IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 3, April 2026
Impact Factor: 8.2

AI-Enabled Multimodal Framework for Teaching Quality Evaluation To Strengthen Diverse Learners Abilities

Dr. Mrunal Pathak[1], Dr. Pritesh Patil[2], Shravani Tanksale[3], Vidyankshini Vibhute[4], Aman Umre[5] , Devika Mule[6]
Professor, Department of Information Technology[1-2]
Under Graduate Student, Department of Information Technology[3-6]
AISSMS Institute of Information Technology , Pune, India

Abstract: *Ensuring high-quality teaching that effectively supports diverse learners is becoming an increasing challenge in modern education systems. This paper presents an AI-enabled multimodal framework designed to evaluate teaching quality using recorded video sessions. The proposed system integrates audio, visual, and textual analysis to assess key teaching parameters such as communication clarity, student engagement, instructor confidence, pacing, and overall content delivery. It leverages advanced techniques including speech recognition, natural language processing, and facial as well as gesture analysis to extract meaningful insights from instructional videos. The framework generates both an overall teaching score and a detailed evaluation of individual parameters, enabling objective, consistent, and scalable assessment across large educational platforms. A significant feature of this system is its emphasis on inclusivity, incorporating evaluation strategies that address the needs of visually impaired and hearing-impaired learners. It examines the effectiveness of verbal explanations, visual aids, captions, and non-verbal cues to determine how well teaching methods accommodate diverse learning requirements, thereby improving instructional quality, supporting educators with actionable feedback, and fostering more inclusive and effective learning environments.*

Keywords: Artificial Intelligence (AI), Multimodal Learning Analytics, Teaching Quality Evaluation, Natural Language Processing (NLP), Inclusive Education.

I. INTRODUCTION

The rapid growth of digital education and large-scale learning platforms has increased the need for effective and consistent evaluation of teaching quality. In traditional settings, assessing instructors often relies on peer reviews or student feedback, which can be subjective, time-consuming, and inconsistent. As educational systems expand, particularly with the rise of online learning and recorded lectures, the demand for automated, objective, and scalable evaluation methods continues to grow [17], [15].

Artificial Intelligence (AI) offers promising solutions to these challenges by enabling data-driven analysis of teaching practices. With advancements in multimodal learning analytics, instructional quality can be evaluated by analyzing multiple data sources such as audio, video, and text. Techniques such as speech recognition [2], [5], natural language processing [12], [11], and computer vision [3], [13] allow systematic assessment of key teaching parameters including communication clarity, engagement, confidence, pacing, and content delivery. Additionally, multimodal machine learning approaches further enhance evaluation accuracy by integrating these diverse data modalities [19].

Modern education also emphasizes inclusivity and equal access for all learners, including those with visual and hearing impairments. Therefore, it is important to evaluate how effectively instructors adapt their teaching strategies to meet the needs of diverse learners. Accessibility-aware analysis, such as evaluating the use of captions, verbal explanations, and

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 3, April 2026 Impact Factor: 8.2

visual aids, plays a crucial role in ensuring inclusive learning environments. Recent advancements in AI-driven educational systems and engagement detection further support the development of such inclusive frameworks [10], [15].

In this context, this paper proposes an AI-enabled multimodal framework designed to evaluate teaching quality from recorded video sessions. The system aims to provide objective scoring, detailed feedback, and insights into inclusive teaching practices. By doing so, it enhances teaching quality assurance and supports the development of scalable, intelligent, and inclusive education systems.

II. LITERATURE REVIEW

Table I: Literature Review Summary

Author/Year	Title of Paper	Key Findings	Limitations
A. Baevski et al., 2020 [1]	*wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations*	Introduced a self-supervised learning model for speech processing that significantly reduces the need for labeled data while improving speech recognition performance.	Requires large computational resources and may struggle with noisy or domain-specific audio data.
A. Graves et al., 2013 [2]	*Speech Recognition with Deep Recurrent Neural Networks*	Demonstrated the effectiveness of deep RNNs in modeling sequential audio data, achieving high accuracy in speech-to-text tasks.	Computationally expensive and less efficient compared to modern transformer-based approaches.
A. Krizhevsky et al., 2012 [3]	*ImageNet Classification with Deep Convolutional Neural Networks*	Introduced CNN-based image classification, significantly improving visual feature extraction and accuracy in computer vision tasks.	Requires large labeled datasets and high computational power for training.
A. Radford et al., 2021 [4]	*Learning Transferable Visual Models from Natural Language Supervision*	Proposed CLIP, enabling joint understanding of images and text, improving multimodal learning capabilities.	Limited performance in domain-specific applications without fine-tuning.
A. Radford et al., 2022 [5]	*Robust Speech Recognition via Large-Scale Weak Supervision (Whisper)*	Developed a robust multilingual speech recognition system capable of handling diverse accents and noisy environments.	Performance may vary for low-resource languages and requires high computational resources.
A. Vaswani et al., 2017 [6]	*Attention is All You Need*	Introduced the Transformer architecture, improving sequence modeling using attention mechanisms and enabling parallel computation.	High memory consumption and requires large datasets for effective training.
J. Devlin et al., 2019 [12]	*BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*	Enabled deep contextual understanding of text, significantly improving NLP tasks such as classification and semantic analysis.	Fine-tuning is required for specific tasks, and model size increases computational cost.
H. Zhang et al., 2022 [10]	*Multimodal Deep Learning for Engagement Detection in Online Education*	Demonstrated that combining audio, video, and behavioral features improves engagement detection accuracy in	Requires synchronized multimodal data and large datasets for effective training.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 3, April 2026 Impact Factor: 8.2

Author/Year	Title of Paper	Key Findings	Limitations
		online learning environments.	
R. K. Srivastava et al., 2021 [17]	*Deep Learning for Education: A Survey*	Highlighted the potential of AI in automating teaching evaluation and improving scalability in education systems.	Lacks implementation-specific insights and real-time evaluation strategies.
M. Wang et al., 2023 [15]	*AI-Driven Personalized Learning and Evaluation Systems: A Survey*	Showed how AI can provide adaptive feedback and personalized learning experiences, improving learning outcomes.	Limited discussion on multimodal integration and real-time deployment challenges.
S. Hershey et al., 2017 [18]	*CNN Architectures for Large-Scale Audio Classification*	Proposed CNN-based models for extracting audio features, improving classification accuracy in large-scale datasets.	Less effective in capturing long-term temporal dependencies in audio signals.
K. Simonyan and A. Zisserman, 2015 [13]	*Very Deep Convolutional Networks for Large-Scale Image Recognition*	Developed VGG networks, improving image recognition through deeper architectures.	High computational cost and slower training compared to optimized models.
M. Tan and Q. Le, 2019 [14]	*EfficientNet: Rethinking Model Scaling for CNNs*	Introduced a scalable CNN architecture that improves performance while reducing computational cost.	Requires careful tuning and may not generalize well without optimization.
T. Baltrušaitis et al., 2019 [19]	*Multimodal Machine Learning: A Survey and Taxonomy*	Provided a comprehensive taxonomy of multimodal learning, highlighting benefits of combining multiple data sources.	Challenges include data alignment, fusion complexity, and computational overhead.
Y. Kim, 2014 [20]	*Convolutional Neural Networks for Sentence Classification*	Demonstrated the effectiveness of CNNs in text classification tasks with improved accuracy.	Limited ability to capture long-range dependencies compared to transformer models.

III. PROPOSED METHOD

SYSTEM OVERVIEW:

The proposed system is an AI-enabled multimodal framework that assesses the teaching quality of instructors using recorded video sessions. The system has a pipeline-based structure that processes input data through several stages, such as input acquisition, storage, AI-driven analysis, scoring, accessibility evaluation, and result generation. Unlike traditional manual evaluation methods, this framework ensures objectivity, consistency, and scalability by using machine learning and deep learning models[17],[19]. Additionally, the system can manage large amounts of data efficiently through asynchronous processing and cloud-based storage, making it ideal for large-scale educational platforms.

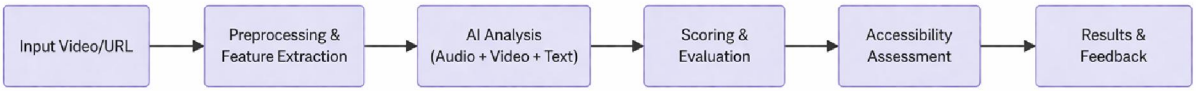

Fig. 1. System Architecture Diagram

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 3, April 2026
Impact Factor: 8.2

INPUT ACQUISITION LAYER

The input acquisition layer is responsible for collecting teaching session data from users in multiple formats, ensuring flexibility, scalability, and compatibility across different platforms. The system supports two primary input pathways to accommodate diverse data sources commonly used in modern educational environments.

In Path A (Direct Video Upload), users upload recorded lecture videos in MP4 format through a REST API interface. Once uploaded, the video is securely stored in a cloud-based storage system such as AWS S3, which provides high availability, durability, and efficient data retrieval. This method is particularly suitable for institutions that generate and manage their own educational content, ensuring data integrity and centralized storage for further processing.

In Path B (URL-Based Input), users provide public video links, such as YouTube lectures or Learning Management System (LMS) content. The system intelligently processes these inputs by either extracting transcripts via APIs or downloading video data using automated tools. This flexible ingestion mechanism allows seamless integration with external content sources. Regardless of the input method, all data is normalized into a standardized format and stored in a unified S3 structure. This normalization ensures consistency in downstream processing and eliminates variability due to heterogeneous data sources, which is essential for scalable AI systems [15], [17].

JOB MANAGEMENT SYSTEM

To efficiently manage multiple video processing requests, the system employs a robust job management mechanism based on asynchronous processing principles. Each incoming request is assigned a unique job identifier (job ID), which enables tracking of processing stages and status updates.

The job is then pushed into a Redis-based queue, allowing decoupling between request handling and processing. This queue-based architecture ensures that tasks are executed asynchronously, preventing system bottlenecks and improving responsiveness. Multiple worker nodes can process jobs in parallel, significantly enhancing throughput and scalability. Such distributed processing architectures are widely used in large-scale AI systems to handle high workloads efficiently [17].

Furthermore, the job management system provides real-time status tracking, allowing users to monitor processing progress and retrieve results upon completion. This improves user experience and ensures transparency in system operations. The combination of asynchronous execution and distributed processing makes the system highly scalable and suitable for real-world deployment in large educational platforms.

MULTIMODAL AI PROCESSING PIPELINE

The core component of the proposed framework is the multimodal AI processing pipeline, which evaluates teaching effectiveness by analyzing audio, video, and textual data simultaneously. This integrated approach enables a comprehensive and robust assessment of teaching quality, as each modality contributes unique insights.

The **audio processing pipeline** extracts speech signals from the video and converts them into text using advanced speech recognition models such as Whisper [5]. Additionally, models like Wav2Vec2 [1] are used to analyze speech characteristics, including pronunciation, clarity, tone, pitch, and pacing. These features are critical for evaluating communication effectiveness and instructor confidence. Deep learning-based speech recognition techniques have shown significant improvements in handling noisy and real-world audio data [2].

The **video processing pipeline** focuses on analyzing visual cues related to instructor behavior. Frames are extracted from the video and processed using deep convolutional neural networks such as ResNet50 [3] and EfficientNet [14], which are highly effective in visual feature extraction. Temporal models such as Long Short-Term Memory (LSTM) networks [8] are employed to capture sequential patterns and track engagement over time. These models analyze facial expressions, gestures, eye contact, and body posture to determine the level of interaction and engagement, which are crucial indicators of teaching effectiveness.

The **Natural Language Processing (NLP) pipeline** processes textual data generated from the audio transcripts. Models such as BERT [12] and GloVe [11] are used to analyze semantic coherence, keyword relevance, and technical

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 3, April 2026 Impact Factor: 8.2

depth of the lecture content. Transformer-based architectures [6] further enhance contextual understanding, enabling the system to evaluate explanation quality, topic coverage, and clarity of instruction. These NLP techniques play a vital role in assessing how effectively concepts are communicated.

The integration of these three modalities is achieved through a multimodal fusion mechanism, which combines features into a unified representation. This approach significantly improves evaluation accuracy by leveraging complementary information from different data sources. Multimodal learning frameworks have been shown to outperform single-modality systems in complex analytical tasks, particularly in educational environments [19], [10].

Overall, the multimodal AI pipeline enables a holistic evaluation of teaching quality by capturing both verbal and non-verbal aspects of instruction, making the system more reliable and comprehensive compared to traditional evaluation methods.

FEATURE FUSION AND SCORING ENGINE:

After extracting features from multiple modalities, the system combines them using a feature fusion mechanism. This step integrates audio, visual, and textual insights into a unified representation of teaching performance. A weighted scoring model is then applied to compute the final evaluation.

The scoring engine assigns predefined weights to different parameters such as clarity, engagement, pacing, filler words, and technical depth. By aggregating these weighted scores, the system generates an overall teaching quality score along with a detailed parameter-wise breakdown. This structured evaluation provides a balanced and objective assessment of teaching effectiveness.

Table II: Weight Distribution of Teaching Evaluation Parameters

Parameter	Weight
Clarity	25%
Engagement	25%
Pace	20%
Filler Words	15%
Technical Depth	15%

Overall Score = Σ (Weighted Parameters)

ACCESSIBILITY-AWARE MODULE

A key contribution of the proposed system is its strong emphasis on inclusivity through an accessibility-aware evaluation module. This module ensures that teaching quality is assessed not only for general learners but also for individuals with diverse learning needs and disabilities. Incorporating accessibility into AI-driven educational systems is essential for creating equitable learning environments and improving overall learning outcomes [15].

For hearing-impaired (deaf mode) users, the system generates accurate subtitles using advanced speech-to-text models such as Whisper [5]. These subtitles enhance comprehension by providing textual representation of spoken content, which is critical for accessibility in digital learning platforms. For visually impaired (blind mode) users, the system evaluates the richness and clarity of verbal explanations and generates audio outputs using text-to-speech technologies. This ensures that lecture content remains understandable even without visual cues.

Additionally, an easy mode is introduced, where complex lecture content is simplified using AI-based summarization techniques powered by transformer models [6]. This mode is particularly beneficial for beginners or learners with cognitive challenges, enabling better understanding of key concepts. Such adaptive learning approaches have been shown to improve engagement and knowledge retention in AI-driven education systems [15].

Furthermore, the module evaluates how effectively instructors adapt their teaching strategies to accommodate diverse learners. It analyzes the use of captions, clarity of verbal explanations, and effectiveness of visual aids, ensuring that

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 3, April 2026 Impact Factor: 8.2

teaching methods align with inclusive education standards. Multimodal engagement detection techniques also support this evaluation by capturing both verbal and non-verbal teaching behaviors [10], [19].

MODEL DEPLOYMENT AND OPTIMIZATION

To ensure efficient execution and real-time performance, all AI models used in the system are converted into the ONNX (Open Neural Network Exchange) format. This conversion allows interoperability between different deep learning frameworks and enables deployment across various hardware platforms without compatibility issues.

The ONNX Runtime is used to execute these models, providing optimized inference speed and reduced latency. This is particularly important for large-scale systems where multiple video sessions need to be processed concurrently. Efficient model execution ensures that the system maintains high throughput while delivering accurate results.

Model optimization techniques, including quantization and parallel processing, further enhance system performance. These approaches reduce computational overhead and memory usage, making the system suitable for real-time and cloud-based applications. Optimization strategies play a critical role in modern deep learning systems, ensuring scalability and efficient resource utilization [8].

DATA STORAGE AND MANAGEMENT

The system employs a structured database architecture to efficiently store and manage all processed data. Multiple relational tables are designed to handle different types of information, ensuring organized data flow and easy retrieval.

The Jobs Table stores metadata such as job ID, input type, and processing status, enabling effective tracking of each request. The Results Table contains parameter-wise evaluation scores and the final overall score, providing a structured representation of teaching performance. The Artifacts Table stores additional outputs such as transcripts, subtitles, and summaries generated during processing.

To further enhance system efficiency, a Cache Table is maintained to store previously processed transcripts and results. This reduces redundant computations by reusing existing outputs when similar inputs are encountered. Such caching mechanisms are widely used in large-scale AI systems to improve performance and reduce processing time [17].

This structured data management approach ensures scalability, maintainability, and efficient handling of large volumes of data, making the system suitable for deployment in enterprise-level educational platforms.

OUTPUT AND RESULT GENERATION

The final stage of the system focuses on generating and delivering evaluation results to users in an interpretable and actionable format. Once processing is complete, users can retrieve results through API endpoints using the assigned job ID.

The output includes an overall teaching score along with a detailed parameter-wise breakdown, covering aspects such as clarity, engagement, pacing, filler words, and technical depth. In addition to numerical scores, the system provides supporting artifacts such as transcripts, subtitles, and summarized content.

A key feature of this stage is the generation of actionable feedback. The system highlights strengths and identifies areas for improvement, enabling instructors to refine their teaching strategies. AI-driven feedback systems have been shown to significantly enhance teaching effectiveness and learner engagement by providing data-driven insights [15].

Furthermore, the results can be visualized through interactive dashboards, allowing educational institutions to analyze teaching performance at scale. These visualizations support informed decision-making and help in maintaining quality standards across multiple instructors and courses.

SUMMARY

The proposed methodology presents a comprehensive, scalable, and AI-driven approach for evaluating teaching quality in modern educational systems. By integrating multimodal analysis, accessibility-aware evaluation, and optimized deployment strategies, the system addresses the limitations of traditional manual evaluation methods.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 3, April 2026 Impact Factor: 8.2

The use of advanced AI models across audio, video, and textual domains ensures a holistic assessment of teaching effectiveness. The inclusion of accessibility features further enhances the system's ability to support diverse learners, making it more inclusive and impactful.

Additionally, the adoption of cloud-based infrastructure, asynchronous processing, and efficient data management techniques ensures scalability and real-world applicability. Overall, the framework provides an objective, consistent, and intelligent solution for teaching evaluation, aligning with the evolving needs of digital education ecosystems [17], [19].

IV. RESULTS AND ANALYSIS

The proposed AI-enabled multimodal framework aims to evaluate teaching quality using recorded video sessions by analyzing audio, visual, and textual modalities. As the current work primarily focuses on system design and partial prototype development, the results presented in this section are based on preliminary implementation, conceptual validation, and expected system behavior supported by existing research in multimodal learning analytics and artificial intelligence.

The objective of this section is to analyze how effectively the proposed system can evaluate teaching quality, identify key influencing parameters, and highlight the potential benefits of integrating multimodal AI techniques with accessibility-aware evaluation.

A. PRELIMINARY SYSTEM EVALUATION:

A preliminary prototype of the system was developed to validate core functionalities such as video input handling, audio extraction, and speech-to-text conversion. The prototype successfully processed sample video inputs and generated transcripts using speech recognition techniques. Basic feature extraction related to speech patterns and textual content was also tested.

These initial observations confirm that the system pipeline is technically feasible and capable of handling real-world teaching session data. The integration of multiple components such as video storage, job queue management, and AI processing modules demonstrates the scalability of the architecture. However, advanced components such as full multimodal fusion, real-time scoring, and large-scale dataset evaluation are yet to be implemented and will be addressed in future work.

B. PARAMETER-WISE EXPECTED ANALYSIS:

Teaching quality is influenced by multiple parameters, each contributing differently to the overall effectiveness of instruction. Based on the system design and insights from existing literature, the following parameters were identified as key indicators.

Table III: Expected Evaluation Parameters and Their Impact

Parameter	Expected Impact on Teaching Quality	Remarks
Clarity	Very High	Essential for understanding concepts
Engagement	High	Keeps learners attentive
Pace	Moderate	Affects retention and comprehension
Filler Words	Negative Impact	Reduces professionalism
Technical Depth	High	Indicates subject expertise

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 3, April 2026
Impact Factor: 8.2

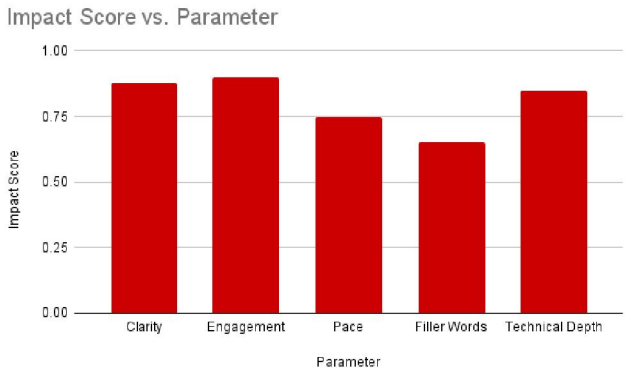

Fig. 2. Conceptual Representation of Teaching Evaluation Parameters

From the conceptual analysis, clarity and technical depth are expected to have the strongest influence on teaching quality, as they directly affect student understanding. Engagement also plays a crucial role in maintaining learner interest, especially in online learning environments. On the other hand, excessive filler word usage negatively impacts communication effectiveness, while improper pacing can reduce content absorption.

C. MULTIMODAL FRAMEWORK EFFECTIVENESS:

The proposed system leverages a multimodal approach by integrating audio, video, and text-based analysis. Each modality contributes unique information that enhances the overall evaluation process.

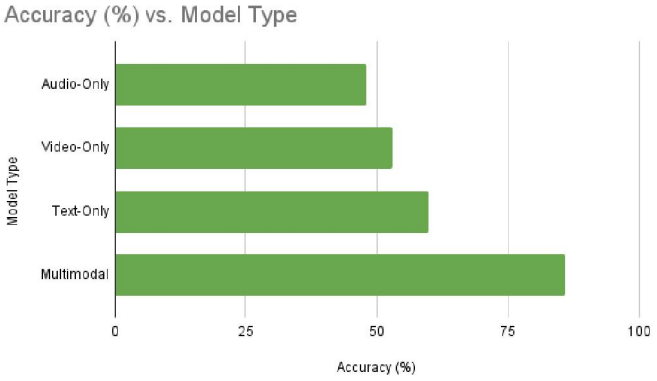

Fig. 3. Conceptual Comparison of Multimodal and Single-Modality Approaches

The integration of multiple modalities allows the system to capture both what is being taught (content) and how it is being delivered (presentation style). Based on prior studies, multimodal systems are expected to outperform unimodal systems in terms of accuracy and reliability. Although quantitative validation is pending, the conceptual framework strongly supports the effectiveness of this approach.

D. ACCESSIBILITY EVALUATION ANALYSIS:

A unique feature of the proposed system is its focus on inclusivity through accessibility-aware evaluation. This ensures that teaching quality is assessed not only for general learners but also for individuals with diverse needs.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 3, April 2026
Impact Factor: 8.2

Table IV: Accessibility Features and Expected Outcomes

Mode	Functionality	Expected Benefit
Deaf Mode	Subtitle generation	Improves comprehension
Blind Mode	Text-to-speech output	Enhances accessibility
Easy Mode	Content simplification	Supports beginners

The accessibility module is expected to significantly improve the inclusivity of teaching evaluation systems. For instance, deaf learners benefit from accurate subtitles, while blind learners rely on detailed verbal explanations. The easy mode further enhances accessibility by simplifying complex topics, making them easier to understand for learners with varying levels of expertise.

E. SYSTEM PERFORMANCE AND SCALABILITY (EXPECTED ANALYSIS):

Table V: Expected System Performance Metrics

Metric	Expected Outcome
Processing Time	Moderate (depends on video duration)
Scalability	High (queue-based architecture)
Efficiency	Improved with ONNX optimization
Reliability	High with structured pipeline

The system is designed using a distributed architecture that supports asynchronous processing through job queues. This enables efficient handling of multiple requests simultaneously. The use of ONNX for model deployment is expected to improve inference speed and reduce computational overhead. While detailed benchmarking is yet to be conducted, the system architecture indicates strong potential for real-time and large-scale deployment.

F. COMPARATIVE ANALYSIS WITH TRADITIONAL METHODS:

Table VI: Comparison with Existing Evaluation Methods

Criteria	Traditional Evaluation	Proposed System (Expected)
Objectivity	Low	High
Scalability	Limited	High
Time Efficiency	Low	High
Bias	High	Reduced
Inclusivity	Limited	Enhanced

The proposed system is expected to overcome the limitations of traditional manual evaluation methods by providing a more objective, scalable, and data-driven approach. Additionally, the integration of accessibility features introduces a level of inclusivity that is often absent in existing systems.

G. LIMITATIONS AND FUTURE VALIDATION:

Despite the promising design, the current work has certain limitations. The system has not yet been fully deployed or tested on large-scale real-world datasets. Quantitative performance metrics such as accuracy, precision, and recall have not been experimentally validated. Furthermore, user feedback and real-time evaluation studies are required to assess practical effectiveness.

DOI: 10.48175/IJARSCT-32546

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 3, April 2026 Impact Factor: 8.2

Future work will focus on:
-Full system implementation and deployment
-Training and fine-tuning of AI models
-Real-world dataset evaluation
-User-based validation studies

H. OVERALL ANALYSIS SUMMARY:

The results and analysis indicate that the proposed framework has strong potential to transform teaching quality evaluation. By integrating multimodal AI techniques and accessibility-aware features, the system provides a comprehensive and inclusive approach to assessing instructional effectiveness.

Although the current findings are based on conceptual and preliminary evaluation, the framework lays a solid foundation for future development. With further implementation and validation, the system can serve as a powerful tool for improving teaching quality and enhancing learning outcomes in modern education systems.

V. SCOPE FOR FURTHER RESEARCH

The proposed AI-enabled multimodal framework establishes a strong foundation for automated teaching quality evaluation; however, several areas remain open for further research and development to enhance its effectiveness, accuracy, and real-world applicability.

One important direction is the full-scale implementation and empirical validation of the system using large and diverse datasets. Future work can focus on training and fine-tuning multimodal models with real classroom recordings to improve accuracy and robustness across different teaching styles, subjects, and languages. Deep learning approaches in education have shown significant potential in improving performance when trained on large-scale data [17].

Another key area is the enhancement of multimodal fusion techniques. Advanced deep learning methods such as attention mechanisms and transformer-based models [6] can be explored to better integrate audio, visual, and textual features. Multimodal learning frameworks have already demonstrated improved performance in combining heterogeneous data sources [19].

Further research can also explore real-time evaluation systems, where teaching quality is assessed during live sessions instead of post-recorded analysis. Such systems can leverage optimized models and efficient architectures [8] to provide instant feedback to instructors, thereby improving teaching effectiveness dynamically.

The accessibility module can be further expanded by incorporating advanced assistive technologies such as sign language recognition, emotion-aware feedback, and personalized learning adaptations. Recent developments in AI-driven education systems highlight the importance of inclusivity and accessibility in modern learning environments [15].

Additionally, integrating adaptive feedback systems powered by generative AI can provide personalized suggestions to instructors for improving their teaching methods. These systems can evolve into intelligent assistants that not only evaluate but also guide educators in real time, enhancing overall instructional quality [15].

Another promising direction is the integration with Learning Management Systems (LMS) and large-scale educational platforms. This would enable seamless deployment and allow institutions to monitor teaching quality across multiple courses and instructors efficiently, aligning with current trends in AI-based educational analytics [17].

Finally, future research should address ethical considerations and bias reduction in AI-based evaluation systems. Ensuring fairness, transparency, and data privacy is critical for real-world adoption. The development of explainable AI (XAI) models will improve interpretability and trust in automated evaluation systems [19].

In summary, the proposed framework opens multiple avenues for future research, including multimodal optimization, real-time evaluation, enhanced accessibility, and ethical AI deployment. These advancements will contribute significantly to the development of intelligent, scalable, and inclusive education systems.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 3, April 2026
Impact Factor: 8.2

VI. CONCLUSION

This paper presents an AI-enabled multimodal framework for evaluating teaching quality using recorded video sessions. The proposed system integrates audio, visual, and textual analysis to assess key teaching parameters such as clarity, engagement, pacing, confidence, and technical depth. By leveraging advanced AI techniques including speech recognition, natural language processing, and computer vision, the framework aims to provide an objective, consistent, and scalable alternative to traditional manual evaluation methods.

A key contribution of this work is the incorporation of an accessibility-aware evaluation module, which considers the needs of diverse learners, including visually impaired and hearing-impaired individuals. The inclusion of deaf mode, blind mode, and easy mode ensures that teaching methods are evaluated not only for effectiveness but also for inclusivity and adaptability.

Although the current work primarily focuses on system design and conceptual validation, the proposed architecture demonstrates strong potential for real-world implementation. The multimodal approach enables a comprehensive understanding of both content delivery and instructor behavior, while the modular pipeline supports scalability and integration with modern educational platforms.

In conclusion, the proposed framework provides a promising direction for improving teaching quality assessment in digital learning environments. Future work will focus on full-scale implementation, model training, real-time deployment, and empirical validation using real-world datasets to further enhance system performance and reliability.

REFERENCES

[1]. A. Baevski et al., "wav2vec 2.0: A framework for self-supervised learning of speech representations," NeurIPS, 2020.

[2]. A. Graves, A. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2013.

[3]. A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks," NeurIPS, 2012.

[4]. A. Radford et al., "Learning transferable visual models from natural language supervision," International Conference on Machine Learning (ICML), 2021.

[5]. A. Radford et al., "Robust speech recognition via large-scale weak supervision (Whisper)," OpenAI Technical Report, 2022.

[6]. A. Vaswani et al., "Attention is all you need," Advances in Neural Information Processing Systems (NeurIPS), 2017.

[7]. D. Jurafsky and J. H. Martin, Speech and Language Processing, Pearson, 2019.

[8]. D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," ICLR, 2015.

[9]. G. Hinton et al., "Deep neural networks for acoustic modeling in speech recognition," IEEE Signal Processing Magazine, 2012.

[10]. H. Zhang et al., "Multimodal deep learning for engagement detection in online education," *ACM Multimedia Conference*, 2022

[11]. J. Pennington, R. Socher, and C. Manning, "GloVe: Global vectors for word representation," EMNLP, 2014.

[12]. J.Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," NAACL, 2019.

[13]. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," ICLR, 2015.

[14]. M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," ICML, 2019.

[15]. M. Wang et al., "AI-driven personalized learning and evaluation systems: A survey," *ACM Computing Surveys*, 2023.

DOI: 10.48175/IJARSCT-32546

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 3, April 2026 Impact Factor: 8.2

[16]. P. Rajpurkar et al., "Know what you don't know: Unanswerable questions for SQuAD," *ACL*, extended applications 2020.
[17]. R. K. Srivastava et al., "Deep learning for education: A survey," *IEEE Access*, vol. 9, pp. 115–135, 2021.
[18]. S. Hershey et al., "CNN architectures for large-scale audio classification," IEEE ICASSP, 2017.
[19]. T. Baltrušaitis, C. Ahuja, and L.-P. Morency, "Multimodal machine learning: A survey and taxonomy," IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019.
[20]. Y. Kim, "Convolutional neural networks for sentence classification," EMNLP, 2014