Available on: http://gjetr.ep-journals.org/ E-ISSN: 3051-3782, P-ISSN: 3051-3774

Global Journal of Engineering and Technology Review

Volume: 02 Issue: 05 May 2026

Conceptual Framework for AI Emotional Recognition in Children with Learning Difficulties for Paediatric Screening

Marcella Peter
DeTeC, School of Computing and Creative Media, University of Technology Sarawak, Sibu, Malaysia, ORCID ID:0009-0005-7349-8625

Jackie Ting Tiew Wei
DeTeC, School of Computing and Creative Media, University of Technology Sarawak, Sibu, Malaysia.

Khairunnisa Ibrahim
ASSET, School of Computing and Creative Media, University of Technology Sarawak, Sibu, Malaysia.

Published Date: May 30, 2026 **Article DOI : 10.65150/EP-gjetr/V2E5/2026-05**

ABSTRACT: Assessing the emotional states of children with learning difficulties (LD) is critical for effective pedagogical and clinical intervention. However, existing diagnostic methods rely on subjective observations which are prone to human bias and lack a localised Asian-centric perspective. This concept paper proposes an AI-based Facial Emotional Recognition (FER) initiative designed to provide objective, real-time assessment of emotional engagement in children with dyslexia during learning activities. The framework utilises a mixed-methods approach, combining a Convolutional Neural Network (CNN) pipeline for video-to-image emotion classification with clinical field validation at the SPARK Child Development Centre. The system leverages 3D depth-sensing technology via Intel RealSense to extract 468 facial landmarks. Preliminary validation of the initiative's expert-rated dataset achieved a Cohen's Kappa of 0.87, indicating excellent inter-rater reliability. By bridging the gap between traditional psychiatric screening and modern affective computing, this digitalisation initiative supports the SDG 4 goals, offering a scalable tool for more precise and timely support for children with learning disabilities.

KEYWORDS: Achievement of Emotion, Affective Computing, CNN, Dyslexia, Facial Emotional Recognition, Machine Learning

INTRODUCTION

The emotional well-being of children with learning difficulties (LD), particularly dyslexia, is a critical yet underserved predictor of academic success. Research in Control-Value Theory developed by Reinhard Pekrun indicates that specific emotions such as frustration, confusion, and anxiety significantly impact the cognitive persistence of students with learning disabilities (Xie & Fan, 2025). In the Malaysian context, the lack of localised, Asian-centric facial stimuli datasets has hindered the development of accurate digital screening tools. (Adi et al ,2024) directly address Malaysian children with dyslexia, showing higher anxiety and behavioural issues. These emotions play a vital role in determining student motivation, task persistence, and learning strategy adoption. Schooling years represent a crucial developmental window during which children must learn to modulate emotional and cognitive functions. However, those struggling with learning differences frequently experience heightened academic distress and anxiety across varied educational settings (Pekrun & Goetz, 2023). Despite massive advancements in computer vision and AI, existing Facial Emotion Recognition (FER) systems are rarely optimized for paediatric or clinical settings.

The majority of deep learning models are trained on adult, Western databases (such as FER2013 and AffectNet), which do not translate well to the subtle, spontaneous, and diverse facial expressions of Asian paediatric cohorts (He et al., 2016; Li & Deng, 2020). While the classic deep learning frameworks are highly powerful, they have historically been trained on adult and Western benchmarks. This structural difference calls for paediatric-focused architectures that account for developmental and cultural variations in expression trajectories (Wang et al., 2024), especially when modeling non-verbal distress cues in child-computer interactions (Choudhary & Prajapati, 2026). To bridge this gap, this paper introduces a conceptual framework for a real-time, non-invasive FER system tailored specifically to children with learning difficulties in Malaysia. Developed through a computational and clinical partnership between University Technology Sarawak (UTS) and SPARK Child Development Centre, this framework aims to transition emotional screening from a subjective clinical art to an objective digital science, aligning directly with United Nations Sustainable Development Goal 4 (Quality Education) and modern healthcare digitalization.

MOTIVATION

Traditional emotional assessments in special education and paediatric psychiatric environments face significant operational challenges. In the Malaysian context, clinicians and educators primarily rely on manual instruments (such as the Smileyometer) and unstructured qualitative observations to monitor a child's frustration or distress during learning activities. These traditional methods present three fundamental limitations:

Subjectivity and Observer Bias

Human evaluation of emotional expressions is inherently subjective. The interpretation of facial cues can vary widely based on the clinician's training, cultural background, and personal biases. Furthermore, children with specific learning difficulties present significantly higher levels of emotional distress and psychological adjustment challenges compared to typically developing peers (Adi et al., 2024). Despite that, they often mask their frustration, expressing it through minute micro-expressions (e.g., fleeting lip tensions or subtle eyebrow furrowing) that are incredibly difficult to capture manually.

Observation Fatigue

Standard diagnostic and special educational sessions typically span 1 to 1.5 hours. Expecting a clinician or teacher to maintain continuous, hyper-focused visual monitoring of a child's facial configurations while simultaneously administering cognitive tests is operationally unrealistic. Consequently, transient micro-expressions signifying extreme cognitive overload are frequently missed.

Lack of Quantifiable Metrics

Subjective notes do not provide a structured, longitudinal baseline. Without a continuous, quantifiable metric, such as an "Engagement Score" or "Frustration Index" tracked frame-by-frame over time, it is difficult for multidisciplinary care teams to measure the progression of a child's emotional regulation abilities or objectively evaluate the efficacy of targeted interventions.

METHODOLOGY

The proposed system addresses these challenges by introducing a passive, high-fidelity digital dashboard overlay that operates alongside standard clinical workflows. By bridging advanced computer vision with paediatric clinical psychiatry, this framework acts as a continuous diagnostic aid rather than a replacement for human expertise.

Figure 1 illustrates a structured, sequential four-stage pipeline designed to process real-time facial geometry and provide quantifiable emotional indices output.

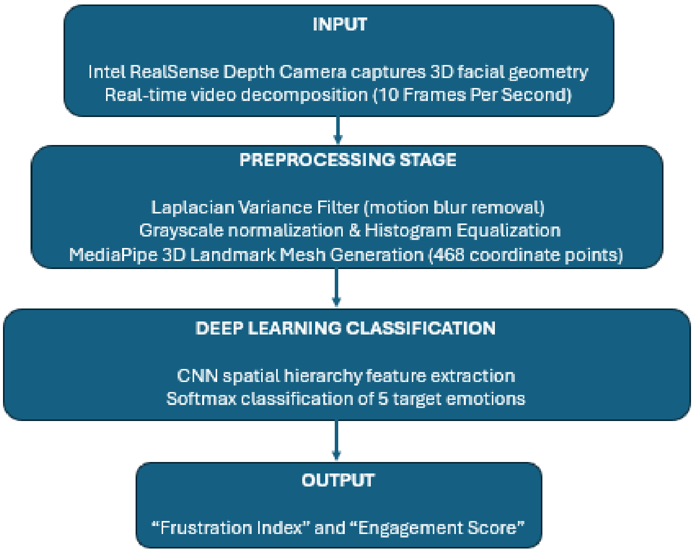

Figure 1. Proposed AI-FER pipeline

Input Stage

This is the entry point of the framework, focusing on high-fidelity data acquisition. We will utilize Intel RealSense Depth Camera, instead of a standard 2D RGB camera. The pipeline begins by capturing 3D facial geometry. This hardware choice is critical for establishing a light-invariant baseline, ensuring that changes in classroom shadows do not interfere with emotion tracking. Then, Real-Time video decomposition will be conducted, where the incoming continuous video stream is decomposed into individual frame sequences at a rate of 10 Frames Per Second (FPS). This balances immediate, real-time diagnostic feedback with low computational overhead.

Preprocessing Stage

This stage prepares the extracted raw frames, removing environmental and physical noise before classification. Restless movements or head-tossing (common in children with learning difficulties) can cause significant motion blur. The Laplacian Variance filter will be used to calculate image sharpness. Any frame falling below a strict threshold is discarded to prevent motion artefacts from degrading classification accuracy. Grayscale Normalization & Histogram Equalization are used to standardize image contrast and eliminate lighting discrepancies across different clinical environments. The preprocessed frame is mapped with a 3D mesh consisting of 468 precise coordinate points. The system tracks the physical distances between muscle groups (e.g., the eyes, brow, and mouth) in three dimensions, ensuring accurate tracking even if the child rotates their head.

Deep Learning Classification Stage

The final stage translates the clean geometric coordinates and image patches into clinical metrics. A CNN (optimized via MobileNetV2) will extract micro-spatial features from the face, identifying extremely subtle facial tensions. The network uses a dense softmax layer to calculate probability distributions across the 5 basic target emotions (Happiness, Sadness, Anger, Fear, and Neutral).

Output Stage

The ultimate output of the AI-FER system is a set of continuous, objective metrics, specifically a computed Frustration Index and an Engagement Score, which are generated frame-by-frame. Unlike traditional classification tools that merely output static, discrete emotion labels (such as 'sad' or 'angry'), this continuous metrics model allows multidisciplinary care teams to observe the rise and fall of a child's cognitive load over time. When plotted against specific timeline benchmarks of academic tests, these metrics empower educators and clinicians to isolate the exact moments of cognitive fatigue, paving the way for highly personalized learning strategies and early, precise paediatric interventions.

TECHNICAL DESIGN AND CNN ARCHITECTURE

The core software pipeline is simulated using Python, OpenCV, MediaPipe, and TensorFlow/Keras, executing a three-stage real-time computer vision process.

Video Decomposition & Filtering: High-resolution video captured via an Intel RealSense depth camera is broken down into frame sequences at 10 frames per second (FPS). This frame rate ensures near-instantaneous feedback without exhausting the computing resources of standard clinical hardware. To manage the physical restlessness common in children with learning differences, each frame undergoes a real-time Laplacian variance filter; frames with a variance score below $T = 100$ (representing significant motion blur) are automatically discarded.

Light-Invariant 3D Face Mesh Tracking: Traditional RGB cameras struggle with varying ambient lighting in classrooms. By using a depth camera, our framework captures 3D point cloud coordinate maps. The MediaPipe framework is utilized to map 468 discrete 3D landmark points across the child's facial structure, tracking Euclidean distance changes between muscle groups (such as eyelids and outer lip corners) independently of ambient shadows or head rotations. Following the recommendation by Romani-Sponchiado et al. (2015), the spatial geometry will then be compared against paediatric expression baselines to ensure precise tracking.

CNN Spatial Feature Extraction: The isolated facial Region of Interest (ROI) is processed through a Convolutional Neural Network (CNN) optimized with a MobileNetV2 architecture. The network extracts deep spatial features and computes a probability distribution across five target emotional classes (happiness, sadness, anger, fear, and neutrality) via a Dense Softmax output layer:

$$E_i = Softmax(W.x + b)$$

where E_i represents the probability of emotion class i, x is the spatial feature vector, and $W.x$ and b are the learned weights and biases.

CLINICAL INTEGRATION

The system is designed to run non-invasively in the background during one-on-one diagnostic sessions at the SPARK Child Development Centre. It integrates alongside standard screening instruments, including the Ages and Stages Questionnaire 3rd Edition (ASQ-3), the Colorado Learning Difficulties Questionnaire (CLDQ), and the Wechsler Fluid Reasoning Index (FRI). By matching the real-time AI emotional data streams (such as a spike in the computed Frustration Index) with specific literacy subtests, clinicians receive an objective mapping of the child's cognitive load threshold. This is supported by Nag et al. (2020), who combined real-time tracking with active cognitive tasks, enabling automated tracking technology to objectively map a child's socio-emotional and cognitive thresholds.

PRELIMINARY RESULTS

To ensure clinical and technological viability, the framework adopts a rigorous, dual-phase validation and implementation strategy:

Phase 1: Paediatric Database Development

To resolve the localized "Asian child facial gap", we conducted a preliminary dataset collection phase.

- **Participants:** Ten children (N=10) aged 4 to 12 years were recruited, alongside five paediatric experts.
- **Procedure:** Children performed cognitive achievement tasks such as reading passages and solving puzzles, while the depth camera captured spontaneous reactions. High-quality image frames were extracted from these streams.
- **Expert Panel Validation:** The five paediatric experts independently rated and annotated the extracted images to establish the ground-truth emotional category.
- **Inter-Rater Reliability:** To evaluate consensus among the human experts and the AI system, Cohen's Kappa (K) was computed. The statistical analysis yielded an average Cohen's Kappa score of 0.87. This score falls within the "Almost Perfect" agreement range (0.81 - 1.00) according to standard biostatistical benchmarks, validating the reliability of our localized child database.

Phase 2: System Field Deployment and Statistical Analysis

Following successful database validation, the system is being prepared for a broader clinical field deployment.

- **Target Sample Size:** Purposive sampling is utilized to recruit a target of 140 participants (N=140), which strategically accounts for a potential 20% clinical attrition rate. According to Van & Morgan (2007), to conduct a linear regression, researchers require a minimum sample size to achieve a statistical power of 0.80 at an alpha level of 0.05. Consequently, this study's strategy for addressing attrition is robust; after accounting for the expected 20% attrition rate, which may result from participants dropping out, relocating, or producing unusable and blurry video frames, the active participant pool is expected to remain at approximately 112 individuals. This number surpasses the necessary threshold of 100 participants, thereby ensuring sufficient statistical power for the analysis.
- **Screening Metrics:** Eligible participants are children aged 4.5 to 11.4 years screened for learning difficulties, maintaining a Fluid Reasoning Index (FRI) score ≥ 90 to control for severe intellectual impairments. Children with comorbid conditions (such as Autism Spectrum Disorder) are excluded to ensure facial muscle data remains unconfounded.
- **Statistical Pipeline:** Collected emotional indices will undergo descriptive and normal distribution analyses. Bivariate correlation will evaluate the relationship between AI-detected emotional states and clinical socio-emotional functioning (via the SDQ), and linear regression models will assess the degree to which transient frustration states predict overall task engagement.

Corresponding Author: Marcella Peter

DISCUSSION

The implementation of this digital framework yields positive multi-dimensional outcomes across several fields:

- **Academic and Scientific Contribution:** This framework addresses the critical "Asian child facial gap" by providing a reliable database of spontaneous paediatric facial expressions specifically tailored to the Malaysian socio-cultural context. This expands the global repository of affective computing datasets.
- **Clinical Diagnostic Advancements:** By delivering a continuous, objective "Frustration Index" and "Engagement Score" alongside traditional evaluations, the system reduces diagnostic latency, minimizes manual observation bias, and supports psychiatrists in designing highly tailored educational therapy plans.
- **Ethical and Privacy Standards:** The framework establishes a secure, split-data architecture in compliance with the Medical Research & Ethics Committee (MREC) approval under reference 22-02144-MS7. Raw video data is securely de-identified and destroyed after 3 years, while numerical coordinate datasets are maintained under double-encryption for 6 years before secure disposal.
- **Social Inclusivity:** Modernizing special education screening ensures that vulnerable children with learning difficulties receive timely, precise interventions, helping them overcome emotional barriers to education.

CONCLUSION

This conceptual framework presents a non-invasive, objective, and culturally representative approach to tracking emotional states in children with learning difficulties. By combining 3D depth-sensing hardware with specialized CNN spatial architectures, the system successfully addresses the limitations of environmental lighting, motion blur, and subjective human bias that have historically constrained paediatric emotional assessment. Future extensions of this work will focus on integrating Explainable AI (XAI) architectures, such as Grad-CAM. By visually highlighting the exact facial region (e.g., brow tension or mouth compression) that led to a specific emotion classification, the system will provide clinical staff with transparent, trustworthy, and actionable clinical data, paving the way for empathetic, digitalized classroom environments.

ACKNOWLEDGMENT

This project is supported by the University Technology Sarawak (UTS) internal research grant (2/2021/02). The authors would like to express their gratitude to the SPARK Child Development Centre (SPARK CDC), Petaling Jaya, Selangor, Malaysia, for collaborating and assisting with the ethical clearance process and clinical data collection. The clinical data application received formal ethical clearance from the Medical Research & Ethics Committee (MREC), Ministry of Health Malaysia, on 8th February 2023. All participations are subject to explicit caregiver written consent and child verbal assent. The authors would also like to thank all participants who were involved during data collection.

REFERENCES

1) Adi, N. S., Kuay, H. S., & Mustafa, Q. M. (2024). A study on the psychological functioning of children with specific learning difficulties and typically developing children. *BMC Psychology*, *12*(1), 1–12. https://doi.org/10.1186/s40359-024-02151-4
2) Banerjee, A., Mutlu, O. C., Kline, A., Surabhi, S., Washington, P., & Wall, D. P. (2023). Training and profiling a paediatric facial expression classifier for children on mobile devices: Machine learning study. *JMIR Formative Research*, *7*, e39917. https://doi.org/10.2196/39917
3) Romani-Sponchiado, A., Sanvicente-Vieira, B., Mottin, C., Hertzog-Fonini, D., & Arteche, A. (2015). Child Emotions Picture Set (CEPS): Development of a database of children's emotional expressions. Psychology & Neuroscience, 8(4), 467–478. https://doi.org/10.1037/h0101430.
4) Choudhary, K., & Prajapati, G. G. L. (2026). Graph-based multimodal affect recognition in children using prototypical networks. *Frontiers in Computer Science*, *8*, 104–118. https://doi.org/10.3389/fcomp.2026.1774796
5) He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 770–778). https://doi.org/10.1109/CVPR.2016.90
6) Li, S., & Deng, W. (2020). Deep facial expression recognition: A survey. *IEEE Transactions on Affective Computing*, *13*(3), 1195–1215. https://doi.org/10.1109/TAFFC.2020.2981446
7) Nag, A., Haber, N., Voss, C., Tamura, S., Daniels, J., Ma, J., ... & Wall, D. P. (2020). Toward continuous social phenotyping: Analyzing gaze patterns in an emotion recognition task for children with autism. *Journal of Medical Internet Research*, *22*(4), e13810. https://doi.org/10.2196/13810
8) Pekrun, R., & Goetz, T. (2023). How universal are academic emotions? A control-value theory perspective. In G. Hagenauer, R. Lazarides, & H. Järvenoja (Eds.), *Motivation and emotion in learning and teaching across educational contexts: Theoretical and methodological perspectives and empirical insights* (pp. 85–99). Taylor & Francis / Routledge.
9) Van Voorhis, W. C., & Morgan, B. L. (2007). Understanding power and rules of thumb for determining sample sizes. *Tutorials in Quantitative Methods for Psychology*, *3*(2), 43–50. https://doi.org/10.20982/tqmp.03.2.p043
10) Wang, Y., Luo, Q., Zhang, Y., & Zhao, K. (2024). Synchrony or asynchrony: development of facial expression recognition from childhood to adolescence based on large-scale evidence. *Frontiers in Psychology*, *15*, 1379652. https://doi.org/10.3389/fpsyg.2024.1379652
11) Xie, J., & Fan, M. (2025). The role of academic emotions in learning: Addressing student motivation, performance, and learning strategies. In *Academic Achievement-Perspectives of Students, Teachers, and Institutions*. IntechOpen. https://doi.org/10.5772/intechopen.1009197