedback in real-time, which results in more effective and scalable learning experiences. Preference and performance metrics were applied to support interface adaptation. Results showed improved adaptability and stability of the MA model. However, limitations included reliance on subjective user evaluations and the need for broader experimental validation.

An AI-based English teaching resource integration aimed to enhance accuracy and reduce processing time. A MA collaborative algorithm was employed [17], where a user interest model was built using implicit feedback and ontology weighting. Results showed 100% integration accuracy with processing time reduced to 3 seconds. However, limitations included reliance on simulation results without extensive real-world validation.

English proficiency by integrating a multiverse technique was enhanced [18] by Personalized Robot (ProBot). It employed fuzzy logic-based linear models and ontology-driven personalization to deliver adaptive lessons and feedback. Results showed improved student proficiency across skills and engagement. Limitations included reliance on simulations and challenges in real-world scalability and 6G infrastructure readiness. Single-agent RL models maximize the learning of individual students by modifying lesson plans in response to ongoing feedback provided by the performance of the student. The system monitors the activities of the learner, reinforces the correct answers or progress and modifies the subsequent lessons to suit the changing needs of the learner. This individualized strategy assists in making sure that students concentrate on the areas that require the most attention and therefore result in more effective and efficient learning.

Challenges in personalized education were addressed [19] by developing the AI in Education at Hochschule für Technik (HFT) Stuttgart framework. It applied multimodal data integration, deep RL and privacy-preserving feedback methods to analyze and adapt teaching strategies. Results demonstrated improved personalization and learning outcomes in a case study. It was limited by the reliance on high-level digital infrastructure, scalability issues, and ethical issues on data security and privacy. A smart ELT system that covers the gap of personalization was presented [20]. It used interactive mobile technology with Deep Q-Network (DQN) to adjust the strategy dynamically and neural collaborative filtering to adapt the interest in the short term. The outcomes indicated enhanced efficiency, streamlined learning journeys and pertinent real-time suggestions. Limitations were difficulties in fine-tuning of the model, scalability to various learners and the need to have continuous data. The effect of AI conversational agents and their influence on selfregulation and knowledge retention of learners in educational institutions was approximated [21]. It was a pretest-posttest study involving 203 students and mixed-method analysis. The findings indicated that there was a significant SRL improvement (133%) and retention, and adaptive feedback was better than non-adaptive groups. Limitations were that they were dependent on controlled experimental settings and difficulties in real world generalization.

2.1. Research Gap

Current AI and multi-agent systems-based ELT models enhance personalization, resource integration and feedback but are constrained by dependence on simulation, reliance on infrastructure, scalability concerns and adaptability to the real world. The suggested R-MAPDEQL framework will solve these disadvantages by providing resource-aware optimization, adaptive feedback in real-time, deep personalization and scalable deployment, which will guarantee effective and engaging English learning results.

3. Results and discussion

The goal is to create a MARL-based ELT system that would be personalized, provide realtime feedback, be scalable and learner-engaging. This methodology section describes the research methodology, data collection of text and audio, preprocessing of text and audio in terms of tokenization, noise removal and normalization, feature extraction and R-MAP-DEQL framework implementation. It describes the design of the MARL agent, rewarding schemes, policy optimization and assessment procedure of personalized ELT. The proposed R-MAP-DEQL framework workflow is depicted in figure 1.

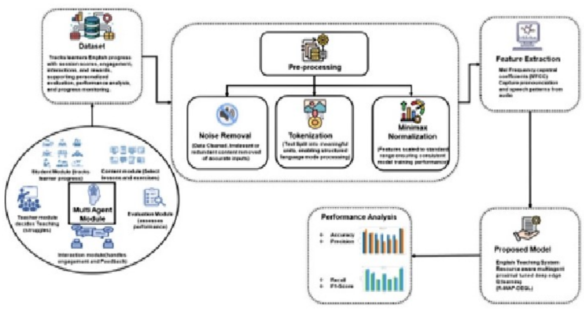

Fig. 1. Visualization of the proposed R-MAP-DEQL framework workflow.

3.1. Learning Resource Collection for the English Teaching System

The English learning progress was collected through the Kaggle platform (https://www.kaggle.com/datasets/zyan1999/english-learning-progress-dataset/data), that contains a dataset that records the learning progress of English learners in several sessions. It contains pre- and post-session vocabulary, grammar, reading and pronunciation, and engagement scores, interaction logs and reward indicators during each session. The rows are the sessions of each learner, which allows studying the learning patterns, tracking the progress, and the individual evaluation of the performance. Table 1 gives a detailed description of the characteristics of the dataset, which helps to analyze the learning process of students and the efficiency of different types of lessons. The data contained engagement metrics from learner interaction logs, such as time spent per session, frequency of activity, response latency and participation in feedback tasks. These metrics were normalized and placed into the reward function as additional modifiers to encourage consistent participation and sustained attention during learning sessions.

3.2. Preprocessing

Preprocessing involves the preparation of raw English text and audio data to be used in the MARL system. Noise removal (removing irrelevant data), tokenization (text splitting) and minmax normalization (scaling inputs) were employed in this study. The readable textual inputs are tokenized to guarantee readable textual inputs to learn features accurately, the scales of normalization balances are used to achieve faster convergence and MFCC extraction incorporates significant acoustic features to pronunciation analysis. All the methods were selected in order to maximize the accuracy of the R-MAP-DEQL model, minimize noise and improve adaptive performance in language learning.

3.2.1. Noise removal

It is the procedure of removing irrelevant, inconsistent, or redundant data in text or audio, enhancing clarity, precision and model performance, making personalized and adaptive ELT effective. This noise removal in research purifies text (vocabulary, grammar, reading) and audio (pronunciation) information, so that it is correctly inputted into R-MAP-DEQL, improving personalized lesson delivery, real-time feedback and performance.

3.2.2. Tokenization

The process of dividing continuous text into small units of words known as tokens is called tokenization. It translates vocabulary, grammar sentences and reading passages into meaningful tokens in an ELT system after eliminating stop words and punctuation marks. This allows proper feature extraction, enhances embedding quality, minimizes noise and improves R-MAP-DEQL performance on personalized learning and real-time feedback. The result of this process is structured word units that are cleaned and enhance feature extraction, minimize noise and allow personalized learning to be accurate.

3.2.3. Min-Max Normalization

This is achieved by dividing the English and audio language input characteristics, such as vocabulary count, grammar error rate, reading comprehension scores and pronunciation signals, into a common $[0, 1]$ range. This guarantees equal contribution of features, less bias due to different data scales and enhanced R-MAP-DEQL training stability. The system removes the differences between heterogeneous inputs, which leads to more efficient learning and allows the system to adapt the lesson precisely and generate feedback in real-time. The mathematical model is given in equation (1).

$$D_{\text{norm}} = \frac{D - D_{\text{min}}}{D_{\text{max}} - D} \tag{1}$$

Where (D) is the unique data value, (D_{min}) is the minimum value, (D_{max}) is the maximum value, (D_{norm}) is the normalized value (between 0 and 1). This normalization creates balanced, noise-free inputs, enabling faster model convergence, higher personalization accuracy and stable adaptive ELT across modules.

3.2.4. Bias Mitigation and Dataset Considerations

While normalization mitigates bias from the various scales of data, further approaches have been applied to mitigate the potential bias from the dataset. This dataset, that is collected from various learners, is pre-processed to ensure equitable representation of various skill levels and learning backgrounds. The generalization of the system and its effect on the demographic disparities or overrepresented groups will also be considered in the subsequent stages and will be addressed in further directions such as data augmentation and fairness-aware training as the way to enhance the generalization of the system and reduce its effect on the dynamics of the demographic disparity or overrepresented groups.

Table 1. Feature Importance for Predicting Student Learning Outcomes

Feature	Description	Uses
student_id	Unique identifier for each student	Tracking individual student progress and performance
lesson_id	Identifier for each lesson	Associating student performance with specific lessons
lesson_type	Type of lesson (e.g., vocabulary, grammar, reading, pronunciation)	Analyzing performance across different lesson types
completion_time	Time taken by the student to complete the lesson	Assessing time efficiency and identifying potential learning obstacles
accuracy	Percentage of correct responses in the lesson	Evaluating student understanding and mastery of lesson content
engagement_level	Measure of student engagement during the lesson (e.g., low, medium, high)	Correlating engagement with learning outcomes
feedback_received	Type of feedback provided to the student (e.g., positive, corrective)	Understanding the impact of feedback on learning progress
post_lesson_score	Score obtained by the student after completing the lesson	Gauging immediate learning outcomes and the effectiveness of the lesson
pre_lesson_score	Score obtained by the student before starting the lesson	Measuring knowledge gain and lesson effectiveness
lesson_difficulty	Difficulty level of the lesson (e.g., easy, medium, hard)	Analyzing performance trends across different difficulty levels

3.3. MFCC Based Speech Feature Extraction for English Learning

Extraction is the process of determining and choosing data characteristics that are to be analyzed or modeled. In the research, MFCC s are extracted from student pronunciation audio to capture speech patterns, pronunciation accuracy, tone and intonation. MFCCs are particularly helpful because they represent the short-term power spectrum of speech that mimics human hearing perception and provide the system with important parameters of pronunciation such as clarity, stress and pitch. The MARL-based system uses these features to assess the pronunciation of students and give them personalized feedback on learning English.

Step 1. Pre-Emphasis

The speech signal is then subjected to a high-pass filter to boost high frequencies, which increases energy in those areas that are significant to speech recognition as shown in equation (2).

$$k(n) = l(n) - b.l(n-1) \tag{2}$$

where $(l(n))$ is the input speech signal, $(k(n))$ is the output signal after pre-emphasis and (b) is the filter coefficient. This measure is necessary to make sure that high-frequency sounds, which are significant in the assessment of pronunciation, are highlighted.

Step 2: Frame Blocking

The speech signal is continuous, which is divided into small frames and each frame is analyzed in local properties. Permit the samples between the number of frames and Adjacent frames to overlap. This allows time-related dynamics of pronunciation.

Step 3: Windowing

Each frame is multiplied by a window function to smooth discontinuities at the edges. Hamming window is a popular window that is described in equation (3).

$$K(n) = L(n) \cdot Z(n) \tag{3}$$

where $L(n)$ is the input frame, $Z(n)$ is the Hamming window function and $K(n)$ is the windowed signal. This guarantees a seamless transition and maintains signal statistics, which is essential in correct feature extraction.

Step 4: Fast Fourier Transform (FFT)

The FFT is used to convert the windowed signal of the time domain to the frequency domain to convert $K(n)$ into $K(f)$, which captures the frequency spectrum of pronunciation and maintains significant speech characteristics at the same time compressing data. This step allows the MARL system to analyze spectral characteristics of pronunciation scoring.

Step 5: Mel-Scale Triangular Filter Bank

Triangular band-pass filters are used to smooth the magnitude spectrum of the Mel scale frequency spectrum, to mimic the human auditory experience and generate (N) mel-scale coefficients.

Step 6: Discrete Cosine Transform (DCT)

Equation (4) gives the logarithm of the filter bank energies and is transformed using DCT to generate MFCC coefficients.

$$D(n) = \sum_{h=1}^{N} C_h \cdot \cos\left[n \cdot (h - 0.5) \cdot \frac{\pi}{N}\right] \quad (4)$$

Where $D(n)$ is the MFCC coefficient (n), energy output of the (h^{th}) filter and (N) is the total number of triangular filters. These MFCCs represent the time-domain features of speech and are used by the Evaluation Module in MARL to assess pronunciation and provide personalized feedback. The MFCC-based approach has high accuracy in capturing pronunciation nuances because the human auditory response is modeled on the MFCC. In noisy environments, MFCC is better at reliable and less error than any other method, such as Perceptual Linear Prediction (PLP) or spectral feature extraction. It is robust and precision-levelled for consistent pronunciation assessment in the R-MAP-DEQL. While MFCCs evaluate pronunciation, textual features like word choice, grammar and sentence structure assess vocabulary and grammar proficiency. This textual information is extracted from written words to help review words for accuracy and comprehension and helps develop a cohesive language skills learning system.

MFCC Feature Extraction Steps:

Pre-emphasis emphasizes high-frequency speech sounds. Frame blocking analyzes short audio segments. Windowing smooths frame edges. FFT converts signals to the frequency domain. Mel filter bank extracts human-audible spectral features. DCT generates MFCC coefficients for pronunciation evaluation.

3.4. R-Map-DEQL Real-Time Adaptive Multi-Agent English System

The proposed R-MAP-DEQL framework is a hybrid MA system combining MARL with Proximal-tuned DEQL. It enables real-time, personalized ELT, where Student, Teacher, Content and Evaluation modules collaboratively optimize lesson selection, difficulty and feedback. R-MAP-DEQL framework is comprised of five agents that collaborate with each other. Student module monitors the performance and progress of learners, Teacher module creates appropriate exercises and feedback and Content module chooses or changes learning materials. The Evaluation module determines the accuracy, pronunciation and comprehension whereas the Interaction module manages the communication between all modules. These modules combine to create a continuous learning cycle in which student performance data are used to make real-time adjustments to lesson difficulty, pacing and feedback to achieve adaptive and personalized learning. Proximal tuning is important to provide efficient adaptation by balancing exploration and exploitation, which is critical in offering timely and relevant feedback to individual learners. The system is dynamic in that it modifies the level of exploration and exploitation according to the progress of the learner so that the new learning strategies are explored and the already discovered successful strategies are reinforced. The method prevents the disadvantages of over exploration, which may slow down learning and over-exploration, which may result in stagnation. Resource awareness also maximizes computational resources, which allows low-latency responses that are important to keep students engaged. Also, the edge-device optimization enables the system to operate efficiently on resource-constrained devices, which guarantees scalability and accessibility on various platforms. MFCC-based pronunciation analysis guides rewards, ensuring adaptive, scalable, low-latency learning for vocabulary, grammar, reading and pronunciation improvement.

3.4.1. Multi-Agent Reinforcement Learning (MARL)

MARL extends RL by allowing multiple agents to learn simultaneously in a shared environment. The agents are independent of each other, but they interact with each other in order to maximize their rewards. Basically, MARL can be thought of as a group of intelligent learners, learning through their own actions and the actions of the other learners.

- Agents: Student Module, Teacher Module, Content Module, Evaluation Module, Interaction Module
- Environment: Virtual classroom or learning platform
- The student interacts with lessons (actions).
- Modules communicate and exchange information.
- The Evaluation Module provides personalized feedback (reward signals).
- Agents update their policies to optimize future lesson selection and feedback.
- The system continuously adapts to improve engagement and learning outcomes.

Personalization in the MARL-based ELT system is used to make sure that lessons are tailored to the needs of each student. Adaptivity allows real-time adjustments based on performance. Collaboration enables modules Teacher, Student, Content, Evaluation and Interaction to work together for optimal teaching. Scalability supports multiple students and lessons simultaneously, enhancing engagement, learning efficiency and overall outcomes.

While a single, centralized RL agent may theoretically handle all the features of personalized ELT, MARL has specific benefits, as it allocates tasks among specialized agents. As an example, the Student Module is only concerned with the progress of individual learners, the Teacher Module is concerned only with the adjustment of teaching strategies and the Content Module is concerned with learning materials. This distribution enables parallel optimization and real-time responsiveness without straining a centralized system. The modules are able to scale and enhance efficiency because each can be adapted independently depending on its functionality.

A single system would not be able to handle the complexity of modifying the difficulty, pace and feedback of lessons among many learners at the same time. The distributed agents of MARL are able to make independent decisions based on learner data and each student is provided with unique content and feedback depending on their progress. This results in better management of multifaceted learning tasks, which are complex and multi-faceted (e.g., personalized feedback, lesson selection, adaptive pacing) and which a single agent would struggle to control.

In addition to engagement metrics, accuracy rates and feedback responses to each learner, the R-MAP-DEQL methodology continuously checks one learner's progress. These indicators give a dynamic perspective on how the system adjusts to difficulty and pace. Both faster learners and failure learners are guided towards higher levels of achievement and struggling students are easily aided by simple exercises or repetition with improved feedback. This adaptive loop will guarantee that every learner, irrespective of their level of proficiency or slowness, is given real-time content according to the development path.

3.4.2. Reinforcement Learning in R-MAP-DEQL

RL is a ML method in which agents are taught to take the best actions through interaction with an environment and rewarding them. The interaction between each agent (Student, Teacher, Content, Evaluation) and the learning environment in the research is in choosing lessons, exercises, or feedback strategies.

- Q-Value: The expected cumulative reward represents the outcome of taking a specific action in a given state. In this research, student performance metrics, including vocabulary accuracy, grammar, reading and pronunciation, serve as reward indicators, while actions correspond to lesson selection or feedback delivery.

- Q-Value Update (DEQL), this function is derived in equation (5).

$$Q\left(r_t, q_t\right) = Q\left(r_t, q_t\right) + \alpha\left[s_{t+1} + \gamma \max_a Q\left(s_{t+1}, q\right) - Q\left(r_t, q_t\right)\right] \tag{5}$$

Where (r_t) is the current state (student performance and engagement), (q_t) is the action taken (lesson or feedback selected), (s_{t+1}) is the immediate reward based on learning outcome, (α) is the learning rate and (γ) is the discount factor.

Reward Design: Rewards are calculated to reflect pronunciation, grammar, reading and vocabulary improvements. Its mathematical representation is provided in equation (6).

$$S_t = z_1 \cdot H_t + z_2 \cdot I_t + z_3 \cdot J_t + z_4 \cdot K_t \tag{6}$$

Where (H_t) is the pronunciation score, (I_t) is the grammar accuracy, (J_t) represents vocabulary correctness, (K_t) is the reading comprehension score and (z_1, z_2, z_3, z_4) are the weight coefficients for each skill. These weight coefficients, (z_1, z_2, z_3, z_4) were empirically adjusted at multiple simulation tests to balance skills importance and emphasized the importance of pronunciation and grammar with the help of experts. This expert-empirical hybrid calibration gave a consistent convergence of learning policy convergence and realistic priority over learning skills in general. This reward informs adaptive lesson choice and individual feedback, enabling agents to dynamically enhance student learning outcomes.

3.4.3. Deep Edge Q-Learning (DEQL) in Personalized English Teaching

Deep Q-Networks (DQN) are a combination of Q-Learning and deep neural networks that operate on large state-action spaces. DEQL builds upon DQN, allowing resource-aware lowlatency execution on edge devices, enabling real-time, personalized learning. In contrast to the traditional Q-learning, which uses Q-value tables, DEQL uses deep neural networks to estimate Q-values, which enables it to work effectively in complex environments. In contrast to DQN that are a combination of Q-learning and deep neural networks to handle large state-action spaces, DEQL is much better than DQN, mainly in the form of proximal tuning and edge-device optimization. DEQL compares to DQN, which uses Q-value tables, but estimates Q-values with deep neural networks with a proximal update constraint to mitigate the unsteadiness of policy optimization. Learning is stabilized by the proximal tuning process to ensure that the changes in the Q -value are not too large and the updates are gradual and controlled and not abrupt. This exploration and exploitation balance makes the model converge more quickly and more steadily.

The proximal update can be mathematically represented in equation (7)

$$\left|Q\left(s_t, a_t\right)-Q'\left(s_t, a_t\right)\right| \leq \epsilon \quad (7)$$

Where $Q\left(s_t, a_t\right)$ is the original Q -value, $Q'\left(s_t, a_t\right)$ is the new Q -value and ϵ is a small constant that regulates the size of updates. This process guarantees that the policy changes are gradual, which results in more stable and quicker convergence.

Moreover, DEQL is not just an optimized DQN on edge devices; it incorporates resourceawareness by supporting local computation on edge devices, minimizing latency and computation load. Unlike the slow convergence of DQN and the high computational cost of PPO, the lowlatency nature of DEQL and the low resource cost of learning, it is well-suited to real-time, adaptive educational tasks with limited computational resources.

In the research, DEQL enables MA personalization, where Student, Teacher, Content and Evaluation modules collaboratively adapt lessons in real time. The system performs actions such as selecting lessons, adjusting difficulty and providing pronunciation feedback, while the environment captures learner performance, engagement and knowledge progression. The system describes positive feedback, which reinforces correct responses and encourages continued engagement and corrective feedback, which directs learners to solving pronouncing or grammatical errors. Thus, the whole feedback mechanism renders the learning process more efficient. While it reinforces what is being mastered, it directs focus onto error work, leading to better-balanced skill development for all proficiency levels. Computed rewards rely on vocabulary, grammar, reading and pronunciation results and inform agents to streamline the teaching plan.

The system ensures ongoing adaptation by training DEQL models on a variety of learner data by balancing exploration of new exercises with exploitation of established learner strengths. RThe accuracy of MAP-DEQL is measured by timing correction of policy changes at a fixed threshold to prevent unstable learning and error correction. The process minimizes the sudden changes in Q-values across iterations to ensure that the convergence is smooth and the coordination of the agents is regular. This is a controlled update scheme that enhances optimization and learning stability in multi-agent interactions. The framework offers multi-agent, real-time adaptive ELT environment with dynamic lesson delivery, interactive feedback and effective collaborative learning. The edge deployment ensures low latency, scalability and resource-efficient operation, which is appropriate in large classrooms and real-time educational deployments. R-MAP-DEQL model is scalable due to its decentralized agent coordination and edge processing. The Student-Teacher-Evaluation modules loop are performed by each learner independently; an interaction agent is synchronized to inform several learners on a central Interaction Agent. Large classroom settings can also be supported in real-time with adaptive flexibility simultaneously with this modular, distributed design without using computational resources. It applies edge-based computing to optimize the system to consume low CPU and memory, i.e. local processing on devices like mobile phones or IoT platforms. This will reduce the load on the central servers and reduce the data transfer requirements and enables effective real-time operation even in resourceconstrained environments. The R-MAP-DEQL framework is resilient even in heterogeneous computational systems because it is capable of dynamically adjusting the processing loads to the capacity of the device. Resource allocation and edge-level optimization are ensured for consistent learning performance regardless of low consumption or heterogeneous hardware settings. The exploration-exploitation balance is also dynamically tuned, which also contributes to the optimization of resource use and high responsiveness. Although the R-MAP-DEQL framework is designed for edge computing, its real-world deployment may require network latency, weak bandwidth and hardware variability. To address these, lightweight model compression, asynchronous updates and adaptive load balance can be applied to maintain real-time responsiveness and reliable performance on a variety of educational devices.

DEQL was selected over DQN and PPO, which was faster to perform in resource-limited realtime contexts. In contrast to DQN's slow convergence and high computational cost of PPO, DEQL provides proximal tuning for low latency. Compared with standard DQN, the proximal-tuned DEQL improves over the standard DQN by including a constraint-based proximal update, which minimizes instability in the optimization of Q-values. DEQL is faster and less computationally intensive than PPO, A2C and DDPG and it is paralleled at edge level with lower computational burden. This means it is particularly useful for a realistic time and resource limited ELT context where stability and adaptive learning are required. It balances exploration and exploitation and allows versatile, adaptive multi-agent English instruction.

4. Results and discussion

In this section, the evaluation of key performance metrics is presented to compare student outcomes before and after applying the R-MAP-DEQL-based personalized ELT system.

The experimental setup for the proposed R-MAP-DEQL framework is designed to enhance personalized learning and ELT effectiveness in adaptive learning environments. It utilizes an Intel® Core $^{\text{TM}}$i7 $-$ 12700 K processor (3.6GHz), 32 GB DDR4 RAM and Windows 11 Pro. Python 3.9.13, along with RL and deep learning libraries, is employed to implement the R-MAP-DEQL framework for personalized English education.

Andrews's curves plot visualizes high-dimensional student performance data by mapping each data point to a single curve (Figure 2). The evident grouping of curves by the level of proficiency proves that the metrics of the system can effectively distinguish between different groups of students in terms of their skills, which proves that the system can effectively model the progress of learners. The labels A1, A2, B1, B2, C1 and C2 represent distinct student proficiency levels.

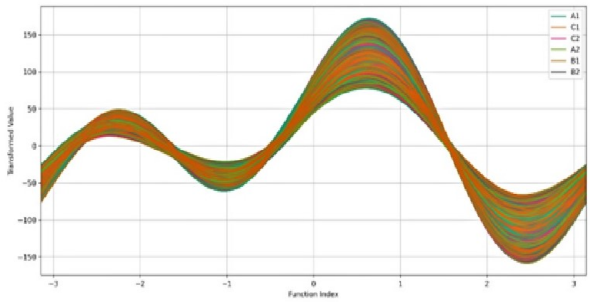

Fig. 2. Andrews Curves: Visualizing Student Proficiency Levels.

Figure 3 (a) uses contour plot to show the concentration of the learner performance. The fact that the high-scoring data points are concentrated in the warm-colored areas proves that the system is effective in general in bringing most of the learners to the higher levels of grammar and vocabulary proficiency. Figure 3 (b) gives a clear individual level perspective of learning gains in the lollipop plot. The lines relate the pre-system score (red) of every learner to their better postsystem score (green), which directly supports the visual proof of the effectiveness of the framework. Figure 3 (c) shows the scatter plot of the relationship between post-system grammarand vocabulary scores of a learner. The positive correlation between the two metrics is indicated by the upward trend of the data points, which shows that the system affects both skills.

A graphical representation of a 3D scatter plot, which illustrates the reward mechanism of the system, is given in Figure 4. It demonstrates a positive correlation with a significant positive value between the increased post-test scores in grammar and vocabulary and the increased "Target Reward," which proves that the reinforcement signal

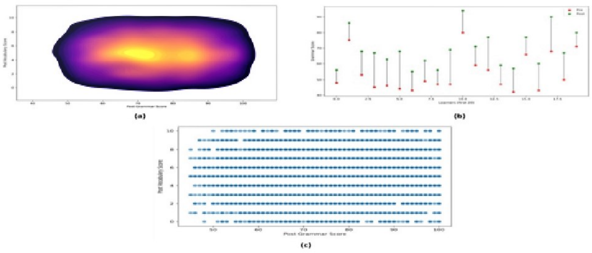

Fig. 3. (a) Grammar vs. Vocabulary Score Distribution and (b) Learner Performance Gain and (c) Post-System Grammar and Vocabulary Scores

of the framework is properly adjusted to its purpose of enhancing the performance of the learner. This relationship is further emphasized by the change of color between purple and yellow.

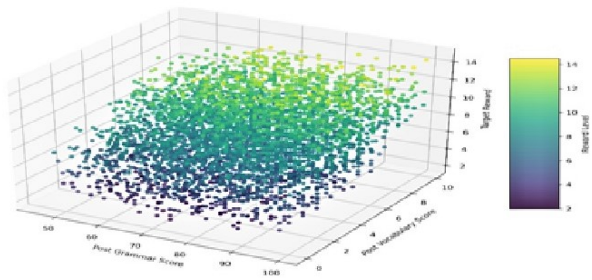

Fig. 4. R-MAP-DEQL: A Multi-Agent Framework for Personalized English Education

Figure 5 (a) represents the bubble chart of the system in terms of reward distribution according to post-test scores. Bigger, yellow-colored bubbles indicate bigger rewards, which are clumped in the top-right, which indicates that the framework works well to associate better performance in both reading and grammar with a bigger reward signal. Figure 5 (b) is a hexbin plot that represents the joint distribution of post-test scores in grammar and reading. The fact that the number of high-count hexagons is concentrated in the upper-right corner indicates a positive correlation, which implies that the adaptive strategies of the system improve the two areas of skills at the same time.

Figure 6 shows a dumbbell plot that gives a clear individual level perspective of learning gains. Every horizontal line links the initial score of a learner in pronunciation with the higher score of the learner after the system. This visualization directly helps to prove the study results that the framework is effective to improve pronunciation skills.

Figure 7 is a pair plot that displays the relationships between all the key performance measures: pre- and post-

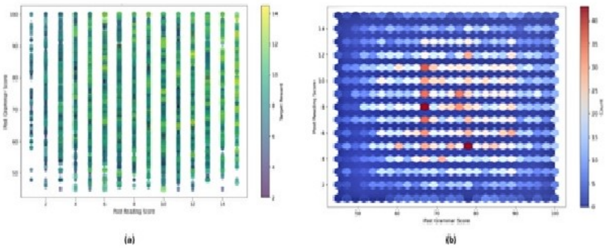

Fig. 5. (a) Reading and Grammar Score Reward and (b) Reading and Grammar Score Correlation

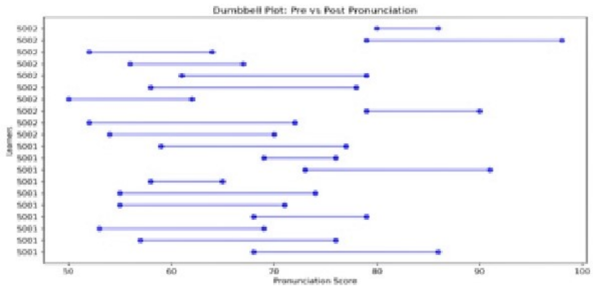

Fig. 6. Visualization of Learner Pronunciation Gains

system scores of vocabulary and grammar and the Target Reward. The positive correlation between pre- and post-test scores is strong, which proves the efficacy of the framework in enhancing the proficiency of learners, and the definite correlation between postscores and reward proves the correspondence of the system reinforcement mechanism to the educational goals.

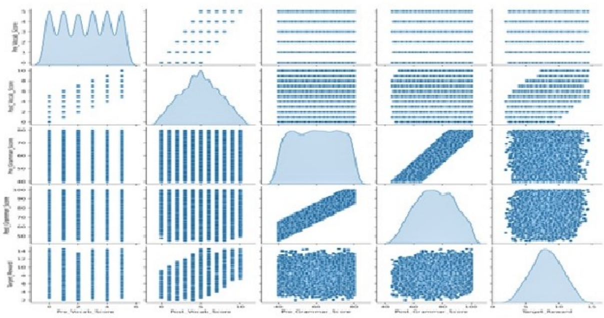

Fig. 7. Graphical Pair Plot representing Learning Metrics and Reward

Figure 8 indicates the target reward of the R-MAP-DEQL framework of five learners in ten sessions. The colored lines are used to represent the learners, and the upward trend shows a positive learning pattern. The fluctuations underscore the dynamism and real-time aspect of the system as it adjusts teaching strategies to individual performance to ensure that learning and engagement are continuous.

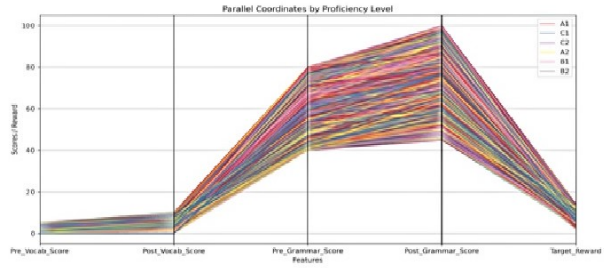

Fig. 8. Graphical Representation of Learning Progress of Learners Over Sessions

Figure 9 indicates that there are strong positive correlations between pre- and post-test scores on all language skills, which implies learning effectiveness. It further points out that the Target_Reward has a strong correlation with Engagement_Score (1.00), which implies that rewards are directly related to engagement measures in the design of the system. This proves the fact that the system is aimed at encouraging user participation.

◇ **Comparative Analysis**

The proposed R-MAP-DEQL model is contrasted with the current approaches, including Random Forest (RF) [22], Decision Tree (DT) [22], Improved Fuzzy Cognitive Network (IFCN) [22], Artificial Neural Network (ANN) [22], Support Vector Machine (SVM) [22] and Extreme Gradient Boosting (XG Boost) [22] to compare and predict the performance and progress of learners in ELT. The comparative models were trained in the same conditions to make the analysis of computational complexity objective and benchmarking fair.

Computational complexity is a measure of the time, memory, or resources required by an algorithm to solve a problem as the input size increases. Figure 6 (b) gives the graphical representation of efficient ELT through R-MAP-DEQL, which has better computational performance.

Figure 10 is a comparison of the computational complexity of various deep RL methods with training epochs. The complexity of proximal policy optimization (PPO) [19], Asynchronous advantage actor critic (A3C) [19] and Deep deterministic policy gradient (DDPG) [19] also increases, whereas the complexity of the proposed Deep Q Network (DQN)-based method is always lower. The proposed R-MAP-DEQL framework in the research further streamlines this by incorporating proximal tuning and resource-aware strategies, which guarantee lower computational cost and high learning performance. This efficiency allows real-time customized ELT interventions without overwhelming resources. The high performance of R-MAP-DEQL demon-

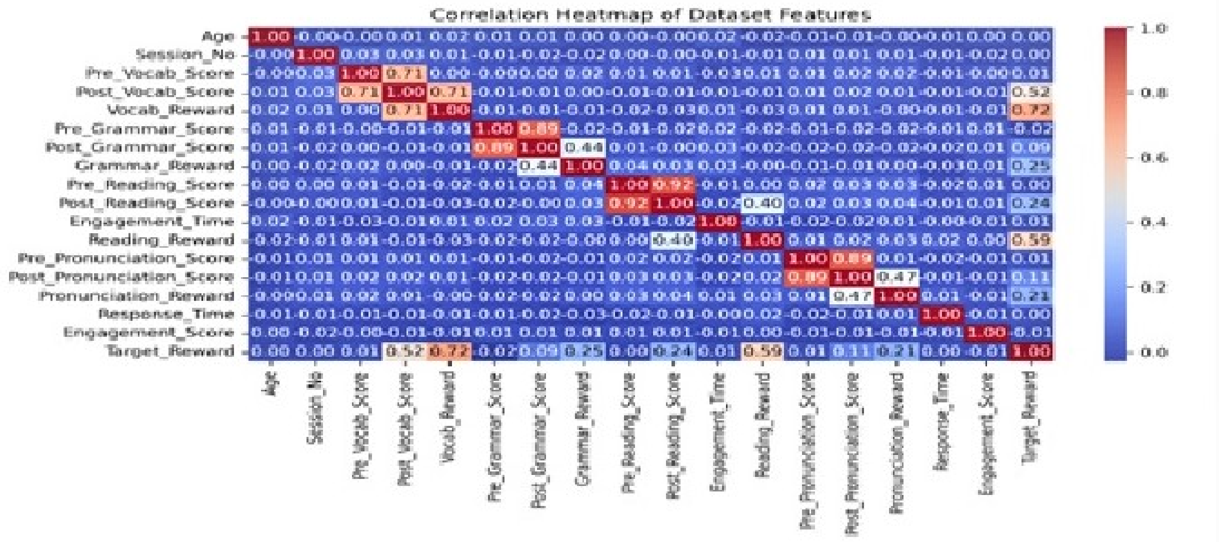

Fig. 9. Correlation Analysis of AI English Teaching System Feature

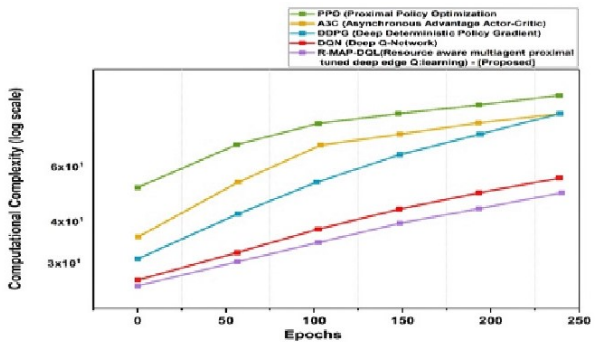

Fig. 10. Computation Performance of R-MAP-DEQL for Personalized English Teaching.

strates its capacity to trade-off between accuracy, adaptability and efficiency, which is why it can be successfully deployed in dynamic classroom settings and scalable AI-based language learning systems.

◇ **Training-Validation Assessment of Proposed R-MAP-DEQL English Teaching Framework**

Training loss and accuracy are used to measure the ability of the model to learn based on the training dataset, where the loss is used to measure the error in prediction and accuracy is used to measure the correct output. Validation loss and accuracy measure the performance on unseen data and indicate the generalization of the model. All these metrics together prove the ability of the system to learn patterns without overfitting, which guarantees the correct operation in the real world.

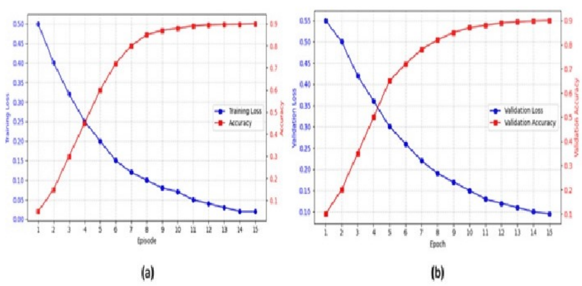

Fig. 11. (a) Training and (b) Validation Performance of R-MAP-DEQL

Figure 11 shows the training and validation outcomes of the proposed R-MAP-DEQL framework that attains stable convergence and high generalization in personalized teaching of English. The above figure indicates that training loss reduces gradually as the accuracy increases beyond 90% and this indicates that the multi-agent system learns effective teaching policies. Equally, it demonstrates that validation loss decreases with accuracy that is similar to training performance, meaning that the model fits unseen learner data without overfitting. These results demonstrate that the agents are able to dynamically optimize vocabulary, grammar, reading and pronunciation activities and resource constraints, thus, guaranteeing scalability, reliability and individual learning in real-time learning settings.

Table 2 shows the evaluation measures that will be applied to determine the performance of the proposed R-MAP-DEQL-based ELT system in terms of prediction

Table 3 provides the comparison result of exiting meth-

Table 2. Proposed R-MAP-DEQL Method Performance Evaluation Metrics

Metric	Formula	Description
Accuracy	$\frac{(P_T + N_T)}{(P_T + N_T + P_F + N_F)}$ $P_T -$ True Positives $N_T -$ True Negatives $P_F -$ False Positives $N_F -$ False Negatives	Measures correctly predicted learning outcomes across vocabulary, grammar, reading and pronunciation.
Precision	$\frac{(P_T)}{(P_T+P_F)}$	Evaluates the proportion of correctly identified teaching interventions among all applied strategies
Recall	$\frac{(P_T)}{(P_T+N_F)}$	Identifies the proportion of true student learning improvements captured by the framework.
F1-Score	$2\times\frac{(\text{ Precision } \times \text{ Recall })}{(\text{ Precision } + \text{ Recall })}$	Harmonic mean of precision and recall for balanced evaluation.

ods with the proposed method Accuracy and Precision outcome values.

Table 3. Comparative Accuracy and Precision of R-MAP-DEQL vs. Existing Methods

Method	Accuracy (%)	Precision (%)
I-FCN [22]	84	93
ANN [22]	78	80
XG Boost [22]	76	54
SVM [22]	75	46
RF [22]	72.5	55
DT [22]	71.5	55
R-MAP-DEQL [Proposed]	97	95

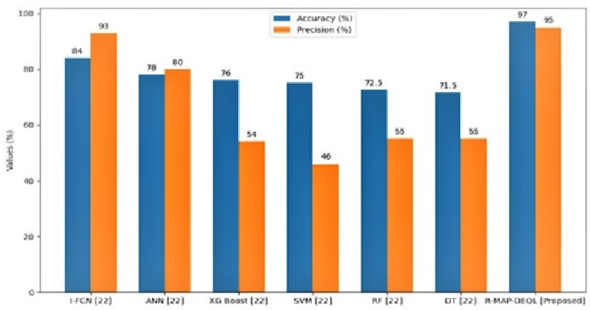

Fig. 12. Comparative Analysis of Accuracy and Precision in English Teaching Models

Figure 12 illustrates the comparative performance highlights the effectiveness of the proposed R-MAP-DEQL framework against existing AI and ML models. Traditional methods such as DT (71.5% accuracy, 55% precision), RF (72.5%, 55%) and SVM (75%, 46%) show relatively low performance, limiting their ability to provide accurate and personalized feedback in ELT. High-tech approaches such as ANN (78%, 80%) and I-FCN (84%, 93%) were obtained but remain not flexible and real-time accurate. Conversely, the R-MAP-DEQL model has an accuracy of 97% and precision of 95% which is a clear indication of its better ability to provide adaptive, personalized and resource conscious ELT interventions. In order to ascertain the statistical significance of the performance difference between R-MAP-DEQL and other approaches, pair t tests were considered to be accurate, precise and recall to validate the findings. Moreover, R-MAPDEQL significantly outperforms DQN, PPO and A2C in all of the measures compared to all three scores ($p < 0.05$). This shows that the performance benefits are not random variance, but because the system has increased the ability to engage in personalized learning. Table 4 gives the result of the compared values of the proposed model and the existing methods.

Table 4. Comparative Analysis of Recall and F1-Score for English Teaching Models

Method	Recall (%)	F1-Score (%)
I-FCN [22]	88	91
ANN [22]	72	76
XG Boost [22]	48	47
SVM [22]	71	71
RF [22]	49	49
DT [22]	57	55
R-MAP-DEQL [Proposed]	93	96

The comparative results of recall and F 1 -score are presented in figure 13, which indicates that the proposed R-MAP-DEQL model performs better than the current methods in the ELT system. The R-MAP-DEQL also has the best recall of 93% and F1-score of 96%, which means that it is highly effective in identifying the relevant learning patterns without compromising on the precision and recall. By contrast, the I-FCN model has a relatively high recall

Table 5. Comparison of R-MAP-DEQL with Traditional ELT Methods

Feature	Proposed R-MAP-DEQL Framework	Intelligent Assisted Teaching System for ChineseEnglish Learners [23]	MARLBased ESL Modules for Tamil Learners [24]
Personalization / real-time adaptation	High - multi-agent + edge-device + proximal tuning	Moderate - adaptive system, less emphasis on edge/MARL	Moderate MARL applied but less edge-optimization
Use of MARL or multi-agent design	Yes - student, teacher, content, evaluation, interaction agents	No - mostly single system	Yes - MARL for ESL but less edge-device focus
Edge-device / resource-constraint focus	Yes - designed for low-latency, resource-aware execution	Limited	Limited
Theoretical contribution (e.g., proximal tuning)	Yes - novel proximal tuned DEQL mechanism	Limited	Limited
Strengths	Scalability, real-time adaptive, personalized feedback, resource-aware	Adaptive support for Chinese learners	MARL for ESL Tamil context
Limitations	Requires infrastructure, tuning complexity, edge deployment challenges	Less emphasis on infrastructure/resource constraints	Context specific to Tamil learners

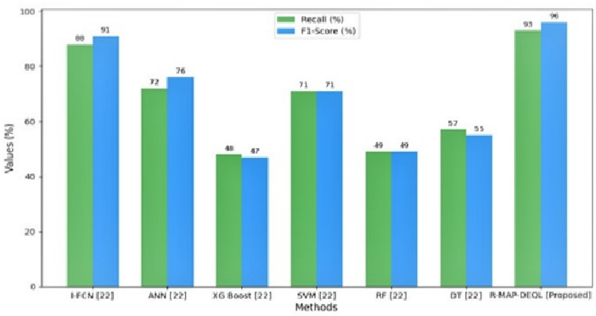

Fig. 13. Comparative Recall and F1-Score Analysis in English Teaching Models

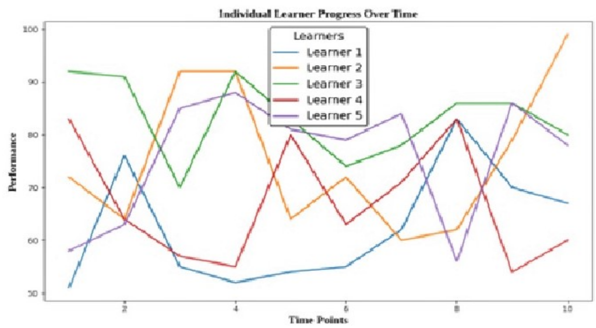

Fig. 14. Individual Learner Progress Over Time

of 88% and F 1 -score of 91% but remains relatively low compared to the proposed approach. ANN presents moderate scores (72% recall, 76% F1-score), SVM (71%, 71%) and DT (57%, 55%) have irregular performance. XG Boost (48%, 47%) and RF (49%, 49%) have the lowest results, as they do not effectively capture key learning characteristics. In general, this discussion proves that R-MAPDEQL is always better than the traditional methods, which guarantee more accurate prediction of student performance, individual feedback and adaptive learning results.

Figure 14 demonstrates the progress of five separate learners through ten time points. The lines are the performance of the individual learners in a personalized learning environment where the x -axis represents the time points and the y -axis represents the performance levels. The different colored lines depict every learner and indicate how their performance varies over time. The figure shows the active flexibility of the system because it shows how the performance of each learner is tracked and adjusted based on his or her progress. This visualization confirms that the R-MAPDEQL model is useful in providing personalized learning experiences.

◇ **Comparison with Traditional ELT Methods**

Table 5 provides the comparison of the R-MAP-DEQL framework with two other systems: the Intelligent Assisted Teaching System of Chinese English Learners and the MARL-Based ESL Modules of Tamil Learners. It underlines such important characteristics as personalization, MARL and edge-device emphasis. The table also shows the pros and cons of each system against the traditional ELT methods with emphasis on the fact that R-MAP-DEQL offers superior real-time adaptation and resource-awareness. This comparison helps to prove the innovations and challenges of R-MAP-DEQL within the framework of adaptive ELT systems

◇ **Sensitivity Analysis of Model Parameters**

Table 6 shows the findings of a sensitivity analysis performed to investigate the effect of various parameter values on the performance of the R-MAP-DEQL framework. The most important parameters to be tested are the learning rate (α), exploration factor (ϵ), and weight on the reward (w_{reward}). The best values of each parameter are determined by the performance of the system in a set of values. The effects of these parameters on the convergence of the system, exploration-exploitation balance and learning efficiency are also summarized in the table that provides the stability and efficient real-time performance of the system.

5. Discussion

The existing methods that are applied in ELT predictive tasks are all of superior deficiency. IFCN [22] is very demanding in terms of training data and is not flexible. ANN [22] is prone to overfitting small data and has low interpretability. XG Boost [22] is inclined towards repeating patterns and cannot capture sequential learning habits. Likewise, SVM [22] demands heavy feature engineering, has poor scalability and produces less accurate feedback. RF [22] and DT [22] models are characterized by high variance, overfitting and poor generalization. PPO [19], A3C [19] and DDPG [19] in DQN face drawbacks like high computational cost, instability, sensitivity to hyperparameters, poor generalization, sample inefficiency and difficulty handling sparse rewards in complex multi-agent environments. These shortcomings render them undesirable for personalized, real-time education systems. The R-MAP-DEQL framework presented here successfully alleviates these shortcomings with fixed resource-aware MARL, self-adjusting reward mechanisms and noise-handling techniques. Further, the proposed R-MAP-DEQL model does show significant advantages over existing ELT models by providing personalization, adaptiveness and real-time responsiveness. In addition to its ability to produce dynamic, studentcentered feedback and adaptive teaching strategies, it is highly accurate and accurate. Therefore, R-MAP-DEQL produces better, individualized and responsive English instruction results than current methods. The R-MAP-DEQL model is validable in various classroom environments through the application in various learner groups and learning materials in a variety of public or private schools. Its modular structure enables the learners to be able to adjust to different learning patterns, culture and level of proficiency. Initial experiments indicate consistent reliability, but the future research is concerned with large-scale testing on heterogeneous populations.

5.1. Ethical Considerations

The ethical issues related to data privacy, consent of students and management of sensitive data are of paramount importance in the design and implementation of the R-MAP-DEQL framework. The system will ensure that any information collected concerning the students, the records of the interactions, performance and audio recordings are anonymized and stored securely to guarantee privacy of users. Informed student consent is provided prior to any data collection and it clearly outlines how their data will be used in education and how their data will be ensured to be confidential. In addition, the framework adheres to the relevant data protection regulations, such as GDPR and provides users with the option of consenting and deleting their data at any time. All this is needed in order to maintain the integrity of the system and ensure that the information of users is handled responsibly.

6. Conclusion

Research developed a personalized, adaptive ELT that confirms that integrating MARL within ELT systems provides significant improvements over traditional methods. The dataset, gathered from Kaggle sources, ensured diverse inputs, including text and audio, for robust model training. Data preprocessing techniques such as tokenization, normalization and MFCC extraction enhanced the quality of input for accurate analysis. By leveraging the R-MAP-DEQL framework, the system effectively delivers personalized, adaptive and real-time feedback across vocabulary, grammar, reading and pronunciation. Experimental results demonstrate that R-MAP-DEQL achieves an accuracy of 97%, a precision of 95%, a recall of 93% and an F 1 - score of 96%. R-MAP-DEQL efficiently delivers superior performance for real-time personalized ELT. These results highlight the framework's robustness in handling noise, balancing recall and precision and providing adaptive interventions. The proposed R-MAP-DEQL system enhances English learning through personalization, scalability and efficiency while boosting learner engagement, representing a significant advancement toward AI-driven intelligent education systems with improved adaptability and effectiveness.

Table 6. Sensitivity Analysis of Key Model Parameters

Parameter	Range Tested	Optimal Value	Impact on Performance
Learning Rate (α)	0.001 to 0.1	0.01 to 0.05	Performance degraded above 0.1 due to overshooting of Q-values. Best performance between 0.01 and 0.05 .
Exploration Factor (ϵ)	0.1 to 0.5	0.2	Lower values (0.1) led to insufficient exploration. Higher values (0.4) caused slower convergence. Best at 0.2 .
Reward Weight (w_{reward})	0.5 to 1.5	1.0	Overfitting occurred above 1.2; underfitting below 0.8 . Best balance at 1.0 , providing effective learning feedback.

◇ Limitations and Future Scope

While R-MAP-DEQL is highly adaptable and efficient, it suffers from sensitivity to hyperparameter tuning, coordination complexity among agents and unstable proximal updates. But this was mitigated by controlled learning rates and edge-level optimization during implementation as well as the accuracy of R-MAP-DEQL in terms of adaptability and efficiency. While the R-MAP-DEQL model is better accuracy and adaptability, it does require structured data, high computational costs and limited testing in many classrooms. In real-world educational environments, practical challenges may include variation in digital infrastructure, teacher readiness for AI-supported systems and ensuring student data privacy during continuous learning analytics. These concerns can be addressed through scalable cloud-based deployment, lightweight model compression and structured teacher-system co-adaptation. Future work will focus on scalability, multilingual adaptability, lightweight architectures and real-time deployment to enhance accessibility and broader educational impact.

Declarations

Funding: Authors did not receive any funding.

Conflicts of interests: Authors do not have any conflicts.

Data Availability Statement: The data generated and analyzed during the current study are available from the author Linzhi Shao upon reasonable request but are not yet publicly available due to ongoing research.

Code availability: Not applicable.

Authors' Contributions: Linzhi Shao is responsible for designing the framework, analyzing the performance, validating the results, and writing the article.

References

[1] Y. H. Jiang, R. Li, Y. Zhou, C. Qi, H. Hu, Y. Wei, B. Jiang, and Y. Wu, (2024) *"AI agent for education: von Neumann multi-agent system framework"* **arXiv preprint *arXiv:2501.00083***: DOI: 10.48550/arXiv.2501.00083.

[2] C. Cai, S. Hong, M. Ma, H. Feng, S. Du, M. Chow, W. L. L. Teo, S. Liu, and X. Fan, (2025) *"Analyzing the teaching and learning environments through student feedback at scale: a multi-agent LLMs framework"* **Education and Information Technologies**: 1–33. DOI: 10.1007/s10639-025-13633-2.

[3] F. Jiang, Y. Peng, L. Dong, K. Wang, K. Yang, C. Pan, D. Niyato, and O. A. Dobre, (2024) *"Large language model enhanced multi-agent systems for 6G communications"* **IEEE Wireless Communications**: DOI: 10.1109/MWC.016.2300600.

[4] A. Lazaridou, A. Potapenko, and O. Tieleman, (2020) *"Multi-agent communication meets natural language: Synergies between functional and structural language learning"* **arXiv preprint *arXiv:2005.07064***: DOI: 10.48550/arXiv.2005.07064.

[5] L. Qian, (2022) *"Research on college English teaching and quality evaluation based on data mining technology"* **Journal of Applied Science and Engineering *26***(4): 547–556.

[6] S. Jha, S. Ahmad, H. A. Abdeljaber, A. A. Hamad, and M. B. Alazzam, (2021) *"A post COVID machine learning approach in teaching and learning methodology to alleviate drawbacks of the e-whiteboards"* **Journal of Applied Science and Engineering *25(2)***: 285–294.

[7] L. Wei, (2023) *"Artificial intelligence in language instruction: impact on English learning achievement, L2 motivation, and self-regulated learning"* **Frontiers in Psychology *14***: DOI: 10.3389/fpsyg.2023.1261955.

[8] Z. Zhang, D. Zhang-Li, J. Yu, L. Gong, J. Zhou, Z. Hao, J. Jiang, J. Cao, H. Liu, Z. Liu, and L. Hou, (2024) *"Simulating classroom education with LLM-empowered agents"* **arXiv preprint *arXiv:2406.19226***: DOI: 10.48550/arXiv.2406.19226.

[9] M. Wang, D. Zhang, J. Zhu, and H. Gu, (2025) *"Effects of incorporating a large language model-based adaptive mechanism into contextual games on students' academic performance, flow experience, cognitive load and behavioral patterns"* **Journal of Educational Computing Research *63(3)***: 662–694.

[10] B. R. Gudivaka, (2021) *"Designing AI-assisted music teaching with big data analysis"* **Current Science & Humanities *9(4)***: 1–14.

[11] O. Hamal, E. Faddouli, and M. H. A. Harouni, (2021) *"Design and implementation of the multi-agent system in education"* **World Journal on Educational Technology: Current Issues *13(4)***: 775–793.

[12] M. Liu, (2022) *"Intelligent integration method of AI English teaching resource information under multi-agent collaboration"* **Advances in Multimedia *2022(1)***: 1104443. DOI: 10.1155/2022/1104443.

[13] J. Wei, (2021) *"Study on the effective mechanism of network English independent learning platform based on multi-agent of big data"* **Information Science and Education (ISE)**: 1073–1076. DOI: 10.1109/ICISE-IE53922.2021.00244.

[14] A. Cristea and T. Okamoto, (2000) *"MyEnglishTeacher: VOD based distance academic English teaching via an adaptive, multi-agent environment"* **KAIS Journal**:

[15] S. Xie and J. Xu, (2023) *"Design and implementation of physical education teaching management system based on multi-agent model"* **International Journal of Computational Intelligence Systems *16(1)***: 172. DOI: 10.1007/s44196-023-00349-9.

[16] A. Fernández-Caballero, V. López-Jaquero, F. Montero, and P. González, (2003) *"Adaptive interaction multi-agent systems in e-learning/e-teaching on the web"* **International Conference on Web Engineering**: 144–153. DOI: 10.1007/3-540-45068-8_27.

[17] B. Wang, (2024) *"An intelligent integration method of AI English teaching resources information under multi-agent cooperation"* **International Journal of Continuing Engineering Education and Life Long Learning *34(1)***: 88–99. DOI: 10.1504/IJCEELL.2024.135266.

[18] Y. Zhai, (2025) *"Computer-assisted English teaching model for improving learning performance of college students"* **Journal of Computational Methods in Sciences and Engineering *14727978251361524***: DOI: 10.1177/14727978251361524.

[19] M. Sharif and D. Uckelmann, (2024) *"Multi-modal LA in personalized education using deep reinforcement learning based approach"* **IEEE Access *12***: 54049–54065. DOI: 10.1109/ACCESS.2024.3388474.

[20] J. Hu and G. Jin, (2024) *"An intelligent framework for English teaching through deep learning and reinforcement learning with interactive mobile technology"* **International Journal of Interactive Mobile Technologies *18(9)***: DOI: 10.3991/ijim.v18i09.49289.

[21] Q. Du, (2025) *"How artificially intelligent conversational agents influence EFL learners' self-regulated learning and retention"* **Education and Information Technologies**: 1–67. DOI: 10.1007/s10639-025-13602-9.

[22] M. Hooda, C. Rana, O. Dahiya, A. Rizwan, and M. S. Hossain, (2022) *"Artificial intelligence for assessment and feedback to enhance student success in higher education"* **Mathematical Problems in Engineering *2022(1)***: 5215722.

[23] X. Li, (2025) *"Development and Implementation of an Intelligent Assisted Teaching System for Chinese English Based on Natural Language Processing and Reinforcement Learning"* **Computers and Education: Artificial Intelligence *100466***:

[24] D. R. Rathinasamy, (2025) *"The Role of Multiagent Reinforcement in Teaching English as a Second Language to Tamil Learners"* **International Journal of English Language Teaching *7(5)***: 728–738.