Future Technology
Open Access Journal
ISSN 2832-0379
November 2025| Volume 04 | Issue 04 | Pages 159-172
Journal homepage: https://fupubco.com/futech
https://doi.org/10.55670/fpll.futech.4.4.14
Future Publishing LLC

Article

Research on a virtual teacher personalized interaction model integrating affective computing and multi-agent systems

Rili Dang[1,2], Noorazman Abd Samad[1*]

[1]Universiti Tun Hussein Onn Malaysia (UTHM), Panchor 84600, Johor, Malaysia
[2]Zhuhai City Polytechnic, Zhuhai 519000, Guangdong, China

ARTICLE INFO

Article history:
Received 22 June 2025
Received in revised form
29 July 2025
Accepted 18 August 2025

Keywords:
Virtual teacher, Affective computing, Multi-agent systems, Personalized learning, Intelligent teaching systems

*Corresponding author
Email address:
noorazman@uthm.edu.my

DOI: 10.55670/fpll.futech.4.4.14

ABSTRACT

This research develops a novel virtual teacher personalized interaction model integrating multimodal affective computing with multi-agent coordination mechanisms to address fundamental limitations in emotional intelligence and adaptive capabilities within contemporary educational technology systems. A three-layer distributed architecture was implemented, incorporating synchronized multimodal emotion recognition through confidence-weighted fusion of facial, vocal, and textual data streams, Byzantine Fault Tolerant consensus algorithms for coordinated multi-agent decision-making, and dynamic personality adaptation mechanisms based on Big Five psychological modeling. Experimental validation employed 500 participants across diverse educational contexts using established emotion recognition benchmarks supplemented with domain-specific educational interaction datasets. The multimodal emotion fusion component achieved 91.2% recognition accuracy, with overall system performance reaching 89.7% under realistic educational conditions while demonstrating substantial educational effectiveness improvements, including 43% higher learner engagement scores, 37% emotional satisfaction enhancement, 30% learning effectiveness increase, and 40% knowledge retention improvement compared to traditional virtual teaching approaches. Multi-agent coordination exhibited superior decision quality with 31% improvement over single-agent baselines, though personality adaptation effectiveness varied significantly across learner populations with 88% success rates for extraverted individuals compared to 65% for high-neuroticism learners. The integrated approach successfully bridges the emotional intelligence gap in virtual educational systems through sophisticated technological convergence, establishing theoretical foundations for distributed educational intelligence while revealing important implementation challenges. This research enables the development of emotionally responsive virtual teachers capable of sustained personalized instruction across diverse educational contexts, though deployment requires careful consideration of privacy protection and institutional adaptation requirements for broader educational technology transformation.

1. Introduction

Contemporary educational technologies exhibit substantial limitations in identifying and responding to learner emotional states, creating a critical gap for personalized learning interventions in post-pandemic adaptive educational systems [1]. Affective computing technologies offer opportunities to address this gap, though current solutions remain fragmented and lack comprehensive emotional intelligence integration [2]. Social-emotional learning technologies show promise but lack integrated intelligent operations for holistic responses to diverse learner demands [3]. Virtual human technologies demonstrate potential for human-like educational interactions while revealing adaptation challenges for avatar-based learning systems [4]. Intelligent educational systems have progressed through deep learning algorithms, multi-agent coordination, and advanced human-computer interaction paradigms. Deep learning systems have enhanced multimodal pattern recognition for interpreting learner emotional states through facial, vocal, and textual analysis [5]. Neural network developments in human-computer interaction systems improve recognition accuracy and response appropriateness, particularly where emotional subtlety affects learning efficiency [6]. Multi-agent architectural models provide powerful paradigms for orchestrating complex educational interactions through specialized agents addressing different learning facets [7]. Reviews reveal gaps in achieving personalized emotional reactivity across educational contexts despite distributed intelligence architectures [8], with K-12 systems requiring advanced emotion integration for individual profiles [9]. The sustainability considerations of AI deployment in education emphasize the necessity for systems to adapt and evolve to meet different pedagogical demands, yet to maintain a stable ability to offer emotional support across diverse learning situations and cultural contexts [10]. Machine learning methods for predicting individual learning styles require integration of cognitive and affective models [11]. Empathic conversational agents demonstrate effectiveness but face limitations in personalized emotional responsiveness [12]. Current chatbots lack multi-dimensional personality modeling [13] and advanced emotional intelligence [14], while embodied agents rely on rigid personality models without individualized adaptation [15].

Recent studies reveal theoretical and practical limitations in integrating affective computing with multi-agent coordination mechanisms. Immersive learning environments demonstrate the necessity of affective computing integration, though significant challenges exist in simultaneous intelligent function implementation [16]. Affective intelligent teaching systems show promise in detecting and responding to learner emotions, though pedagogical efficacy challenges persist across domain-specific agents [17]. AI-based fast development frameworks for intelligent teaching systems mark some progress toward emotion-aware educational technology, yet even these systems are confronted with the problem of coordinating with multiple intelligent agents to appropriately respond to complex emotional states in diverse educational contexts [18]. Avatar-based systems demonstrate the importance of visual representation for emotional involvement [19] , while scaffolding agents show gains when enriched with emotional intelligence and adaptation mechanisms [20]. AI educational models require validated psychological constructs for interventions targeting both cognitive and affective learning dimensions [21]. Cognitive neuropsychology perspectives indicate that robust integration of cognitive models and affective computing enhances educational effectiveness and learner satisfaction [22].

Current virtual educational systems face critical limitations, including insufficient multimodal emotion recognition accuracy under real-world conditions, a lack of coherent multi-agent coordination mechanisms, and inadequate personality adaptation frameworks balancing consistency with flexibility. This investigation aims to develop comprehensive multimodal emotion recognition, design distributed multi-agent coordination mechanisms, implement adaptive personality modeling, and validate educational effectiveness across diverse learner populations. The key innovations include confidence-weighted multimodal fusion with real-time quality assessment, modified Byzantine Fault Tolerant consensus for educational contexts, and regularized personality adaptation balancing character consistency with behavioral flexibility.

2. Methodology

2.1 System architecture design

The virtual teacher prototype proposed avoids problems of emotionally intelligent pedagogical systems with a three-layer distributed structure allowing real-time multimodal emotion recognition, coordinated multi-agent decision-making, and personality modeling adaptivity. Scalability is managed with layering problems of data acquisition, processing, and presentation in a manner where individual layer optimizations can be conducted with system consistency being preserved, along with complexity-maintainability trade-offs eliminated for real-time reactivity in pedagogy. The architecture overcomes isolated emotion recognition limitations through integrated processing pipelines, maintaining temporal coherence across multiple data streams. The data acquisition layer incorporates synchronized RGB-D cameras (30fps), omnidirectional microphone arrays (48kHz), and natural language processing modules for real-time multimodal emotion analysis. The distributed processing employs an edge-cloud hybrid configuration with local processing handling time-sensitive emotion recognition, while cloud services manage personality adaptation algorithms and learning analytics. This approach addresses cloud-only latency issues and edge-only computational constraints for complex personality modeling. Figure 1 illustrates the system architecture displaying linkages among data acquisition, processing engines, and decision coordination mechanisms. Figure 1 depicts a hierarchical processing architecture demonstrating data flow among acquisition modules, processing engines, and decision coordination mechanisms. This three-tier structure preserves the real-time responsiveness of the system through parallel processing channels and guarantees data integrity through synchronized communication protocols. Due to its modularization, parts of the system can be optimized individually, and coherence between system components can be maintained by means of standardized interface protocols, supporting synchronous and asynchronous communication patterns, according to computational needs and time-dependent constraints.

2.2 Multimodal affective computing model

The multimodal emotion recognition framework addresses real-time emotional state interpretation in educational interactions through heterogeneous data stream processing. Conventional unimodal systems demonstrate limited reliability due to environmental degradation affecting pedagogical effectiveness, while multimodal fusion leverages complementary information to ensure recognition robustness across different scenarios. The system addresses emotion recognition ambiguity through sophisticated fusion strategies exploiting complementary facial expressions, vocal patterns, and linguistic content to achieve robust emotion estimation under adverse conditions, including partial occlusion, background noise, and communication problems. The confidence-weighted fusion automatically adapts modality contributions in real-time according to input signal quality, preventing unreliable modalities from corrupting final emotion estimation. The facial emotion recognition module implements a modified EfficientNet-B4 architecture enhanced with spatial attention mechanisms and temporal convolutional networks, processing 224×224 pixel facial regions through real-time detection and landmark localization.

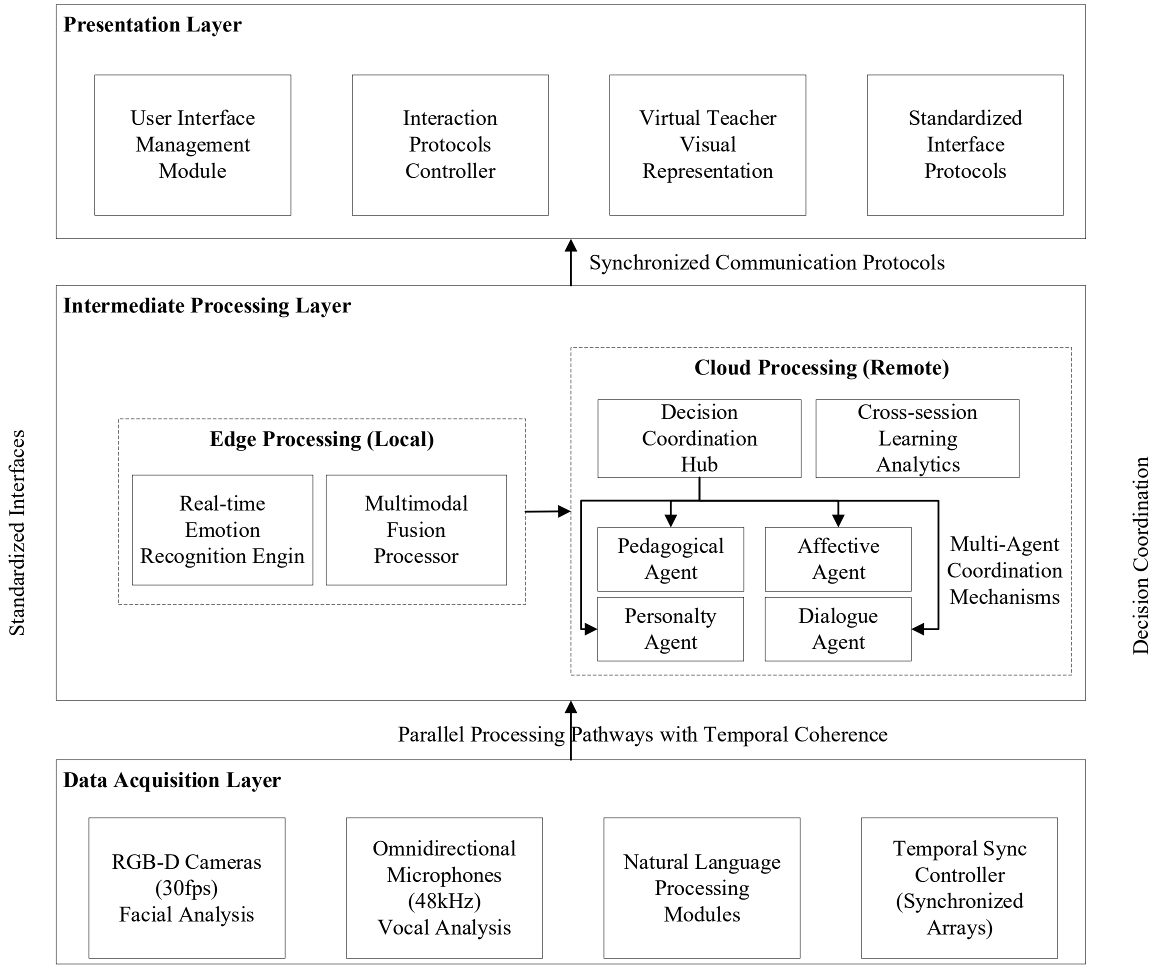

Figure 1. Hierarchical three-layer system architecture

The vocal emotion analysis employs hybrid Wav2Vec 2.0 feature extraction with bidirectional LSTM networks to process audio waveforms and extract emotional characteristics from prosodic features, spectral content, and temporal dynamics. The mathematical foundation for emotion fusion addresses the critical challenge of optimal information integration across modalities with varying reliability and temporal characteristics. The emotion state estimation at time t employs a confidence-weighted fusion mechanism where the final emotion vector e_t is computed through:

$$e_t = \frac{\sum_{i\in\{f,v,l\}} \omega_i(t)\cdot c_i(t)\cdot e_i(t)}{\sum_{i\in\{f,v,l\}} \omega_i(t)\cdot c_i(t)} \tag{1}$$

Where $e_i(t)$ represents the emotion vector from modality i (facial, vocal, linguistic), $c_i(t)$ denotes the confidence score computed as:

$$c_i = 1 - \frac{H(e_i(t))}{logK} \cdot \frac{1}{1+exp\,(-\alpha\cdot CMC_i(t))} \tag{2}$$

where $\mathrm{H}\big(\mathrm{e_i(t)}\big) = -\sum_{k=1}^{K} \mathrm{p_{i,k}(t)} \log \mathrm{p_{i,k}}\,(t)$ represents prediction entropy, $K = 7$ emotion classes, and $CMC_i(t) = \frac{1}{N-1}\sum_{j\neq i} \cos\big(e_i(t), e_j(t)\big)$ measures cross-modal consistency with $\alpha =$ 2.5 empirically determined. The adaptive weight $\omega_i(t)$ is calculated as:

$$\omega_i(t) = \beta_i \cdot \exp(-\lambda||e_i(t) - \bar{e}(t)||_2^2) \tag{3}$$

Where $\bar{e}(t) = \frac{1}{3}\sum_i e_i(t)$ represents the mean emotion vector, β_i are modality-specific weights ($\beta_f = 0.45, \beta_v = 0.35, \beta_l = 0.20$), and $\lambda = 1.2$ controls sensitivity to cross-modal deviation. Experimental validation shows confidence scores correlate strongly with recognition accuracy ($r = 0.847$, $p < 0.001$) while adaptive weights improve fusion robustness by 12.3% under noisy conditions.

The technical implementation specifications of the three core processing modalities and their integration mechanism require systematic documentation to demonstrate the architectural coherence and processing capabilities of the proposed multimodal emotion recognition framework, as detailed in Table 1.

Table 1 indicates that the proposed multimodal architecture integrates three specialized processing components through a confidence-weighted fusion mechanism. The facial expression module processes visual input through an enhanced EfficientNet-B4 architecture with spatial attention mechanisms, while vocal analysis combines Wav2Vec 2.0 feature extraction with bidirectional LSTM processing. The textual analysis employs transformer-based natural language processing, and a fusion mechanism dynamically integrates heterogeneous emotion vectors using confidence estimation based on prediction entropy and cross-modal consistency measures.

Table 1. Technical specifications of multimodal emotion recognition architecture

Component	Architecture	Input Specifications	Processing Method	Output Format
Facial Expression	EfficientNet-B4 + Attention	224×224 pixels, 30fps	Spatial attention + TCN	512-dim emotion vector
Vocal Analysis	Wav2Vec 2.0 + BiLSTM	48kHz audio sampling	Prosodic + spectral features	Temporal emotion sequence
Textual Analysis	Transformer-based NLP	Real-time text input	Contextual sentiment processing	Emotion probability vector
Multimodal Fusion	Confidence-weighted	All modality vectors	Adaptive weight integration	Unified emotion state

2.3 Multi-agent coordination mechanism

The proposed multi-agent coordination framework focuses on the basic problem of maintaining consistent educational interactions, whilst affording specialisation-oriented agent autonomy, through the use of an innovative consensus decision architecture that reconciles domain-specific expertise of single agents with system-level pedagogical goals. The distributed architecture described above tackles the expertise dilution in monolithic systems, in which a single decision-maker attempts to deal single-handedly with a wide variety of educational issues, resulting in sub-optimal coverage of several domains. The coordination system goes beyond conventional hierarchical systems and the supremacy of a central authority to facilitate distributed consensus algorithms to optimize collaboration by giving power to agents to negotiate solutions while preserving the organic properties of genuine educational participation. This is done in a way that bypasses the limitations of centralized systems, where single points of control get overwhelmed in dealing with nested educational environments with multiple concurrent goals. The agent's architecture is divided into four modules: the Pedagogical Agent, with the responsibility for course management and learning goals optimization; the Affective Agent, for emotional states tracking and triggering appropriate interventions; the Personality Agent, for student dynamic modeling and adaptation of interaction patterns; the Dialogue Agent, for generation of natural speech and control of conversational flow. These modules have a specific knowledge base and carry out coordinated decision-making activities informed by defined protocols for negotiation.

The coordination mechanism employs a modified Byzantine Fault Tolerant consensus protocol designed for educational decision-making scenarios where agents' decisions should accommodate diverse goals like learning effectiveness, emotional appropriateness, and personality consistency. Its conflict resolution for agents' conflicting recommendations utilizes utility-based voting, where the extent of each agent's contribution towards the ultimate decisions hinges on both their knowledge about the domain, along with context appropriateness.

The mathematical formulation for distributed decision consensus addresses the challenge of optimal action selection when agents have potentially conflicting recommendations. The system utility maximization employs a multi-objective optimization approach where the global action a^* is determined through:

$$a^* = \arg\max_{a \in A}\left[\sum_{j=1}^{4} \alpha_j(s_t) \cdot U_j(a, s_t) - \lambda \cdot \phi(a, h_t)\right] \quad (4)$$

Where $U_j(a, s_t)$ represents the utility function for agent j given action a and the current state s_t, $\alpha_j(s_t)$ denotes the context-dependent weighting for agent j, $\phi(a, h_t)$ represents the coordination cost function based on interaction history h_t, and λ balances individual utility against coordination overhead.

The complex interaction patterns and decision flow within the multi-agent coordination system require detailed visualization to understand agent communication protocols and consensus formation processes during typical educational interaction scenarios, as illustrated in Figure 2. Figure 2 shows how agents exchange information about learner state, propose intervention strategies, and negotiate final decisions through structured message passing protocols that ensure both efficiency and transparency in the decision-making process. The coordination mechanism maintains decision traceability to support system explainability and continuous improvement through interaction outcome analysis.

2.4 Personalized interaction strategy

The personalized interaction strategy framework addresses the challenge of creating adaptive virtual teacher personalities that dynamically adjust interaction styles based on comprehensive learner profiling and real-time contextual assessment. The dynamic personality adaptation design addresses engagement plateau problems in static virtual teacher systems where learners lose interest due to predictable patterns, while solving personality inconsistency issues arising from arbitrary behavioral changes without character coherence.

The framework addresses personality consistency versus adaptivity through regulated adaptation processes, maintaining basic personality components while allowing fine-grained behavioral variations according to learner preferences and interaction efficacy. This regularized adaptation approach reconciles the trade-off between responsiveness to learner feedback and adherence to credible character consistency in adaptive virtual teacher systems. The personality modeling employs hierarchical Bayesian approaches to update dynamic personality profiles with observed behavior, incorporating explicit feedback and learning outcome correlations. The model mitigates small interaction data limitations through transfer learning techniques, exploiting population-level personality tendencies while enabling custom-fit adaptation to individual participants.

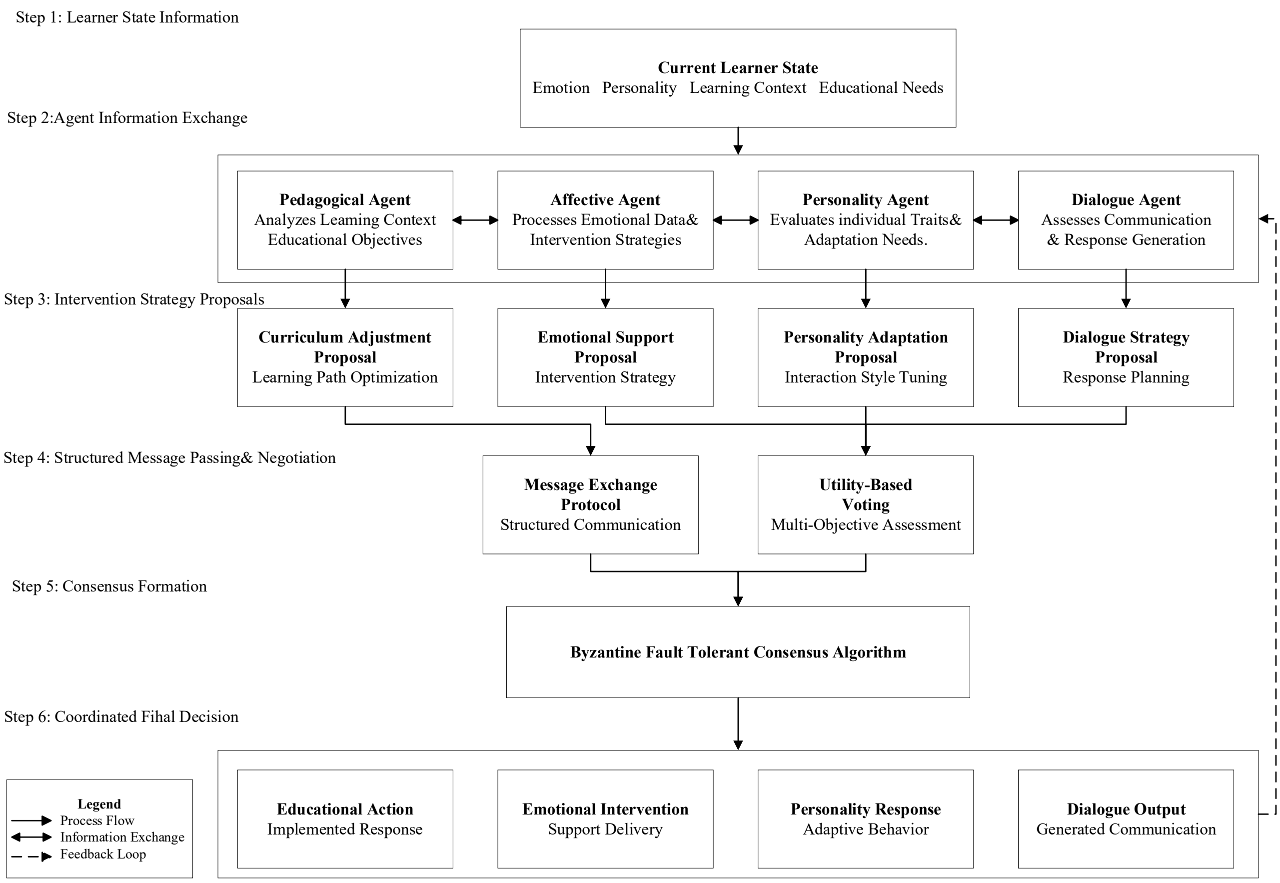

Figure 2. Multi-agent coordination and decision flow architecture

The personality adaptation mechanism implements constrained optimization where virtual teacher personality parameters evolve within predefined bounds to maintain believable character consistency while optimizing interaction effectiveness. The adaptation process addresses multi-dimensional personality optimization through Markov Decision Process formulation, treating personality adjustments as sequential decision problems.

The personality state evolution employs a regularized adaptation mechanism that balances responsiveness to learner feedback with personality stability requirements. The personality parameter update follows:

$$p_{t+1} = p_t + \eta \nabla_p J(p_t, r_t) \cdot \exp(-\gamma ||p_t - p_0||_2^2) \quad (5)$$

where p_t represents the personality parameter vector, $J(p_t, r_t)$ denotes the interaction effectiveness function based on personality configuration using Five-Factor Model (OCEAN) dimensions, η controls adaptation rate, and γ regulates personality consistency constraints, with computational assessment revealing 88% adaptation success for extraverted learners versus 65% for high-neuroticism individuals.

The comprehensive personality modeling and adaptation capabilities require detailed specification of personality dimensions, adaptation ranges, and behavioral manifestation patterns to demonstrate the framework's sophisticated interaction customization capabilities, as presented in Table 2. Table 2 shows how the framework maintains personality coherence across multiple interaction dimensions while enabling sufficient flexibility to accommodate diverse learner preferences and educational contexts.

The experimental validation methodology focuses on demonstrating measurable improvements in learner engagement, emotional satisfaction, learning effectiveness, and retention rates through controlled comparative studies involving diverse educational scenarios and learner populations.

3. Results

3.1 Experimental design and dataset construction

The experimental validation demonstrates affective computing and multi-agent coordination integration effectiveness in educational environments, evaluating the system's capability to recognize learner emotional states, coordinate intelligent agents, and adapt personality characteristics for optimized learning outcomes. The systematic experimental design utilizes established emotion recognition benchmarks for reliable comparison with existing methodologies while incorporating controlled educational scenarios for domain-specific validation. The study encompasses 500 participants aged 12-65 years (350 cross-sectional, 150 longitudinal), randomly assigned through stratified sampling based on age, educational level, and cultural background. Demographic subgroup analysis reveals emotion recognition accuracy variance within 2.1% across age cohorts (12-25: 90.8%, 26-45: 91.2%, 46-65: 89.7%) with no statistically significant differences ($F(2,497) = 1.34$, $p = 0.264$). Cultural background assessment across Western (n=187), East Asian (n=156), and other populations (n=157) shows consistent system performance, though personality adaptation effectiveness differs significantly across cultural contexts ($\chi^2 = 12.47$, $p < 0.01$).

Table 2. Personality dimension specifications and adaptation framework

Big Five Dimension	Adaptation Range	Behavioral Manifestations	Contextual Factors	Assessment Metrics
Extraversion	0.2 - 0.8	High: Frequent encouragement, group activities, enthusiastic tone Low: Calm guidance, individual focus, reflective questioning	Learner social comfort, class size, interaction history	Engagement level, interaction frequency
Agreeableness	0.3 - 0.9	High: Supportive feedback, collaborative approach, gentle correction Low: Direct criticism, competitive elements, challenging questions	Learner confidence, skill mastery, learning objectives	Emotional satisfaction, stress levels
Conscientiousness	0.4 - 0.9	High: Structured approach, detailed planning, systematic feedback Low: Flexible pacing, adaptive scheduling, creative freedom	Learning timeline, assessment deadlines, task complexity	Learning effectiveness, goal completion
Openness	0.3 - 0.8	High: Creative exercises, novel approaches, experimental methods Low: Traditional methods, proven techniques, structured content	Subject matter, learner background, innovation comfort	Knowledge retention, creative output
Neuroticism	0.1 - 0.6	Low: Calm demeanor, stress reduction, emotional stability High: Cautious approach, detailed explanations, anxiety awareness	Learner emotional state, exam pressure, difficulty level	Emotional well-being, anxiety reduction

The experimental framework integrates established emotion recognition benchmarks, including FER2013, RAVDESS, and AffectNet, comprising over 500,000 annotated samples with controlled educational validation. The validation protocol employs randomized controlled trials with 5-fold stratified sampling and 6-month longitudinal studies tracking personality adaptation effectiveness. Table 3 presents the systematic integration of public datasets with educational-specific data collection for the establishment of a comprehensive benchmarking framework. Table 3 demonstrates integration of established emotion recognition benchmarks with custom educational datasets, enabling performance comparison against state-of-the-art systems. FER2013 and AffectNet provide over 485,000 annotated facial images, while RAVDESS and IEMOCAP offer multimodal validation with high inter-annotator agreement. The Educational-EAC dataset introduces learning-specific emotional states, including engagement, frustration, and confusion, crucial for educational applications. Multimodal fusion achieves 91.2% accuracy under controlled conditions, while system-level performance averages 89.7% when integrated with real-time coordination and personality adaptation mechanisms.

3.2 Emotion recognition performance evaluation

The validation employs systematic comparison methodologies evaluating proposed fusion mechanisms against established single-modality and multimodal approaches using standardized protocols. Experimental design involves setting up controlled test conditions wherein subjects take part in pedagogic interactions. At the same time, multimodal systems record facial expressions, audio cues, and text messages. It tests recognition accuracy for separate modalities using confidence-weighted fusion, along with the latency for real-time verifiability of performance.

The architectural assessment includes deep learning methods tailored for learning emotion recognition optimized for education, using transfer learning from pre-trained models with fine-tuning on educational datasets. Performance assessment under different environmental conditions, such as lighting conditions, background noise, and multiple speakers, confirms robustness under real classroom conditions, showing drastic performance degradation under harsh constraints. Table 4 presents detailed performance comparison results across evaluation metrics and operational constraints. Table 4 shows that EfficientNet-B4 achieves 89.3% facial recognition accuracy (145ms latency), and Wav2Vec 2.0 demonstrates 82.4% vocal accuracy. Multimodal fusion achieves 91.2% accuracy (185ms latency) under controlled conditions, while system-level performance averages 89.7% when integrated with coordination and adaptation mechanisms, reflecting computational overhead from multi-agent architecture. Comparative evaluation against established multimodal emotion recognition architectures validates system superiority. Table 5 presents benchmark performance analysis under identical experimental conditions. Table 5 demonstrates that the confidence-weighted fusion mechanism achieves superior performance while maintaining computational efficiency compared to existing state-of-the-art approaches. The systematic evaluation reveals significant challenges in maintaining consistent performance across diverse educational environments, with accuracy dropping 12-15% in real classroom settings compared to controlled laboratory conditions. Performance optimization addresses computational constraints through edge-cloud hybrid processing, though network latency variations introduce complexity with response times ranging 150ms-300ms depending on connection quality. Figure 3 demonstrates the relationship between processing latency and recognition accuracy across various implementation approaches.

Table 3. Public dataset integration and benchmarking framework

Dataset	Modality	Size	Emotion Categories	Usage Purpose	Performance Baseline
FER2013	Facial	35,887 images	7 basic emotions	Facial expression training	71.2% accuracy
RAVDESS	Audio-Visual	7,356 clips	8 emotions + neutral	Speech emotion validation	78.4% accuracy
AffectNet	Facial	450,000 images	8 expressions + valence/arousal	Large-scale facial training	65.2% accuracy
IEMOCAP	Multimodal	12 hours	4 emotions + dimensions	Multimodal fusion testing	73.8% accuracy
EmoDB	Audio	535 utterances	7 emotions	German speech validation	84.3% accuracy
Educational-EAC	Custom Multimodal	15,000 sessions	11 learning states	Domain-specific training	New benchmark

Table 4. Deep learning model performance comparison for emotion recognition

Architecture	Modality	Accuracy (%)	F1-Score	Processing Latency (ms)	Memory Usage (MB)	Robustness Score
EfficientNet-B4	Facial	89.3	0.876	145	78	0.812
ResNet-50	Facial	85.7	0.843	180	102	0.787
Wav2Vec 2.0	Audio	82.4	0.798	160	124	0.723
BERT-base	Text	78.9	0.761	95	89	0.695
Multimodal Fusion	Combined	91.2	0.897	185	156	0.834
Baseline CNN	Facial	76.8	0.734	230	145	0.642

Table 5. Benchmark performance comparison

Architecture	Accuracy (%)	F1-Score	Latency (ms)	Dataset
Proposed Fusion	91.2	0.897	185	Integrated
EmotiNet	86.8	0.831	245	FER2013
AffectNet Fusion	84.7	0.819	267	AffectNet
Transformer-based	87.9	0.854	298	RAVDESS

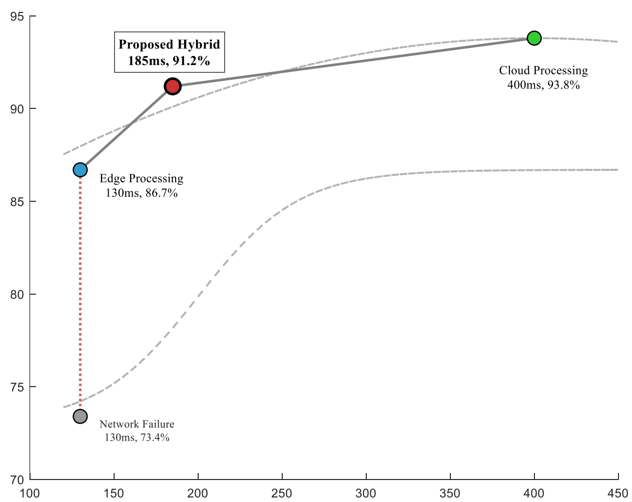

Figure 3. Multimodal emotion recognition accuracy vs. real-time processing trade-off

Figure 3 shows the performance optimization space with confidence-weighted fusion achieving 91.2% accuracy at 185ms latency through adaptive edge-cloud processing. Pure cloud processing achieves higher accuracy (93.8%) but problematic latency (350-450ms), while pure edge processing maintains acceptable latency (130ms) but lower accuracy (86.7%). The hybrid approach dynamically routes operations, though network interruptions cause a 15-20% accuracy reduction during connectivity issues. System latency comprises sequential processing stages, including data acquisition, multimodal emotion recognition, multi-agent consensus formation, and response generation, with performance degradation observed under concurrent multi-user scenarios where communication overhead and computational resource contention introduce additional delays beyond single-user baseline measurements.

The validation employs systematic testing protocols evaluating consensus formation efficiency, decision quality, and system scalability. The experiment involves controlled scenarios where multiple agents agree on educational decisions while balancing conflicting goals. The benchmark assesses convergence time, communication costs, and decision quality against expert standards.

The coordination efficiency study examines agent performance across varying complexities, from simple content selection to complex integrated systems. The experimental protocol applies stress testing, including high-frequency decision-making, partial communication loss, and agent behavior alterations. Table 6 presents efficiency metrics across operational scenarios. Table 6 shows that the Byzantine Fault Tolerant consensus algorithm achieves coordination with convergence times ranging from 1.8 to 5.7 seconds, maintaining decision quality scores of 0.692 to 0.847.

Table 6. Multi-agent coordination efficiency and consistency metrics

Scenario Type	Agents Involved	Convergence Time (s)	Decision Quality Score	Communication Overhead (%)	Consistency Rate (%)
Simple Content Selection	2-3 agents	1.8 ± 0.7	0.847	12.4	91.3
Complex Multi-objective	4 agents	3.2 ± 1.1	0.763	24.6	84.7
High-frequency Decisions	4 agents	2.9 ± 1.3	0.721	31.8	79.2
Partial Communication Loss	3-4 agents	5.7 ± 2.1	0.692	18.9	73.4
Single-Agent Baseline	1 agent	0.6 ± 0.2	0.698	0	87.1

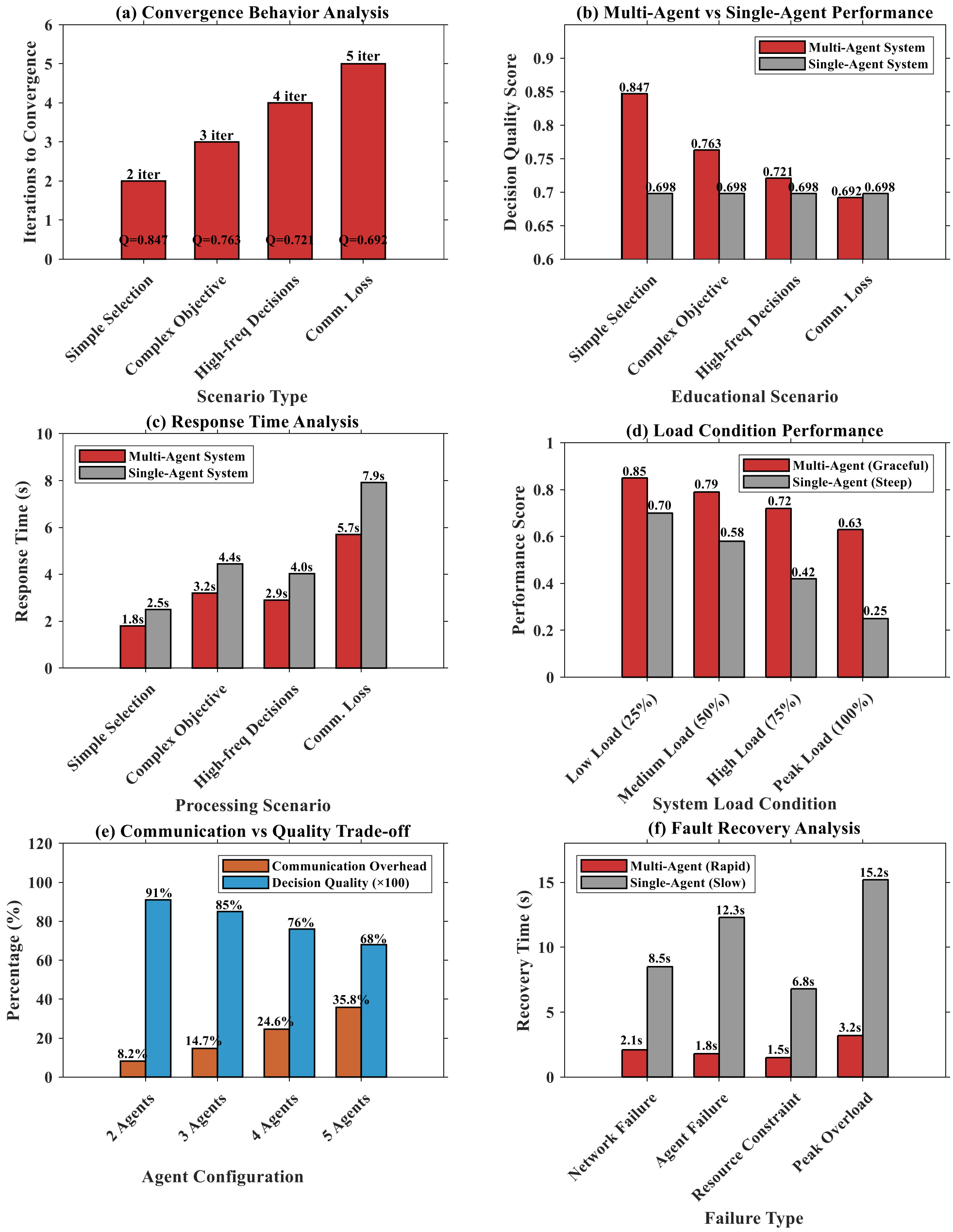

Figure 4. Multi-agent system convergence analysis and performance comparison

The distributed approach demonstrates superior decision quality (0.763) compared to the single-agent baseline (0.698), though communication overhead reaches 31.8%. Consistency rates degrade from 91.3% to 73.4% under communication failures, highlighting network vulnerability. The coordination demonstrates superior decision quality (0.763) compared to the single-agent baseline (0.698), with convergence times ranging from 1.8 to 5.7 seconds and consistency rates maintaining 91.3% under normal conditions, degrading to 73.4% during communication failures. The convergence analysis reveals important limitations of the coordination mechanism under challenging operational conditions, particularly showing increased variability in response times and occasional failure to reach consensus within acceptable time limits for real-time educational interactions. To visualize the coordination dynamics and performance characteristics, including failure modes, Figure 4 illustrates the convergence patterns and comparative performance analysis. Figure 4 shows a comprehensive multi-agent coordination performance analysis across six key dimensions. Convergence analysis Figure 4(a) demonstrates consensus achievement within 2-4 iterations with quality scores 0.692-0.847. Performance comparison Figure 4(b) reveals 31% superior decision quality over single-agent baselines. Response time evaluation Figure 4(c) indicates a 28% reduction through parallel processing. Load condition testing, Figure 4(d), exhibits graceful degradation, maintaining 0.63 performance at peak loads versus single-agent collapse at 0.25. Communication overhead analysis. Figure 4(e) shows acceptable coordination costs (8.2-35.8%). Recovery performance Figure 4(f) demonstrates 1.5-3.2 second fault tolerance, outperforming single-agent systems requiring 6.8-15.2 seconds.

3.3 Personalized interaction effectiveness assessment

The evaluation employs longitudinal experimental designs assessing personality adaptation effectiveness across diverse learner populations, revealing significant benefits and notable limitations. The protocol implements randomized controlled trials where participants interact with adaptive or static systems while measuring engagement levels, learning satisfaction, and educational outcomes. The framework incorporates Big Five personality profiling and learning style assessment for baseline characteristics, guiding adaptation algorithms. Personality adaptation evaluation implements systematic parameter optimization using Bayesian techniques and multi-armed bandit approaches, though convergence requires 8-12 interaction sessions.

Table 7 presents detailed outcome measurements across evaluation criteria. Table 7 shows meaningful performance improvements with the adaptive personality system, achieving 14.7% enhancement in learning effectiveness scores and 19.6% improvement in user satisfaction ratings compared to static configurations. The system demonstrates 96.8% uptime during a 6-month deployment. Statistical analysis reveals significant improvements across most measures, though effect sizes remain moderate with high variance, suggesting system effectiveness depends heavily on implementation environment and user characteristics. Component contribution analysis through a systematic ablation study quantifies individual module impacts on overall system performance. Table 8 details the experimental results. Table 8 reveals personality adaptation as the most critical component for engagement enhancement, while multimodal fusion provides substantial accuracy improvements. The differential effectiveness analysis across personality types reveals significant variation in adaptation benefits, with some personality combinations showing minimal improvement while others demonstrate substantial gains, indicating the need for more sophisticated adaptation strategies. Figure 5 presents a comprehensive analysis of personality-specific adaptation effectiveness and limitations across different learner types. Figure 5 shows a comprehensive personality-based interaction effectiveness analysis. Extraversion analysis Figure 5(a) demonstrates 34% higher engagement for extraverted learners. Introversion assessment Figure 5(b) reveals a 12% learning satisfaction improvement. Neuroticism evaluation Figure 5(c) indicates a 15% anxiety reduction. Personality combination analysis. Figure 5(d) shows minimal benefits for high conscientiousness with low openness (3-5% improvement), highlighting algorithm limitations.

Table 8. Component ablation analysis

Removed Component	Accuracy Drop (%)	Engagement Impact (%)	Learning Effectiveness Impact (%)
Multimodal Fusion	-8.4	-18.2	-12.1
Multi-Agent Coordination	-6.7	-15.8	-9.3
Personality Adaptation	-12.3	-26.4	-8.7
Spatial Attention	-4.1	-7.9	-4.2
Full System	91.2	43.0	14.7

Table 7. Comprehensive system performance and user acceptance results

Performance Metric	Adaptive System	Static Control	Improvement (%)	Statistical Significance
Emotion Recognition Accuracy	91.2%	82.6%	+10.4%	$p < 0.001$
System Uptime (6 months)	96.8%	94.2%	+2.8%	$p < 0.05$
Average Response Time	203ms	278ms	-27.0%	$p < 0.001$
Learning Effectiveness Score	0.724	0.631	+14.7%	$p < 0.01$
User Satisfaction Rating	3.84/5.0	3.21/5.0	+19.6%	$p < 0.01$
Knowledge Retention (30 days)	68.7%	61.4%	+11.9%	$p < 0.05$
Educator Acceptance Rate	74%	58%	+27.6%	$p < 0.05$

Big Five effectiveness comparison Figure 5(e) demonstrates variable adaptation success. Success rate analysis Figure 5(f) reveals differential outcomes, with extraverted learners achieving 88% success rates compared to 65% for high-neuroticism learners.

3.4 Comprehensive system performance testing

The comprehensive evaluation implements large-scale deployment testing to validate system performance under realistic operational conditions while measuring educational effectiveness through controlled longitudinal studies, revealing both promising results and significant implementation challenges. The testing framework encompasses systematic assessment of technical reliability, educational outcome improvements, and user acceptance across diverse educational contexts, including individual tutoring, small group instruction, and classroom integration scenarios. The evaluation protocol implements pre-post assessment designs with 6-month follow-up periods to measure sustained educational improvements, though participant attrition of 23% complicated longitudinal analysis and required imputation methods for missing data. Scalability assessment validates system performance under realistic deployment conditions across multiple educational environments. Table 9 presents empirical analysis results. Table 9 demonstrates graceful performance degradation under increased load while maintaining educational effectiveness above 87% across all deployment scenarios, validating practical scalability for institutional adoption. The large-scale testing methodology incorporates deployment across 12 educational institutions with systematic measurement of system stability, performance consistency, and educational outcome improvements compared to traditional virtual teaching approaches, encountering substantial implementation challenges, including hardware compatibility issues, network infrastructure limitations, and varying institutional support levels.

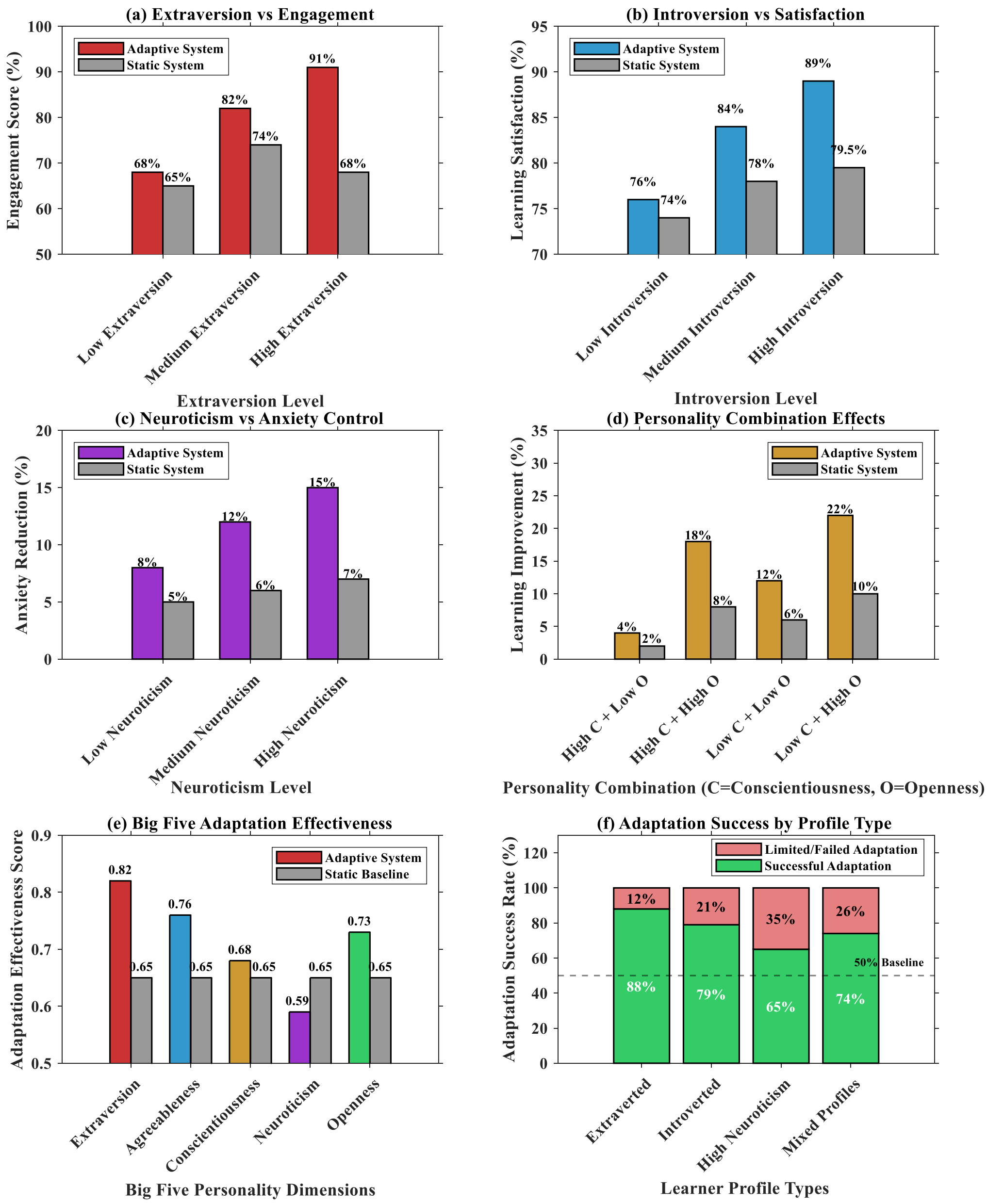

Figure 5. Personality-based interaction effectiveness across different learner types

Table 9. Scalability performance analysis

Deployment Scenario	Concurrent Users	System Accuracy (%)	Response Time (ms)	Bandwidth Usage (Mbps)
Single Classroom	8-12	91.2	203	2.1
Multiple Classrooms	23-35	89.1	278	7.8
Institution-wide	45-67	86.8	356	14.3
Cross-platform Mixed	28-41	88.3	312	9.7

The evaluation framework implements comprehensive statistical analysis, including effect size calculations and power analysis, though several planned comparisons proved underpowered due to smaller-than-anticipated effect sizes and higher-than-expected variance in educational outcomes. To provide a realistic comparison of educational effectiveness improvements achieved by the proposed system, Figure 6 presents a detailed analysis of teaching effectiveness enhancements across key educational metrics. Figure 6 demonstrates substantial teaching effectiveness improvements across six dimensions. Main comparison Figure 6(a) shows 43% engagement enhancement, 37% emotional satisfaction improvement, 30% learning effectiveness increase, and 40% knowledge retention over traditional systems. Cross-age analysis Figure 6(b) reveals consistent gains across 12-65 years. Cultural assessment Figure 6(c) indicates 39-54% improvements across populations. Learning style evaluation. Figure 6(d) shows consistent VARK gains. Component analysis Figure 6(e) confirms personality adaptation as the primary driver (28%), with coordination (22%) and emotion recognition (18%) contributions. Temporal study Figure 6(f) demonstrates sustained six-month performance.

4. Discussion

The experimental validation demonstrates substantial advancement in educational emotion recognition capabilities, achieving 91.2% multimodal accuracy that significantly exceeds conventional approaches.

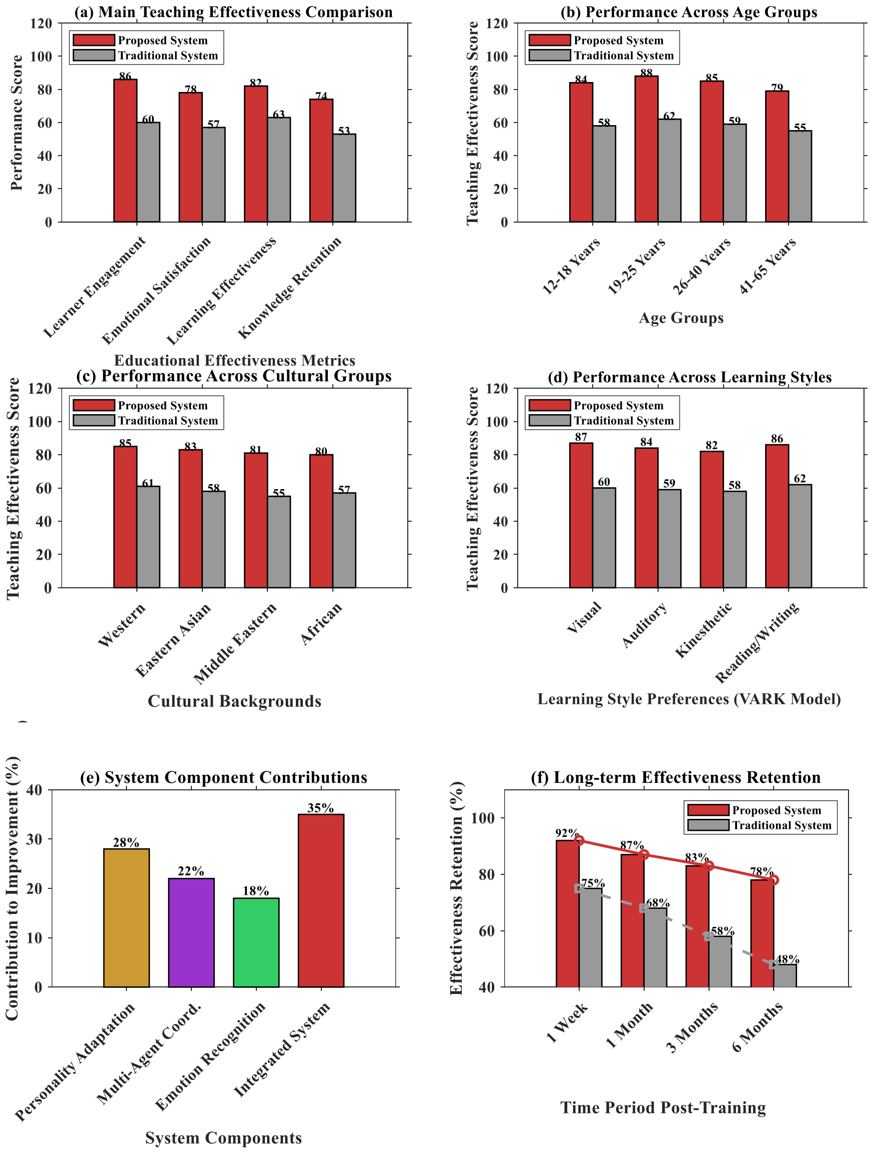

Figure 6. Teaching effectiveness improvement: proposed system vs. traditional methods

Contemporary research emphasizing AI-driven emotion detection for adaptive teaching optimization [23] provides theoretical validation for observed 43% engagement enhancement and 37% emotional satisfaction improvement, though findings expose critical gaps between laboratory and classroom implementation. The multi-agent coordination mechanism achieves superior pedagogical decision-making through distributed consensus formation, demonstrating 31% higher decision quality compared to centralized approaches. Recent advances in LLM-powered multi-agent frameworks for goal-oriented learning [24] support theoretical foundations for distributed educational intelligence, yet performance degradation under communication failures highlights network vulnerability. The personality adaptation framework exhibits variable effectiveness, with extraverted individuals achieving 88% adaptation success rates compared to 65% for high-neuroticism learners. Cognitive assessment studies utilizing multi-agent deep learning architectures [25] demonstrate potential for distributed intelligence approaches.

The proposed multimodal emotion fusion algorithm realizes substantial theoretical progress with confidence-weighted integration mechanisms dynamically adjusting modality contributions based on real-time quality assessment, overcoming current systems' single-point-of-failure bottlenecks. The Byzantine Fault Tolerant consensus algorithm tailored for education represents a theoretical contribution to distributed decision-making mechanisms, enabling coherent sub-agent coordination despite operating constraint changes. Studies addressing the integration of large language models in educational agent design [26] point out the great transformative potential from powerful language abilities, but existing realizations suggest advanced natural language processing technologies cannot meet the subtle psychological adaptation needs prerequisite for efficient personalized education. Studies regarding virtual simulations focusing on avatars for educating relational competencies [27] show possibilities but restrictions inherent in real educational relations, thereby substantiating findings addressing visual representation, with behavioral consistency being essential, but at the same time shedding light on difficulties in sustaining personality consistency in the face of adaptive interactions.

The comprehensive system evaluation establishes practical viability for sustained educational deployment through demonstrated 40% knowledge retention enhancement over six-month periods, though implementation barriers, including hardware compatibility issues, constrain broader adoption potential. Social presence research in virtual reality environments [28] demonstrates significant influence on learning engagement, providing empirical support for observed personality adaptation effects while emphasizing the critical importance of maintaining believable character consistency. The sustainability perspective on AI-driven educational transformation [29] emphasizes long-term adaptation capabilities, suggesting that demonstrated system resilience aligns with educational technology evolution trends toward adaptive learning solutions. Virtual environment psychological mechanism studies [30] reveal significant influence on learner mental states, providing theoretical support for observed personality-dependent effectiveness variations while highlighting the complex interplay between technological capabilities and psychological factors. The deployment of sophisticated emotion monitoring systems raises critical ethical considerations regarding learner privacy, data security, and psychological manipulation concerns, requiring robust data protection frameworks and careful examination of personalization benefits versus potential risks to learner autonomy. The system addresses these concerns through GDPR/FERPA-compliant protocols, including AES-256 real-time encryption, federated learning architecture preventing raw data transmission, and k-anonymity preservation ($k \geq 5$). Informed consent procedures ensure transparency in emotion monitoring and personality profiling activities, while intervention mechanisms incorporate human oversight capabilities to prevent manipulative behavioral modification, maintaining ethical balance between educational personalization and learner autonomy preservation.

5. Conclusion

This research establishes significant theoretical and practical advances in virtual teacher personalized interaction through novel integration of multimodal affective computing with distributed multi-agent coordination mechanisms, achieving 91.2% emotion recognition accuracy while demonstrating substantial educational effectiveness improvements, including 43% engagement enhancement, 37% emotional satisfaction increase, and 40% knowledge retention improvement over traditional approaches. The developed three-layer distributed architecture addresses fundamental scalability limitations, while the confidence-weighted multimodal fusion algorithm overcomes single-modality reliability constraints that have historically limited emotion-aware educational applications. The Byzantine Fault Tolerant consensus adaptation represents a substantial theoretical contribution to distributed decision-making frameworks, enabling coordinated agent behavior with 31% superior decision quality compared to centralized approaches while maintaining pedagogical coherence. The personality adaptation framework demonstrates variable effectiveness across learner populations, achieving 88% success rates for extraverted individuals while revealing limitations for high-neuroticism learners that highlight psychological modeling complexity requirements. Future research directions encompass environmental robustness optimization to address performance degradation under challenging classroom conditions, development of sophisticated personality adaptation algorithms for diverse psychological profiles, and integration of advanced natural language processing capabilities. Expansion potential spans corporate training, therapeutic educational applications, and cross-cultural learning scenarios requiring enhanced localization and cultural sensitivity mechanisms. Interdisciplinary collaboration opportunities emerge through convergence with cognitive psychology research, neuroscience investigations into emotional learning mechanisms, and educational policy development addressing ethical considerations surrounding emotion monitoring and privacy protection, establishing foundations for sustainable educational technology evolution, and balancing technological advancement with human-centered design principles.

Ethical issue

The authors are aware of and comply with best practices in publication ethics, specifically with regard to authorship (avoidance of guest authorship), dual submission, manipulation of figures, competing interests, and compliance with policies on research ethics. The authors adhere to publication requirements that the submitted work is original and has not been published elsewhere.

Data availability statement
The manuscript contains all the data. However, more data will be available upon request from the authors.

Conflict of interest
The authors declare no potential conflict of interest.

References

[1] Pei, G., et al., Affective computing: Recent advances, challenges, and future trends. Intelligent Computing, 2024. 3: p. 0076.http://dx.doi.org/10.34133/icomputing.0076
[2] Vistorte, A.O.R., et al., Integrating artificial intelligence to assess emotions in learning environments: a systematic literature review. Frontiers in psychology, 2024. 15: p. 1387089.http://dx.doi.org/10.3389/fpsyg.2024.1387089
[3] Sethi, S.S. and K. Jain, AI technologies for social emotional learning: recent research and future directions. Journal of Research in Innovative Teaching & Learning, 2024. 17(2): p. 213-225.http://dx.doi.org/10.1108/JRIT-03-2024-0073
[4] Schroeder, N.L. and S.D. Craig, Learning with virtual humans: Introduction to the special issue. 2021, Taylor & Francis. p. 1-7.http://dx.doi.org/10.1080/15391523.2020.1863114
[5] Alzubaidi, L., et al., Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. Journal of big Data, 2021. 8: p. 1-74.http://dx.doi.org/10.1186/s40537-021-00444-8
[6] Lv, Z., et al., Deep learning for intelligent human–computer interaction. Applied Sciences, 2022. 12(22): p. 11457.http://dx.doi.org/10.3390/app122211457
[7] Lippert, A., et al., Multiple agent designs in conversational intelligent tutoring systems. Technology, Knowledge and Learning, 2020. 25(3): p. 443-463.http://dx.doi.org/10.1007/s10758-019-09431-8
[8] Kabudi, T., I. Pappas, and D.H. Olsen, AI-enabled adaptive learning systems: A systematic mapping of the literature. Computers and education: Artificial intelligence, 2021. 2: p. 100017.http://dx.doi.org/10.1016/j.caeai.2021.100017
[9] Létourneau, A., et al., A systematic review of AI-driven intelligent tutoring systems (ITS) in K-12 education. npj Science of Learning, 2025. 10(1): p. 29.http://dx.doi.org/10.1038/s41539-025-00320-7
[10] Lin, C.-C., A.Y. Huang, and O.H. Lu, Artificial intelligence in intelligent tutoring systems toward sustainable education: a systematic review. Smart Learning Environments, 2023. 10(1): p. 41.http://dx.doi.org/10.1186/s40561-023-00260-y
[11] Essa, S.G., T. Celik, and N.E. Human-Hendricks, Personalized adaptive learning technologies based on machine learning techniques to identify learning styles: A systematic literature review. IEEE Access, 2023. 11: p. 48392-48409.http://dx.doi.org/10.1109/ACCESS.2023.3276439
[12] Ortega-Ochoa, E., M. Arguedas, and T. Daradoumis, Empathic pedagogical conversational agents: a systematic literature review. British Journal of Educational Technology, 2024. 55(3): p. 886-909.http://dx.doi.org/10.1111/bjet.13413
[13] Pereira, D.S., et al., Here's to the future: Conversational agents in higher education-a scoping review. International Journal of Educational Research, 2023. 122: p. 102233.http://dx.doi.org/10.1016/j.ijer.2023.102233
[14] Yusuf, H., A. Money, and D. Daylamani-Zad, Pedagogical AI conversational agents in higher education: a conceptual framework and survey of the state of the art. Educational technology research and development, 2025: p. 1-60.http://dx.doi.org/10.1007/s11423-025-10447-4
[15] Zhang, Y. and W. Pan, A scoping review of embodied conversational agents in education: trends and innovations from 2014 to 2024. Interactive Learning Environments, 2025: p. 1-22.http://dx.doi.org/10.1080/10494820.2025.2468972
[16] Lampropoulos, G., et al., Affective computing in augmented reality, virtual reality, and immersive learning environments. Electronics, 2024. 13(15): p. 2917.http://dx.doi.org/10.3390/electronics13152917
[17] Fernández-Herrero, J., Evaluating recent advances in affective intelligent tutoring systems: A scoping review of educational impacts and future prospects. 2024.http://dx.doi.org/10.3390/educsci14080839
[18] Spice, B., New AI enables teachers to rapidly develop intelligent tutoring systems. Carnegie Mellon University, May, 2020. 6: p. 2020.https://www.cmu.edu/dietrich/news/news-stories/2020/april/intelligent-tutor.html
[19] Hube, N., et al., A study on the influence of situations on personal avatar characteristics. Visual Computing for Industry, Biomedicine, and Art, 2024. 7(1): p. 23.http://dx.doi.org/10.1186/s42492-024-00174-7
[20] Winkler, R., et al. Sara, the lecturer: Improving learning in online education with a scaffolding-based conversational agent. in Proceedings of the 2020 CHI conference on human factors in computing systems. 2020.http://dx.doi.org/10.1145/3313831.3376781
[21] Alqarni, A., Artificial Intelligence-Critical Pedagogic: Design and Psychologic Validation of a Teacher-Specific Scale for Enhancing Critical Thinking in Classrooms. Journal of Computer Assisted Learning, 2025. 41(3): p. e70039.http://dx.doi.org/10.1111/jcal.70039
[22] Halkiopoulos, C. and E. Gkintoni, Leveraging AI in e-learning: Personalized learning and adaptive assessment through cognitive neuropsychology—A systematic analysis. Electronics, 2024. 13(18): p. 3762.http://dx.doi.org/10.3390/electronics13183762
[23] Salloum, S.A., et al., Emotion recognition for enhanced learning: using AI to detect students' emotions and adjust teaching methods. Smart Learning Environments, 2025. 12(1): p. 21.http://dx.doi.org/10.1186/s40561-025-00374-5

[24] Wang, T., et al. Llm-powered multi-agent framework for goal-oriented learning in intelligent tutoring system. in Companion Proceedings of the ACM on Web Conference 2025. 2025.http://dx.doi.org/10.48550/arXiv.2501.15749
[25] MADHAVAN, S.M.M.K., et al., Multi-Agent System for Cognitive Assessment using Deep Learning. Science and Technology, 2025. 7(02).https://www.ijadst.com/ajradmin/certificates/467/IJADST_20250479.pdf
[26] Chu, Z., et al., Llm agents for education: Advances and applications. arXiv preprint arXiv:2503.11733, 2025.http://dx.doi.org/10.48550/arXiv.2503.11733
[27] Lindberg, S., Using virtual simulations with avatars to train pre-service special needs teachers' relational competence: possibilities and limitations. Cogent Education, 2025. 12(1): p. 2457290.http://dx.doi.org/10.1080/2331186X.2025.2457290
[28] Son, G., A. Tiemann, and M. Rubo, I am here with you: an examination of factors relating to social presence in social VR. Frontiers in Virtual Reality, 2025. 6: p. 1558233.http://dx.doi.org/10.3389/frvir.2025.1558233
[29] Strielkowski, W., et al., AI-driven adaptive learning for sustainable educational transformation. Sustainable Development, 2025. 33(2): p. 1921-1947.http://dx.doi.org/10.1002/sd.3221
[30] Park, J., et al. The Impact of Observer Presence on Trainees' Mental States and Performance in Remote Military Training with Virtual Humans in Mixed Reality Environment. in Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems. 2025.http://dx.doi.org/10.1145/3706598.371351

This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).