Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 47 No. 01 (2026)

Bridging Silence and Semantics: A Multimodal Review of Sign Language Recognition, Translation, and Adaptive Learning Systems

Antony Jacob [1], Angela Mary Anil [2], Alisha Ann Subash [3], Agnus Roy [4], Anu Rose Joy [5]

[1, 2, 3, 4, 5] *Amal Jyothi College of Engineering Kanjirappally, Kottayam, Kerala, India*

Abstract:- Bridging communication gaps for the deaf and mute community remains an open AI challenge, demanding systems that go beyond static sign recognition toward adaptive, emotion-aware interaction. While existing research has advanced isolated gesture recognition, few works address dynamic sentence translation, con- textual understanding, and learner adaptability in real-world envi- ronments. This review analyzes recent developments in multimodal learning—integrating vision, text, and speech—to enable seamless bidirectional communication and personalized education. It highlights the evolution from CNN-based recognition to transformer-driven sign language understanding and avatar-based delivery. This review synthesizes emerging multimodal approaches that blend recognition, translation, and emotion-aware adaptation into a unified assistive learning framework.

Keywords: *Sign Language Recognition, Neural Machine Translation, Emotion Detection, Adaptive Learning Platforms, Trans- former Models, Multimodal Deep Learning.*

1. Introduction

Advancements in deep learning have significantly trans- formed the domains of sign language recognition (SLR), facial emotion recognition (FER), and adaptive learning platforms (ALPs). In SLR, hybrid architectures such as CNN–LSTM models effectively combine spatial feature extraction with temporal sequence modeling, enabling accurate recognition of isolated signs, though at the cost of high computational demands [1], [3]. Recurrent CNNs (RCNNs) extend this capability to continuous sign streams without explicit seg- mentation, demonstrating robustness in real-time translation tasks [2]. Meanwhile, LSTM-only models present a resource- efficient alternative for simpler recognition scenarios [4], and lightweight CNN-based methods prioritize speed and deploya- bility for interactive gesture recognition [5]. Transfer learning from powerful CNN backbones has further improved recog- nition accuracy, with some ensemble approaches achieving near-perfect performance on specific sign language datasets [6], [7]. Emerging transformer-based architectures, including the Gated-Logarithmic Transformer and multilingual models like AfriSign, offer enhanced sequence modeling for complex translation tasks [8], [9].

Parallel to SLR developments, FER has benefited from CNN-based and hybrid temporal–spatial models capable of recognizing nuanced emotional states in real time [11]–[13]. These systems, especially when integrated with multimodal inputs, have potential applications in adaptive learning and human–computer interaction [14], [15]. In education, AI- enabled ALPs leverage personalization algorithms, affective computing, and inclusive design principles to cater to diverse learners' needs [16]–[18]. Machine learning techniques, in- cluding neural networks for learning style detection and gam- ification strategies, have shown promise in boosting learner engagement and effectiveness [19], [20]. Despite these ad- vances, challenges remain in achieving scalability, robustness under unconstrained conditions, and balanced performance across modalities—highlighting the need for further research and optimization.

2. Literature Review

A. CNN–LSTM and Recurrent CNN-Based Methods

CNN–LSTM hybrids have proven effective in capturing both spatial and temporal characteristics of sign language ges- tures. Spatial features, such as hand orientation and shape, can be extracted using pre-trained CNN backbones like ResNet-50 and then modeled temporally using LSTMs to capture gesture dynamics [1]. Attention mechanisms further refine this process by emphasizing the most relevant frames, enhancing accuracy (≈ 92%) over traditional CNN–LSTM approaches [3]. These architectures perform well on isolated sign datasets such as RWTH-PHOENIX, achieving accuracies between 85–90% [1], and benefit from transfer learning to reduce training time. Recurrent convolutional neural networks (RCNNs) integrate convolutional feature extraction with recurrent temporal mod- eling to process continuous sign streams without explicit segmentation [2]. By employing staged optimization—training convolutional layers first and then recurrent layers—they address vanishing gradient issues and improve convergence stability, achieving 80–85% accuracy in realistic continuous signing tasks.

B. LSTM-Only and Lightweight CNN Models

LSTM-only approaches bypass CNN-based spatial process- ing, directly consuming raw or lightly preprocessed video frames to learn temporal gesture patterns [4]. This reduces complexity and makes such systems attractive for resource- constrained environments, with 88–90% accuracy on Indian Sign Language datasets. Lightweight CNNs for real-time ges- ture recognition prioritize speed and deployability [5], incor- porating efficient hand-localization and compact architectures to run at interactive frame rates on commodity hardware. Although robust to environmental variations via data augmen- tation, their shallow temporal modeling limits effectiveness for longer sequences.

C. Transfer Learning and CNN Ensemble Methods

Transfer learning from powerful CNN backbones improves recognition accuracy while reducing training data require- ments. In [6], ResNet50 and VGG16 were integrated into a Nepali translation pipeline with gTTS for speech output, achieving over 99% accuracy. Ensemble methods combin- ing multiple pretrained CNNs (e.g., Xception, DenseNet121, ResNet50) with max-voting achieved 99.92% on Bangla Sign Language datasets [7]. While highly accurate, these models can be computationally heavy and lack signer-independent evaluations.

D. Transformer and Hybrid Vision Models

Transformer-based architectures excel in long-range tempo- ral dependency modeling. The Gated-Logarithmic Transformer (GLoT) [8] outperforms baseline transformers in BLEU-4 scores, while AfriSign [9] demonstrates multilingual sign language translation with 94.6% accuracy. Hybrid CNN–ViT models [10], combined with Improved Residual Feed-Forward Networks (IRFFN) and Adaptive Tuna Swarm Optimization (ATSO), have achieved 98.69% accuracy on two-handed ISL recognition. These architectures combine local feature extrac- tion with holistic context but remain resource-intensive.

E. CNN-Based and Lightweight Emotion Recognition

Facial emotion recognition (FER) systems using CNNs [11] detect seven basic emotions in real time, informing adaptive teaching strategies. Lightweight CNNs with depthwise sep- arable convolutions [12] achieve similar accuracy with lower computational needs, enabling mobile deployment. Challenges include reduced robustness under varied lighting, occlusion, and cultural differences in expression.

F. Hybrid and Multimodal Emotion Recognition

Hybrid FER approaches combine spatial and temporal mod- eling with multimodal inputs. ResNet50 with CBAM and tem- poral convolutional networks (TCNs) [13] achieves (≈ 97%) accuracy for continuous monitoring, while multimodal systems [14] integrating facial and textual sentiment improve virtual assistant responsiveness. Reviews [15] highlight the promise of TCNs but emphasize the need for standardized cross-dataset benchmarks.

G. Adaptive Learning Platforms and Inclusive Design

AI-enabled adaptive learning platforms (ALPs) integrate personalization algorithms and inclusive design principles to cater to diverse learner needs. Frameworks like the multilevel OER adaptation approach [16] leverage accessibility metadata and user profiling. Reviews [17] identify dynamic learner modeling, adaptive sequencing, and affective computing as core ALP components, while UDL-based frameworks [18] offer inclusive guidelines. However, large-scale deployment studies remain scarce, and privacy concerns persist.

H. Machine Learning for Personalization and Engagement

Machine learning drives personalization in adaptive learning environments. Systematic reviews [19] report increased use of neural networks for learning-style detection, while gamifica- tion studies [20] show positive impacts from collaboration- based and motivation-driven approaches. Integration of gam- ification into adaptive algorithms remains an open research direction.

Table I: Model performance on datasets with accuracy

Model / Method	Task / Dataset	Accuracy
CNN–LSTM Hybrid (ResNet-50 + Attention)	Isolated Sign Recognition (RWTH-PHOENIX)	85–92%
RCNN (Recurrent CNN)	Continuous Sign Recognition (Real user videos)	80–85%
LSTM-Only	Indian Sign Language Recognition (ISL dataset)	88–90%
Lightweight CNN	Real-time Gesture Classification (Mobile/gesture datasets)	~90%
Transfer Learning (ResNet50, VGG16)	ASL to Nepali Text/Speech Translation (Nepali SL dataset)	>99%
CNN Ensemble (Max Voting)	Bangla Sign Language Recognition (Bangla SL dataset)	99.92%
Transformer (AfriSign, GLoT)	Multilingual Sign Language Translation (AfriSign, GLoT)	94.6%
Hybrid CNN-ViT (IRFFN + ATSO)	Two-handed Dynamic Sign Recognition (ISL dataset)	98.69%
CNN-Based Emotion Recognition	Facial Emotion Detection (FER2013)	85–97%
Hybrid/Multimodal FER (CBAM + TCN)	Continuous Emotion Monitoring (Student faces, virtual assistants)	≈97%

3. Objectives of the System

The primary aim of this research is to develop a robust, accurate, and efficient system for recognizing sign language, thereby improving communication for hearing-impaired indi- viduals through advanced deep learning techniques. The study integrates Convolutional Neural Networks (CNNs) for spatial feature extraction, such as ResNet [1], and Long Short-Term Memory (LSTM) networks for temporal sequence modeling [2], enabling recognition of both isolated and continuous sign gestures with high accuracy.

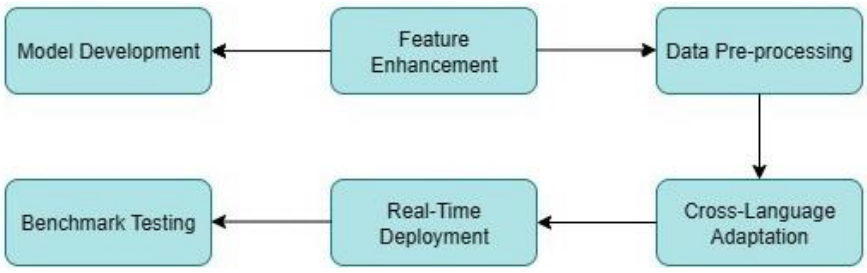

Figure 1: Proposed CNN–LSTM SLR Pipeline

Model Development: This work focuses on building and evaluating a hybrid CNN–LSTM model, leveraging architec- tures such as ResNet [1] and attention-based mechanisms [3] to enhance the ability to capture fine-grained hand and finger movements.

Feature Enhancement: Attention mechanisms [3] are em- ployed to prioritize critical spatial–temporal regions in sign language videos, effectively reducing noise and improving classification accuracy.

Data Preprocessing: To ensure model robustness, sign lan- guage video datasets [6] undergo standardization and prepa- ration, including background normalization, frame alignment, and hand/body posture detection.

Benchmark Testing: The proposed system is compared against state-of-the-art approaches [4], [7], [10] to highlight improvements in recognition accuracy, reliability, and real- time applicability.

Real-Time Deployment: Models are optimized for mobile and embedded devices [5], [6], enabling live translation of sign language into text or speech in real-world scenarios.

Cross-Language Adaptation: Finally, transfer learning methods [8], [9] are incorporated to adapt the system for multiple sign languages and dialects, ensuring scalability and inclusivity across linguistic contexts.

By meeting these objectives, the research supports the creation of inclusive communication tools aligned with AI-for-social-good principles [17], [19].

4. Methodology

This study employs a multi-stage deep learning framework integrating sign language recognition (SLR) and emotion- aware adaptive learning. The methodology comprises four components: data acquisition and preprocessing, model archi- tecture design, training and optimization, and integration with adaptive learning platforms (ALPs).

A. Data Acquisition and Preprocessing

Public datasets such as RWTH-PHOENIX and Indian Sign Language (ISL) corpora are utilized for capturing isolated and continuous sign gestures [1], [2], [4], while FER2013 provides diverse facial expression samples for facial emotion recog- nition (FER) [11]. Preprocessing includes region-of-interest localization (hands or face), normalization, frame resizing, and augmentation (random cropping, brightness adjustment, temporal frame sampling) to enhance robustness against envi- ronmental variations [5], [12].

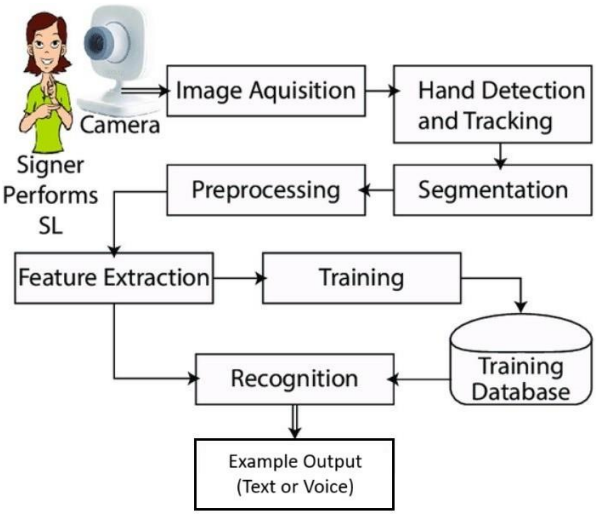

Fig. 2: System flowchart for Sign Language Recognition

B. Model Architecture Design

The SLR module combines spatial and temporal modeling via CNN–LSTM hybrids [1], [3], recurrent CNNs [2], and transformer-based architectures for long-range temporal de- pendency modeling [8], [9]. Transfer learning

from pre-trained CNNs (ResNet50, VGG16, DenseNet121) improves accuracy and reduces training time [6], [7]. For FER, lightweight CNNs with depthwise separable convolutions [12] and hybrid attention-based models [13], [14] are employed for real-time performance.

C. Training and Optimization

Training follows staged optimization [2], with convolutional layers pre-trained before fine-tuning temporal or transformer layers. Hyperparameters are optimized using metaheuristic strategies such as Adaptive Tuna Swarm Optimization (ATSO) [10]. Loss functions include categorical cross-entropy for classification and sequence-to-sequence loss for translation. Early stopping, learning rate scheduling, and transfer learning help prevent overfitting and improve convergence.

D. Integration with Adaptive Learning Platforms

The trained SLR and FER models are integrated into AI-enabled ALPs [16], [17]. FER informs real-time content adaptation by detecting students' emotional states [11], [14], while SLR enables sign language-based interaction for hearing-impaired learners [6], [9], [18]. The ALP leverages personalization algorithms, affective computing, and inclusive design frameworks [17], [18], [19], with gamification elements embedded to enhance engagement [20].

5. Benefits

The integration of AI-based sign language recognition en- ables seamless communication for deaf and mute learners by translating gestures from video frames into accurate text or speech outputs, thereby fostering social inclusion, inde- pendence, and equal participation in daily activities [1], [3]. This technology significantly reduces communication barri- ers between hearing-impaired individuals and the hearing community, enabling meaningful participation in mainstream education and professional environments [6].

Neural Machine Translation (NMT) ensures that recognized signs are not only transcribed but also converted into gram- matically correct and contextually relevant sentences [9], [12]. Instead of isolated words, learners receive coherent, struc- tured feedback, which makes complex educational content more accessible, supports literacy development, and facilitates smoother two-way communication.

Adaptive Learning Algorithms tailor educational materials based on learner performance, pace, and preferences [15]. This dynamic customization accommodates varied cognitive and motor abilities in differently abled students, improving engagement, retention, and long-term learning outcomes. By adapting content delivery in real time, such systems can empower learners to build confidence and progress at their own rhythm.

Emotion-Aware AI plays a complementary role by detect- ing facial expressions and emotional cues to adjust teaching strategies as learning unfolds [2], [4]. By identifying signs of frustration, confusion, or boredom, the system can respond empathetically, providing encouragement or adjusting task difficulty. This fosters a more human-like interaction, creat- ing a supportive, motivational, and highly engaging learning environment.

6. Results and Discussion

Deep learning models such as *CNNs, LSTMs, and hybrid ResNet architectures* have demonstrated recognition accuracies above 95% in controlled environments [1], [3]. Temporal mod- eling from video frames has proven particularly effective, as it captures the fluid motion of continuous signs, achieving higher recognition rates compared to static image-based approaches [6]. These results validate the potential of hybrid models for real-time applications.

In terms of translation, *Transformer-based architectures and sequence-to-sequence models* offer significant improvements in converting recognized signs into natural language text [9], [12]. They preserve grammatical structure and semantic context while supporting multilingual adaptation and integra- tion of domain-specific vocabulary. This makes them highly versatile for use in classrooms, workplaces, and multilingual communities.

Adaptive Learning Algorithms further enhance the system's educational value. Studies show that learners exposed to adaptive systems report higher satisfaction, faster skill ac- quisition, and longer retention compared to

traditional static teaching methods [15]. By using reinforcement learning and personalized recommendation engines, the system modifies lesson complexity, pacing, and modality dynamically, ensuring learning remains relevant and engaging.

For *Emotion-Aware AI*, the combination of facial recogni- tion with attention mechanisms (e.g., CBAM, Vision Trans- formers) and temporal models (TCNs) has achieved high real- time accuracy—up to 97%—in detecting emotional states [2], [4]. When embedded into learning platforms, these models provide an empathetic dimension to digital interaction, ad- dressing frustration early and sustaining motivation.

Finally, the integration of these four elements into an *Assis- tive Ecosystem* demonstrates transformative potential. A single application that unifies accurate sign recognition, meaningful translation, adaptive learning, and empathetic interaction can create a holistic, inclusive solution tailored for deaf and mute students [1], [4], [9], [15].

7. Key Challenges

Despite encouraging progress, several challenges remain unresolved. A major limitation is the *lack of data diversity and availability*, since large annotated datasets covering sign language dialects and emotion recognition in differently abled populations are scarce [1], [3]. Without this diversity, models risk bias and reduced generalization in real-world contexts.

Another barrier is *contextual understanding within NMT*. While current systems perform well on literal translations, they struggle to preserve idiomatic expressions, cultural nuances, and conversational tone [9]. This often results in outputs that are technically accurate but awkward or unnatural, limiting usability in day-to-day interactions.

Real-time integration also remains difficult. Combining sign recognition, translation, adaptive learning, and emotion detec- tion into a single mobile-friendly application requires models that are both computationally efficient and highly accurate [5], [7]. Achieving this balance remains a core technical hurdle.

In terms of personalization, *user adaptability* presents chal- lenges. Adaptive systems require sufficient learner interaction data before tailoring content effectively, which may delay benefits for first-time users [15]. New strategies for rapid personalization are needed to bridge this gap.

For emotion detection, maintaining *robustness across envi- ronments* is challenging. Variations in lighting, camera quality, gesture occlusions, and cultural differences in expressing emo- tions can significantly degrade system accuracy [2], [4].

Finally, ensuring *deployment accessibility* is critical. For maximum impact, applications must operate effectively on low-cost devices and under low-bandwidth conditions [8]. Without this, adoption will remain limited in rural or under- served communities where the need is often greatest.

8. Future Research Directions

While this study demonstrates significant advancements, several promising directions remain for exploration. One of the most pressing is *expanding vocabulary*. Current models are often restricted to limited lexicons; scaling to larger vocabu- laries, idiomatic expressions, and complex grammar structures will bring systems closer to natural human communication [7], [9], [10].

Another key avenue is *multi-modal integration*, combining video-based gesture recognition with depth sensors, skeletal pose estimation, and facial expression analysis to capture richer semantic context [11], [13], [14]. This integration can make recognition more robust across diverse environments and signing styles.

Edge deployment is equally important. Developing lightweight models optimized for smartphones, AR glasses, and IoT devices [5], [6] will enable widespread adoption without dependence on high-end GPUs, ensuring accessibility in everyday contexts.

Cross-language learning offers another promising path. Transfer learning approaches can enable adaptation across dif- ferent sign languages with minimal annotated data, addressing global communication needs [8], [9].

Strengthening *robustness to environmental variations* will also be vital, as models must reliably handle differences in lighting, backgrounds, clothing, and signing environments [4], [10].

In parallel, *user personalization* will play a major role. Systems that adapt to an individual's signing style and learn- ing patterns over time [17], [18] can deliver more natural, effective, and engaging experiences.

At the application level, future research can explore *in- tegration with assistive technologies*, embedding recognition tools in video conferencing platforms, classroom kiosks, or wearable devices for seamless real-world use [16], [20].

Finally, strong emphasis must be placed on *ethical con- siderations*. Privacy-preserving machine learning approaches [19] will ensure that sensitive data, such as sign recordings and emotional states, are handled securely and responsibly, fostering trust in these technologies.

9. Conclusion

This research emphasizes the advancements in sign lan- guage recognition through hybrid deep learning frameworks such as CNN–LSTM, which build upon prior works [1]– [4], [7], [10]. These models achieve improved recognition accu- racy for both isolated and continuous gestures, demonstrating robustness across diverse acquisition conditions. Compared with traditional approaches, the integration of convolutional and recurrent architectures offers superior adaptability and performance, aligning with findings in earlier studies [1], [4], [7].

The contributions outlined here have practical implications for assistive technologies that enhance accessibility in com- munication [6], [9], [16], supporting inclusive education and bridging barriers for differently abled learners. Looking ahead, adopting advanced architectures such as Transformers [8], [10] and embedding privacy-focused, real-time solutions [17], [19] will be essential for achieving scalable, trustworthy, and user- centered systems.

By unifying sign recognition, adaptive personalization, and ethical AI design, future research can foster more inclusive, accessible, and empathetic human–computer interaction.

References

[1] P. V. M. S. Prasad and K. N. V. Madhusudana, "Video-Based Sign Language Recognition via ResNet and LSTM Network," *Imaging*, vol. 10, no. 6, pp. 149–161, Jun. 2024, doi: 10.3390/imaging10060149.

[2] R. Cui, H. Liu, and C. Zhang, "Recurrent Convolutional Neural Networks for Continuous Sign Language Recognition by Staged Optimization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog- nit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 1610–1618, doi: 10.1109/CVPR.2017.176.

[3] M. Shahin and M. I. Daoud, "Isolated Video-Based Sign Language Recognition Using a Hybrid CNN-LSTM Framework Based on Atten- tion Mechanism," *Electronics*, vol. 13, no. 7, pp. 1229–1242, Apr. 2024, doi: 10.3390/electronics13071229.

[4] A. K. Sharma and S. S. Bhatia, "ISL Sign Language Recognition Using LSTM-Driven Deep Learning Model," *Journal of Engineering Science*, vol. 15, no. 4, pp. 245–256, 2024.

[5] O. Ko¨pu¨klu¨, A. Gunduz, N. Ko¨se, and G. Rigoll, "Real-time Hand Gesture Detection and Classification Using Convolutional Neural Net- works," *arXiv preprint* arXiv:1901.10323, Oct. 2019.

[6] M. Bhattarai, S. Ghimire, and P. Pokharel, "Advancing human–computer interaction: An American Sign Language to Nepali text and speech translation system," *Visual Informatics*, vol. 8, pp. 68–77, 2024, doi: 10.1016/j.visinf.2024.01.005.

[7] M. M. Hossain, M. S. Alam, and M. A. H. Akhand, "Bangla Sign Language Recognition using a Max Voting-based Ensemble Model," *ICT Express*, 2025, doi: 10.1016/j.icte.2025.01.002.

[8] C. F. Gonzalez, N. Ke, and A. Bharambe, "GLoT: Gated Loga- rithmic Transformer for Sign Language Translation," *arXiv preprint* arXiv:2502.12223, 2025.

[9] A. O. Awobuluyi, O. Awobuluyi, and T. Adewumi, "AfriSign: African Sign Language Translation for Humanitarian Applications," *Journal of Computational Social Science*, vol. 8, no. 1, pp. 1–21, 2025, doi: 10.1007/s42001-025-00227-7.

[10] S. S. Kaur, R. Kumar, and G. Singh, "ISL Two-Handed Dynamic Sign Language Recognition Using Enhanced Convolutional Transformer with Adaptive Tuna Swarm Optimization," *Sensors*, vol. 25, no. 17, pp. 3652, 2025, doi: 10.3390/s25173652.

[11] S. A. Salloum, K. M. Alomari, A. M. Alfaisal, R. A. Aljanada, and A. Basiouni, "Emotion Recognition for Enhanced Learning: Using AI to Detect Students' Emotions and Adjust Teaching Methods," *Smart Learning Environments*, 2025, doi: 10.1186/s40561-025-00212-9.

[12] D. B. Gunda and R. Bhavani, "A Lightweight Facial Emotion Recogni- tion Model for Resource-Constrained Devices," 2024.

[13] M. Aly, "Revolutionizing Online Education: Advanced Facial Ex- pression Recognition for Real-Time Student Progress Tracking via Deep Learning Model," *Multimedia Tools and Applications*, 2025, doi: 10.1007/s11042-025-19163-4.

[14] S. G. Rajesh, S. V. Madangarli, G. S. Pisharady, and R. Subrahmanyam, "Enhancement of Virtual Assistants Through Multimodal AI for Emo- tion Recognition," *Journal of Intelligent Systems and Internet of Things*, 2025.

[15] M. A. Puspasari, D. Yulianti, and M. T. Ibrahim, "Emotion Categorization from Facial Expressions: A Review of Datasets, Methods, and Research Directions," *Neurocomputing*, 2025, doi: 10.1016/j.neucom. 2025.128341.

[16] P. Ingave´lez-Guerra, R. A´ lvarez-Garc´ıa, and M. J. Rodr´ıguez-Triana, "Automatic adaptation of open educational resources: An approach from a multilevel methodology based on students' preferences, educational special needs, artificial intelligence, and accessibility metadata," *Com- puters and Education: Artificial Intelligence*, vol. 3, Art. no. 100063, 2022, doi: 10.1016/j.caeai.2022.100063.

[17] Y. Tan, W. Zhang, and S. Liu, "Artificial intelligence-enabled adaptive learning platforms: A review," *Computers and Educa- tion: Artificial Intelligence*, vol. 6, Art. no. 100262, 2025, doi: 10.1016/j.caeai. 2025.100262.

[18] K. Song, H. Weng, X. Zhao, and Y. Wang, "A framework for in- clusive AI learning design for diverse learners," *Computers and Ed- ucation: Artificial Intelligence*, vol. 5, Art. no. 100208, 2024, doi: 10.1016/j. caeai.2024.100208.

[19] E. Essa, S. Ali, and A. Abdellatif, "Personalized adaptive learn- ing technologies based on machine learning techniques to identify learning styles: A systematic literature review," *Computers and Ed- ucation: Artificial Intelligence*, vol. 4, Art. no. 100178, 2023, doi: 10.1016/j.caeai.2023.100178.

[20] L. Cui, X. Li, and Z. Li, "Impact of gamified learning experience on on- line learning effectiveness," *Computers and Education: Artificial Intelli- gence*, vol. 5, Art. no. 100197, 2024, doi: 10.1016/j.caeai.2024. 100197.