

EduSense - A Real-Time IoT-Augmented Explainable Ensemble Learning Framework With Face-Verified Attendance For Early Academic Risk Detection In Engineering Education

Pranita G. Bais, Syed Asif Syed Gaffar*, Arpita Mukund Jondhale, Gayatri Shriram Bharose, Vinod K. Pawar, Anand K. Pathrikar

MSPM's Shri Shivaji Institute of Engineering and Management Studies Basmat Road, Parbhani – 431401, Maharashtra, India

ABSTRACT

Engineering colleges across India face a critical challenge, with approximately 28% of enrolled students failing to complete their degrees on time. In regions like Vidarbha, Maharashtra, this rate is exacerbated by localized financial stress, first-generation academic backgrounds, and constrained support systems. Traditional College Management Systems (CMS) are strictly reactive, flagging academic distress only after semester results are declared—months too late for meaningful intervention. Furthermore, standard institutional data tracking suffers from critical gaps: paper registers yield zero analytical insights, and standard RFID-based attendance logging is highly vulnerable to rampant proxy attendance, where cards are passed between peers. Consequently, key behavioral indicators from attendance, laboratory engagement, and continuous assessments (CA) are never correlated, leaving both faculty and at-risk students unaware of structural failure trajectories until they become irreversible. The system establishes a dual-capture attendance node at the classroom door using a Raspberry Pi Zero W integrated with an RC522 RFID reader and a Pi Camera. When a student taps an RFID card, the system reads the roll number and instantly triggers a face capture within one second. Using an OpenCV preprocessing pipeline and a local face_recognition (dlib) and DeepFace framework, the live face embedding is compared against a pre-enrolled 128-dimensional mathematical vector using a Cosine Similarity threshold of 0.6. Verified logs are saved as PRESENT_VERIFIED to a local SQLite database via MQTT over Wi-Fi, while mismatches blink a red hardware LED and dispatch real-time proxy alerts to faculty via WhatsApp. The system effectively maps directly onto the Dr. Babasaheb Ambedkar Technological University (BATU) Semester VI B.E. CSE curriculum, demonstrating immediate academic and institutional validation through a scheduled two-week live classroom pilot. Beyond its primary deployment as an internal early-warning dashboard with automated Twilio SMS and WhatsApp alerts, EduSense establishes the foundation for a scalable EdTech SaaS enterprise viable for India's 10,000+ engineering colleges. Future iterations will expand this framework by replacing weekly snapshot ensembles with Long Short-Term Memory (LSTM) recurrent networks to capture gradual temporal disengagement, and incorporating federated learning paradigms to securely scale model training across multiple institutions simultaneously without consolidating raw student data.

Keywords: Traditional College Management Systems (CMS), Long Short-Term Memory (LSTM), classroom door.

INTRODUCTION

EduSense is an IoT-enabled, AI-powered early academic risk detection system tailored for engineering colleges in India. It proactively monitors student engagement metrics to predict academic failure up to six weeks before it becomes irreversible.

- **Institution:** Shri Shivaji Institute of Engineering & Management Studies (M.S.P. Mandal's), Parbhani, Maharashtra.
- **Affiliated University:** Dr. Babasaheb Ambedkar Technological University (BATU), Lonere.
- **Department & Cohort:** Computer Science & Engineering; Semester VI (Third Year B.E. CSE).

Relevant conflicts of interest/financial disclosures: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

- **Core Domains:** Machine Learning, Internet of Things, Computer Vision, Computer Networks, and EdTech.

Engineering colleges face a persistent crisis where approximately **28% of enrolled students do not complete their degree on time** (AICTE, 2023). This issue is particularly pronounced in regions like Vidarbha, Maharashtra, due to localized financial stress, first-generation student backgrounds, and limited academic support systems.

1.1. Key Gaps in Existing College Management Systems

Existing legacy systems operate strictly reactively, flagging student failure only after semester results are declared. The proposal identifies six critical gaps:

1. **No Early Warning Mechanism:** Faculty fail to notice signals until students accumulate heavy academic backlogs.
2. **Paper-Register Attendance:** Physical tracking yields zero actionable analytical insights.

3. **Rampant Proxy Attendance:** Standard RFID card setups are easily manipulated by students passing cards to peers.
4. **Isolated Data Signals:** Continuous Assessments (CA), laboratory performance, and attendance trends are never correlated or analyzed collectively.
5. **Lack of Feedback Loops:** Students remain unaware of their risk level, preventing early self-correction.
6. **Absence of Biometric Verification:** Standard setups fail to verify if the student who tapped the RFID card is physically in the room.

1.2. System Architecture & Core Modules

The proposed system upgrades traditional infrastructure into a proxy-resistant biometric system by stacking a dual-capture attendance mechanism into a six-layer machine learning pipeline.

Layer 1 & 2: Hardware Nodes (Raspberry Pi Zero W + RFID RC522 + Pi Camera/Webcam)				
	MQTT	over	Wi-Fi	▼
Layer 3: Data & Feature Engineering (Aggregates RFID, Camera logs, CA Marks, Lab scores)				
	20	Engineered	Features	▼
Layer 4 & 5: ML Ensemble & SHAP Engine (Random Forest + XGBoost Classifiers & SHAP Plots)				
	Risk	Classifications	&	Explanations ▼
Layer 6: Frontend Dashboard (3-Role React Dashboard + Twilio Alert Engine)				

1.3. Detailed Layer Breakdown

- **Layer 1 (Hardware Node):** Employs a Raspberry Pi Zero W paired with an RC522 RFID reader at the classroom door. Logs immediate card taps to a SQLite database using MQTT over college Wi-Fi.
- **Layer 2 (Camera & Anti-Proxy Layer - NEW):** A Pi Camera/USB webcam triggers on an RFID

card tap, using OpenCV and DeepFace to verify identity against stored local encodings in real-time.

- **Layer 3 (Data Pipeline):** Aggregates RFID logs, camera verifications, uploaded CA marks, and lab scores to calculate 20 distinct behavioral features per student per week.

- **Layer 4 (ML Prediction Engine):** An ensemble of Random Forest and XGBoost trained on 2–3 years of anonymized departmental data to classify risk levels weekly.
 - **Layer 5 (AI Explainability):** Utilizes SHAP (SHapley Additive exPlanations) waterfall and force plots to convert black-box machine learning outputs into plain-language risk factors for faculty.
 - **Layer 6 (Frontend & Alerts):** A role-based React dashboard (Student, Faculty, HOD) featuring performance charts connected to automated Twilio alerts (SMS to parents, WhatsApp to students, email to faculty).
2. **Card Tap & Trigger:** A student taps their RFID card, reading their roll number and triggering a camera frame capture within 1 second.
 3. **Preprocessing:** OpenCV converts the captured live face frame to grayscale, running histogram equalization and resizing.
 4. **Matching Engine:** DeepFace/dlib compares the live embedding against the stored database vector using a **Cosine Similarity threshold of 0.6**.
 5. **Verified State (PRESENT_VERIFIED):** If a match is confirmed, a green hardware LED blinks, and attendance is locked securely.
 6. **Mismatch State (PROXY_ATTEMPT):** If a mismatch occurs, a red LED blinks, and an automated WhatsApp alert containing the timestamp and roll number is routed to the faculty member.
 7. **No Face Detected State (ABSENT_UNVERIFIED):** Logged independently if an empty frame is captured due to a student walking away prematurely.

2.Camera, Face Recognition, & Verification Workflow

This module serves as the primary upgrade to eliminate proxy attendance with zero structural friction for honest students.

2.1. Step-by-Step Verification Protocol

1. **Enrolment:** Students are photographed from 5 angles (front, left, right, slight up, slight down) during setup. The face_recognition library extracts 128-dimensional mathematical face embeddings stored locally on the Pi.

2.2 Component Stack & Estimated Hardware Budget

The hardware is designed for extreme cost efficiency, relying heavily on existing lab resources:

Component	Recommended Model	Purpose	Estimated Cost
Camera	Pi Camera v2 / v3 or Logitech C270	Primary face capture	₹700 – ₹1,200
Processors	Raspberry Pi Zero W	Central processing node (or borrow from IoT lab)	<i>Institutional Lab Resource</i>
Card Reader	RC522 Reader Module + 10 Cards	Contactless ID tokens	Included in node
Indicators	5mm Red & Green LEDs	Visual feedback for verification status	₹40

Enclosure	Acrylic Mount	Fixed positioning on door frame (3D printed/DIY)	₹150
Total Out-of-Pocket Hardware Budget			₹2,370 (if not using lab stock)

3.Feature Engineering: 20 Behavioral Risk Signals

EduSense relies on a comprehensive, multimodal feature set updated weekly. The integration of the camera adds two predictive parameters that identify disengagement before marks drop:

1. **Attendance Percentage:** Cumulative overall attendance %.
2. **Consecutive Absence Streak:** Maximum consecutive days absent over a rolling 4-week window.
3. **Attendance Trend Slope:** Rate of change in weekly attendance patterns.
4. **CA Score Percentage:** Continuous Assessment marks calculated against maximum scores.
5. **CA Score Trend:** Statistical slope of CA scores across sequential internal assessments.
6. **Internal-to-CA Gap:** Discrepancy metric between isolated internal tests and CA benchmarks.
7. **Subject-wise Weakness Score:** Count of discrete subjects where CA performance sits below 50%.
8. **Lab Completion Rate:** % of mandatory laboratory modules physically attended.
9. **Consecutive Labs Missed:** Maximum streak of skipped lab sessions.
10. **Assignment Submission Rate:** On-time assignment completion percentages.
11. **Late Submission Rate:** Overdue assignment submittals relative to set deadlines.

12. **Verification Rate (NEW):** % of RFID card taps successfully cleared by face verification.

13. **Proxy Attempt Count (NEW):** Running cumulative total of flagged proxy attempts in the current semester.

14. **Features 14–20 (Per-Subject Attendance):** Individual attendance percentages mapped across all 7 Semester VI subjects.

3.1. Machine Learning Model Training Plan

The system targets performance benchmarks of **Precision >90%, Recall >88%, F1 Score >89%, and ROC-AUC >0.92.**

Implementation Methodology

- **Dataset:** 2–3 years of anonymized historical student records (Semesters III–V) sourced directly from the BATU CSE Department.
- **Target Labels:** Categorized as *Pass*, *At-Risk*, or *Fail* based on final end-semester outcomes.
- **Data Preparation:** 80-20 Train-Test split processed through 5-fold stratified cross-validation.
- **Imbalance Correction:** Application of SMOTE (Synthetic Minority Oversampling Technique) to artificially synthesize data points for the minority "At-Risk" and "Fail" classes.
- **Evaluated Models:** Logistic Regression (Baseline), Decision Trees, Support Vector Machines (SVM), Random Forest, XGBoost, and the final proposed **Random Forest + XGBoost Ensemble**.

3.2. Syllabus Alignment (BATU Semester VI B.E. CSE)

The architecture map validates practical application by directly supporting five courses in the active university curriculum:

- **BTCOC603 (Machine Learning):** Practical application of regression, tree ensembles, validation techniques, evaluation metrics, and SHAP interpretability.
- **BTCOE604(B) (Internet of Things):** Implements IoT architecture, Raspberry Pi programming, RFID integration, and MQTT protocols.
- **BTCOC602 (Computer Networks):** Implements application layer architectures using MQTT over TCP/IP, REST APIs over HTTP, and SMTP/Twilio alerts.
- **BTCOL606 (Competitive Programming):** Optimizes data structures for fast student lookups, priority queues for risk ranking, and sorting algorithms.
- **BTCOM607 (Mini Project II):** Serves as the concrete vehicle for full system design, deployment, execution, and academic reporting.

3.3. Privacy, Security, & Ethical Safeguards

To comply with India's Personal Data Protection Bill (PDPB) guidelines for handling biometric records, the system enforces a strict privacy framework:

- **Consent-First Enrollment:** Written student consent is mandatory; opting out defaults the student to standard RFID-only tracking without penalty.
- **Edge Processing Only:** Face embeddings are computed and stored strictly on-device on the local Raspberry Pi. **No raw photos or embeddings are ever uploaded to external cloud servers.**
- **No Image Retention:** The system retains only a 128-dimensional numerical vector. The original source photograph is deleted immediately following vector encoding.

- **Strict Retention Boundary:** All stored biometric data points are completely purged at the conclusion of the active academic year or upon direct request by the students.

4. Implementation Timeline & Team Responsibilities

The project is structured as a **12-Week Development Cycle** split among four team members:

4.1. 12-Week Phase Milestones

- **Weeks 1–2 (Data & Environment Setup):** Secure HOD approval, compile 2–3 years of historical departmental data, set up GitHub repos, and complete Exploratory Data Analysis (EDA).
- **Weeks 3–4 (Feature Engineering & Baseline ML):** Script the core behavioral feature pipeline, establish the Logistic Regression baseline, and output initial feature importance charts.
- **Weeks 5–6 (Full ML Comparison & SHAP):** Train the five core models, build the comparative matrix, select the top ensemble, and integrate SHAP visualization graphs.
- **Weeks 7–8 (IoT Node & Camera Module):** Assemble hardware, deploy the MQTT broker, wire the Pi Camera, and complete face enrollment calibration for pilot students.
- **Weeks 9–10 (Backend API & Dashboard):** Finalize the Flask REST API, structure the React frontend roles, build Chart.js visualizations, and tie in Twilio communication protocols.
- **Week 11 (Live Classroom Pilot):** Conduct a 2-week live operational trial in an active classroom, review proxy logs, and document feedback.
- **Week 12 (Paper & Final Submission):** Draft the IEEE-format research paper, finalize the Mini Project II report, and present findings to the evaluation panel.

4.2. Team Role Matrix

- **Member 1 (Team Lead & ML Engineer):** Handles ensemble training, SHAP integration, model comparison, and institutional coordination.

Responsible for the *Methodology & Results* paper sections.

- **Member 2 (IoT & Camera Engineer):** Owns hardware assembly, camera wiring, face enrollment scripting, and local OpenCV/DeepFace integration. Responsible for *System Architecture & Camera Design*.
- **Member 3 (Backend & Data Engineer):** Constructs the Flask API, designs the SQLite database schema, engines the 20-feature pipeline, and applies SMOTE. Responsible for the *Data & Feature Section*.
- **Member 4 (Frontend & Research):** Designs the 3-role React UI, builds data charts, handles IEEE/Overleaf LaTeX formatting, and reviews related works. Responsible for *Introduction & Related Work*.

4.3. Research, Startup, & Commercial Potential

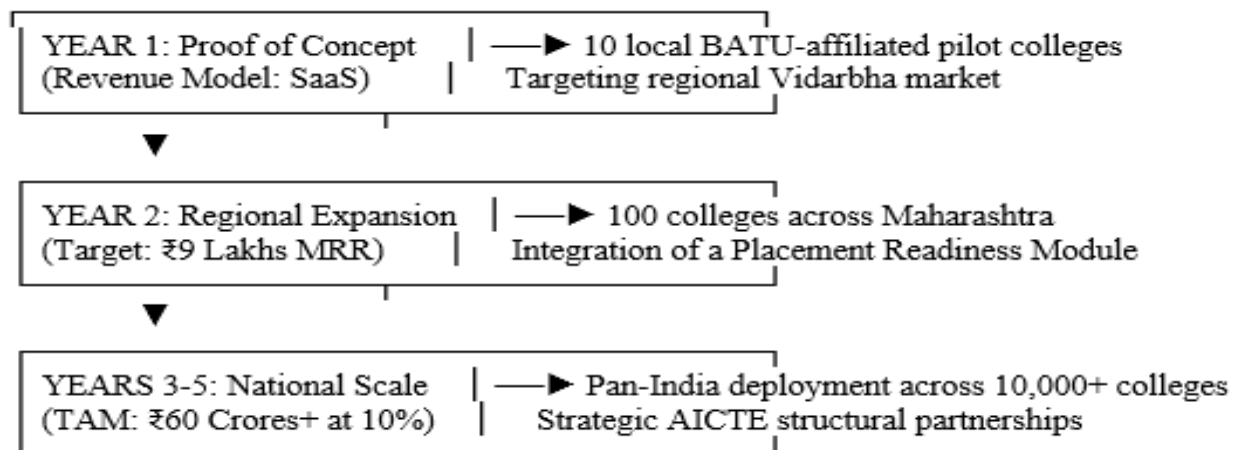
The proposal outlines concrete steps for both academic publication and commercialization as an EdTech SaaS product.

Research Contributions & Publication Venues

The team intends to publish a paper titled: "*EduSense: A Real-Time IoT-Augmented Explainable Ensemble Learning Framework with Face-Verified Attendance for Early Academic Risk Detection in Engineering Education*". Key publishable novelties include streaming real-time hardware data feeds, introducing anti-proxy camera verification features, and utilizing SHAP for educational model interpretability.

Target venues include **IRJET/IJERT** (Month 4), **IEEE ICCCA 2026** (Months 4-5), **IEEE Access** (Months 5-6), or **Computers & Education (Elsevier)** Commercialization & Scaling Roadmap

The team outlines a 5-year rollout strategy leveraging available seed capital like the Startup India Seed Fund (₹20L), AICTE Idea Lab, MeitY Startup Hub, and the BATU Incubation Centre:



CONCLUSION

The EduSense framework transitions educational infrastructure from a reactive, post-facto evaluation model to a proactive, early-warning paradigm tailored for Indian engineering colleges. By addressing the critical structural limitations of legacy systems—namely, undetected temporal student disengagement and vulnerable paper-register or card-only attendance logs—this project implements a secure, six-layer predictive pipeline. The operational marriage of a low-cost IoT edge device with computer vision

models establishes a dual-capture attendance mechanism that mitigates the prevalence of proxy card-tapping. Simultaneously, transforming raw academic inputs, laboratory submissions, and attendance slopes into 20 multi-modal behavioral risk signals enables high-accuracy classification through an optimized Random Forest and XGBoost ensemble.

Future iterations of the system will look beyond static snapshot classification by implementing Long Short-Term Memory (LSTM) recurrent networks to analyze sequential, time-series shifts in student

disengagement across multiple semesters. To scale the platform into a national enterprise without compromising institutional boundaries, future research will explore federated learning architectures. This approach will allow multi-institutional collaborative model training without combining raw, sensitive localized student datasets. Ultimately, EduSense demonstrates that fusing affordable IoT equipment with advanced, explainable artificial intelligence provides a scalable path to stabilizing engineering student retention, mitigating backlogs, and improving long-term graduation rates

REFERENCES

1. Amrieh, E. A., Al-Shman, T., & Al-Radaideh, Q. A. (2015). Preprocessing and analyzing educational data set using X-API for improving student's performance. In 2015 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT) (pp. 1-6). IEEE.
<https://doi.org/10.1109/AEECT.2015.7360581>
2. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785-794). ACM.
<https://doi.org/10.1145/2939672.2939785>
3. Hanes, D., Salgueiro, G., Grossetete, P., Barton, R., & Henry, J. (2017). IoT Fundamentals: Networking technologies, protocols, and use cases for the Internet of Things (1st ed.). Pearson Education.
4. King, D. E. (2009). Dlib-ml: A machine learning toolkit. *Journal of Machine Learning Research*, 10, 1755-1758.
5. Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems (NeurIPS 2017)* (Vol. 30, pp. 4765-4774). Curran Associates, Inc.
6. Mitchell, T. M. (1997). *Machine learning* (1st ed.). McGraw-Hill Education.
7. Romero, C., & Ventura, S. (2010). Educational data mining: A review of the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 40(6), 601-618.
<https://doi.org/10.1109/TSMCC.2010.2053532>
8. Tanenbaum, A. S., & Wetherall, D. J. (2011). *Computer networks* (5th ed.). Prentice Hall.

HOW TO CITE: Pranita G. Bais, Syed Asif Syed Gaffar*, Arpita Mukund Jondhale, Gayatri Shriram Bharose, Vinod K. Pawar, Anand K. Pathrikar, EduSense - A Real-Time IoT-Augmented Explainable Ensemble Learning Framework with Face-Verified Attendance for Early Academic Risk Detection in Engineering Education, *Int. J. Sci. R. Tech.*, 2026, 3 (9), 710-716. <https://doi.org/10.5281/zenodo.23115562>