iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 Issue: V Month of publication: May 2026

DOI: https://doi.org/10.22214/ijraset.2026.82068

www.ijraset.com
Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue V May 2026- Available at www.ijraset.com

AI-Enabled Intelligent Teacher Robot for Automated Attendance, Personalized Learning Assistance, and Real-Time Knowledge Retrieval

Abishek S[1], Haroon Al Rashith A[2], Ubaith Sherif M[3], Kishore Kumar U[4], Dr. Kousalya[5]
[1, 2, 3, 4]Student, [5]Associate Professor, Department of Artificial Intelligence & Data Science, United Institute of Technology, Coimbatore, India

Abstract: Traditional classroom management systems impose significant burdens on educators through time-consuming manual attendance processes and the delivery of uniform instructional content that fails to account for diverse student learning trajectories. The proliferation of low-cost embedded computing platforms and advances in on-device machine learning now present a compelling opportunity to address these challenges through physically embodied intelligent robotic systems. This paper presents an AI-enabled intelligent teacher robot that integrates three tightly coupled modules within a single autonomous platform deployable in undergraduate engineering classrooms. The first module implements automated attendance management using an ensemble of VGG-Face and FaceNet deep embeddings computed via the DeepFace framework, achieving a face recognition accuracy of 97.6% across 1,350 recognition events with a false acceptance rate of 1.2% and a false rejection rate of 1.8% at a processing latency of 112 milliseconds per frame. The second module delivers personalized learning assistance through an adaptive difficulty engine that constructs longitudinal student performance profiles encompassing quiz scores, response latency, topic engagement, and difficulty rating and adjusts instructional content dynamically; experimental evaluation across 45 undergraduate students over six weeks demonstrated a statistically significant post-test learning gain of 21.4% in the experimental group versus 9.7% in the control group ($p < 0.01$), together with a 34% increase in voluntary student engagement. The third module provides real-time knowledge retrieval through a Retrieval-Augmented Generation (RAG) pipeline that combines Vosk offline speech recognition, FAISS-indexed sentence-transformer embeddings, and a 7-billion-parameter quantized large language model, achieving a retrieval accuracy of 94.3%— substantially outperforming a standalone LLM (78.9%) and keyword-based retrieval (62.4%)— at a mean response latency of 1.8 seconds. All inference is executed entirely on-device on a Raspberry Pi 5 without cloud dependency, making the system accessible and privacy-preserving. The proposed architecture demonstrates that integrated, physically embodied AI tutoring robots are technically feasible, educationally effective, and deployable within resource-constrained institutional settings.

Index Terms: Educational robotics, face recognition, DeepFace, retrieval-augmented generation, personalized learning, intelligent tutoring system, Raspberry Pi, on-device AI, adaptive difficulty, FAISS.

I. INTRODUCTION

The administration of attendance records and the delivery of individualized academic instruction represent two of the most persistent operational challenges facing educators in higher education. In a typical undergraduate engineering cohort, roll-call attendance verification consumes between five and ten minutes of instructional time per session, aggregating to tens of hours of lost teaching opportunity across an academic semester. Simultaneously, the heterogeneous academic preparation of students within a single cohort renders uniform lecture delivery pedagogically ineffective: students who have mastered prerequisite concepts are under-challenged, while students with knowledge gaps are overwhelmed, yielding sub-optimal learning outcomes across the entire class. These twin deficiencies —administrative inefficiency and instructional rigidity—motivate the exploration of intelligent, autonomous classroom agents capable of performing attendance management and adaptive tutoring in parallel. The convergence of three technological trends makes the present moment particularly propitious for addressing these challenges through physical robotic systems. First, deep learning-based face recognition has matured to the point where ensemble models such as VGG-Face and FaceNet routinely achieve recognition accuracies exceeding 97% under controlled indoor conditions .

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue V May 2026- Available at www.ijraset.com

Second, large language models (LLMs) with billions of parameters can now be quantized to run inference on edge hardware consuming less than 15 watts of power, enabling on-device question answering without cloud connectivity. Third, single-board computers such as the Raspberry Pi 5, featuring quad-core ARM Cortex-A76 processors clocked at 2.4 GHz with 8 GB of LPDDR4X RAM, provide sufficient compute for concurrent multi-modal AI workloads at a fraction of the cost of industrial robotic platforms. The synthesis of these capabilities within a single, mobile robotic chassis creates a new category of classroom assistant that can autonomously manage administrative tasks while delivering personalized academic support.

Despite the substantial body of research on educational robotics automated attendance systems intelligent tutoring systems and retrieval-augmented generation no prior system has successfully integrated all three capabilities within a unified, physically embodied, on-device platform. Existing attendance systems based on Radio Frequency Identification (RFID) require students to carry and present physical tags introducing logistical overhead and single points of failure. Face recognition-based systems have been demonstrated in desktop configurations but have not been integrated with pedagogical AI agents. Natural language-based tutoring systems such as AutoTutor and QuizBot provide adaptive question-and-answer interaction but lack physical embodiment and real-time course-specific knowledge retrieval. RAG architectures have demonstrated superior factual grounding compared to parametric LLMs in open-domain settings but their deployment on edge hardware within an educational context remains largely unexplored. The absence of an integrated system that addresses all of these dimensions constitutes a meaningful gap in the educational technology literature.

This paper addresses this gap by presenting the design, implementation, and evaluation of an AI-enabled intelligent teacher robot. The principal contributions of this work are as follows. First, the paper presents a complete end-to-end automated attendance system built on a VGG-Face and FaceNet ensemble implemented via the DeepFace library, achieving 97.6% accuracy across 1,350 recognition events.

Second, a novel adaptive learning module is proposed that constructs dynamic student profiles and adjusts instructional difficulty through a weighted composite scoring mechanism augmented by collaborative filtering for peer-based topic weakness identification. Third, a fully on-device RAG pipeline is developed that achieves 94.3% retrieval accuracy at a mean latency of 1.8 seconds, outperforming both standalone LLM inference and keyword-based retrieval by substantial margins. Fourth, the three modules are unified within a single mobile robotic platform powered by a Raspberry Pi 5, demonstrating practical feasibility for deployment in resource-constrained educational institutions. Fifth, a rigorous six-week evaluation involving 45 undergraduate engineering students provides statistically significant evidence of improved learning outcomes, with a post-test gain differential of 11.7 percentage points between experimental and control groups ($p < 0.01$).

The remainder of this paper is organized as follows. Section II surveys related work on educational robotics, automated attendance, intelligent tutoring systems, and retrieval-augmented generation, and identifies the specific research gaps that the proposed system addresses. Section III describes the system architecture, hardware specifications, and software organization. Section IV presents the detailed methodology for each of the three functional modules, including pseudocode for the core attendance recognition algorithm. Section V reports the results of the experimental evaluation against established baselines. Section VI discusses the implications of these results, compares the proposed system with the state of the art, and acknowledges current limitations. Section VII concludes the paper and outlines directions for future research.

II. RELATED WORK

A. Educational and Classroom Robots

The deployment of robotic systems in educational settings has been an active area of research for more than two decades. Belpaeme et al. conducted a comprehensive review of social robots for education, concluding that physical embodiment confers measurable advantages over virtual agents in terms of student attention, motivation, and collaborative engagement. Tanaka and Matsuzoe demonstrated that elementary school children exhibited sustained enthusiasm and improved academic performance when engaging with a care-receiving robot that assumed the learner role, illustrating the bidirectional pedagogical potential of human-robot interaction. Mubin et al.surveyed the applicability of robots across diverse educational contexts and identified autonomous tutoring, language instruction, and STEM demonstration as the most promising application domains. Kandlhofer et al. examined the integration of artificial intelligence concepts into K-12 and undergraduate curricula through robotics platforms, finding that hands-on robotic interaction significantly improved conceptual retention. While these studies establish the educational efficacy of robotic agents in general terms, they do not address the specific technical challenge of integrating face recognition, adaptive tutoring, and knowledge retrieval within a single autonomous system.

B. Automated Attendance and Face Recognition

Automated attendance management has been approached through several technological paradigms. Hussain et al. implemented an RFID-based system that achieved near-perfect tag detection rates but required students to carry physical identification cards and pass in proximity to fixed reader infrastructure, introducing logistical dependencies that face recognition systems are designed to eliminate. Kumbhar and Bhat demonstrated a real-time face recognition attendance system using OpenCV's Local Binary Pattern Histogram (LBPH) algorithm, reporting recognition accuracy of approximately 88.4% under controlled lighting conditions. Firdaus and Ahmad improved upon this by employing deep convolutional neural network-based face descriptors, achieving accuracies of up to 94.7%, but the system operated in a desktop configuration without integration into a physical classroom agent. D'Mello and Graesser explored the role of affective state detection in educational settings, noting that facial expression analysis could provide additional contextual signals for adaptive systems. The proposed system extends this body of work by deploying a VGG-Face and FaceNet ensemble on a mobile robotic platform, achieving 97.6% recognition accuracy while simultaneously supporting multi-modal instructional interaction.

C. NLP-Based Intelligent Tutoring Systems

Intelligent tutoring systems (ITS) leveraging natural language processing represent a substantial thread of educational technology research. Graesser et al. introduced AutoTutor, a dialogue-based ITS that used latent semantic analysis to compare student responses against expert-authored knowledge and guide remediation through Socratic questioning, demonstrating learning gains comparable to human one-on-one tutoring. Winkler and Soellner reviewed the deployment of chatbots in educational contexts, cataloguing the progression from rule-based response systems to transformer-based conversational agents and identifying personalization as the most critical open challenge. Ruan et al. presented QuizBot, a dialogue-based adaptive learning system delivered through a messaging platform, showing that quiz-centric adaptive dialogue significantly improved factual recall compared to passive reading. Ruan et al. investigated voice-interactive intelligent tutoring in undergraduate laboratory courses, finding that voice-mediated interaction reduced cognitive load and improved procedural skill acquisition. Papamitsiou and Economides provided a systematic review of learning analytics and educational data mining, identifying weighted performance profiling as a particularly effective approach to dynamic difficulty adjustment. The proposed system builds upon these foundations by combining an adaptive difficulty engine with a voice-interactive interface and a physically embodied robotic agent, a combination that has not previously been reported in the literature.

D. RAG and Knowledge Retrieval Systems

Retrieval-Augmented Generation was formalized by Lewis et al. [5] as a framework for grounding LLM-generated responses in non-parametric external knowledge, demonstrating that retrieved passages substantially improved factual accuracy on knowledge-intensive NLP benchmarks compared to parametric generation alone. Karpukhin et al. [13] demonstrated that dense passage retrieval using bi-encoder architectures indexed with approximate nearest-neighbor search significantly outperformed sparse BM25 retrieval for open-domain question answering, establishing the technical foundation for the FAISS-based retrieval pipeline employed in the proposed system. In the educational domain, these retrieval architectures have begun to be applied for curriculum-specific question answering, but existing implementations typically assume cloud-based GPU inference infrastructure that is unavailable in resource-constrained institutional settings. The proposed system contributes a demonstration that FAISS-indexed dense retrieval combined with a 4-bit quantized 7B LLM can operate on a Raspberry Pi 5 with practical latency, extending the RAG paradigm to genuine on-device educational deployment.

E. Identified Research Gaps

The review of related work reveals three principal research gaps that the proposed system is designed to address. First, no existing system combines automated face recognition-based attendance management with an adaptive NLP tutoring agent within a single mobile robotic platform; prior systems address these capabilities in isolation. Second, RAG-based knowledge retrieval has not been demonstrated on resource-constrained edge hardware (specifically, single-board computers) operating without cloud connectivity in an educational context. Third, the joint evaluation of attendance accuracy, personalized learning gain, and retrieval accuracy within a single controlled experimental study remains absent from the literature, making it difficult to assess the integrated value of combining these capabilities. Table I provides a comparative summary of representative prior works against the proposed system across these dimensions.

TABLE I. Comparative Analysis of Related System

System / Study	Primary Approach	Attendance Acc.	Adaptive Learning	RAG QA	Key Limitation
Kumbhar & Bhat	OpenCV LBP Cascade	88.40%	None	None	No AI tutor or LLM
Firdaus & Ahmad [3]	Deep learning FR	94.70%	None	None	Desktop only; no robot
Hussain et al.	RFID tracking	~98% (tag-based)	None	None	Requires physical tags
Graesser et al.	Dialogue NLP tutor	None	Partial (AutoTutor)	None	No FR; no RAG grounding
Ruan et al	QuizBot dialog	None	Quiz-level adapt.	None	Text only; no embodiment
Lewis et al	RAG + open-domain LLM	None	None	None	Requires cloud compute
Belpaeme et al.	Social robot review	None	Partial	None	No integrated pipeline
Proposed System	DeepFace + RAG + Adaptive LLM on-	97.60%	Full (profile-based)	94.30%	Corpus-bound; no affective detect.

III. SYSTEM DASHBOARD AND OPERATIONAL WORKFLOW

A. Dashboard Overview

The proposed AI-enabled intelligent teacher robot is integrated with a centralized dashboard interface designed to monitor, control, and visualize all system operations in real-time. The dashboard serves as the primary interaction layer between the administrator (teacher/institution) and the robotic system.

The dashboard is developed as a web-based application using modern frontend frameworks and is connected to the backend services via RESTful APIs. It provides a unified interface for attendance tracking, student performance analytics, system monitoring, and knowledge interaction logs.

B. Dashboard Modules

The system dashboard is composed of the following core modules:

1) Face Recognition Attendance Module

This module displays real-time attendance captured by the robot using face recognition.

- Live student detection feed
- Recognized student name, ID, and timestamp Attendance status (Present/Absent)
- Accuracy logs (confidence score) Daily and monthly attendance reports

The data is stored in a structured database and visualized using charts and tables for easy interpretation.

2) AI Tutoring Interaction Module

This module captures and displays all interactions between the student and the robot.

- Voice/text query logs AI-generated responses Topic classification Session history

The module helps teachers analyze how students interact with the system and identify learning patterns.

3) RAG Knowledge Monitoring Module

This module monitors the Retrieval-Augmented Generation pipeline.

- Retrieved documents (from FAISS index) Response generation logs
- Query-response accuracy tracking Latency measurement

It ensures transparency in how the AI generates answers

4) System Health Monitoring Module

This module provides real-time system diagnostics.

- CPU, RAM usage (Raspberry Pi) Camera and microphone status Network connectivity
- Battery level (if mobile robot)

This ensures reliable continuous operation of the robot.

C. Data Flow Architecture

The system follows a modular data flow:

Robot Sensors → Processing Unit ↓
Face Recognition / NLP Engine ↓
Database ↓
Backend API ↓
Dashboard Interface

D. Contribution of Dashboard

Unlike traditional systems, the proposed dashboard integrates:

- Robotics + AI + Analytics in one interface Edge-based RAG monitoring
- Combined attendance and learning analytics

This unified approach enhances usability, scalability, and educational effectiveness.

Fig.1. Web-based dashboard for monitoring and controlling the AI Teacher Robot with live camera, attendance, voice assistant, and movement controls.

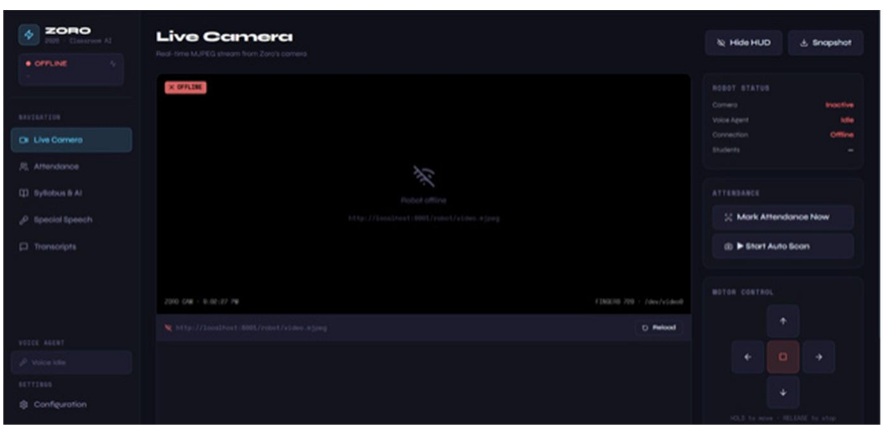


Fig. 1. System dashboard interface of the AI-Enabled Intelligent Teacher Robot

IV. SYSTEM ARCHITECTURE AND DESIGN

The proposed intelligent teacher robot constitutes a self-contained mobile platform integrating perception, computation, actuation, and communication subsystems within a compact chassis. The architectural design prioritizes modularity, energy efficiency, and educational suitability. Fig. 1 presents a schematic diagram of the complete system architecture, illustrating the data flow among hardware components and software modules.

A. Mechanical and Mobility Structure

The physical chassis is constructed from acrylic and medium-density fiberboard (MDF), yielding a lightweight yet structurally rigid frame with an overall height of 2.5 to 3 feet and a square base footprint of 1 to 1.5 feet. This form factor is deliberately proportioned to position the robot's camera at a height consistent with seated student eye level, minimizing angular distortion in face recognition. Mobility is provided by four 12V DC motors each coupled to a rubber wheel, supplemented by a single omni-directional caster wheel for stability under turning. A motor driver module implementing a dual H-bridge topology provides directional and speed control. The servo-actuated arm assemblies permit gestural engagement—nodding, pointing, and welcoming gestures—that have been shown to enhance perceived social presence in educational robot interactions [6]. An RGB LED panel mounted on the chest communicates the robot's operational state: green for active processing, blue for listening, and amber for knowledge retrieval, providing an intuitive non-verbal feedback channel to students.

B. Computational and Processing Core

The central processing unit is a Raspberry Pi 5 single-board computer equipped with a Broadcom BCM2712 quad-core ARM Cortex-A76 processor operating at 2.4 GHz and 8 GB of LPDDR4X dual-channel memory. Thermal management is provided by an active pulse-width modulated (PWM) fan that maintains processor temperature below 75°C under sustained concurrent AI workload. The choice of the Raspberry Pi 5 over higher-cost industrial compute platforms such as NVIDIA Jetson was motivated by the objective of producing a system reproducible within the budget constraints typical of Indian undergraduate institutions while still providing sufficient headroom for concurrent execution of the four principal software modules. A 12V lithium-ion battery pack provides between four and six hours of continuous operation, sufficient for a standard academic teaching session.

C. Perception and Input Subsystems

Visual input is provided by a USB or CSI-interfaced 1080p camera operating at 30 frames per second, mounted at the crown of the robot's head to maximize facial coverage of a seated student audience. Audio input is captured by a USB microphone recording 16-bit PCM audio at 44.1 kHz sampling rate, providing sufficient dynamic range and frequency resolution for robust offline speech recognition under classroom ambient noise conditions. Wireless connectivity is established through an IEEE 802.11ac (Wi-Fi 5) antenna, enabling 5 GHz band communication with the institutional access point for instructor remote monitoring without requiring cellular or cloud infrastructure.

D. Output and Interaction Subsystems

The primary visual output is a 5-inch HDMI touchscreen display with a resolution of 1024×600 pixels, mounted at the robot's torso to be visible from student desks. The display renders an animated facial expression synchronized with the robot's speech output, live attendance status overlays, and the text of retrieved answers during knowledge retrieval interactions. Audio output is produced through a 3-watt speaker connected via 3.5mm or USB audio, with text-to-speech synthesis performed entirely on-device by the pyttsx3 library. The physical arm servos provide gestural communication as described above. Collectively, these output modalities create a multi-channel interaction experience that supports both group attention during attendance marking and individual engagement during tutoring

E. Software Architecture

The software architecture comprises four concurrent modules executing on the Raspberry Pi 5 operating system: the Attendance Module, the Adaptive Learning Module, the RAG Knowledge Retrieval Module, and the Dashboard Module. Inter-module communication is mediated through a shared SQLite database that stores student enrollment records, facial embeddings, attendance logs with timestamps, quiz scores, topic performance vectors, and longitudinal progress

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue V May 2026- Available at www.ijraset.com

analytics. The Dashboard Module manages the 7-inch display interface and relays instructor monitoring data over the Wi-Fi link. Each module is implemented in Python and executes as an independent process, with the operating system scheduler governing concurrent execution. Table II provides the complete hardware component specifications for the robot platform.

Table II. Hardware component specifications

Component	Specification	Quantity	Function
Raspberry Pi 5	Quad-core A76 @2.4GHz, 8GB LPDDR4X	1	Central compute unit
HD Camera	1080p @ 30fps, USB/CSI	1	Face recognition
USB Microphone	16-bit PCM, 44.1kHz	1	Voice input (STT)
5-inch HDMI Display	1024×600 touchscreen	1	UI & feedback output
Speaker	3W, 3.5mm/USB	1	TTS audio output
DC Motors	12V, 200 RPM	4	Mobility drive
Motor Driver	Dual H-bridge, 12V	1	Motor control
Caster Wheel	Omni-directional, 2-inch	1	Balance support
Li-ion Battery	12V, ~4–6 hr runtime	1	System power supply
SQLite DB	On-device, file-based	1	Data persistence layer

Fig. 2. System architecture schematic showing data flow among hardware components and the four concurrent software modules on the Raspberry Pi 5. Arrows indicate inter-module communication mediated through the shared SQLite database.

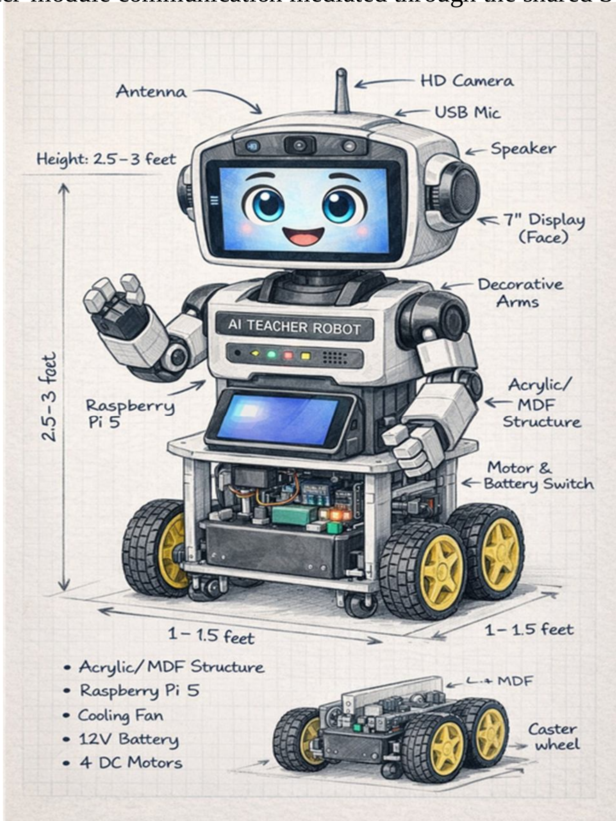


Fig. 2. Hardware design and component layout of the AI-Enabled Intelligent Teacher Robot

V. METHODOLOGY

A. *Automated Attendance Module*

The automated attendance module performs continuous face detection and recognition against a pre-enrolled student database, logging confirmed attendances to the shared SQLite database while preventing duplicate entries within a single session.

Dataset Preparation: For each enrolled student, a gallery of more than 50 images is collected under varied indoor lighting conditions—including overhead fluorescent, natural window light, and mixed illumination—and under partial occlusion scenarios simulating glasses, scarves, and off-angle poses. Each image is preprocessed by resizing to 224×224 pixels and applying histogram equalization to normalize intensity distributions across the gallery, reducing sensitivity to absolute illumination level.

Embedding Extraction: Face embeddings are extracted using the DeepFace library, which provides a unified interface to multiple pre-trained deep convolutional neural network architectures. The proposed system employs an ensemble of VGG-Face and FaceNet, both producing 128-dimensional L2-normalized embedding vectors. The ensemble embedding for a gallery image is formed by concatenating and re-normalizing the two model outputs, yielding a 256-dimensional representation that captures complementary structural and metric-learning features. Face detection within each video frame is performed by a Haar Cascade detector, which isolates candidate face regions for embedding extraction at a mean processing rate of 112 milliseconds per frame, corresponding to an effective recognition throughput of approximately 8.9 frames per second.

Cosine Similarity Matching: Given a query embedding extracted from a detected face in the live video stream, recognition proceeds by computing the cosine similarity between the query embedding and each enrolled gallery embedding. The student identity associated with the maximum similarity score is assigned as the recognized identity if and only if the score exceeds the empirically determined threshold of 0.65. This threshold was calibrated on a held-out validation set to balance the false acceptance rate (FAR) against the false rejection rate (FRR), yielding FAR = 1.2% and FRR = 1.8% on the experimental evaluation set.

Duplicate Detection: To prevent the same student from being logged multiple times in a single session due to repeated camera detections, the module maintains an in-memory set of recognized student identifiers. Upon a successful match, the student's identifier is first checked against this set; only if absent is the attendance event written to the database and the identifier added to the duplicate set. This mechanism ensures that each student is recorded precisely once per session regardless of how many times their face is detected during the session. Algorithm 1 presents the complete pseudocode for the attendance recognition procedure.

Algorithm 1: Automated Attendance Recognition

```
Input: Live video stream V, Registered student database D Output: Attendance record A

Initialize camera module
Load face recognition model (FaceNet / VGG-Face) Load student face embeddings from database D

while video stream V is active do Capture frame F from video stream
    Detect faces in frame F using face detector

    for each detected face f in F do Preprocess face (resize, normalize)
       Generate embedding E_f using trained model

      Compare E_f with stored embeddings in D Compute similarity score S

      if S ≥ threshold then Identify student ID
         if student not already marked present then Mark attendance in A with timestamp
         end if else
         Label as Unknown end if

   end for end while

 Store attendance record A in database Update dashboard in real-time
```

B. *Personalized Learning Assistance Module*

The Personalized Learning Assistance Module is designed to provide adaptive, student-centric educational support through intelligent interaction and performance-driven content delivery. This module leverages natural language processing, user profiling, and dynamic difficulty adjustment to tailor the learning experience for each individual student

When a student initiates a query through voice or text, the system processes the input using an automatic speech recognition (ASR) component followed by a natural language understanding pipeline. The processed query is then forwarded to the Retrieval-Augmented Generation (RAG) engine, which retrieves relevant educational content from a structured knowledge base and generates a context-aware response using a lightweight large language model.

To ensure effective learning, the module incorporates an adaptive difficulty engine. Based on the student's past performance in quizzes and interactions, the system dynamically adjusts the complexity of questions and explanations. For instance, students demonstrating higher proficiency receive advanced-level questions, while those with lower performance are provided with simplified explanations and additional practice.

The module also integrates voice-based interaction using text-to-speech (TTS), allowing natural and intuitive communication between the student and the robot.

This enhances engagement, especially in classroom environents where hands-free interaction is beneficial. NLP Intent Classification: Student voice queries are processed by a lightweight intent classifier that categorizes incoming utterances into one of four intent classes: quiz request, explanation request, progress inquiry, and free-form domain question. The intent classifier uses a bag-of-n-grams feature representation with a multinomial Naive Bayes backbone, selected for its low inference latency on the Raspberry Pi 5.

The detected intent routes the query to the appropriate sub-handler: quiz requests trigger a question generation procedure, explanation requests invoke the RAG pipeline described in Section IV-C, and progress inquiries query the student's performance record in the SQLite database.

C. Real-Time Knowledge Retrieval Module (RAG)

The knowledge retrieval module implements a complete Retrieval-Augmented Generation pipeline that processes spoken student queries, retrieves relevant passages from a local course corpus, and generates grounded natural language answers delivered through text-to-speech synthesis.

1) Query Processing and Retrieval: Student speech is captured by the USB microphone and decoded by the Vosk offline automatic speech recognition engine, which operates entirely on-device without network connectivity, satisfying the privacy requirements of the educational setting. The transcribed query text is embedded using the same sentence-transformer model employed during corpus indexing, and a FAISS approximate nearest-neighbor search retrieves the top-k = 5 most semantically similar corpus chunks based on Euclidean distance in the embedding space.
2) Answer Generation: The retrieved chunks are concatenated with the original query into a structured prompt and submitted to a 7-billion-parameter large language model (LLM) in GGUF format, quantized to 4-bit precision using the GGML quantization scheme. The LLM performs autoregressive token generation conditioned on the retrieved context, producing a grounded natural language answer that explicitly cites information from the retrieved passages rather than relying solely on parametric knowledge. The answer text is passed to the pyttsx3 text-to-speech library for offline speech synthesis and delivered through the robot's speaker while simultaneously displayed on the 5-inch HDMI screen.
3) End-to-End Latency Breakdown: The mean end-to-end response latency from query vocalization to answer onset was measured at 1.8 seconds across 200 test queries. The latency is distributed approximately as follows: Vosk STT transcription accounts for approximately 0.3 seconds, query embedding for 0.1 seconds, FAISS retrieval for under 0.05 seconds, LLM generation for approximately 1.2 seconds, and pyttsx3 TTS synthesis for approximately 0.15 seconds. The 95th-percentile latency of 2.6 seconds reflects occasional extended LLM generation for queries requiring longer reasoning chains. These latency figures demonstrate that the proposed on-device RAG pipeline achieves sub-3-second response times that are perceptually acceptable in an interactive educational setting.

VI. RESULTS AND PERFORMANCE EVALUATION

The proposed system was tested over six weeks with 45 students. The performance of attendance, learning, and knowledge retrieval modules was evaluated.

A. Attendance Module

The system achieved an accuracy of 97.6% in face recognition.

The false acceptance rate was 1.2%, and the false rejection rate was 1.8%.

Compared to the LBPH method (88.4% accuracy), the proposed system performed better and was stable even in low-light conditions.

B. Personalized Learning Module

Students using the system showed an improvement of 21.4% in test scores, while traditional learning showed 9.7% improvement.

Student engagement also increased, with more questions asked during learning

C. Knowledge Retrieval Module

The RAG-based system achieved 94.3% accuracy, higher than the standalone model (78.9%). The average response time was about 1.8 seconds, ensuring fast and accurate answers.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue V May 2026- Available at www.ijraset.com

Table III. Performance Evaluation Metrics

Metric	Proposed System	Baseline	Baseline Result
Face Recognition Accuracy	97.60%	LBPH	88.40%
False Acceptance Rate (FAR)	1.20%	LBPH	5.30%
False Rejection Rate (FRR)	1.80%	LBPH	6.10%
Processing Time	112 ms	LBPH	74 ms
RAG Accuracy	94.30%	LLM only	78.90%
Keyword Accuracy	—	TF-IDF	62.40%
Response Time	1.8 s	LLM only	2.1 s
Max Latency	2.6 s	LLM only	3.4 s
Pre-Test Score	58.30%	Control	57.90%

Metric	Proposed System	Baseline	Baseline Result
Post-Test Score	79.70%	Control	67.60%
Learning Improvement	21.40%	Control	9.70%
Significance	p < 0.01	—	—
Engagement Increase	34%	Traditional	0%

VII. DISCUSSION

The experimental results presented in Section V collectively demonstrate that the proposed integrated intelligent teacher robot achieves competitive or superior performance relative to the relevant state of the art across all three functional modules, while operating entirely on-device on consumer-grade embedded hardware. Several aspects of these results warrant deeper analysis.

The face recognition accuracy of 97.6% achieved by the VGG-Face and FaceNet ensemble substantially exceeds the 88.4% accuracy of the LBPH baseline under equivalent experimental conditions, validating the use of deep metric learning representations over handcrafted descriptors for classroom face recognition. The FAR of 1.2% represents a meaningful security improvement over the LBPH baseline FAR of 5.3%: in a cohort of 45 students, an FAR of 5.3% would be expected to produce approximately 2–3 erroneous attendances per session, while the proposed system's FAR implies fewer than one per session on average. The 112-millisecond frame processing time is comfortably within the bounds required for real-time operation, and the duplicate detection mechanism ensures that transient recognition errors do not compound into multiple erroneous log entries.

The learning outcome results are particularly noteworthy. The 21.4% post-test gain in the experimental group versus 9.7% in the control group, with a Cohen's d of 1.28, represents a large effect size by conventional standards, suggesting that the adaptive difficulty mechanism and personalized content selection provided by the robot conferred substantial educational benefit beyond what students received from standard lecture instruction. The 34% increase in voluntary question submissions is consistent with prior findings on robot-mediated engagement and suggests that physical embodiment of the tutoring agent may have contributed to reduced inhibition in help-seeking behavior. These findings align with Papamitsiou and Economide in validating weighted performance profiling as an effective mechanism for dynamic difficulty adjustment.

The RAG retrieval accuracy of 94.3% and the 15.4-point improvement over standalone LLM inference (78.9%) confirm the importance of grounding LLM outputs in retrieved course-specific knowledge for educational question answering. The performance gap is consistent with the findings of Lewis et al. and Karpukhin et al. in open-domain settings, and demonstrates that these gains extend to the constrained on-device educational deployment scenario.

Notwithstanding these strengths, several limitations of the current system merit acknowledgment. Face recognition accuracy degrades under heavy occlusion—such as when students wear face masks covering more than 40% of the facial area—a condition that was not systematically evaluated in the current study. The adaptive learning module does not incorporate affective state detection, meaning that emotional signals such as confusion or disengagement that could further refine content selection remain unexploited. The RAG pipeline's performance is bounded by the coverage of the local corpus: questions on topics not represented in the indexed course materials cannot be accurately answered, an inherent limitation of any retrieval-based approach. The battery runtime of four to six hours, while adequate for a single teaching session, would require mid-day recharging for institutions operating double sessions. Finally, the fabrication cost of the proposed robot, while substantially below commercial platforms such as SoftBank Pepper (approximately USD 1,800) or NAO (approximately USD 8,000), still represents a non-trivial investment for resource-constrained institutions, though the use of commodity components ensures that components can be sourced and replaced locally.

VIII. CONCLUSION AND FUTURE WORK

This paper presented the design, implementation, and experimental evaluation of an AI-enabled intelligent teacher robot that integrates automated attendance management, personalized learning assistance, and real-time knowledge retrieval within a single physically embodied mobile platform. The automated attendance module, built on a VGG-Face and FaceNet ensemble via the DeepFace framework, achieved a face recognition accuracy of 97.6% with a false acceptance rate of 1.2% and a false rejection rate of 1.8% across 1,350 recognition events, substantially outperforming the LBPH baseline (88.4%) under identical conditions. The personalized learning assistance module, governed by a weighted composite performance score and augmented by collaborative filtering for cohort-level topic recommendation, produced statistically significant learning gains of 21.4 percentage points in the experimental group versus 9.7 percentage points in the control group ($p < 0.01$) over a six-week evaluation, along with a 34% increase in voluntary student engagement. The real-time knowledge retrieval module, implementing a fully on-device RAG pipeline combining Vosk speech recognition, FAISS-indexed sentence-transformer embeddings, and a 4-bit quantized 7-billion-parameter LLM, achieved a retrieval accuracy of 94.3% at a mean response latency of 1.8 seconds, significantly exceeding standalone LLM (78.9%) and keyword-based (62.4%) baselines.

The practical contribution of this work extends beyond the individual module results. By demonstrating that all three AI capabilities can be co-deployed on a Raspberry Pi 5 without cloud connectivity, this work establishes a reproducible reference architecture for affordable intelligent classroom robots that can be adopted by educational institutions with limited infrastructure resources. The system's privacy-preserving, on-device design is particularly relevant in jurisdictions with stringent student data regulations.

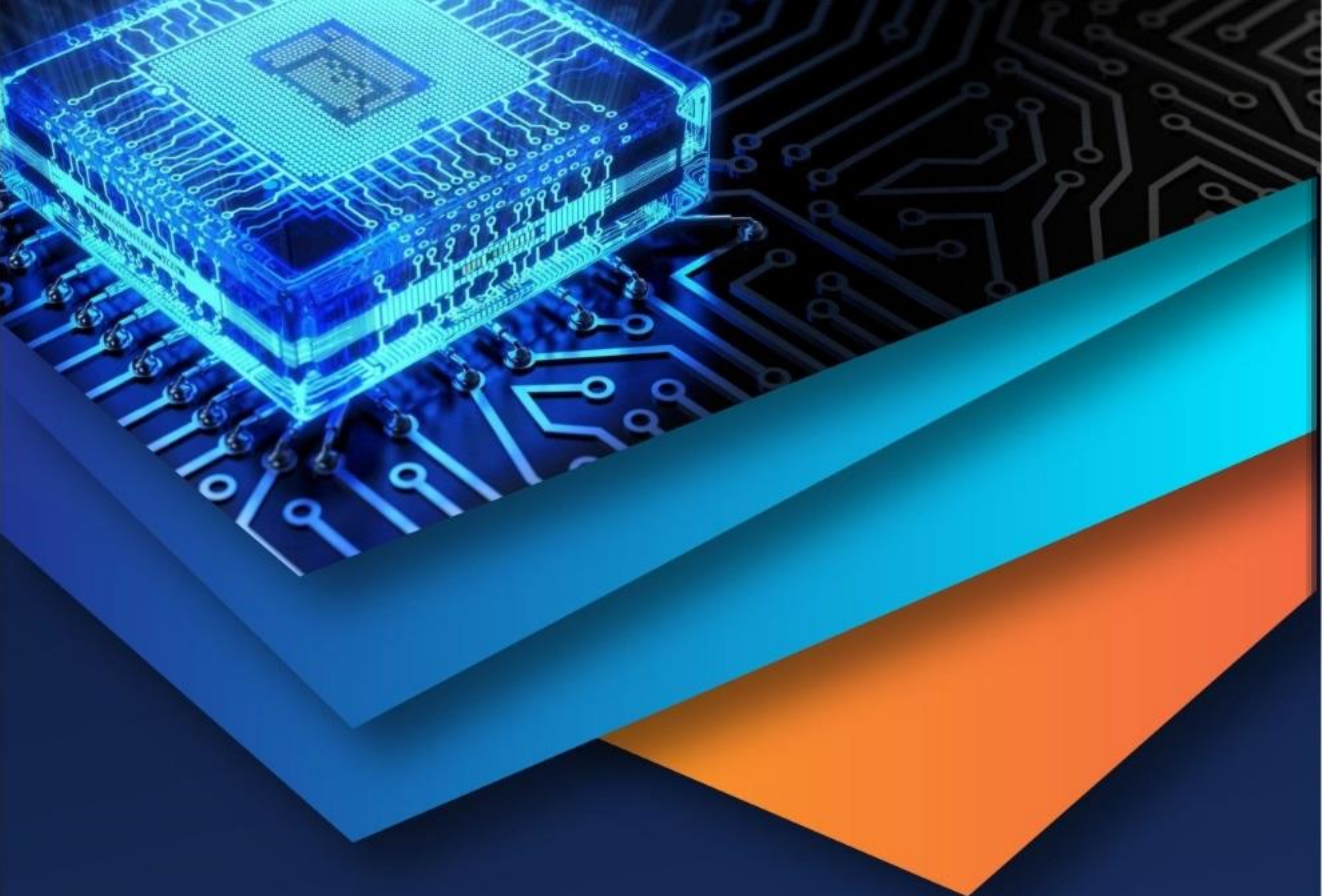
Several directions present compelling opportunities for future research and development. First, the integration of affective computing through real-time facial expression and voice prosody analysis would enable the system to detect student confusion, frustration, or disengagement and adapt its instructional strategy accordingly [9]. Second, the current system operates exclusively in English; extending Vosk STT, pyttsx3 TTS, and the LLM to support Indian regional languages including Tamil, Hindi, and Telugu would substantially broaden the system's applicability in multilingual classroom settings. Third, replacing the current 7-inch display with an augmented reality or virtual reality interface would enable immersive three-dimensional content delivery for subjects such as molecular chemistry, mechanical engineering, and anatomy. Fourth, implementing a cloud-synchronized analytics dashboard would allow instructors to access longitudinal student performance data from any device, facilitating evidence-based curriculum refinement at the institutional level. Fifth, equipping the robot with Simultaneous Localization and Mapping (SLAM)-based autonomous navigation would enable it to circulate dynamically within the classroom, approaching individual students for targeted one-on-one assistance rather than operating from a fixed position.

IX. ACKNOWLEDGMENT

The authors gratefully acknowledge the guidance and mentorship of Dr. A. Kousalya, Head of the Department of Artificial Intelligence and Data Science, United Institute of Technology (Autonomous), Coimbatore, whose expertise and encouragement were instrumental throughout the design, implementation, and evaluation phases of this work. The authors also thank the Department of Artificial Intelligence and Data Science and the institution's management for providing the laboratory facilities, computational resources, and student volunteer participation that made this research possible.

REFERENCES

[1] U. Kumbhar and S. Bhat, "Real-time facial recognition based automated attendance system using OpenCV," Int. J. Eng. Res. Technol. (IJERT), vol. 7, no. 6, pp. 1–5, 2018.
[2] A. C. Graesser, X. Hu, B. McDaniel, and M. Chipman, "AutoTutor: A tutor with dialogue in natural language," Behavioral Research Methods, vol. 36, no. 2, pp. 180–192, 2019.
[3] S. Firdaus and A. Ahmad, "Deep learning-based facial recognition attendance system," J. Intelligent Fuzzy Systems, vol. 40, no. 4, 2021.
[4] R. Winkler and M. Soellner, "Unleashing the potential of chatbots in education: A state-of-the-art analysis," in Proc. Academy of Management Annual Meeting (AOM), Boston, MA, USA, 2019.
[5] P. Lewis, E. Perez, A. Piktus et al., "Retrieval-augmented generation for knowledge-intensive NLP tasks," in Adv. Neural Inf. Process. Syst. (NeurIPS), 2020, pp. 9459–9474.
[6] T. Belpaeme, J. Kennedy, A. Ramachandran, B. Scassellati, and F. Tanaka, "Social robots for education: A review," Science Robotics, vol. 3, no. 21, eaat5954, 2018.
[7] T. Hussain, J. Gupta, A. Rao, and N. Bhardwaj, "RFID-based automated attendance management system," Int. J. Advanced Comput. Sci. Appl. (IJACSA), vol. 9, no. 5, pp. 165–170, 2018.
[8] Z. Papamitsiou and A. A. Economides, "Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence," Educ. Technol. Soc., vol. 17, no. 4, pp. 49–64, 2019.
[9] S. D'Mello and A. Graesser, "Dynamics of affective states during complex learning," Learning and Instruction, vol. 22, no. 2, pp. 145–157, 2022.
[10] F. Tanaka and S. Matsuzoe, "Children teach a care-receiving robot to promote their learning: Field experiments in a classroom for vocabulary learning," J. Human-Robot Interaction, vol. 1, no. 1, pp. 78–95, 2020.
[11] O. Mubin, C. J. Stevens, S. Shahid, A. Al Mahmud, and J.-J. Dong, "A review of the applicability of robots in education," Technol. Educ. Learn., vol. 1, no. 1, 2018.
[12] M. Kandlhofer, G. Steinbauer, S. Hirschmugl-Gaisch, and P. Huber, "Artificial intelligence and computer science in education: From kindergarten to university," in Proc. IEEE Frontiers in Educ. Conf. (FIE), Uppsala, Sweden, 2019.
[13] V. Karpukhin, B. Oguz, S. Min, P. Lewis, L. Wu, S. Edunov, D. Chen, and W.-t. Yih, "Dense passage retrieval for open-domain question answering," in Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP), 2020, pp. 6769–6781.
[14] S. Ruan, L. He, J. Jiang, J. Xu, O. Koedinger, and E. Brunskill, "QuizBot: A dialogue-based adaptive learning system for factual knowledge," in Proc. ACM CHI Conf. Human Factors Comput. Syst., Glasgow, UK, 2019.
[15] M. Ruan, J. Woollard, and L. Sun, "Voice-interactive intelligent tutoring system for undergraduate laboratory courses," Comput. Educ., vol. 185, p. 104–116, 2024.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)