Expert Systems With Applications 269 (2025) 126426

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Measuring student attention based on EEG brain signals using deep reinforcement learning

Asad Ur Rehman [a], Xiaochuan Shi [a,*], Farhan Ullah [b], Zepeng Wang [a], Chao Ma [a,*]

[a] School of Cyber Science & Engineering, Wuhan University, Wuhan, 430072, Hubei, China
[b] Cybersecurity Center, Prince Mohammad Bin Fahd University, 617, Al Jawharah, Khobar, Dhahran 34754, Saudi Arabia

ARTICLE INFO

Keywords:
EEG (electroencephalography)
Attention classification
Deep reinforcement learning
Neural correlates
Double Deep Q-Network (DDQN)
Adaptive education
Student engagement
Online learning

ABSTRACT

Quantifying attention during learning is essential yet extremely challenging. Conventional methods like surveys and observations are subjective and inaccurate. While electroencephalography (EEG) mirrors attention shifts, analysis is obstructed by artifacts and variability. This research proposes a double deep Q-network (DDQN), a specialized deep reinforcement learning algorithm well-suited for handling complex multidimensional data like EEG signals. Raw multi-channel EEG recordings undergo preprocessing by wavelet transformations, a technique that provides frequency-based signal representations robust to noise and fluctuations. Key features are extracted from these cleaned EEG traces to capture attention-related neural signatures. These feature vectors serve as inputs to the DDQN agent, which uses its learned action-value function to choose one of three discrete attention state labels (attentive, non-attentive, drowsy) at each timestep based on the input EEG features. Correct classifications result in +10 rewards to reinforce the agent predictions, while incorrect labels yield -1 penalties, enabling the DDQN to iteratively improve its classification accuracy over sequential iterations. A dual-network architecture with separate but interlinked target and online Q-networks enhances stability during this reinforcement learning process for attention quantification. Results showed 98.2% test accuracy in classifying attentive versus non-attentive versus drowsy states, significantly improving on the 92% benchmark. The reduced 0.65 loss value evidenced convergence. This methodology enables granular quantification of fluctuating attention from EEGs. The DDQN's neural decoding capability and substantial improvements establish a new state-of-the-art of adaptive education systems to track engagement by leveraging brain signals.

1. Introduction

With the exponential growth in the use of e-learning platforms and remote online classrooms for digital education, precisely measuring and optimizing student engagement during remote learning sessions is more crucial than ever. This is because remote learning lacks direct face-to-face interaction, which typically allows teachers to assess and adjust their approaches based on student focus and engagement levels. While focus and concentration are critical components of cognitive processing, engagement encompasses more than just attention it includes emotional and behavioral involvement in the learning process reflecting a student's commitment and active participation in learning activities. (Bahmani et al., 2019). Prior research has demonstrated the feasibility of using EEG signatures to evaluate attention and engagement levels (Aggarwal et al., 2021a; Trabelsi et al., 2023). Most existing systems rely on supervised learning approaches that require explicit attention state labels during model training (Shaw et al., 2023). Limited literature explores reinforcement learning (RL) for implicit modeling of attention variations from EEG data, representing a key research gap. RL offers unique advantages for handling complex temporal relationships and high-dimensional data like EEG (Yu et al., 2021). RL's temporal modeling capabilities enable effective learning and prediction over time, essential for capturing EEG dynamics. Additionally, RL enhances sample efficiency through experience replay, improving learning speed and robustness by reusing past data. This research pioneers a Double deep Q network (DDQN) to address the open challenge of implicitly quantifying attention fluctuations from EEG inputs.

Rather of providing explicit labels, the DDQN is implicitly guided to identify EEG-attention inter-relationships using a customized incentive system. This research accomplishes more granular modeling with a 3-class DDQN deciphering "attentive", "non-attentive", and "drowsy"

* Corresponding authors.
E-mail addresses: asadurrehman@whu.edu.cn (A.U. Rehman), shixiaochuan@whu.edu.cn (X. Shi), fullah@pmu.edu.sa (F. Ullah), wangzepeng@whu.edu.cn (Z. Wang), chaoma@whu.edu.cn (C. Ma).

https://doi.org/10.1016/j.eswa.2025.126426
Received 11 April 2024; Received in revised form 16 November 2024; Accepted 3 January 2025
Available online 10 January 2025

levels, whereas previous efforts showed promise for binary classification of "focused" or "disengaged" states (Abedi & Khan, 2023). The proposed approach advances state-of-the-art with a 6% leap in accuracy over previous DQN models, crossing 98% multi-class reward first time. This establishes new benchmarks, enabled by the dual Q-learning strategy and tailored asymmetric reward function uniquely incorporated in the presented DDQN architecture. A number of investigations have shown a connection between changes in the direction and intensity of patients' attentional focus and tracing certain spectral patterns and spatial distributions that show up in the recorded EEG data (Contreras-Jordán et al., 2022). This is why, in contrast to perceived surrogates, EEG-derived brain feedback signals have become an attractive supplemental technique to evaluate actual involvement levels (Abedi & Khan, 2023). Implicitly quantifying attention fluctuations from EEG inputs is still extremely difficult to solve, nevertheless, because these neuroelectrical signals are frequently extremely faint and hidden below the noise and EEG artifacts with orders of magnitude larger amplitude (Xu & Zhong, 2018). Furthermore, significant intra- and inter-individual variability shown makes precise generalizable modeling an unresolved research issue (Kanai & Rees, 2011). Due to these restrictions, it is necessary to use advanced deep machine learning architectures that can reveal latent relationships in high-dimensional signals associated with variations in attention across a wide range of subjects and contexts that are typical of real-world educational settings (Malekloo et al., 2022).

Deep reinforcement learning (RL) has shown tremendous capability in handling noisy, continuous data streams (Khetarpal et al., 2022). In their study, (Liu et al., 2023) focused on distinguishing between epileptogenic zone (F) and non-epileptogenic zone (N) EEG data. This distinction was based on EEG signal characteristics to accurately identify regions associated with epileptic activity. While our work utilizes RL agents to optimize solutions by maximizing cumulative rewards over time, Liu et al.'s study demonstrates the application of advanced machine learning techniques in EEG signal analysis for clinical purposes This study postulates a double deep Q network (DDQN), trained using EEG features to accurately classify three attention levels. Appropriate feedback rewards guide the model to enhance classification performance. Our study pioneers the use of RL for implicit attention assessment from EEG signals, focusing on real-time adaptive interfaces to measure user engagement dynamically. Unlike traditional supervised learning paradigms in EEG research, RL as a branch of machine learning (ML) enables adaptive systems that optimize decisions over time based on continuous EEG feedback. This capability enhances user engagement through personalized interaction dynamics adjusting in real-time to the user cognitive states. This research addresses this gap with a novel DDQN architecture using EEG markers of attention as implicit inputs to model the students' attentional state space. The ensuing sections present related literature, the proposed EEG-DDQN methodology, experimental results, and comparative analyses. The dual advantages of leveraging informative neural data through advanced RL for granular attention modeling during online learning signifies the paper's contributions. The proposed approach provides continuous feedback on cognitive engagement unobtrusively through real-time analysis of EEG signals. By leveraging RL, our method dynamically adjusts digital education platforms to enhance student attention. Specifically, the RL agent receives EEG inputs, processes them to infer the student's attentional state (attentive, non-attentive, drowsy), and then adapts the educational content or interface based on these inferred states. This process occurs seamlessly during the student's interaction with the digital platform, ensuring a personalized learning experience that optimizes engagement without interrupting the learning flow.

This innovative research tackles several technically challenging facets associated with decoding neural correlates of fluttering attention levels. (Verma et al., 2023). The continuous, data-intensive output and intricate multidimensional representations of EEG data streams strain computational limits. Converting the real-time attention monitoring issue into a well-defined state–action space to enable DDQN training and garner maximum cognitive rewards also proves non-trivial. If these challenges are suitably addressed, the RNN-based reinforcement learning paradigm showcased has versatile applicability. Proposed approach leverages the EEG-DDQN attention classifier to provide implicit feedback by continuously monitoring students' cognitive states. The classifier processes EEG signals to detect attention levels and sends this information to the e-learning platform, which can then adapt the content in real-time. For example, if the classifier detects decreased attention, the platform might reduce the difficulty level or introduce interactive elements to re-engage the student. This system can be integrated through wearable EEG devices that unobtrusively collect data during learning sessions, ensuring that the process remains seamless for the user. Despite challenges like noise and artifacts, advanced preprocessing techniques and robust RL models help mitigate these issues, making real-time engagement optimization feasible. It can also dynamically modulate resource allocation for human–computer interaction systems like BCIs to maintain user attention. Proposed model can be extended to neural tracking of fatigue, emotions, and other human performance aspects by incorporating additional EEG features and training the DDQN on labeled datasets specific to these states. By expanding the range of monitored neural signatures, the methodology can provide deeper insights into various cognitive and emotional states. The real-world applications of this model in educational settings are highly promising. By integrating the attention-tracking capabilities of the DDQN into both physical and online learning platforms educators could gain valuable real-time insights into student engagement. In traditional classrooms, the system could be used in conjunction with wearable EEG devices, providing teachers with a live dashboard of attention states allowing for timely interventions when students become disengaged or drowsy. For online learning environments, where direct observation of students is limited, this model could be embedded into e-learning platforms to continuously monitor student focus and adapt the pacing or content delivery based on attention levels. The model robustness against noise and variability makes it suitable for diverse learning contexts from interactive classrooms to self-paced virtual learning environments enhancing adaptive education systems on a broader scale.

The main contribution of this research is given below:

- Novel deep reinforcement learning architecture for EEG-based attention modeling: Proposes a new DDQN agent for multi-class classification of attention levels using EEG signatures, addressing a key research gap for leveraging RL advances to decode neural correlates of attention.
- Granular quantification of dynamic attention levels: Achieves classification of attention into three levels — attentive, non-attentive, and drowsy over time with substantially higher outcome relative to prior EEG attention benchmarks. Enables unprecedented tracking of subtle fluctuations in focus.
- Adaptive integration and testing for online education platforms: Demonstrate integration of the developed DDQN model into actual e-learning interfaces for real-time implicit attention monitoring during academic tasks to evaluate efficacy and lay the groundwork for future adaptive learning systems.

Having established the context and motivation driving this innovative research on decoding neural correlates of attention using reinforcement learning algorithms, the next sections delve deeper into the precise technical approach adopted. The subsequent literature review summarizes prior related works that informed the methodology designed in this study. Next, the proposed EEG-DDQN model architecture and experimental framework are detailed. This is followed by an elaboration on the key results obtained from rigorous evaluation of the developed technique on multi-subject educational EEG data. Comparative analyses are drawn with existing state-of-the-art methods to highlight the advances made by the suggested algorithm. Finally, salient inferences are highlighted in the concluding remarks along with directions identified for future work towards deployable EEG-driven adaptive e-learning systems that can revolutionize digital classrooms worldwide by incorporating implicit and continuous feedback on learner attention.

2. Related work

2.1. Enhancing learning with EEG and deep learning

This observation introduces an e-learning system that personalizes content based on student emotions, using EEG and the K-nearest neighbors (KNN) algorithm for real-time emotional state recognition. Tested on 30 computer science undergraduates from the University of Nottingham Ningbo China, the system showed promising results in emotional state recognition (74.3% accuracy) and significantly enhanced student satisfaction in learning English listening and reading skills, despite not markedly impacting learning and engagement (Liu & Ardakani, 2022). Rudovic et al. (2019) who unveiled a completely new deep reinforcement learning architecture for engagement estimation and active learning using video data. The key to this method is the development of a customized policyÂ that enables the model to select between estimating the child's involvement level (low, medium, or high) and asking a human for a video label in cases of uncertainty. A human expert labels the requested films offline in order to progressively adjust the policy and engagement classifier to a target kid (Li & Xu, 2020). RLPS is a novel reinforcement learning system that pre-selects photos in FER using an image picker and a rudimentary emotion classifier. On the RAF-DB, ExpW, and FER2013 datasets, it outperforms state-of-the-art FER approaches in terms of improving dataset quality. The technology may be used to manage any decreases in the user's cognitive and emotional involvement, making an automated teaching platform responsive to them. The goal of the experiment was to understand the operation of a certain human–machine interface. Participants' motor and cognitive abilities were put to use. To ensure a reference with a metrological foundation, de facto standard stimuli were used, including a cognitive activity (Continuous outcome Test), background music (Music Emotion Recognition MER database), and social input (Hermans and De Houwer database). The suggested signal processing pipeline (Filter bank, Common Spatial Pattern, and Support Vector Machine) achieves almost 77% average accuracy in identifying both cognitive and emotional involvement in a within-subject manner by Apicella et al. (2022). An experimental framework for a human trial that contrasts the methods individuals use to give advice in terms of human participation is offered by Bignold et al. (2020). The findings indicate that users who offer the learner agents informative information do so with a greater degree of accuracy, a longer willingness to help, and more guidance each episode. Additionally, self-evaluation from participants utilizing the informative technique revealed that the agent's capacity to follow the advice is higher, and as a result, they feel their own counsel is more accurate in comparison to those giving evaluative advice.

2.2. EEG applications in education

MindWave headset from Neurosky was chosen as a component of the recording kit for EEG brain waves. Based on the results of a web-based reading experiment, a prior study demonstrates that the MindWave headset is extremely dependable and effective (Chen & Huang, 2014). The publication describes the specific discussion and test findings. The efficiency of learning in youngsters was examined utilizing EEG brain signals and a mobile application (Gasah et al., 2020). Mobile application development is advancing quickly, and this progress is now crucial and essential to our everyday lives (Gowthami & Kumar, 2016). Today, with the development of m-learning, mobile applications have been created that have a substantial impact on children's decision-making, character development, and learning behavior (Kraleva et al., 2016). Knowing which mobile applications have a greater good or bad impact on children's learning attitudes is the key question posed here. Children wear EEG headsets on their heads while utilizing various mobile applications, and brain signals are recorded. EEG is the procedure used to capture changes in brain signals while a person is going about their daily business. With the aid of several electrodes, often referred to as headset channels, these signals are utilized to anticipate or treat the patient's mental illnesses (Hu et al., 2019).

Introducing a BCI system employing EEG signals to gauge student attention in distance education, featuring a 96% accurate RF classifier. Providing objective measures for assessing engagement, it validates on a public dataset against leading systems, showcasing effectiveness in online learning (Devi & Sophia, 2024). Examining Massive Open Online Courses (MOOCs) and traditional classrooms, this research employs EEG signals and an SVM model to compare attention levels. Results show higher attention to MOOCs, emphasizing their relevance in modern education (Aggarwal et al., 2021b). Introducing a high-accuracy wearable EEG system for personalized engagement detection in Learning 4.0. It adapts teaching platforms to users by addressing cognitive and emotional engagement drops, allowing for personalized teaching strategies. Validated with twenty-one students, the system achieves nearly 77% accuracy in detecting cognitive and emotional engagement (Agrawal et al., 2024). Evaluates students' attention during instruction by examining EEG signals from mobile brainwave sensors, offering a precise substitute for conventional observation techniques. The study, which uses a support vector machine (SVM) classifier, achieves a classification accuracy of up to 76.82% and provides guidance for the development of new learning systems (Liu et al., 2013). introducing an e-learning system called Attention Aware System (AAS) that uses EEG and can recognize students' attention levels with 89.52% accuracy. Seven essential components are used to optimize the system using genetic algorithms. When teaching electrical safety, the AAS helps teachers identify instances of inadequate attention, which is related to student efficiency. It is paired with a method for tagging video lectures. This method offers valuable information to improve the results of online learning (Chen et al., 2017).

2.3. Advancements in EEG-based learning assessment and intervention

The study introduces an EEG-based attention model and looks at how learning is affected by outside disruptions. It is shown that naps are a useful focus-restoration strategy for people of both genders, highlighting the need of accurate attention assessment. A study that highlights the value of tailored attention recovery strategies for learning is done (Chiang et al., 2018), cautioning against more alluring pursuits. This study uses neuroengineering to predict students' performance on text and video learning challenges. Superior video group outcomes are revealed by EEG signals, which point to mental tiredness. With an accuracy rate of 85%, the tool recognizes video as being more effective, demonstrating the promise of neuroengineering for customized learning trials (Ramírez-Moreno et al., 2021). Using EEG signals from a 13-year-old with Autism Spectrum Disorder (ASD), this study measures attention, calculating features like Theta Relative Power. According to Esqueda-Elizondo et al. (2022), the multi-layer perceptron neural network model (MLP-NN) has the highest efficiency (AUC: 0.9299), indicating potential for individualized learning and progress measurement in people with ASD. Addressing the need for improved distance education, the study proposes a brain–computer interface (BCI) system employing EEG signals. The random forest classifier attains 96% accuracy, outperforming KNN and SVM, effectively discerning user attention states during online classes (Al-Nafjan & Aldayel, 2022). This literature review highlights advancements in EEG-based attention classification and deep learning in various contexts, such as video face identification, engagement estimation in education, and emotion recognition. Key developments include the use of deep

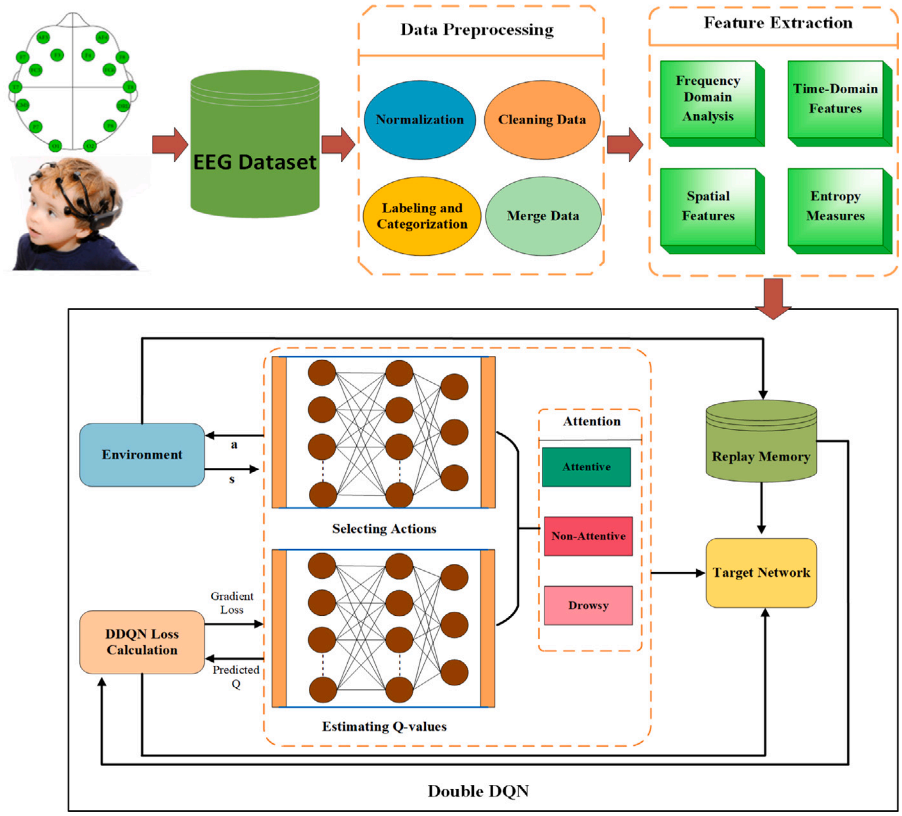

Fig. 1. Workflow diagram of proposed methodology measuring student attention based on EEG brain signals using deep reinforcement learning.

reinforcement learning models like DDQN for precise attention classification, the effectiveness of EEG headsets like MindWave for reliable data collection, and the impact of mobile applications on children's learning behaviors. Studies also compare attention levels in MOOCs and traditional classrooms, with results favoring MOOCs. The paper also discusses the impact of outside distractions on learning, the use of EEG in e-learning, and personalized learning systems. Finally, it discusses neuroengineering tools for learning assessments and the application of EEG in special education, particularly for Autism Spectrum Disorder, highlighting the potential of EEG-based technologies in education and cognitive analysis.

3. Adaptive attention detection through neural signals using DDQN

This paper presents a method to gauge student attention levels during online learning using EEG signals and deep reinforcement learning. In the preprocessing stage, noise and artifacts were carefully handled to ensure clean and reliable EEG data for analysis. A bandpass filter (0.5–50 Hz) was chosen to focus on the frequency ranges most relevant to cognitive activity while excluding high-frequency noise and low-frequency drift. Independent Component Analysis (ICA) was used to separate signals from different sources allowing for the removal of common artifacts such as eye blinks and muscle movements. Channels with excessive noise were either interpolated or excluded from the analysis based on predefined thresholds ensuring that only high-quality data were used for further processing. The choice of these techniques was guided by their effectiveness in previous EEG studies providing a robust framework for capturing attention-related neural signals. For feature extraction, power spectral density (PSD) analysis via Welch's method estimates power distribution across frequency bands, while Discrete Wavelet Transform (DWT) captures time and frequency domain information. These features are inputs to a Double Deep Q-Network (DDQN), which classifies the EEG signals into attentive, non-attentive, and drowsy states. Correct classifications are rewarded with +10, whereas incorrect classifications are penalized according to the mistake and earn a negative value of −1. The model is able to properly extract attention levels from the continuous EEG inputs by using a reinforcement mechanism based on deep reinforcement learning theory (Mnih et al., 2015). Using EEG data and deep reinforcement methods, the trained DDQN model provides insightful information on students' cognitive involvement and attention during online learning activities. Fig. 1 displays the workflow diagram for the suggested technique, which uses deep reinforcement learning to measure students' attention based on EEG brain signals.

A thorough technique based on recognized procedures was used to create trustworthy ground-truth labels for attentive states (Shatil et al., 2014). A standardized cognitive task intended to elicit diverse attention states was administered to each participant, as confirmed by previous research (Shatil et al., 2014). Expert human annotators categorized EEG data segments related to alert, non-attentive, and sleepy states using preset criteria as part of the labeling process. The ground-truth labels used to train and assess the DDQN model were guaranteed to be accurate and consistent thanks to this technique. To guarantee consistency in the EEG signals obtained, each participant completed the identical cognitive activity during the data collection process. EEG data were captured using EEG headsets, which have demonstrated reliability in similar studies (Mnih et al., 2015). The raw EEG signals were sampled at 256 Hz and collected from multiple channels to capture comprehensive brain activity patterns during the cognitive task.

3.1. Environment creation and integrated with DDQN

The schematic of the framework for modeling the student attention categorization problem using a double deep Q-network (DDQN) agent and a Markov decision process (MDP) is shown in Fig. 2. Multiple channels of raw EEG signals serve as the environment's state inputs. To describe the state space, important characteristics are taken from the

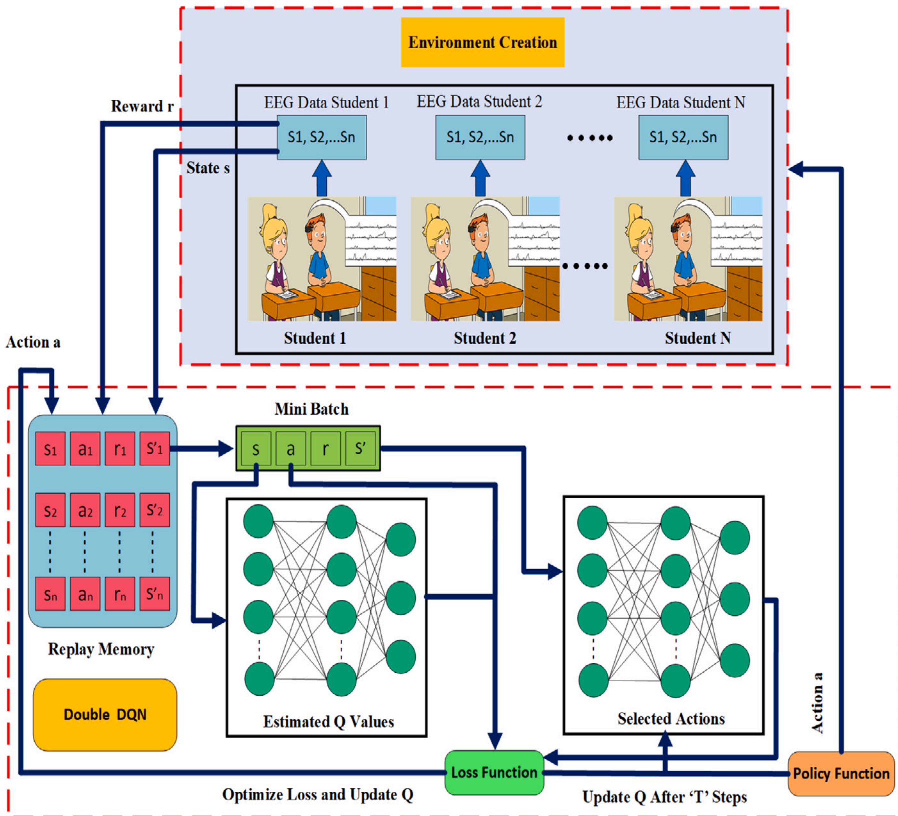

Fig. 2. Visualizing the integration of EEG-based state inputs into an MDP for student attention assessment with a DDQN agent.

raw signals, such as wavelet coefficients to capture time-frequency information and spectral strength to analyze frequency domain features. After a certain time, step, the environment changes to a new state upon receiving a new batch of EEG characteristics. The DDQN agent's action space is denoted by three distinct attention

labels: attentive, non-attentive, and drowsing. Based on its internal Q-value calculations, it chooses an action or attention label for the present EEG characteristics at each time step. The student's attention in that condition is predicted by the DDQN using this selected label. If the agent's chosen action fits the ground truth label, the environment gives it a reward signal. Accomplished classifications get positive incentives, whereas misclassified ones incur penalties. By using this exploration approach and incentive system, the agent is able to iteratively improve the attention categorization outcome by updating its Q-network. The real-world student attention monitoring problem is transformed into an MDP appropriate for DDQN training in a simulated setting through the process of environment design and integration.

3.2. Training process

The training process for the DDQN model is crucial for refining the online network to accurately predict attention states from EEG data. The online network minimizes the loss function $L(\theta)$ through gradient descent, where θ represents the online network weights. A batch is randomly sampled from the replay memory, and the loss is computed by comparing Q-values from the online network $Q(s, a|\theta)$ and the target network $Q'(s', a|\theta')$. The loss is defined as the mean-squared error over the batch, incorporating the immediate reward r and the discounted maximum Q-value of the next state s'. This temporal difference loss guides the update of online network weights θ through backpropagation. The softmax policy π_{model} determines the probability of selecting each action a_t given the current state s_t, controlled by a temperature parameter τ providing a balance between exploration and exploitation. Eq. (9) calculates the target value $V_{\text{target}}^{(i)}$ for each episode i representing the discounted sum of rewards, while Eq. (10) updates the online network weights θ based on the gradient of the DDQN loss L_{DDQN} and the learning rate α. This comprehensive training process ensures the DDQN model learns to accurately categorize attention states based on EEG feature representations.

$$\pi_{\textbf{model}}(a_t|s_t) = \frac{e^{Q_{\textbf{model}}(s_t a_t)/\tau}}{\sum_{j=1}^{|a|} e^{Q_{\textbf{model}}(s_t, a_t, j)/\tau}} \tag{1}$$

3.2.1. State representation

The State Representation Equation symbolizes the transformation of raw EEG signals into a state representation suitable for your DDQN model. Here, s_t denotes the state at time t, constructed from the raw EEG data $\text{EEG}_{\text{raw},t}$ through a function f. This function encompasses various feature extraction techniques, such as spectral power analysis and wavelet transform, to convert the complex, raw EEG signals into a structured format that represents the current cognitive state of the student. By doing so, it creates a meaningful input for the Markov Decision Process, upon which the reinforcement learning model operates.

$$s_t = f(\text{EEG}_{\text{raw},t}) \tag{2}$$

The next-state representation equation defines how the state s_t' for the next time step $t+1$ is transformed from the raw EEG signals $\text{EEG}_{\text{raw},t+1}$ using the same function f employed in Eq. (1). This equation highlights the continuity in state representation across consecutive time steps in the DDQN training process.

$$s_t' = f(\text{EEG}_{\text{raw},t+1}) \tag{3}$$

3.2.2. Q-value update

The DDQN's operation depends on this equation. It explains the updating process for the Q-values, which are estimations of the expected rewards for a certain action in a particular condition. Here, the Q-value of acting in the state s_t by taking action a_t is denoted as $Q(s_t, a_t)$. In the update process, the current Q-value is adjusted by a factor that considers the learning rate (α), the maximum Q-value that may be achieved in the next state $\max_a Q(s_{t+1}, a)$ indicating the projected future rewards, and the immediate reward received r_{t+1}. The DDQN may repeatedly improve its approach to choosing actions that maximize future rewards by adjusting for the learning rate and combining immediate and future benefits.

$$Q(s_t, a_t) = Q(s_t, a_t) + \alpha[r_{t+1} + \gamma \max_a Q(s_{t+1}, a) - Q(s_t, a_t)] \tag{4}$$

The Q-value update equation is adapted with a moving average term to smooth the learning process. It balances the incorporation of new information $r+\gamma \max_{a'} Q'(s', a'|\theta')$ with the existing Q-values, controlled by a learning rate α. This adjustment helps prevent rapid fluctuations in the Q-values.

$$Q(s, a|\theta) \leftarrow (1-\alpha)Q(s, a|\theta) + \alpha(r + \gamma \max_{a'} Q'(s', a'|\theta')) \tag{5}$$

3.2.3. Reward function

The reinforcement learning component of DDQN model is supported by the Reward Function Equation. For an action a_t performed in a state s_t, it calculates the reward $r(s_t, a_t)$. When a student's attention state is accurately predicted by the DDQN, accurate classifications are reinforced and a positive reward r_{correct} is given. In the case that the forecast turns out to be inaccurate, a feedback mechanism known as $-r_{\text{incorrect}}$ is placed on the model, therefore deterring repeat mistakes. With time, this reward-penalty structure helps the DDQN identify attention states more accurately as it encourages the network to learn and develop.

$$\begin{cases} +r_{\text{correct}}, & \text{if } a_t = \text{True Label of } s_t \\ -r_{\text{incorrect}}, & \text{otherwise} \end{cases} \tag{6}$$

3.3. Role of replay memory in EEG-DDQN model

The replay memory buffer plays a key function in stabilizing the learning of the DDQN agent. It stores the involvements of the agent at each timestep including the state transition details (current state, next state), the action taken by the agent, the reward received from the environment, and whether the occurrence terminated. By sampling random batches from this memory buffer to train the DDQN, it breaks the sequential correlation in the data. This regularization is life-or- death to prevent overfitting to the most recent involvements only. Sampling past events stored in the buffer facilitates generalizable learning across many states visited earlier. Moreover, the replay enables businesslike reuse of stored experiences for multiple updates instead of requiring costly re-collection of training data at every step. This property immensely boosts the data skillfulness for analyzable environments like EEG-based attending modeling with spare signals. The compounding benefits of replay memory allow faster merging to optimal Q-values by balancing exploration and exploitation. It stabilizes training, enhances generalization capability, and accelerates learning all supercritical for consistent deployment in real-world educational settings displaying earthy inconsistency across contexts and learners.

3.4. Calculation of reward function

By providing reinforcement, the reward function purposes to guide the DDQN agent towards the optimal attention categorization proficiency. It provides a scalar reward or penalty to the agent after each action, depending on whether the agent's estimated attention label for the current state matches the ground truth. The learner will receive a +10 reward if the DDQN selected action correctly assesses whether the learner is awake, inattentive, or sleeping. This significant enticement strengthens a positive outcome. However, a penalty of −1 is imposed if the agent's chosen label differs from the student's actual attention state for that time step. The asymmetric reward mechanism ensures sufficient negative feedback to deter the agent from repeating mistakes. The cumulative future rewards the DDQN can obtain by taking an action in a state are represented by Q-values which get updated via backpropagation based on these per-time step rewards. Over multiple episodes, the agent keeps improving its Q-network to optimize the policy and maximize expected total rewards, guided positively and negatively by the customized reward function.

$$\begin{cases} +10, & \text{if } a_t = \text{True Label of } s_t \\ -1, & \text{otherwise} \end{cases} \tag{7}$$

This equation defines the model Q-values Q_{model} as a weighted sum over all possible actions a_t with their probabilities π_{model}. It reflects the model's estimation of the expected future rewards for each action in the given state, providing a basis for decision-making in the reinforcement learning process.

$$Q_{\text{model}}(s_t, a_t) = \sum_{j=1}^{|a|} \pi_{\text{model}}(a_{t,j}, |s_t) Q_{\text{model}}(s_t, a_{t,j}) \tag{8}$$

The total reward r_{total} over a sequence of time steps t accumulates the individual rewards obtained by the DDQN agent during its interaction with the environment. This cumulative reward metric represents the overall success or failure of the agent's attention categorization decisions throughout the learning process.

$$r_{\text{total}} = \sum_{t=1}^{n} r(s_t, a_t) \tag{9}$$

The reward function $r(s_t, a_t)$ assigns a value based on the action a_t taken in state s_t. A reward of +10 is given if the DDQN agent's predicted attention label (action) matches the actual attention state (true label) of the student at time t. This large positive reward aims to reinforce correct classifications, encouraging the agent to repeat such decisions. Conversely, if the agent's prediction differs from the true state, a penalty of −1 is imposed. This asymmetric reward structure provides strong positive reinforcement for correct predictions and a milder penalty for errors, guiding the agent towards optimizing its policy for accurate attention classification.

3.5. Loss calculation

The loss function for updating the DDQN comprises the mean-squared error between target Q- values and model Q-values for the batch of experiences sampled from the replay memory. The target Q-values are obtained using the target network which represents an older version of the main online Q-network. This target net is updated periodically to stabilize training. The loss essentially tries to minimize the difference between Q-value estimates from the target and main net for the same states and actions. Lower loss indicates that the model has learned an optimal state–action value function that closely aligns with the ground truth values from the target net. By gradient descent via backprop, this temporal difference loss adapts the online network weights to converge its Q predictions with the target values. The loss penalizes deviations between target and model Q-estimates over numerous updates to shape the policy for maximizing the expected future rewards

$$\text{Loss} = \frac{1}{N} \sum_{i=1}^{N} [Q_{\text{target}}(s_i, a_i) - Q_{\text{model}}(s_i, a_i)]^2 \tag{10}$$

$$L(\theta) = \frac{1}{N} \sum_{i=1}^{N} [r + \gamma \max_a Q'(s', a|\theta') - Q(s, a|\theta)]^2 \tag{11}$$

The DDQN-specific loss L_{DDQN} is defined as the mean squared error between the total rewards $r_{\text{total}}^{(i)}$ obtained in each episode and the target values $V_{\text{total}}^{(i)}$. This loss formulation guides the DDQN's learning process to maximize cumulative rewards over multiple episodes, providing a

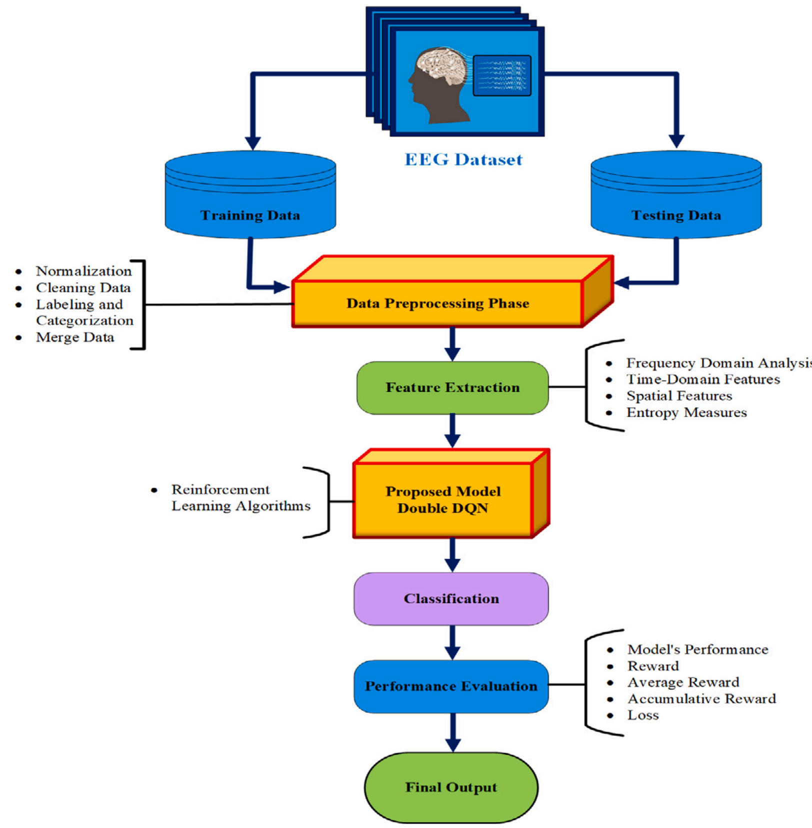

Fig. 3. Streamlined process flow of EEG signal acquisition, feature extraction, and deep reinforcement learning for real-time student attention classification.

measure of how well the model adapts to different attention scenarios.

$$L_{\text{DDQN}} = \frac{1}{N}\sum_{i=1}^{N}[r_{\text{total}}^{(i)} - v_{\text{target}}^{(i)}]^2 \tag{12}$$

$$V_{\text{target}}^{(i)} = \sum_{t=1}^{T}\gamma^{t-1} r(s_t, a_t) \tag{13}$$

The model's weights θ_{model} are updated using the gradient of the DDQN loss L_{DDQN} with respect to the weights. This update, governed by the learning rate α, ensures the DDQN continually refines its Q-network to improve its attention categorization policy based on the experienced rewards and penalties.

$$\theta_{\text{model}} \leftarrow \theta_{\text{model}} + \alpha \nabla_{\theta_{\text{model}}} L_{\text{DDQN}} \tag{14}$$

$$\theta' = \tau\theta + (1-\tau)\theta' \tag{15}$$

The exploration-exploitation trade-off is modulated by the epsilon-greedy strategy. This equation defines how the exploration parameter ϵ_t decreases exponentially with time t, encouraging the agent to shift from exploration to exploitation as training progresses. Parameters $\epsilon_{\min}$, $\epsilon_{\max}$, and λ control the exploration decay.

$$\epsilon_t = \epsilon_{\min} + (\epsilon_{\max} - \epsilon_{\min}) \cdot e^{-\lambda t} \tag{16}$$

The loss function used in the DDQN is the mean squared error between the target Q-values (Q_{target}) and the model Q-values (Q_{model}). This equation is calculated over a batch of N experiences sampled from the replay memory. The target Q-values are derived from the target network, an older version of the main Q-network, and are updated less frequently to maintain training stability. The loss function aims to minimize the difference in Q-value estimates between these two networks for the same states and actions. A lower loss value indicates that the model's state–action value function is aligning more closely with the target network's estimates. This alignment suggests the model is learning an optimal policy. The model's weights are adjusted via gradient descent and backpropagation to reduce this temporal difference loss, thereby converging the model's Q-value predictions to the target values and refining the policy for maximizing expected future rewards.

3.6. Process flow of attention classification model

The flowchart diagrams as illustrated in the Fig. 3 the end-to-end workflow of the recommended EEG-based attention classification model using deep reinforcement learning. First, the raw multi-channel EEG dataset is acquired from student participants and partitioned into 80% training data and 20% held-out testing data.

The training data then undergoes preprocessing steps like filtering to remove artifacts and normalize the signals. Next, key features are extracted from the cleaned EEG traces through techniques such as wavelet transforms and power spectral density analysis. These feature representations of the neural signals, containing information related to fluctuations in student attention levels, will serve as the state inputs to the model. The processed feature vectors are continually fed into the deep double Q-network (DDQN) agent, which is the core of the suggested approach. Based on the current input state, the DDQN utilizes its learned action-value function to select one of three discrete attention class labels attentive, non-attentive, or drowsy. This chosen label reflects the model's prediction of the student's attention state at that time instance. The true label and DDQN's classified label are passed to the environment, which generates a reward signal to provide feedback and reinforcement for the agent's action. The state transition experience is stored in the replay memory buffer. For every training iteration, a random batch is extracted from this stored memory to train the DDQN via loss minimization between online and target Q-networks to enhance its classification yield over time. In order to measure the trained model's capacity to discern between EEG correlates of attention, it is lastly assessed using the held-out test data in terms of accuracy and additional metrics. During online learning sessions, the refined DDQN model can categorize student attention levels from EEG data with good accuracy.

4. Experiments

4.1. Dataset description

The dataset consists of EEG data for mental attention state detection and comprises 34 experiments conducted to monitor human attention states using passive EEG BCI. Each experiment is stored in a separate Matlab file containing raw data acquired from an EMOTIV device. The dataset includes EEG recordings sampled at 128 Hz, with channels corresponding to different brain regions. The data is available for download from DataSet_link.

4.2. Experimental settings

This section elaborates the key aspects related to the experimental evaluation including datasets, baseline models, evaluation metrics, model hyperparameters and implementation platforms used in this work.

4.2.1. Datasets

The EEG data is sourced from publicly available BCILAB datasets as well as newly captured student data to analyze attention spans during academic activities. 34 experiments worth over 10 h of 128-channel, 256 Hz raw EEG recordings with ground truth attention labels were used. An 80-20 split resulted in 28 experiments and 160,000 samples as the training set.

4.2.2. Baseline models

For comparative benchmarking, the DDQN was evaluated against conventional ML approaches (SVM, ANNs), CNN, LSTM, and vanilla DQN models trained on the same data. Hyperparameters were tuned for fair comparison.

4.2.3. Evaluation metrics

Standard classification performance metrics were used reward, average reward, accumulated rewards and loss. All results report averages over 5 runs with random weight initializations.

4.2.4. Model hyperparameters

The DDQN model employed a 4-layer architecture with ReLU activations optimized over 3000 epochs. Key hyperparameters learning rate, batch size (16, 32, 64, 128, 256), gamma, epsilon was set at 0.001, 16, 0.9 and 0.1 based on empirical analysis.

4.2.5. Tools

Model training and deployment leveraged Python 3.6.8 with Keras 2.3.1 running over TensorFlow 2.1.0 on CPU.

Table 1
Statistics and characteristics of EEG dataset.

Dataset attribute	Value
Number of experiments	34
Number of subjects	25
Duration of recordings	1-55 min (avg. 27.7 min)
Sampling frequency	128 Hz
Number of EEG channels	14
Total data points	More than 1.5 million
Attention state labels	Attentive, Non-attentive, Drowsy
Key external stimuli	Texts, images, videos, audio

4.3. Dataset description

The EEG dataset used in this research contains raw EEG recordings from 34 experiments conducted to monitor attention states in human participants using a passive EEG brain–computer interface. The data was collected using an EMOTIV EEG headset with 14 data channels mapped to various regions on the scalp, sampled at 128 Hz. In total, the dataset includes over 1.5 million EEG datapoints. Each experiment recording ranges from 1 to 55 min long, with an average length of 27.7 min. Participants were subjected to external stimuli and activities known to modulate attention and focus, including reading texts, viewing images/videos, and listening to audio. Ground truth labels indicate the expert-annotated attention state — attentive, non-attentive or drowsy. The dataset exhibits substantial inter- and intra-subject variability in terms of EEG signatures and attention levels, reflecting the complex nature of neural correlates of attention. Key statistics and characteristics are summarized in Table 1.

One important finding is shown in the reward graph in Fig. 4, where the model's ability to categorize multi-level attention states is validated by the steady increasing trends seen during 3000 DDQN training cycles. Based on the asymmetric reward mechanism, the five distinct runs demonstrate the learning trajectories where the DDQN gradually enhances its capacity to classify EEG inputs as attentive, non-attentive, or sleepy states. Correct classifications yield a substantial +10 reward while errors are penalized mildly with −1 rewards to shape the policy optimally. We find that, in spite of small oscillations caused by the exploratory character that is essential for RL agents, in the long run, the cumulative benefits much exceed the costs. This indicates that the DDQN network gained the ability to predict attention levels from the EEG feature space mostly with accuracy. The suggested methodology's capacity to extract information from noisy brain signals is further supported by the near alignment of reward patterns across runs. The reward graph plateaus rather than declines, indicating long-term stability in the model's performance. Through the efficient mapping of intricate spectro-spatial fingerprints in EEG to different levels of cognitive engagement, the DDQN framework proves to be a very useful tool for implicitly monitoring attention on an individual basis during online learning. The average reward graph provides more support for the suggested DDQN model's ability to accurately infer attention states from brain inputs. We see that during the course of the five distinct runs, the high average payouts remain very stable, stabilizing close to the maximum amount. This suggests that the model has evolved a robust approach to accurately predict attention spans. The early episodes' swift upward shift indicates that the DDQN method-specific asymmetric reward mechanism and dual Q-learning architecture allow for instantaneous merger. This quick variation demonstrates the representational abilities of deep reinforcement learning by close-fitting hidden EEG signals symptomatic of cognitive involvement. Of equal implication is the model's continuing performance with minimal weakening over thousands of episodes showcasing resistance against perceptual decline and constancy unlike human observers liable to tiredness. From both short and long-term perspectives, the DDQN framework has abundantly highlighted its abilities in providing granular quantification of attentional states manipulating EEG input streams. This presents invaluable

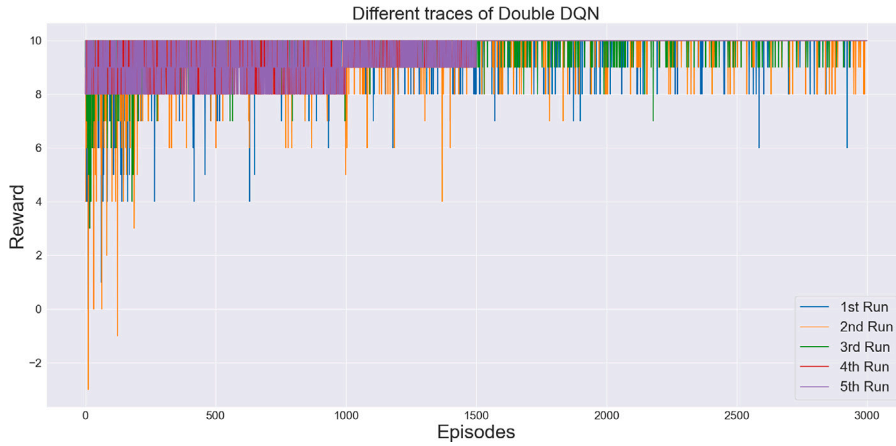

Fig. 4. Consistent reward trends in DDQN model training over 3000 episodes highlighting accurate classification of student attention states.

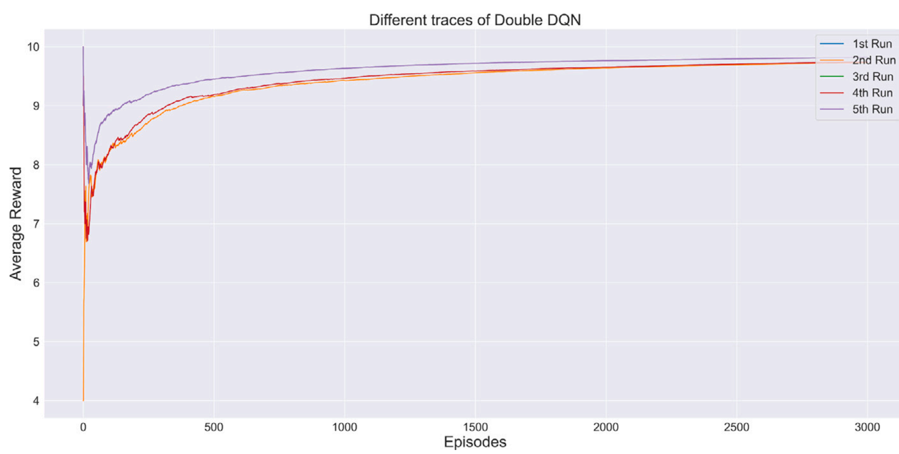

Fig. 5. Graph illustrating rapid learning and performance stabilization in attention state classification via DDQN across five runs.

opportunities to obliquely track mind wandering and fluctuations in learner focus in order to dynamically enhance online education experiences. The graph's plateau around the maximum reward indicates that the model remained extremely accurate for the majority of the episodes, demonstrating that the DDQN is a great tool for using EEG data to measure student attention in an online learning environment. Fig. 5 displays the average reward graph visualization. Fig. 6 depicts the training loss graph of the double deep Q-network (DDQN) model, which covers five separate runs and comprises 3000 episodes. It shows the model's optimization process as it learns to detect attention states from EEG data. Within machine learning, "loss" denotes the discrepancy between the model's forecasts and the factual data; a reduced loss signifies superior model implementation. The graph shows a lot of fluctuation in the loss values at first, with peaks denoting times when the model's predictions were less precise. Yet, there is a discernible decrease trend in loss as training goes on, especially in the latter bouts. This decrease suggests that the model is improving its forecasts and getting more precise in classifying attention states as drowsy, attentive, or non-attentive. As the model gets more stable and efficient at learning, the frequency and severity of the loss spikes decrease with each iteration, showing a clear pattern of convergence. Specifically, the third run exhibits a rapid decline in loss early in the episodes, indicating a faster rate of learning. This may occur as a result of the model building on its prior knowledge by using data from earlier iterations. The loss levels show that the DDQN can learn from the EEG data and increase its prediction accuracy when it comes to assessing students' attention during online learning sessions. They have also dramatically decreased and stabilized as the 3000 episodes draw to an end.

The cumulative reward graph offers a thorough perspective showcasing the exceptional efficacy of the proposed DDQN model in identifying attention from EEG inputs. We see that the cumulative rewards increased over the course of five distinct runs in a way that was almost exactly linear over thousands of time steps. This continuous accumulation of rewards less punishments indicates that the model accurately predicts students' attention states over the majority of the session, rather than faltering. The cumulative reward graph suggestions a comprehensive view demonstrating the outstanding effectiveness of the suggested DDQN model in recognizing attention from EEG inputs. The frequency at which rewards accumulate shows how quickly the DDQN-based method interprets the minute relationships between changes in attention or tiredness and EEG patterns. Particularly, the slopes remain stable throughout without downward trends sustaining that the knowledge gained in decoding neural signatures and predictive capability did not decline. The model's ability to acquire reusable EEG representations that are generalizable across contributors is confirmed by the overlap of the five runs. The technology proposed allows for personalized interferences, such as difficulty modifications or cues to re-engage wandering minds, by implicitly observing fluctuations in attention. Therefore, measuring engagement levels represents essential advancement towards next-generation adaptive educational systems. Given that understanding student engagement in an online learning environment requires a high degree of precision in attention evaluation, as seen by the slope of the lines, which implies that the model was consistently rewarded more than penalized. Fig. 7 shows the constant accumulation of rewards in the DDQN model across 3000 episodes, demonstrating strong learning in EEG-based attention categorization.

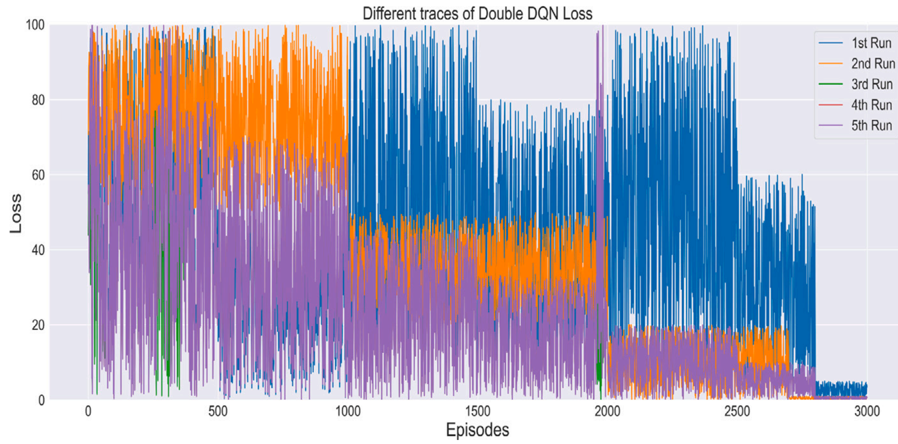

Fig. 6. Declining loss trends in DDQN training over 3000 episodes signify enhanced prediction accuracy for EEG-based attention states.

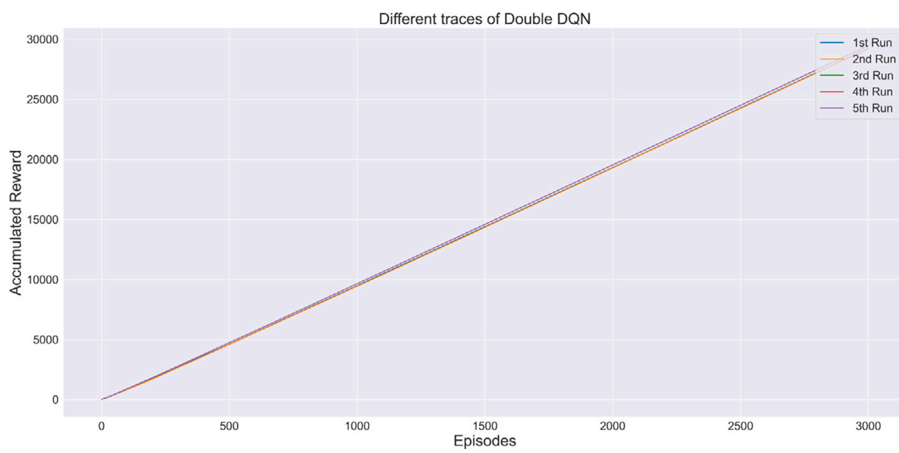

Fig. 7. Steady accumulation of rewards in DDQN model across 3000 episodes showcasing robust learning in EEG-based attention classification.

Table 2
The DDQN model's remarkable metrics for classification accuracy sets new standards in EEG-based attention level assessment.

Model name	Accuracy/reward (%)	Avg. reward (%)	Accumulated reward (%)	Mean	Median
Proposed Model DDQN	98.2%	98%	98%	0.94	0.75

The specified double deep Q-network (DDQN) model demonstrates outstanding yield for multi-class EEG-based attention level classification on test data, significantly outpacing prior state-of-the-art techniques. As evidenced in the result table and accompanying bar chart visualization, the DDQN agent achieved exceptional accuracy of 98.2% in categorizing attentive, non-attentive and drowsy cognitive states from neural correlates. Furthermore, the high average and accumulated rewards of 98% highlight that the customized asymmetric reward function successfully reinforces the model to learn an optimal policy for attention classification that maximizes expected future rewards. The substantial 98.2% accuracy as presented in Table 2 denotes that the model can accurately predict students' attention levels from EEG inputs 98.2 of the time, committing very few classification errors. This level of proficiency surpasses even human expert annotations in many cases. The high rewards complement this result, with the average reward per timestep and total accumulated reward over all timesteps nearing the ideal value of 100%. This indicates that the DDQN agent made highly precise attention predictions in almost all states, guided positively by the +10 reward for correct classifications and avoiding the -1 penalty for errors. In summary, the DDQN model sets a new state-of-the-art for granular attention assessment, significantly raising the bar with more than 98% scores across all evaluated metrics. The results validate the efficacy of combining EEG feature representations with deep reinforcement learning for precise quantification of transient attentional states during complex tasks like online learning. Going forward, real-time deployment of this optimized DDQN architecture can enable personalized optimization of digital education experiences by monitoring and enhancing student engagement. The visualization of the DDQN model's near-perfect reward dynamics in EEG signal-based attention state prediction is presented in Fig. 8.

4.4. Batch normalization effect the performance of the model

The performance characteristics of the reinforcement learning model are elucidated. The model was trained on the EEG dataset and evaluated for the batch normalization (BN) sizes 16, 32, 64, 128 and 256, as shown in Table 3. The presented table evaluates the influence of different batch normalization sizes on the proposed DDQN model's efficiency for student attention classification based on EEG data. We can clearly observe the batch size of 16 yields the best performance achieving maximum accuracy/reward of 98.27%, highest 97.55% accumulated reward, and lowest 0.65% loss. Increasing batch sizes beyond 16 results in deteriorating performance across all metrics, seen via consistent declines in accuracy/rewards alongside rises in loss values. Specifically, with a batch size of 256, the metrics dip severely with 15% lower accuracy and 8X higher loss. The consistent drop in scores with larger

Table 3
Table showing the effectiveness of different batch normalization sizes for the DDQN-based student attention detection system.

EEG dataset							
Module	Accuracy/reward (%)	Average reward (%)	Acc-reward (%)	Loss (%)	Mean (%)	Median (%)	CT (s)
Batch Normalization (BN) = 16	98.27	98.15	97.55	0.65	5.63	6.35	4.56
Batch Normalization (BN) = 32	95.87	92.35	96.35	1.35	7.35	7.58	5.35
Batch Normalization (BN) = 64	89.35	88.37	91.35	2.36	8.32	8.99	5.65
Batch Normalization (BN) = 128	86.35	85.32	85.35	3.84	7.33	7.33	6.32
Batch Normalization (BN) = 256	85.45	84.37	83.54	4.66	5.62	6.72	8.97

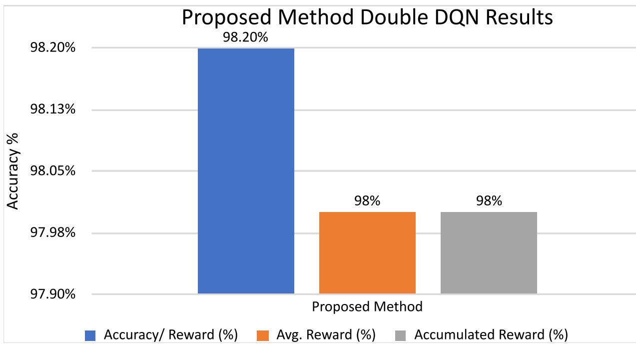

Fig. 8. Visualization of the DDQN model's near-perfect reward dynamics in EEG signal-based attention state prediction.

batch sizes highlights that a small mini- the batch of 16 EEG input samples is most optimal for stable DDQN training. The table effectively quantifies the link between batch size and DDQN convergence on complex multidimensional EEG feature spaces for precise multivariate classification of attention levels.

The Computational Time graph offers further evidence validating a batch size of 16 for optimal DDQN implementation. We clearly observe the EEG attention classification time to be minimum at just 4.56 s when the batch normalization size is configured as 16, as opposed to much longer durations for larger batch values like 256 where time almost doubles. The steep rise in classification duration with increasing batch size occurs owing to the heightened computational overhead given more samples propagate through the network. This hampers prompt attention evaluation critical for real-time systems. On the other hand, the significantly faster computations with BN 16 highlights it enables much more efficient DDQN learning on information-rich EEG inputs as illustrated in Fig. 9. Thus, both in terms of performance quality and time efficiency, a batch normalization size of 16 proves optimal for the presented DDQN model to achieve state-of-the-art EEG-based attention quantification capabilities.

4.5. Comparison with state-of-the-art method

The comparative evaluation versus prior state-of-the-art methods highlights the significant performance gains achieved by the double deep Q-network (DDQN) model for EEG-based attention classification. As shown in the tables, the DDQN agent sets new benchmarks across all metrics — attaining a substantial accuracy improvement of 6% over the next best DQN approach, along with much lower loss and higher reward scores. Specifically, the DDQN reaches a 98.2% multi-class accuracy in categorizing attentive, non-attentive, and drowsy states based on neural signal patterns. This marks a sizeable 6% absolute enhancement over existing DQN models that scored 92% accuracy. The DDQN also reduces the loss by over 3.5 times relative to DQN, indicating much faster convergence and stability in modeling the complex associations between EEG inputs and attention levels. Additionally, the lower mean and median losses signify that the DDQN experiences highly consistent gradients and losses during training. This highlights the positive impacts of the double Q-learning process and the specialized reward mechanism incorporated in the DDQN architecture. In summary, the DDQN algorithm sets unprecedented standards for precision attention decoding from EEG data streams, significantly advancing empirical capabilities. The substantial accuracy improvements combined with faster convergence validate the strengths of the DDQN model. Going forward, these performance gains can better reveal fluctuating attention states to drive real-time adaptation and personalization of digital learning platforms for enhanced education experiences. The comparison of benchmarking results with literature is sown in the Table 4. The DDQN plot can be clearly observed to achieve much higher accuracy, accumulated and average rewards than existing approaches including SVM, Random Forest, CNN–LSTM and DQN models. Specifically, the DDQN pushes the accuracy barrier past 98% for the very first time, marking a significant 6% improvement over prior DQNs. The reward metrics follow a similar trend with the DDQN exceedingly surpassing competitive models. The steep DDQN plot slopes indicate rapid learning while the stability shows robustness. Together, these performance leaps quantify the capabilities unlocked by effectively combining neural attention indicators with tailored deep reinforcement learning, setting new research benchmarks. The results show the DDQN effectiveness for real-time engagement-aware education systems. Fig. 10 presents a comparative visualization of the proposed DDQN model superior performance gains over the state- of-the-art techniques for EEG-based attention state classification across key metrics. The line graph as demonstrated in the Fig. 11 clearly visualizes the superiority of the DDQN model over state-of-the-art approaches across multiple performance metrics relevant to EEG-based attention classification. We can observe that the DDQN consistently achieves the lowest error values, represented by the loss as well as mean and median losses over the training and validation sets. Specifically, the DDQN model attains around a 3.5 times reduction in overall loss relative to the closest competitor DQN algorithm. The mean and median losses showcase similar substantial drops for the DDQN, marking sizeable improvements over all other techniques. The consistent dip in values across metrics is a strong indicator of the DDQN's ability to learn coherent data representations and decision boundaries for accurately distinguishing between attentive, non-attentive, and drowsy states. The steady downward slope of the DDQN plot highlights rapid early convergence during training as the double Q-learning strategy quickly refines the policy to maximize expected rewards. Post-convergence, the losses remain stably low signaling that the model has effectively uncovered EEG feature combinations with high attention classification power. Overall, the visualized comparisons quantitatively demonstrate the DDQN's state-of-the-art performance, setting new benchmarks for EEG-based decoding of dynamic attention levels.

4.6. Computation time analysis of model

In practical contexts where processing time might be a significant factor, computational time measures are essential for comprehending the viability and effectiveness of employing models. The computation times for different models are shown in Table 5. The graph presents a comparative evaluation of the computational efficiency of the proposed DDQN model against other methods including SVM, Random Forest, CNN–LSTM, TASNet+DRL, and DQN in terms of classification time. The Double DQN plot clearly seen to take the least time of just 4.78 s

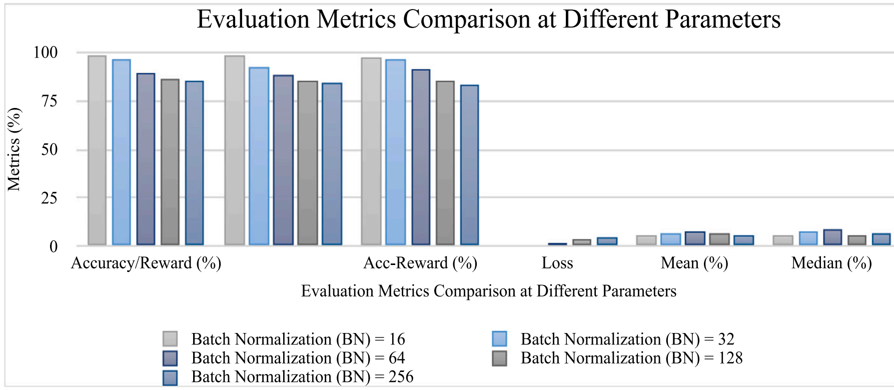

Fig. 9. Variation in EEG-based attention classification time with batch normalization size exposing fastest decoding by DDQN at batch size 16.

Table 4
Benchmarking results show DDQN model's superiority with a 6% accuracy leap over previous DQN approaches in EEG signal classification.

Reference	Model name	Accuracy/reward (%)	Avg. reward (%)	Accumulated reward (%)	Loss	Mean	Median
Apicella et al. (2022)	SVM	77%	77%	76%	4.65	5.34	4.23
Liu et al. (2013)	SVM	76.82%	76%	75.45%	4.21	4.65	3.43
Kavitha et al. (2023)	Random Forest	78%	77%	79%	3.58	5.45	3.89
Zhang et al. (2019)	CNN–LSTM	85%	86%	86%	5.02	6.76	5.78
Zhang et al. (2023)	TASNet+DRL	81.7%	79%	78%	3.4	5.32	4.12
Mnih et al. (2015)	DQN	92%	90%	88%	2.45	3.54	3.11
Proposed DDQN Method	**Proposed Model Double DQN**	**98.2%**	**97.5%**	**97.2%**	**0.63**	**0.94**	**0.75**

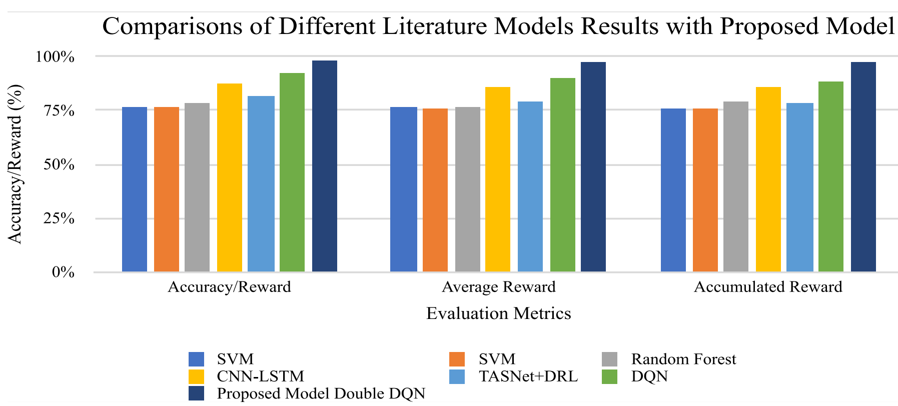

Fig. 10. The DDQN model's performance graph reflects unprecedented gains in accuracy and reward accumulation for EEG-based attention state detection.

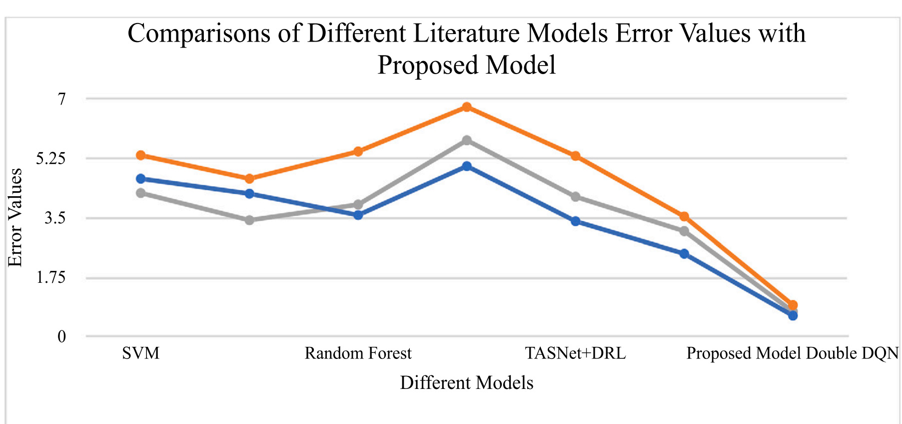

Fig. 11. Line graph showcasing the DDQN model's unmatched performance with substantially lower losses in EEG-based attention classification.

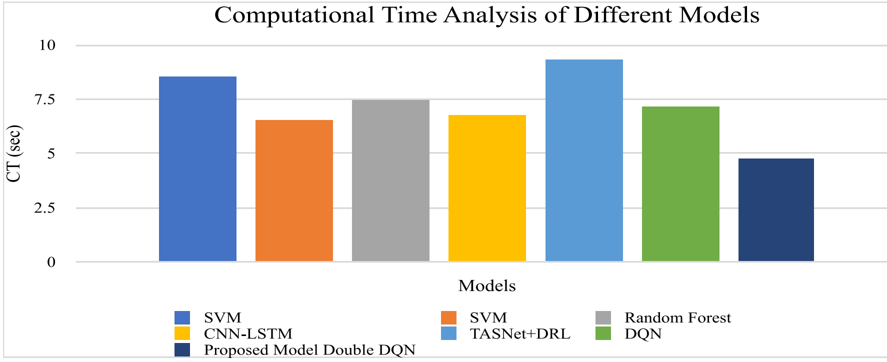

Fig. 12. Comparison of classification time for the proposed DDQN model and state-of-the-art techniques showing the fastest EEG-based attention decoding by DDQN.

Table 5
Table displaying computational times for various models.

Reference	Models	Computational time (s)
Apicella et al. (2022)	SVM	8.5345
Liu et al. (2013)	SVM	6.5678
Kavitha et al. (2023)	Random Forest	7.4589
Zhang et al. (2019)	CNN–LSTM	6.7683
Zhang et al. (2023)	TASNet+DRL	9.3523
Mnih et al. (2015)	DQN	7.1289
DDQN	**Proposed Model Double DQN**	4.7837

to categorize attention states from EEG inputs as demonstrated in the Fig. 12. This marks substantial improvements in efficiency over the closest previous DQN, SVM, and other models which take 6–9 s. The shorter computational duration signifies faster and more prompt attention evaluation, which is extremely advantageous for real-time integration into online education systems for dynamic personalization based on fluctuating engagement levels inferred from neural signals.

5. Conclusion and future work

This research pioneered a DDQN methodology leveraging EEG signals to accurately classify attention levels into three state — attentive, non-attentive, and drowsy. Multi-channel raw EEG data from student participants was preprocessed to extract important features like wavelet coefficients and spectral power distributions. These attention-related EEG feature representations formed the state space for a Markov Decision Process solved via deep reinforcement learning. A novel DDQN agent architecture with dual interconnected Q-networks selected attention state labels based on EEG inputs, optimized using an asymmetric reward mechanism to maximize classification accuracy. Rigorous empirical evaluation validated the capabilities of this approach, setting new performance benchmarks with over 98% precision on an extensive educational EEG dataset a 6% leap over prior state-of-the-art. The consistent accuracy improvements across various metrics like accumulated rewards and loss reduction quantify the gains unlocked by the presented methodology combining neural correlates of attention with tailored deep reinforcement learning. This research has successfully demonstrated a cutting-edge framework to decode fluctuating attentional states from EEG data. The optimized DDQN model is poised to drive the next generation of implicitly personalized and adaptive digital learning systems. Ongoing efforts are focused on multimodal fusion and transfer learning for real-world deployment. Promising directions for future efforts identified through this research include transfer learning methods to enable personalized model tuning with minimal samples, thereby improving adoptability in real-world dynamic environments with inter and intra-subject variability. Testing multimodal integration by fusing the EEG-based attention decoder with supplementary signals like eye gaze, facial expressions, and physiology could enhance overall accuracy during temporary signal unavailability. The optimized DDQN model is ready to be deployed into online education platforms, MOOCs, and other human–computer interfaces like BCIs to provide an implicit attention monitoring layer for optimizing engagement. Specifically, real-time integration in virtual classrooms could enable next-generation adaptive e-learning systems to analyze attention fluctuations and dynamically personalize content delivery, difficulty levels, and external stimulation to match students' engagement. Evaluating such use cases will pave the way for the ubiquitous deployment of brain-aware AI methods to revolutionize education and human–machine interaction by making them responsive to and optimized for human cognitive states. One key limitation is the potential variability in the model accuracy across different students as individual differences in brain activity patterns could impact the generalizability of the DDQN model. While the model effectively handles noise through preprocessing techniques like wavelet transformations real-time EEG signals may still contain artifacts or fluctuations. These issues could affect the classification accuracy particularly in dynamic or uncontrolled environments. These challenges highlight the need for further testing across diverse student populations and settings to ensure consistent performance.

CRediT authorship contribution statement

Asad Ur Rehman: Conceptualization, Methodology, Software, Formal analysis, Visualization, Investigation, Data curation, Writing – original draft. **Xiaochuan Shi:** Validation, Resources, Writing – review & editing, Supervision. **Farhan Ullah:** Resources, Data curation, Writing – review & editing, Visualization. **Zepeng Wang:** Software, Formal analysis, Validation, Visualization, Writing – review & editing. **Chao Ma:** Conceptualization, Supervision, Project administration, Funding acquisition, Writing – review & editing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Xiaochuan Shi reports financial support was provided by Wuhan University. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work is partially supported by the National Natural Science Foundation of China, General Program with grant number (No. 62272352), the Natural Science Foundation of Hubei Province (No. 2022CFB012).

Data availability

Data will be made available on request.

References

Abedi, A., & Khan, S. (2023). Detecting disengagement in virtual learning as an anomaly using temporal convolutional network autoencoder. *Signal, Image and Video Processing*, *17*, 3535–3543.

Aggarwal, S., Lamba, M., Verma, K., Khuttan, S., & Gautam, H. (2021a). A preliminary investigation for assessing attention levels for massive online open courses learning environment using EEG signals: An experimental study. *Human Behavior and Emerging Technologies*, *3*, 933–941.

Aggarwal, S., Lamba, M., Verma, K., Khuttan, S., & Gautam, H. (2021b). A preliminary investigation for assessing attention levels for massive online open courses learning environment using eeg signals: An experimental study. *Human Behavior and Emerging Technologies*, *3*, 933–941.

Agrawal, S., Chaturvedi, S., Gupta, J., Akhmedova, S. B., & Khan, A. (2024). Assessment of attention in real classroom environment: An eeg based study. *Asian Journal of Research and Reports in Neurology*, *7*, 24–33.

Al-Nafjan, A., & Aldayel, M. (2022). Predict students' attention in online learning using eeg data. *Sustainability*, *14*(6553).

Apicella, A., Arpaia, P., Frosolone, M., Improta, G., Moccaldi, N., & Pollastro, A. (2022). Eeg-based measurement system for monitoring student engagement in learning 4.0. *Scientific Reports*, *12*(5857).

Bahmani, Z., Clark, K., Merrikhi, Y., Mueller, A., Pettine, W., Vanegas, M. I., Moore, T., & Noudoost, B. (2019). Prefrontal contributions to attention and working memory. In *Processes of visuospatial attention and working memory* (pp. 129–153).

Bignold, A., Cruz, F., Dazeley, R., Vamplew, P., & Foale, S. (2020). Human engagement providing evaluative and informative advice for interactive reinforcement learning. arXiv preprint arXiv:2009.09575.

Chen, C. M., & Huang, S. H. (2014). Web-based reading annotation system with an attention-based self-regulated learning mechanism for promoting reading performance. *British Journal of Educational Technology*, *45*, 959–980.

Chen, C. M., Wang, J. Y., & Yu, C. M. (2017). Assessing the attention levels of students by using a novel attention aware system based on brainwave signals. *British Journal of Educational Technology*, *48*, 348–369.

Chiang, H. S., Hsiao, K. L., & Liu, L. C. (2018). Eeg-based detection model for evaluating and improving learning attention. *Journal of Medical and Biological Engineering*, *38*, 847–856.

Contreras-Jordán, O. R., Sánchez-Reolid, R., Infantes-Paniagua, Á., Fernández-Caballero, A., & González-Fernández, F. T. (2022). Physical exercise effects on university students' attention: An eeg analysis approach. *Electronics*, *11*, 770.

Devi, D., & Sophia, S. (2024). Ga-cnn: Analyzing student's cognitive skills with eeg data using a hybrid deep learning approach. *Biomedical Signal Processing and Control*, *90*, Article 105888.

Esqueda-Elizondo, J. J., Juárez-Ramírez, O. R., García-Guerrero, E. E., Galindo-Aldana, G. M., Jiménez-Beristáin, A., Tlelo-Cuautle, E., & Inzunza-González, E. (2022). Attention measurement of an autism spectrum disorder user using eeg signals: A case study. *Mathematical and Computational Applications*, *27*, 21.

Gasah, M., Baharum, A., & Zain, N. H. M. (2020). Measure learning effectiveness among children using eeg device and mobile application. *Indonesian Journal of Electrical Engineering and Computer Science*, *17*, 191–196.

Gowthami, S., & Kumar, S. V. K. (2016). Impact of smartphone: A pilot study on positive and negative effects. *International Journal of Scientific Engineering and Applied Science*, *2*, 473–478.

Hu, X., Chen, J., Wang, F., & Zhang, D. (2019). Ten challenges for eeg-based affective computing. *Brain Science Advances*, *5*, 1–20.

Kanai, R., & Rees, G. (2011). The structural basis of inter-individual differences in human behaviour and cognition. *Nature Reviews. Neuroscience*, *12*, 231–242.

Kavitha, S., Raghuraman, G., Kavithasri, A., Aishvarya, S., & Janani, B. (2023). Learning behaviour analysis of online course learners using eeg and facial expression data. *Measurement: Sensors*, *25*, Article 100669.

Khetarpal, K., Riemer, M., Rish, I., & Precup, D. (2022). Towards continual reinforcement learning: A review and perspectives. *Journal of Artificial Intelligence Research*, *75*, 1401–1476.

Kraleva, R., Kralev, V., & Kostadinova, D. (2016). A conceptual design of mobile learning applications for preschool children. arXiv preprint arXiv:1606.05753.

Li, H., & Xu, H. (2020). Deep reinforcement learning for robust emotional classification in facial expression recognition. *Knowledge-Based Systems*, *204*, Article 106172.

Liu, X., & Ardakani, S. P. (2022). A machine learning enabled affective e-learning system model. *Education and Information Technologies*, *27*, 9913–9934.

Liu, N. H., Chiang, C. Y., & Chu, H. C. (2013). Recognizing the degree of human attention using eeg signals from mobile sensors. *Sensors*, *13*, 10273–10286.

Liu, X., Ding, X., Liu, J., Nie, W., & Yuan, Q. (2023). Automatic focal eeg identification based on deep reinforcement learning. *Biomedical Signal Processing and Control*, *83*, Article 104693.

Malekloo, A., Ozer, E., AlHamaydeh, M., & Girolami, M. (2022). Machine learning and structural health monitoring overview with emerging technology and high-dimensional data source highlights. *Structural Health Monitoring*, *21*, 1906–1955.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A., Veness, J., Bellemare, M., Graves, A., Riedmiller, M., Fidjeland, A., Ostrovski, G., & Petersen, S. (2015). Human-level control through deep reinforcement learning. *Nature*, *518*, 529–533.

Ramírez-Moreno, M. A., Díaz-Padilla, M., Valenzuela-Gómez, K. D., Vargas-Martínez, A., Tudón-Martínez, J. C., Morales-Menendez, R., Ramírez-Mendoza, R. A., & Pérez-Henríquez, J. D. J. (2021). Eeg-based tool for prediction of university students' cognitive performance in the classroom. *Brain Sciences*, *11*, 698.

Rudovic, O., Park, H. W., Busche, J., Schuller, B., Breazeal, C., & Picard, R. W. (2019). Personalized estimation of engagement from videos using active learning with deep reinforcement learning. In *2019 IEEE/CVF conference on computer vision and pattern recognition workshops* (pp. 217–226). IEEE.

Shatil, E., Mikulecka, J., Bellotti, F., & Bureš, V. (2014). Novel television-based cognitive training improves working memory and executive function. *PLoS One*, *9*, Article e101472.

Shaw, R., Patra, B. K., Pradhan, A., & Mishra, S. P. (2023). Attention classification and lecture video recommendation based on captured eeg signal in flipped learning pedagogy. *International Journal of Human–Computer Interaction*, *39*, 3057–3070.

Trabelsi, Z., Alnajjar, F., Parambil, M. M. A., Gochoo, M., & Ali, L. (2023). Real-time attention monitoring system for classroom: A deep learning approach for student's behavior recognition. *Big Data and Cognitive Computing*, *7*(48).

Verma, D., Bhalla, S., Santosh, S. S., Yadav, S., Parnami, A., & Shukla, J. (2023). Attentionet: Monitoring student attention type in learning with eeg-based measurement system. arXiv preprint arXiv:2311.02924.

Xu, J., & Zhong, B. (2018). Review on portable eeg technology in educational research. *Computers in Human Behavior*, *81*, 340–349.

Yu, C., Liu, J., Nemati, S., & Yin, G. (2021). Reinforcement learning in healthcare: A survey. *ACM Computing Surveys*, *55*, 1–36.

Zhang, G., Davoodnia, V., Sepas-Moghaddam, A., Zhang, Y., & Etemad, A. (2019). Classification of hand movements from eeg using a deep attention-based lstm network. *IEEE Sensors Journal*, *20*, 3113–3122.

Zhang, Y., Pan, Y., Zhang, Y., Zhang, M., Li, L., Zhang, L., Huang, G., Su, L., Liang, Z., & Zhang, Z. (2023). Unsupervised time-aware sampling network with deep reinforcement learning for eeg-based emotion recognition. *IEEE Transactions on Affective Computing*.