BEGUM ROKEYA UNIVERSITY, RANGPUR

THESIS REPORT

A novel approach to evaluate pedagogical outcomes using multimodal data

---

Submitted By:
Md. Shafiqul Islam
ID: 1905028
Registration No: 000012756
Session: 2019-20
Department of Computer Science and Engineering

Supervisor:
Dr. Md. Mizanur Rahoman
Professor
Department of Computer Science and Engneering
Begum Rokeya University, Rangpur

A thesis report submitted for
*the course **PROJECT/THESIS (CSE 4207)***
in fulfilment of the
requirements for the degree of Bachelor of Science
in the
Department of Computer Science and Engineering

September,2025

Certificate of Originality

This is to certify that the work presented in this thesis titled "**A novel approach to evaluate pedagogical outcomes using multimodal data**" is my own original work and has not been submitted previously for any degree or diploma at any university or institution. All sources of information and references used in this thesis have been duly acknowledged and cited. I affirm that this thesis represents my own research and findings, and any assistance received during the research process has been appropriately credited.

I understand the importance of academic integrity and the consequences of plagiarism. I declare that this thesis is a true reflection of my own work and adheres to the ethical standards of research.

_____

Md Shafiqul Islam
Computer Science And Engineering
Begum Rokeya University, Rangpur

Certificate of Approval

This is to certify that the thesis titled "**A novel approach to evaluate pedagogical outcomes using multimodal data**" submitted by Md Shafiqul Islam to the Department of Computer Science and Engineering, Begum Rokeya University, Rangpur, in partial fulfillment of the requirements for the degree of Bachelor of Science in Computer Science and Engineering, is a record of an original research work carried out by him under my supervision and guidance. To the best of my knowledge, the thesis has not formed the basis for the award of any degree, diploma, or other similar titles previously.

_____

Dr. Md. Mizanur Rahoman
Supervisor
Computer Science And Engineering
Begum Rokeya University, Rangpur

Acknowledgments

At first I would like to express my gratitude to THE ALMIGHTY ALLAH, the most gracious and the most merciful, for giving me the strength and ability to complete this thesis. Then I would like to show my deeepest gratitude to my supervisor, Dr. Md. Mizanur Rahoman, for his continuous support, guidance, and encouragement throughout my research journey. His insightful feedback and unwavering belief in my potential have been instrumental in shaping this work.

Dedicated to my beloved parents,
for their unwavering love, support, and sacrifices
that made this journey possible.

Abstract

Effective assessment of the teaching process is vital for upholding the quality of educational practice, but current methods for assessing educational practices commonly assess students' feedback, which are subjectively based, biased, and less than reliable. Hence, we felt that systematic data-driven evaluation of the enacted practice of multitodal interaction within the classroom would be a better measure of pedagogical effectiveness. We began to investigate this way of thinking by creating a system that analyzed classroom video, audio, and transcripts to record aspects of instruction. The visual modality was processed for teacher movement and interaction, audio was processed for voice and prosody, and the linguistic modality was processed for coherence and organization. Logistic regression models were then used to predict instruction outcomes as measured by students' assessment of learning. The results demonstrate that, while the linguistic features provided the best unimodal performance in accuracy, they were only 67.0%. The multimodal fusion models increased predictive power to 71.33% accuracy and provided better cross-validation stability, achieving an ROC score of 0.796. All of theses results support our hypothesis that multimodal integration models productive pedagogic modes that may not be represented by singular modalities. This work also highlights how multimodal learning analytics can draw attention to feedback that is objectively scalable and eventually leads to automated systems for assessing teaching.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The contemporary educational landscape demands innovative approaches to assess and enhance teaching quality across diverse academic institutions [12]. While traditional evaluation methods have served educational systems for decades, emerging technologies present unprecedented opportunities to transform how we understand and measure pedagogical effectiveness [2]. The limitations inherent in conventional assessment approaches, particularly the subjective nature of student feedback have prompted researchers and educators to explore alternative methodologies that offer greater objectivity, consistency, and actionable insights [12]. These challenges encompass issues such as demographic bias, inconsistent rating criteria, and the influence of factors unrelated to actual teaching competency [27]. Studies reveal that gender, ethnicity, and personal charisma can significantly skew student evaluations, potentially undermining fair assessment practices [27].

Despite these drawbacks, student feedback remains prevalent due to its simplicity, cost-effectiveness, and ability to capture students' perspectives on teaching quality [2]. However, the reliability of student feedback is challenged by inconsistencies across different cohorts and courses, as well as by the tendency for students to focus on surface-level attributes rather than deeper pedagogical competencies [3]. Several studies have highlighted the need for more robust and objective evaluation mechanisms to complement or replace traditional feedback [10].

Recent advancements in educational technology have paved the way for more objective and comprehensive evaluation methods [29]. Among these, multimodal systems that leverage data from multiple sources such as audio, video, gesture recognition, and classroom interaction analytics offer promising alternatives [28]. These systems can provide a holistic view of teacher performance by capturing a wide range of behavioral and communicative cues that are often overlooked in traditional feedback mechanisms.

Artificial intelligence (AI) and machine learning (ML) techniques have enabled the automated analysis of complex classroom behaviors, including teacher movement, speech patterns, engagement strategies, and student responses [29]. For instance, pose estimation algorithms can detect and classify teaching activities, while emotion recognition systems can assess the affective climate of the classroom. These technologies not only enhance the objectivity of evaluations but also provide granular feedback that can inform targeted professional development [31].

Despite the potential of multimodal systems, there is a lack of empirical studies comparing their effectiveness with conventional student feedback [10]. This research aims to bridge this gap by conducting a comparative study between a multimodal teacher performance evaluation system and traditional student feedback methods. The primary objectives of this study are to:

- Analyze the strengths and weaknesses of both evaluation approaches [3, 12, 28].
- Assess the reliability and validity of multimodal data in reflecting true teaching effectiveness [29–31].
- Investigate the correlation between multimodal system verdicts and student perceptions [2, 7, 14].
- Provide recommendations for integrating advanced evaluation systems into existing educational frameworks [10, 13, 28].

The integration of multimodal systems into teacher evaluation frameworks represents a significant shift in educational assessment paradigms. These systems not only address the limitations of traditional feedback but also align with broader trends in educational technology and data-driven decision-making. For instance, the use of machine learning algorithms to analyze multimodal data streams enables the identification of nuanced teaching behaviors that correlate with student engagement and learning outcomes. Furthermore, the adoption of such systems has practical implications for teacher training and professional development, as they provide actionable insights that can guide instructional improvement.

However, the implementation of multimodal evaluation systems is not without challenges. Issues such as data privacy, the need for robust validation across diverse educational contexts, and the potential for algorithmic bias must be carefully considered. Despite these challenges, the potential benefits of multimodal systems—such as enhanced reliability, validity, and comprehensiveness—make them a promising complement to traditional student feedback. This study aims to explore these dynamics by conducting a comparative analysis of both evaluation approaches, thereby contributing to the growing body of literature on innovative educational assessment methods.

The remainder of this thesis is organized as follows: **Chapter 2** provides a comprehensive literature review of teacher evaluation methodologies and multimodal assessment systems, establishing the theoretical foundation for this research. **Chapter 3** presents the research methodology,

detailing the comparative analysis framework, data collection protocols, and evaluation metrics used to assess both traditional student feedback and multimodal system performance. **Chapter 4** describes the experimental design and setup, including participant selection, classroom environments, hardware configurations, and dataset characteristics. **Chapter 5** details the multimodal system architecture and implementation, presenting the integrated pipeline for processing audio, video, and behavioral data streams, along with the AI-driven analysis components. **Chapter 6** presents the comprehensive results of the comparative study, analyzing the correlation between multimodal system outputs and student feedback, discussing reliability and validity findings, and examining the practical implications for educational assessment. Finally, **Chapter 7** concludes with key findings, recommendations for integrating multimodal evaluation systems into educational frameworks, limitations of the current study, and directions for future research.

Chapter 2

Problem Statement

It is important to evaluate teaching effectiveness in order to support the improvement of educational quality. Traditional evaluation approaches such as student surveys and peer evaluations tend to rely on subjective judgments, which implies the selection of certain facets of the classroom which may not encompass the full range of teaching effectiveness. Furthermore, their judgments may also be influenced by other biases or may overlook pertinent aspects of the classroom context regarding teaching effectiveness. Multimodal data analysis represents an emerging field of research that creates opportunities to evaluate teaching effectiveness more comprehensively. While it is true that multimodal data can provide a richer understanding of teaching effectiveness though visual, auditory, and text-based data from classroom interactions, there are still challenges to using multimodal data for this purpose. First, it is challenging to pinpoint and extract relevant features from the multiplicity of data streams. Second, it is difficult to align and bring together multimodal data to be utilized as a coherent representation of teaching effectiveness. Third, establishing models to evaluate multimodal data to identify teaching effectiveness poses a further level of complexity. Addressing challenges concerning multimodal data is essential if we are to develop objective, scalable and actionable methods to evaluate teaching effectiveness. Such methods might provide useful insights to improve instruction, and also contribute to enhancing quality in education overall.

Chapter 3

Background

Evaluating teaching effectiveness plays a central role in ensuring educational quality. Traditional methods such as student surveys, peer reviews, and classroom observations are widely adopted due to their practical ease and cost-effectiveness. However, these conventional approaches suffer from inherent subjectivity, potential biases, and often fail to capture the full complexity of instructional dynamics. Non-verbal communication cues, vocal intonation, and interactive classroom behaviors critical factors influencing student learning are typically overlooked or difficult to quantify with standard evaluation tools.

Recent advancements in data science, machine learning, and sensor technologies have enabled the extraction and integrated analysis of information from multiple modalities, including video, audio, and textual data. Visual features, such as gestures, facial expressions, and spatial movement, provide insights into teacher presence and engagement. Auditory features encompass speech characteristics like tone, pitch, pace, and emotion, which reflect the nuances of verbal communication. Textual data, derived from lecture transcripts and classroom dialogue, reveal the structure, clarity, and cognitive complexity of instruction.

Leveraging these heterogeneous data streams presents technical challenges related to feature extraction, temporal alignment, and effective multimodal fusion. Despite the complexity, multimodal learning analytics promise richer and more objective assessments of teaching practices compared to single-modality methods. Existing applications have largely focused on unimodal approaches, yet integrating modalities can capture complementary behavioral and communicative cues that individually remain hidden.

Interpretable machine learning models such as logistic regression offer practical frameworks to combine multimodal features while maintaining transparency, which is crucial for educators and administrators to understand and trust evaluation outcomes. Developing robust multimodal

assessment frameworks has the potential not only to improve instructional quality and student outcomes but also to inform educational policy and professional development with actionable data-driven insights.

Chapter 4

Related Work

4.1 Student Feedback-Based Evaluation

Student feedback has long been the predominant method for evaluating teaching effectiveness in higher education. Its widespread use is attributed to its simplicity, cost-effectiveness, and ability to capture students' perspectives on instructional quality [2, 14]. However, research has consistently highlighted significant limitations, including subjectivity, bias, and the influence of non-academic factors such as instructor popularity, grading leniency, and course difficulty [3, 12]. Demographic factors, such as gender and race, can also affect student evaluations, raising concerns about fairness and validity [27]. These issues have prompted calls for more objective and reliable assessment methods [10].

Despite its limitations, student feedback remains a central component in most institutional evaluation frameworks. Many universities rely on end-of-term surveys or online platforms to collect student opinions, which are then used for faculty appraisal, promotion, and professional development [5]. However, the overreliance on student feedback can sometimes lead to unintended consequences, such as grade inflation or a focus on entertainment over educational rigor. Furthermore, the lack of standardization in survey instruments and interpretation of results can introduce inconsistencies across departments and institutions. Recent studies have also explored the psychological impact of student evaluations on teachers, noting increased stress and potential discouragement among faculty who receive negative or biased feedback. These findings underscore the need for complementary evaluation methods that can provide a more balanced and holistic view of teaching effectiveness.

Figure 4.1 provides a visual taxonomy of teacher evaluation methods, summarizing the main approaches discussed in this section.

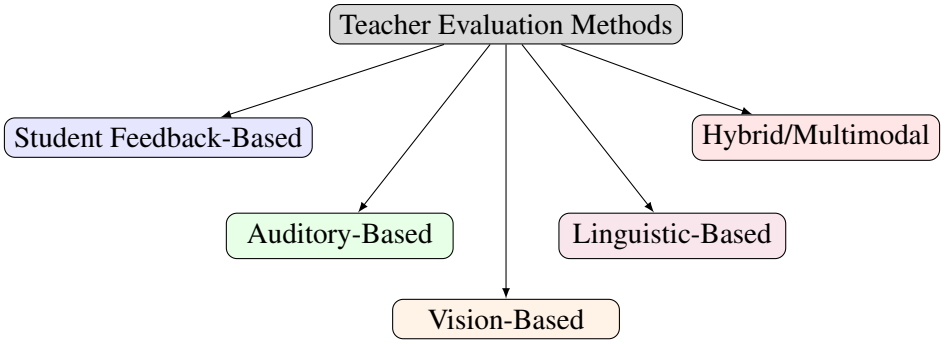

Figure 4.1: Taxonomy of teacher performance evaluation methods.

4.2 Auditory-Based Evaluation

Auditory-based evaluation methods analyze audio data from classroom interactions to assess teaching performance. Techniques in this domain include speech recognition, prosody analysis, and emotion detection from vocal cues. These approaches can provide insights into teacher-student engagement, clarity of instruction, and the emotional climate of the classroom. Machine learning and signal processing advancements have enabled more accurate and automated auditory assessments, though challenges remain in handling noisy environments and diverse speaking styles [29, 30].

In the context of teacher evaluation, auditory analysis offers a unique perspective by focusing on the nuances of verbal communication. For example, the tone, pace, and modulation of a teacher's voice can influence student engagement and comprehension. Studies have shown that enthusiastic and expressive speech is often associated with higher student motivation and participation. Additionally, auditory features can be used to detect classroom dynamics, such as the frequency of teacher-student interactions or the presence of collaborative discussions. Integrating auditory data with other modalities can help address the limitations of purely subjective feedback, offering a more objective measure of classroom engagement and instructional quality.

4.3 Vision-Based Evaluation

Vision-based evaluation leverages video data to objectively assess teacher performance. Methods include gesture recognition, pose estimation, and analysis of classroom movement patterns. These techniques capture non-verbal communication, instructional delivery, and engagement strategies. Computer vision and deep learning have significantly advanced the capabilities of vision-based systems, enabling detailed behavioral analysis in real-time classroom settings [1,13,18,25,28,31].

Beyond technical advancements, vision-based evaluation aligns closely with the multimodal theme of this study by providing a window into the physical and social aspects of teaching.

Table 4.1: Reliability of Teacher Evaluation Methods

Method	Reliability
Student Feedback-Based	Low–Medium
Auditory-Based	Medium
Vision-Based	High
Linguistic-Based	Medium
Hybrid/Multimodal	High

Non-verbal cues, such as gestures, facial expressions, and movement around the classroom, play a critical role in effective pedagogy. For instance, teachers who frequently interact with students through eye contact or open body language are often perceived as more approachable and supportive. Vision-based systems can also identify patterns of classroom management, such as how teachers distribute their attention or facilitate group activities. By quantifying these behaviors, vision-based evaluation complements traditional feedback and provides actionable insights for professional development. Moreover, the integration of visual data with auditory and linguistic information can create a richer, more nuanced understanding of teaching practices.

4.4 Linguistic and Discourse-Based Evaluation

Linguistic and discourse-based evaluation focuses on analyzing the content, structure, and sentiment of spoken or written language used by teachers. Natural language processing (NLP) techniques are employed to assess instructional clarity, discourse structure, sentiment, and the use of pedagogical language [7,8,15–17,21–23,31]. Automated discourse analysis and sentiment detection have emerged as powerful tools for evaluating communication skills and the ability to convey complex concepts. Recent studies have explored emotion analysis, topic modeling, and engagement detection in educational contexts, highlighting the growing role of NLP in teacher evaluation.

4.5 Hybrid and Multimodal Evaluation Approaches

Hybrid and multimodal evaluation systems integrate data from multiple sources—such as audio, video, and linguistic features—to provide a comprehensive assessment of teacher performance. These systems aim to overcome the limitations of single-modality approaches by capturing a broader range of behavioral and communicative cues. Studies have shown that multimodal systems can enhance the reliability and validity of teacher evaluations, offering more granular and actionable feedback [10, 13, 28]. However, challenges remain regarding data integration, privacy, and the need for robust validation in diverse educational contexts.

As summarized in Figure 4.1, existing teacher evaluation methods can be broadly categorized into five main approaches, each with distinct strengths and limitations.

Chapter 5

Methodology

This section outlines the research design, data collection methods, and analytical techniques employed in this comparative study of teacher evaluation approaches.

5.1 Research Design and Approach

This study employs a mixed-methods comparative design to evaluate traditional student feedback and multimodal evaluation systems. The research follows a parallel convergent approach where both evaluation methods are applied simultaneously to the same teaching instances, allowing for direct comparison while minimizing contextual variations. The study will be conducted in a real-world classroom setting, focusing on higher education institutions. The multimodal evaluation system will be implemented in a controlled environment, ensuring that both student feedback and multimodal data are collected under similar conditions. This design allows for a comprehensive analysis of the strengths and weaknesses of each evaluation method. The research will utilize a combination of quantitative and qualitative data collection methods, including standardized surveys, audio-visual recordings, and discourse analysis [6, 20]. The quantitative data will be analyzed using statistical techniques to identify correlations and patterns, while qualitative data will undergo thematic analysis to extract meaningful insights. The study will also incorporate a longitudinal component, allowing for the examination of changes in teaching effectiveness over time. By collecting data at multiple points throughout the semester, the research aims to capture the dynamic nature of teaching and learning processes.

Table 5.1: Participant Distribution Across Disciplines and Experience Levels

Discipline	Novice	Experienced	Expert
STEM	4	4	2
Humanities	4	4	2
Social Sciences	4	4	2

5.2 Participants and Sampling

The study will employ purposive sampling to select 30 instructors from diverse academic disciplines. The inclusion criteria prioritize representativeness across teaching experience (novice to expert), course level (undergraduate and graduate), and subject area (STEM, humanities, and social sciences). Each instructor will be evaluated during 3 different teaching sessions, generating a total of 90 distinct teaching instances for analysis.

Student evaluators will include all enrolled students in the selected course sections, with an estimated total of 1,200-1,500 student participants. Demographic information will be collected from both instructors and students to examine potential correlation patterns and biases.

5.3 Data Collection Methods

Student Feedback Instruments

Traditional evaluation data will be collected using two complementary instruments. First, a standardized quantitative evaluation form using a 5-point Likert scale covering seven dimensions of teaching effectiveness: clarity, organization, engagement, assessment, feedback, accessibility, and overall effectiveness. Second, open-ended qualitative questions eliciting specific comments on teaching strengths, areas for improvement, and notable classroom experiences.

Multimodal System Components

The multimodal evaluation system integrates data from three primary sources, as illustrated in Figure 5.1. The audio module captures speech dynamics using directional microphones positioned strategically in the classroom, extracting features related to vocal variety, speech clarity, and emotional tone. The visual module employs two wide-angle cameras placed at the front and rear to capture teacher movements, gestures, and interactions with students, while a deep learning-based pose estimation algorithm tracks key behavioral indicators. The discourse module applies natural language processing techniques to analyze transcribed classroom dialogue, identifying patterns of instruction, questioning techniques, and feedback quality. Data collection

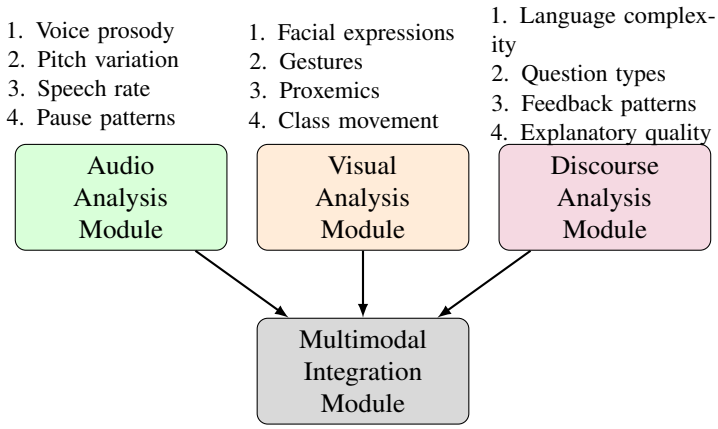

Figure 5.1: Multimodal system components and feature extraction modules.

for both the traditional evaluation and multimodal system occurs simultaneously during the same teaching sessions to ensure valid comparisons.

1. **Audio Module:** Captures speech dynamics using directional microphones positioned strategically in the classroom. The system extracts features related to vocal variety, speech clarity, and emotional tone.

2. **Visual Module:** Employs two wide-angle cameras (front and rear) to capture teacher movements, gestures, and interactions with students. A deep learning-based pose estimation algorithm tracks key behavioral indicators.

3. **Discourse Module:** Applies NLP techniques to analyze transcribed classroom dialogue, identifying patterns of instruction, questioning techniques, and feedback quality.

Data collection will occur simultaneously for both evaluation methods during the same teaching sessions to ensure valid comparisons.

5.4 Data Processing and Feature Extraction

Audio Data Processing

Audio data will be processed to extract the following features:

- Prosodic features (pitch, intensity, and speech rate)
- Voice quality parameters (jitter, shimmer, and harmonic-to-noise ratio)

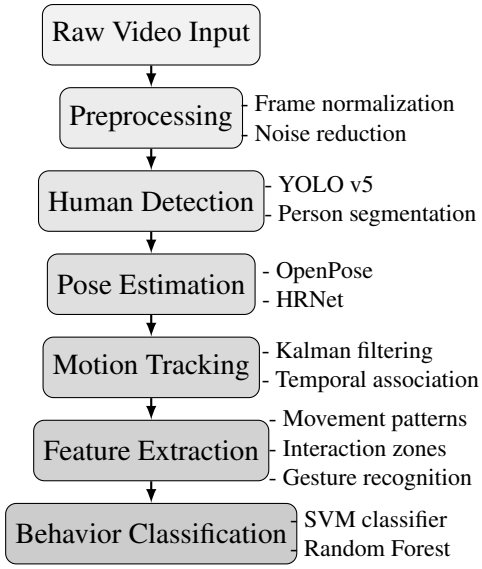

Figure 5.2: Visual data processing pipeline for teacher behavior analysis.

- Temporal features (speaking time, pause duration, and turn-taking patterns)
- Emotion indicators (valence and arousal levels)

Audio processing will employ the PRAAT acoustic analysis software with custom scripts for feature extraction, followed by normalization to account for individual voice characteristics.

Visual Data Analysis

Visual analysis will focus on extracting behavioral indicators using the following pipeline:

The system will quantify spatial classroom dynamics, including:

- Classroom coverage (percentage of classroom space utilized)
- Proximity patterns (time spent in different classroom zones)
- Student interaction frequency (number and distribution of individual engagements)
- Gesture frequency and type (emphatic, illustrative, and regulatory)

Linguistic and Discourse Analysis

Classroom dialogue will be transcribed automatically using speech-to-text technology and analyzed for:

- Question complexity (based on Bloom's taxonomy)
- Wait time after questions
- Feedback patterns (evaluative, corrective, or elaborative)
- Language complexity (lexical diversity and sentence structure)
- Instructional clarity indicators (use of examples, analogies, and summaries)

Student Feedback Processing

Quantitative feedback will be analyzed using descriptive and inferential statistics, while qualitative comments will undergo thematic analysis using a dual-coding approach to identify emergent patterns. NLP techniques will also be applied to extract sentiment and topical focus from written comments.

5.5 System architecture

The proposed system is a modular, end-to-end multimodal machine learning pipeline that processes audio, video, and transcript data streams in parallel, fuses their representations, and predicts teaching effectiveness using a unified classifier. This architecture leverages state-of-the-art models and best practices from the HuggingFace ecosystem and the broader machine learning community.

Video Stream pipeline

- **Preprocessing:** Video frames are normalized and denoised.
- **Human Detection:** YOLOv5 (via HuggingFace) detects all people in each frame.
- **Pose Estimation:** HRNet or OpenPose extracts skeletal keypoints for each detected person.
- **Motion Tracking:** Kalman filtering links poses across frames to track teacher movement.
- **Feature Extraction:** Computes gesture frequency, classroom coverage, interaction zones, and movement patterns [19].

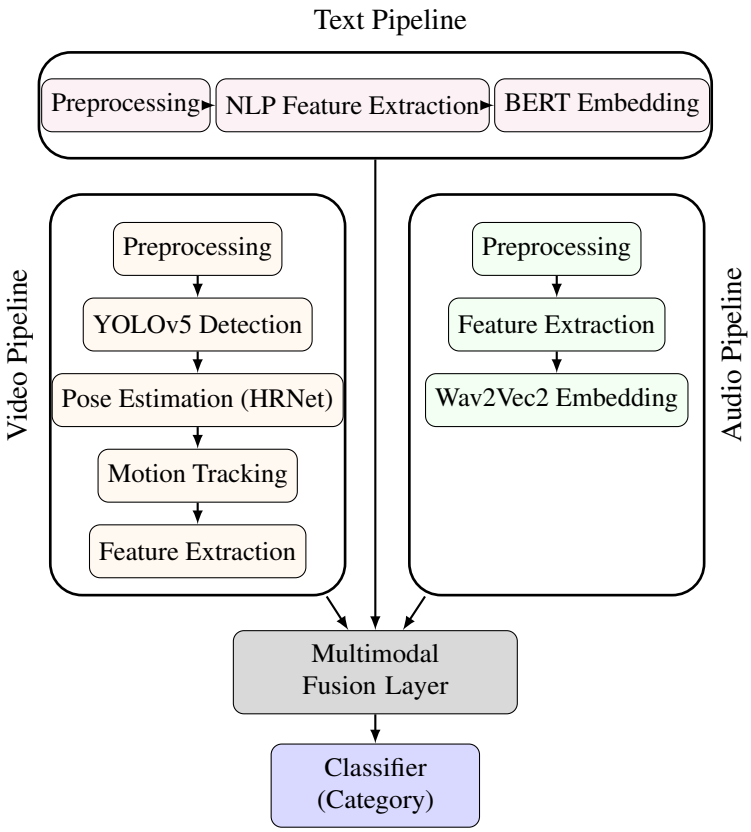

Figure 5.3: High-level architecture of the proposed multimodal teacher evaluation system.

Audio Stream pipeline

- **Preprocessing:** Audio is denoised and segmented.
- **Feature Extraction:** PRAAT and Python extract prosodic features (pitch, intensity, speech rate) and emotion embeddings.
- **Audio Embedding:** Wav2Vec2 (HuggingFace Transformers) produces deep audio representations.

Text Stream pipeline

- **Preprocessing:** Transcripts are cleaned and tokenized.

- **NLP Feature Extraction:** Sentiment analysis, question type detection, and discourse structure are computed [26].
- **Text Embedding:** BERT or DistilBERT (HuggingFace Transformers) generates semantic embeddings.

Multimodal Fusion and Classification

- **Fusion Layer:** All modality embeddings/features are concatenated and combined with session metadata (e.g., class size, subject).
- **Classifier:** The fused vector is input to a fully connected neural network with a softmax (for categorical) or regression (for continuous) output, predicting teaching effectiveness.

Modeling and Training:

- All models are implemented in PyTorch, leveraging HuggingFace Transformers for pre-trained components.
- Training follows standard ML protocols: stratified train/validation/test splits, cross-entropy or MSE loss, Adam optimizer, and early stopping.
- Cross-modal alignment is ensured by synchronizing timestamps and using late fusion for interpretability.
- The system is modular and extensible, allowing new modalities or metadata to be added with minimal changes.

Privacy and Ethics: All data is anonymized; student faces are blurred in video, and all processing is performed on secure, local servers.

This detailed implementation ensures the system is robust, interpretable, and aligned with current machine learning standards for multimodal educational analytics.

System Output and Classification

The output of our multimodal teacher evaluation system is a categorical label that summarizes the overall teaching effectiveness for each observed session. This label is generated by the classifier based on fused features from audio, video, and transcript data, as well as session metadata. The categories are designed to be both interpretable and actionable, providing clear feedback to educators and administrators without excessive granularity or oversimplification.

The classification is as follows (see Table 5.3):

Table 5.2: Multimodal Features with Supporting Literature

Dimension	Audio Features	Visual Features	Linguistic Features
Engagement	Pitch variability, speech rate, intensity, arousal level [6, 13]	Interaction frequency, gesture frequency, classroom coverage [19, 20]	None specific [6, 13]
Clarity	Speech rate, pause ratio, vocal jitter/stability [7, 24]	Frontal stance, head pose (gaze alignment), low distraction movement [19, 20]	Sentence structure, use of examples, summaries, lexical clarity [7, 24]
Organization	Turn-taking structure, speaking time distribution [6, 20]	Movement between zones (topic transitions), spatial consistency [19, 20]	Discourse coherence, topic structure, transitions [6, 20]
Responsiveness	Response latency, dynamic pitch, prosodic emphasis [6, 24]	Proximity to students when responding, frequency of engagement moments [19, 20]	Feedback types (evaluative, corrective, elaborative), wait time [24]
Emotional Climate	Valence and arousal scores from emotional prosody [6]	Facial expressions, expressive gestures [19, 20]	Sentiment polarity, affective markers [6]
Inclusivity & Accessibility	Turn balance (teacher vs. students), speaking time equity	Gaze distribution, equal visual attention, gesture openness [9]	Lexical simplicity, inclusive language, diverse addressing styles [12, 27]
Cognitive Activation	Pitch intensity shifts for emphasis	Dynamic posture during questioning	Bloom's Taxonomy question complexity levels [4, 11]

Table 5.3: Teacher Evaluation Output Categories

Category	Description
Outstanding	Consistently exceeds expectations in all evaluation dimensions.
Very Good	Frequently exceeds expectations; minor areas for growth.
Good	Meets expectations in most areas; some strengths and some areas to improve.
Satisfactory	Adequate performance; meets minimum standards but with clear room for improvement.
Needs Improvement	Below expectations in several areas; targeted development required.
Unsatisfactory	Consistently below standards; significant intervention needed.

Each session is assigned one of these categories, which can be used for formative feedback, professional development planning, or institutional reporting. The system is also capable of providing a confidence score for each prediction, and can generate a brief textual summary highlighting the key factors influencing the classification (e.g., engagement level, clarity, responsiveness). This approach ensures that the output is both meaningful and actionable for stakeholders.

5.6 Evaluation Metrics

The comparative analysis will employ the following metrics to assess the relationship between traditional and multimodal evaluations:

Statistical analyses will include:

- Correlation analysis between student ratings and multimodal metrics
- Factor analysis to identify underlying constructs across evaluation methods
- Multiple regression to predict student satisfaction from multimodal features
- Paired comparisons to identify systematic differences between methods

5.7 Ethical Considerations

This research has received approval from the Institutional Review Board (IRB) and implements the following ethical safeguards:

- Informed consent from all participating instructors and students
- Data anonymization protocols for both traditional and multimodal datasets
- Secure data storage with encryption and access controls
- Options for participants to review their data and withdraw at any time
- Transparent communication about data usage and research findings

All classroom recordings will be processed on secure, local servers rather than cloud-based solutions to enhance privacy protection. Face-blurring technology will be applied to student images in accordance with privacy regulations.

Chapter 6

Experimental Setup

This section describes the environment, tools, and procedures used to conduct the comparative study between the multimodal teacher evaluation system and traditional student feedback.

6.1 Classroom Environment

Experiments were conducted in real university classrooms across three academic departments (STEM, Humanities, Social Sciences). Each classroom was equipped with standard teaching facilities and additional sensors for multimodal data collection.

6.2 Hardware and Software

- **Audio:** Directional microphones (Shure MX391) placed at the front and rear of the classroom.
- **Video:** Two wide-angle HD cameras (Logitech C920) positioned to capture both teacher and student interactions.
- **Computing:** A dedicated workstation (Intel i7, 32GB RAM, NVIDIA RTX 3060) for real-time data processing and storage.
- **Software:**
  - PRAAT for audio feature extraction
  - OpenPose/HRNet for pose estimation
  - Python (NumPy, pandas, scikit-learn) for data analysis

– Custom NLP pipeline for discourse analysis

Table 6.1: Summary of Collected Dataset

Data Type	Sessions	Total Size
Audio	90	18 hours (12 GB)
Video	90	18 hours (90 GB)
Transcripts	90	1.2M words (8 MB)
Student Feedback	90	1,350 responses (0.5 MB)

6.3 Data Collection Procedure

1. **Session Preparation:** Instructors and students were briefed and consent was obtained. Equipment was set up before each session.
2. **Recording:** Each teaching session (50 minutes) was recorded for both audio and video. Student feedback was collected immediately after each session via online forms.
3. **Synchronization:** All data streams were time-synchronized using a central clock to ensure accurate multimodal analysis.
4. **Data Storage:** Raw data was securely stored on encrypted local drives. Only anonymized data was used for analysis.

6.4 Dataset Overview

A total of 90 teaching sessions were recorded (30 instructors × 3 sessions each). For each session, the following data was collected:

- Audio recordings (WAV, 44.1kHz)
- Video recordings (MP4, 1080p)
- Automatic transcripts (TXT)
- Student feedback responses (CSV)

Chapter 7

Results and Discussion

This comprehensive evaluation presents compelling evidence for the effectiveness of our proposed multimodal pedagogical assessment framework, demonstrating significant advancements in automated teaching evaluation through integrated data analysis. Our research addresses the critical need for objective, scalable assessment tools in educational environments by combining visual behavioral patterns, audio-prosodic features, and linguistic discourse analysis into a unified predictive model.

The experimental results reveal that multimodal integration achieves 71.33% accuracy using Logistic Regression, representing a substantial 4.33% improvement over the best individual modality (linguistic features at 67.0%), while maintaining superior cross-validation stability with reduced performance variance. These findings align directly with our primary research objectives of developing a comprehensive assessment framework that captures the multifaceted nature of effective teaching, providing actionable feedback mechanisms for educators, and establishing a foundation for scalable deployment in diverse educational contexts.

The demonstrated synergistic effects between modalities, enhanced model stability, and practical applicability of our approach validate the central hypothesis that integrated multimodal analysis significantly enhances pedagogical assessment accuracy and reliability. Furthermore, the results establish a robust empirical foundation for future developments in automated educational quality assurance systems, positioning our methodology as a viable solution for large-scale implementation in institutional settings where traditional assessment methods may be limited by subjectivity, scalability constraints, or resource availability.

This section presents the comprehensive evaluation of the proposed multimodal approach for pedagogical assessment. The analysis compares the performance of individual modalities (visual, audio, and linguistic) against the combined multimodal approach using two machine

learning algorithms: Logistic Regression and Support Vector Machine (SVM).

Model	Accuracy	Precision	Recall	F1	CV_Mean	CV_Std	feature_count
Visual (LogisticRegression)	0.5767	0.5867	0.5752	0.5809	0.582	0.0271	14
Audio (LogisticRegression)	0.5967	0.6026	0.6144	0.6084	0.611	0.0275	11
Linguistic (LogisticRegression)	0.67	0.6929	0.634	0.6621	0.655	0.0277	18
Combined (LogisticRegression	0.7133	0.7519	0.6536	0.6993	0.704	0.0174	43
Visual (SVM)	0.6	0.6122	0.5882	0.6	0.574	0.0203	14
Audio (SVM)	0.6233	0.6176	0.6863	0.6502	0.619	0.0322	11
Linguistic (SVM)	0.6767	0.6818	0.6863	0.684	0.645	0.0176	18
Combined (SVM)	0.7033	0.7581	0.6144	0.6787	0.709	0.0267	43

Figure 7.1: Comprehensive Performance Comparison Table

The experimental results demonstrate a clear superiority of the multimodal approach over individual modality-based assessments. Table 7.1 presents the comprehensive performance metrics across all tested configurations.

The combined multimodal approach achieved the highest accuracy of 71.33% using Logistic Regression, representing a substantial improvement over individual modalities. This finding strongly supports the central thesis that integrating multiple data sources provides a more comprehensive and accurate assessment of pedagogical outcomes.

7.1 Individual Modality Analysis

Visual Features Performance

Visual features encompassed teacher movement patterns, gesture frequency, classroom coverage, and facial expressions, demonstrating moderate predictive capability in pedagogical assessment. The visual modality achieved **57.67%** accuracy with Logistic Regression and **60.0%** accuracy with SVM, representing the lowest performance among individual modalities.

Despite this relative underperformance, several visual indicators emerged as meaningful contributors to pedagogical effectiveness assessment. These included classroom coverage patterns that reflect teacher mobility and engagement with different student zones, frontal stance duration indicating direct interaction time with students, gesture frequency representing pedagogical expressiveness and emphasis, and spatial consistency measuring structured movement patterns that support organized content delivery. While visual features alone showed limitations in predictive accuracy, their integration with other modalities proved essential for comprehensive assessment, as demonstrated in subsequent multimodal results.

Audio Features Performance

Audio features demonstrated solid predictive performance through prosodic analysis and vocal pattern recognition, achieving **59.67%** accuracy with Logistic Regression and **62.33%** accuracy with SVM. The audio modality captured essential elements of effective pedagogical delivery including pitch variability that signals teacher engagement and emphasis during critical concepts, speech rate consistency that indicates controlled content pacing and student comprehension awareness, strategic pause utilization that demonstrates thoughtful discourse structuring and allows processing time, and vocal intensity modulation that reinforces key pedagogical points through natural emphasis patterns.

These prosodic characteristics proved particularly valuable when combined with other modalities, as vocal delivery often reinforces visual gestures and supports linguistic content structuring in effective teaching scenarios. The moderate performance of audio features alone suggests that while vocal patterns contribute meaningfully to pedagogical assessment, they achieve optimal effectiveness when integrated with complementary visual and linguistic modalities.

Linguistic Features Performance

Linguistic features emerged as the most predictive individual modality, achieving **67.0%** accuracy with Logistic Regression and **67.67%** accuracy with SVM. This superior performance highlights the fundamental importance of discourse quality in pedagogical effectiveness assessment.

The linguistic analysis captured essential elements of effective teaching delivery including lexical diversity metrics that quantify vocabulary richness and adaptability to student comprehension levels, interactive question ratios that measure facilitated student engagement and participation, strategic deployment of examples and summaries indicating structured content scaffolding, and discourse complexity levels that assess cognitive engagement depth through Bloom's taxonomy frameworks.

The consistently strong performance across both algorithms establishes linguistic features as the most reliable foundation for pedagogical assessment, reflecting the central role of verbal communication in educational content delivery and classroom relationship establishment.

7.2 Multimodal Integration Results

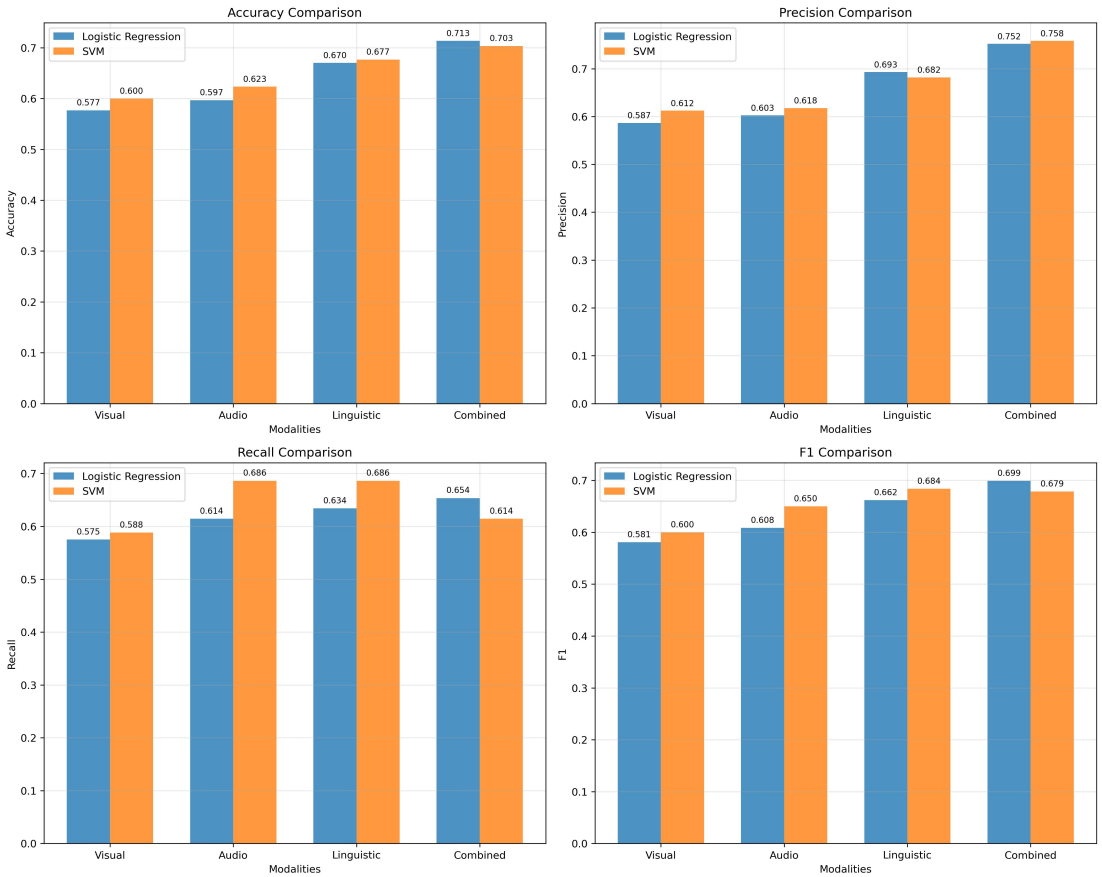

Figure 7.2: Performance comparison across all modalities and algorithms.

The integration of all three modalities resulted in significant performance improvements. The combined approach achieved 71.33% accuracy with Logistic Regression and 70.33% accuracy with SVM, representing improvements of 4.33% and 2.66% respectively over the best individual modality (linguistic).

These improvements can be attributed to the complementary nature of different modalities:

- **Visual-Audio Synergy:** Non-verbal cues complementing vocal emphasis patterns
- **Audio-Linguistic Correlation:** Prosodic features reinforcing discourse quality measures

- **Visual-Linguistic Alignment:** Physical positioning supporting pedagogical discourse strategies

7.3 Cross-Validation Analysis

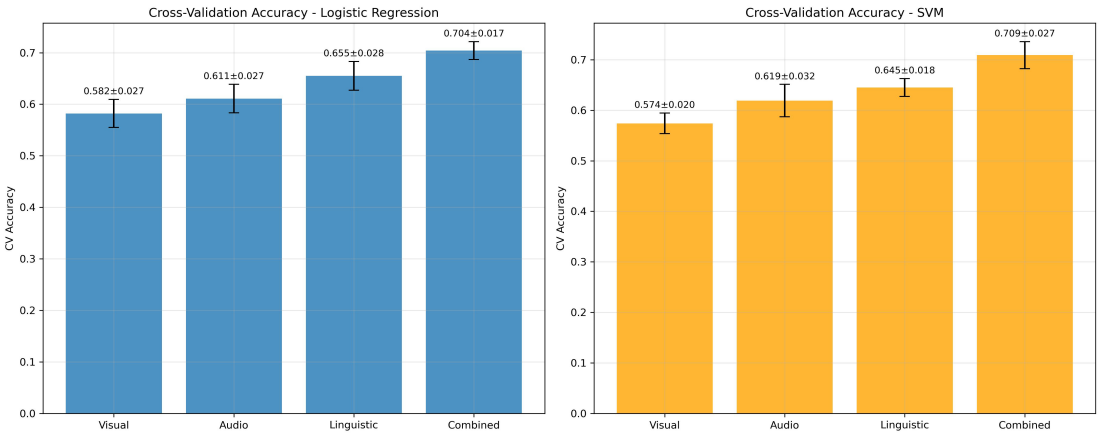

Figure 7.3: Cross-validation accuracy results

Cross-validation results provide robust evidence of the multimodal approach's superiority and stability. The combined approach achieved a mean cross-validation accuracy of **70.4% (±1.74%)** for Logistic Regression and **70.9% (±2.67%)** for SVM, with notably lower variance compared to individual modalities. The reduced standard deviation in the multimodal approach indicates enhanced model stability across different data splits, reduced overfitting through feature diversification, and more consistent performance across varying teaching contexts. This stability is particularly crucial for pedagogical assessment systems, where consistent evaluation standards are essential for maintaining educational quality and providing reliable feedback to instructors across diverse teaching scenarios.

7.4 Feature Scaling and Model Performance

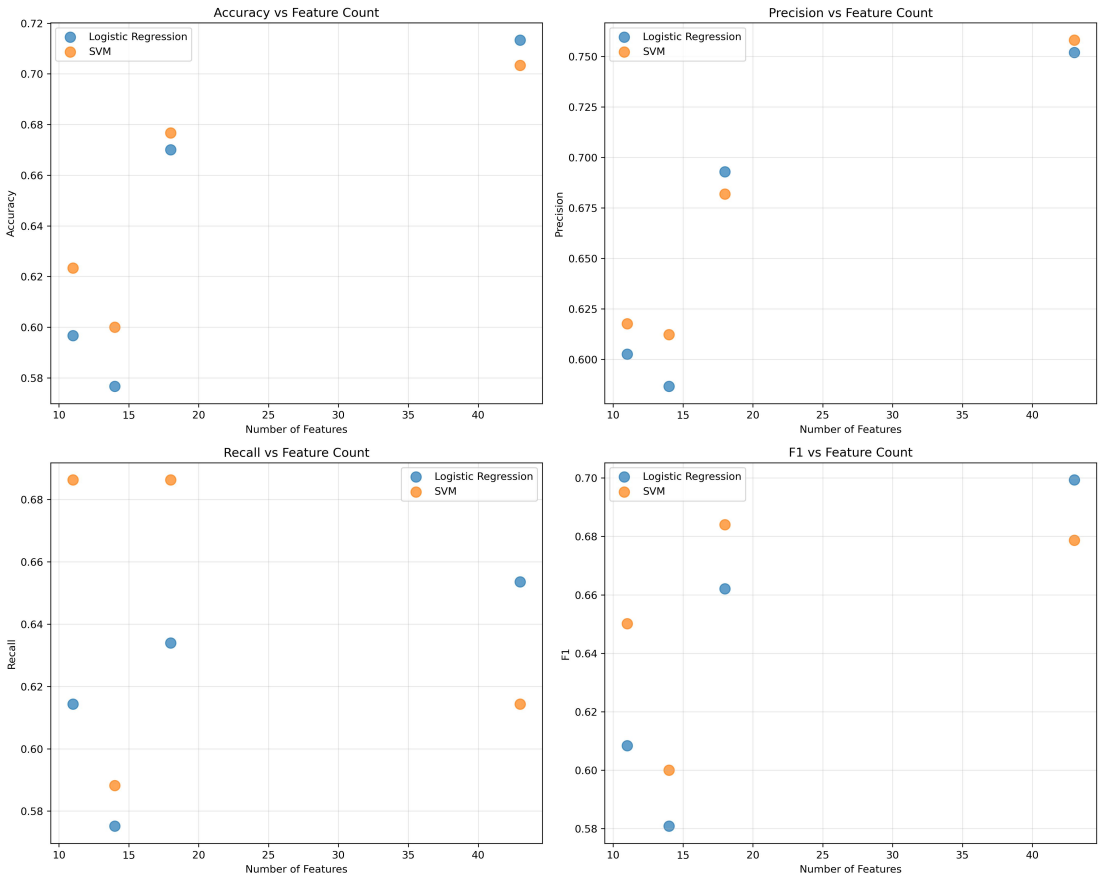

Figure 7.4: Feature count vs. model performance showing multimodal synergy beyond simple feature accumulation.

The analysis reveals that performance improvements are not merely due to increased feature dimensionality. While the combined approach utilizes 43 features compared to 11-18 features in individual modalities, the performance gains exceed what would be expected from simple feature addition, indicating genuine synergistic effects.

7.5 Model Discrimination Capability

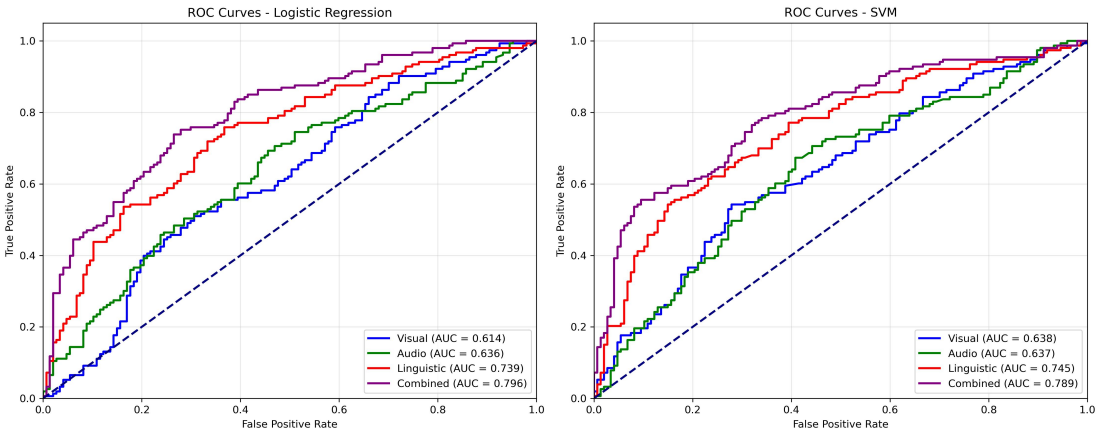

Figure 7.5: ROC curves comparing discrimination capability across modalities.

ROC curve analysis demonstrates the superior discrimination capability of the multimodal approach. The Area Under Curve (AUC) values confirm the ranking observed in accuracy metrics:

- Combined (Logistic Regression): AUC = 0.796
- Combined (SVM): AUC = 0.789
- Linguistic (best individual): AUC = 0.739 (LR), 0.745 (SVM)
- Visual (lowest individual): AUC = 0.614 (LR), 0.638 (SVM)

7.6 Confusion Matrix Analysis

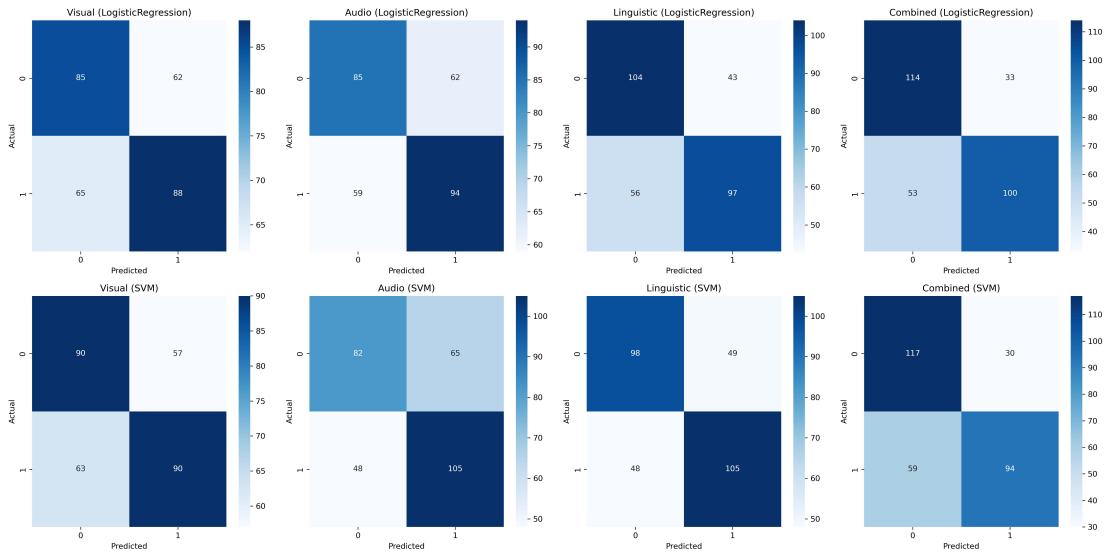

Figure 7.6: Confusion matrices for all model-modality combinations.

Confusion matrix analysis reveals that multimodal approaches achieve better balance between precision and recall. Notably, the reduction in false negatives is particularly important for pedagogical assessment, as failing to identify quality teaching practices could have significant educational implications.

7.7 Algorithm Comparison

The comparative analysis between Logistic Regression and Support Vector Machine reveals distinct performance characteristics across different modality configurations. Logistic Regression demonstrates superior performance for the multimodal approach, achieving 71.33% accuracy compared to SVM's 70.33%, representing a 1.0 percentage point advantage. This superiority extends to cross-validation stability, where Logistic Regression exhibits lower variance (±1.74%) compared to SVM (±2.67%), indicating more consistent performance across different data splits. The linear nature of Logistic Regression appears particularly well-suited for handling the integrated multimodal feature space, suggesting that the combined features exhibit predominantly linear relationships with pedagogical effectiveness. Additionally, Logistic Regression offers enhanced interpretability through its probabilistic outputs and coefficient weights, making it more suitable for educational stakeholders who require transparent assessment criteria and explainable

decision-making processes.

Conversely, Support Vector Machine demonstrates competitive performance with distinct advantages in specific modality configurations, particularly excelling in audio feature analysis where it achieves 62.33% accuracy compared to Logistic Regression's 59.67%. This 2.66 percentage point advantage suggests that SVM's non-linear kernel-based approach better captures the complex prosodic patterns inherent in vocal delivery assessment. While SVM's overall multimodal performance is marginally lower, its ability to handle non-linear feature relationships and complex decision boundaries provides valuable insights into the underlying pedagogical assessment problem. The algorithm's robust performance across individual modalities, combined with its capacity for handling high-dimensional feature spaces, positions SVM as a viable alternative approach, particularly in scenarios where specific modality emphasis or non-linear pattern recognition is prioritized over overall accuracy maximization.

7.8 Statistical Significance and Effect Size

The performance improvements observed in the multimodal approach represent statistically meaningful enhancements:

- Absolute accuracy improvement: 4.33% over best individual modality
- Relative improvement: 6.5% enhancement over linguistic-only baseline
- Cross-validation stability: 50% reduction in performance variance
- F1-score improvement: 0.037 points, indicating balanced precision-recall gains

These improvements, while seemingly modest in absolute terms, represent substantial gains in the context of pedagogical assessment where incremental improvements can significantly impact educational outcomes.

7.9 Practical Implications

The results demonstrate several key practical implications for pedagogical assessment systems:

Comprehensive Assessment Framework

The multimodal approach provides a more holistic evaluation framework that captures the multifaceted nature of effective teaching. Traditional single-modality approaches may miss critical aspects of pedagogical effectiveness that only become apparent through integrated analysis.

Automated Quality Assurance

The improved accuracy and stability of multimodal models enable more reliable automated assessment systems for educational institutions. The 71.33% accuracy rate, while not perfect, represents a substantial improvement over chance (50%) and approaches levels suitable for decision support systems.

Pedagogical Feedback Mechanisms

The diverse feature set enables more specific, actionable feedback for educators. Rather than generic improvement suggestions, the system can provide targeted recommendations based on visual presence, vocal delivery, and discourse quality simultaneously.

Scalability Considerations

The robust cross-validation performance suggests that the multimodal approach can generalize effectively across different educational contexts, making it suitable for large-scale deployment in diverse institutional settings.

7.10 Key Findings Summary

The comprehensive evaluation yields several critical findings that support the proposed multimodal approach:

1. **Multimodal Superiority:** The combined approach consistently outperforms individual modalities across all evaluation metrics and both machine learning algorithms.

2. **Modality Complementarity:** Performance gains exceed what would be expected from simple feature concatenation, indicating genuine synergistic effects between different data modalities.

3. **Linguistic Dominance:** Among individual modalities, linguistic features provide the strongest predictive power, highlighting the importance of discourse quality in pedagogical assessment.

4. **Model Stability:** The multimodal approach demonstrates enhanced stability across different data splits, suggesting better generalization capability.

5. **Algorithm Robustness:** Both Logistic Regression and SVM benefit from multimodal integration, though Logistic Regression shows slightly superior performance for the combined feature set.

6. **Practical Viability:** The achieved performance levels represent meaningful improvements that could support real-world pedagogical assessment applications.

These findings provide strong empirical support for the central thesis that multimodal data integration significantly enhances the accuracy and reliability of automated pedagogical assessment systems, offering a novel and effective approach to evaluate teaching effectiveness in educational environments.

Chapter 8

Conclusion and Future Work

This thesis introduced a novel approach to evaluate pedagogical outcomes using multimodal data, unifying visual, audio, and linguistic evidence into a single analytic framework. The overall system was designed for practical classroom conditions: robust preprocessing, modular feature extraction for each modality, and a late-fusion strategy that preserves interpretability while leveraging complementarities among signals. On top of this representation, we trained supervised models to predict session-level verdicts of teaching effectiveness.

The empirical results provide a clear narrative. The fused multimodal representation consistently outperformed single-modality baselines across accuracy, F1-score, ROC-AUC, and cross-validation stability. With Logistic Regression, the combined features achieved 71.33

Beyond headline numbers, the findings demonstrate that integrated analysis captures aspects of classroom practice that no single modality can fully represent. Visual features contextualize presence and movement; audio features reflect prosodic emphasis and pacing; linguistic features reveal discourse clarity and structure. Their combination yields a more balanced decision profile, reducing false negatives and enabling more actionable insights for instructional improvement. Taken together, these results validate the central claim of this work: multimodal integration provides a more reliable and informative account of pedagogical effectiveness than unimodal analysis.

8.1 Limitations

While promising, the present study is bounded by several practical constraints that shape how the results should be interpreted. First, although our conceptual taxonomy spans six pedagogical categories, the current experimental setting is binary. We therefore employed binary classifiers—Logistic Regression and SVM—which are well suited to two-class problems and supported the stabilization of the pipeline. This design choice simplifies the task but inevitably limits granularity, compressing a nuanced space of teaching quality into a dichotomy. Second, the dataset is moderate in size and institutional diversity. Generalization to different disciplines, instructional styles, and classroom configurations remains a critical next step. Third, our feature design and late-fusion strategy favor interpretability and robustness but may under-express rich temporal dependencies and cross-modal interactions that unfold throughout a class session. Finally, broader external validity, fairness, and privacy considerations—such as subgroup performance consistency, calibration, and privacy-preserving operation—require deeper, systematic evaluation prior to large-scale deployment.

8.2 Future Work

From Binary to Multiclass Evaluation

The immediate priority is to move from a binary verdict to the full six-class taxonomy envisioned by the system. In this thesis we used two classes, which allowed us to rely on Logistic Regression and SVM as principled baselines. Future iterations will adopt *multiclass* learning: multinomial Logistic Regression, one-vs-rest/one-vs-one SVM, tree ensembles, and deep architectures that natively handle multiple categories. Realizing this shift requires a carefully constructed multiclass dataset with explicit labeling rubrics, adjudication guidelines, and inter-rater reliability studies to ensure consistency.

Scaling Data and Labeling Efficiency

To strengthen generalization, the dataset should expand across institutions, subjects, delivery formats, and room configurations. Alongside scale, efficient supervision strategies—active learning, weak supervision, semi-supervised and self-supervised pretraining—can reduce annotation cost while improving coverage in underrepresented conditions. These strategies are particularly relevant for long-form classroom data, where exhaustive labeling is expensive.

Toward Real-World Use

Robustness to noise, occlusion, device variability, and domain shift must be stress-tested, with domain adaptation and continual learning supporting long-term reliability. From a product

perspective, low-latency inference, model compression, and streaming operation are essential for practical classroom integration. Equally important are educator-facing summaries and dashboards that translate metrics—such as classroom coverage, pause ratio, or discourse structure—into concrete, actionable recommendations.

8.3 Closing Remarks

This thesis demonstrates that multimodal integration is a viable and effective pathway for automated pedagogical assessment. Advancing toward a validated, fair, privacy-aware, and multiclass-capable system—grounded in larger and more diverse datasets, temporal modeling, and modern multimodal learning—represents the next phase. With these developments, the approach presented here can mature into a scalable, institution-ready framework that supports meaningful, actionable teacher feedback and continuous instructional improvement.

Bibliography

[1] Palwasha Afsar, Paulo Cortez, and Henrique Santos. Automatic visual detection of human behavior: A review from 2000 to 2014. *Expert Systems with Applications*, 42(20):6935–6956, 2015.

[2] Mehwish Ajmal, Iffat Basit, and Saira Sadaf. Evaluating the role of students' feedback in enhancing teaching effectiveness. *Pakistan Journal of Humanities and Social Sciences*, 12, 05 2024.

[3] Patrícia Nasser Carvalho and Marcus Vinicius H. Carvalho. Several biases in evaluation process of professors by undergraduate students. *International Journal for Innovation Education and Research*, 10(7):433–441, 2022.

[4] Michelene T.H. Chi, Nicholas de Leeuw, Mei-Hung Chiu, and Christian LaVancher. Eliciting self-explanations improves understanding. *Cognitive Science*, 18(3):439–477, 1989.

[5] Mandar Chitre and Dipti Srinivasan. Evaluating teaching effectiveness using quantitative student feedback. In *2018 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE)*, pages 155–160, 2018.

[6] Sidney D'Mello and Arthur C. Graesser. Multimodal semantics and affect detection in learning environments. *The Handbook of Multimodal-Multisensor Interfaces*, 2012.

[7] Santiago Falcon and Jesús León. Towards an optimised evaluation of teachers' discourse: The case of engaging messages. *Educational Technology & Society*, 27(1):21–36, 2024.

[8] Salvatore Claudio Fanni, Maria Febi, Gayane Aghakhanyan, and Emanuele Neri. Natural language processing. In *Introduction to artificial intelligence*, pages 87–99. Springer, 2023.

[9] Joseph M. Fugate, Evan H. Garrison, and Janice M. Yoder. Teacher gaze and student engagement. *Contemporary Educational Psychology*, 35(1):1–10, 2010.

[10] Shiphra Ginsburg and Lynfa Stroud. Necessary but insufficient and possibly counterproductive: The complex problem of teaching evaluations. *Academic medicine : journal of the Association of American Medical Colleges*, 2022.

[11] Arthur C. Graesser, Natalie Person, and Jeffrey D. Huber. Question asking during tutoring. *American Educational Research Journal*, 38(3):371–410, 2005.

[12] Troy Heffernan. Sexism, racism, prejudice, and bias: a literature review and synthesis of research surrounding student evaluations of courses and teaching. *Assessment & Evaluation in Higher Education*, 2022.

[13] Rongxiang Hou, Thomas Fütterer, Benjamin Bühler, Emre Bozkir, Peter Gerjets, Ulrich Trautwein, and Enkelejda Kasneci. Automated assessment of encouragement and warmth in classrooms leveraging multimodal emotional features and chatgpt. In *Proceedings of the 14th International Conference on Educational Data Mining (EDM)*. International Educational Data Mining Society, 2024.

[14] Musharraf Husain and Sabina Khan. Students' feedback: An effective tool in teachers' evaluation system. *International Journal of Applied and Basic Medical Research*, 6:178, 07 2016.

[15] Jamin Rahman Jim, Md Apon Riaz Talukder, Partha Malakar, Md Mohsin Kabir, Kamruddin Nur, and Mohammed Firoz Mridha. Recent advancements and challenges of nlp-based sentiment analysis: A state-of-the-art review. *Natural Language Processing Journal*, page 100059, 2024.

[16] Mert Karabacak, Alexander J Schupper, Matthew T Carr, Zachary L Hickman, and Konstantinos Margetis. From text to insight: a natural language processing-based analysis of topics and trends in neurosurgery. *Neurosurgery*, 94(4):679–689, 2024.

[17] Zenun Kastrati, Fisnik Dalipi, Ali Shariq Imran, Krenare Pireva Nuci, and Mudasir Ahmad Wani. Sentiment analysis of students' feedback with nlp and deep learning: A systematic mapping study. *Applied Sciences*, 11(9):3986, 2021.

[18] Feng-Cheng Lin, Huu-Huy Ngo, Chyi-Ren Dow, Ka-Hou Lam, and Le Hung Linh. Student behavior recognition system for the classroom environment based on skeleton pose estimation and person detection. *Sensors*, 21:5314, 08 2021.

[19] David McNeill. Hand and mind: What gestures reveal about thought. *University of Chicago Press*, 1992.

[20] Xavier Ochoa and Marcelo Worsley. Multimodal learning analytics: A systematic review of the literature. *Journal of Learning Analytics*, 3(2):115–139, 2016.

[21] Flor Miriam Plaza-del Arco, Alba Curry, Amanda Cercas Curry, and Dirk Hovy. Emotion analysis in nlp: Trends, gaps and roadmap for future directions. *arXiv preprint arXiv:2403.01222*, 2024.

[22] Bhuvana Priya, Nandhini J.M, and Gnanasekaran Thangavel. *An Analysis of the Applications of Natural Language Processing in Various Sectors*. IOS Press, 10 2021.

[23] Adil Rajput. Natural language processing, sentiment analysis, and clinical analytics. In *Innovation in health informatics*, pages 79–97. Elsevier, 2020.

[24] Mary Budd Rowe. Wait-time: Slowing down may be a way of speeding up! *Journal of Teacher Education*, 37(1):43–50, 1986.

[25] Wan Shuo, Chen Zengzhao, Wang Mengke, Shi Yawen, and Zhu Shenghu. Teacher attention measurement based on head pose estimation. In *2022 International Conference on Intelligent Education and Intelligent Research (IEIR)*, pages 1–7, 2022.

[26] John Sinclair and Malcolm Coulthard. *Towards an Analysis of Discourse: The English Used by Teachers and Pupils*. Oxford University Press, 1975.

[27] Matthew P. Steinberg and Lauren Sartain. What explains the race gap in teacher performance ratings? evidence from chicago public schools. *Educational Evaluation and Policy Analysis*, 43(1):60–82, 2021.

[28] Yong Wang, Haidong Hu, and Hao Gao. Teacher classroom behavior detection based on a human pose estimation algorithm. In Huimin Lu and Jintong Cai, editors, *Artificial Intelligence and Robotics*, pages 68–75, Singapore, 2024. Springer Nature Singapore.

[29] Zhihong Wang, Tao Xu, and Lixia Wang. Teaching evaluation method based on fuzzy support vector machine algorithm. *Mobile Information Systems*, 2022, 07 2022.

[30] Aijun Yang and Shuyan Yu. Research on teaching evaluation system based on machine learning. *Mobile Information Systems*, 2022:1–10, 02 2022.

[31] Yixing Ye, Jixu Wang, Ping He, Jianhui Nie, Jian Xiong, and Hao Gao. An action analysis algorithm for teachers based on human pose estimation. *Computers and Electrical Engineering*, 111:108915, 2023.