PeerJ Computer Science

Empowering cognitive disabilities in transit: an explainable, emotion-aware ITS framework

Malik Almaliki[1,2], Amna Bamaqa[2,3], Tamer Ahmed Farrag[2,4], Hossam Magdy Balaha[5,6], Mahmoud Badawy[2,3,5] and Mostafa A. Elhosseini[2,5,7]

[1] Department of Computer Science, College of Computer Science and Engineering, Taibah University, Yanbu, Saudi Arabia
[2] King Salman Center for Disability Research, Riyadh, Saudi Arabia
[3] Computer Science and Information Department, Applied College, Taibah University, Medinah, Saudi Arabia
[4] Department of Electrical Engineering, College of Engineering, Taif University, Taif, Saudi Arabia
[5] Faculty of Engineering, Computers and Control Systems Engineering Department, Mansoura University, Mansoura, Egypt
[6] Bioengineering Department, University of Louisville, Louisville, Kentucky, United States
[7] Department of Information Systems, College of Computer Science and Engineering, Taibah University, Yanbu, Saudi Arabia

ABSTRACT

People with disabilities need ongoing support and a balanced lifestyle. Smart cities like NEOM are emerging worldwide. The Saudi government has implemented several disability accessibility programs in public spaces and transportation. This article addresses a critical yet often neglected challenge: accurately recognizing and interpreting facial emotions in individuals with cognitive disabilities to foster better social integration. Current emotion detection systems frequently overlook the unique needs of this demographic (slower response times, difficulty interpreting subtle cues, and varied attention spans) and provide limited transparency, undermining trust and hindering real-time applicability in complex, dynamic contexts. To overcome these limitations, we present a novel, comprehensive framework that utilizes the Internet of Things, fog computing, and advanced You Only Look Once (YOLO)v8-based deep learning models. Our approach incorporates adaptive feedback mechanisms to tailor interactions to each user's cognitive profile, ensuring accessible, user-centric guidance in diverse real-world scenarios. Besides, we introduce EigenCam-based explainability techniques, which offer intuitive visualizations of the decision-making process, enhancing interpretability and trust for both users and caregivers. Seamless integration with assistive technologies, including augmented reality devices and mobile applications, further supports real-time, on-the-go interventions in therapeutic and educational contexts. Experimental results on benchmark datasets (RAF-DB, AffectNet, and CK+48) demonstrate the framework's robust performance, achieving up to 95.8% accuracy and excelling under challenging conditions. The EigenCam outputs confirm that the model's attention aligns with meaningful facial features, reinforcing the system's interpretability and cultural adaptability. By delivering accurate, transparent, and context-aware emotion recognition tailored to cognitive disabilities, this research sets a promising step for inclusive artificial intelligence (AI)-driven solutions, ultimately promoting independence, reducing stigma, and improving quality of life.

Submitted 28 March 2025
Accepted 25 September 2025
Published 4 November 2025

Corresponding author
Mahmoud Badawy,
engbadawy@mans.edu.eg

Academic editor
Luigi Di Biasi

Additional Information and Declarations can be found on page 28

DOI **10.7717/peerj-cs.3301**

How to cite this article Almaliki M, Bamaqa A, Ahmed Farrag T, Magdy Balaha H, Badawy M, A. Elhosseini M. 2025. Empowering cognitive disabilities in transit: an explainable, emotion-aware ITS framework. **PeerJ Comput. Sci. 11:e3301 DOI 10.7717/peerj-cs.3301**

Subjects Artificial Intelligence, Computer Vision, Data Mining and Machine Learning
Keywords Cognitive disabilities, Deep learning (DL), Facial emotion detection, Intelligent transportation, You only look once (YOLO)

INTRODUCTION

Smart cities increasingly focus on developing inclusive environments that address the needs of all residents, including individuals with limitations in utilizing everyday technologies (*Makkonen & Inkinen, 2024*). Several nations are preparing to implement the concept of smart cities in their municipalities. Public transportation planning and analytical methodologies are insufficient when addressing the needs of the disability community. Attaining disability justice necessitates that all individuals have access to the transportation resources essential for leading a meaningful, dignified, and fulfilling life (*Levine & Karner, 2023*). The convergence of the Internet of Things (IoT), cloud computing, big data, and advanced artificial intelligence (AI) revolutionizes transportation provision. This revolution is aimed at improving city infrastructure (*Bennett & Vijaygopal, 2024*).

The Saudi Arabian government is focused on ensuring that all citizens and residents can lead decent lives, especially people who experience disability. Disabilities affect nearly every individual at some point, whether temporarily or permanently. Persons with disabilities may require continuous supportive services and a balanced lifestyle (*Bindawas & Vennu, 2018*). NEOM City is one of the forthcoming smart cities globally. Saudi Arabia has garnered attention due to its economic diversification and modernization program, which are associated with this significant project. Besides, the Saudi government has launched several initiatives to improve disability access to public spaces and transportation. The government has created the "Accessible Public Transportation Program" to make public transit more accessible for disabled people by providing specialized buses and other facilities (*Makkonen & Inkinen, 2024*).

Emotions play a pivotal role in human communication. *Keltner & Cordaro (2017)* categorized emotions into six emotions. There are fear, disgust, anger, happiness, sadness, and surprise (*Keltner & Cordaro, 2017*). Integrating these emotions into social interactions is integral to human life (*Razzaque, 2020*; *Miller & Wallis, 2011*). It enables people to express intentions, interpret social cues, foster relationships, and build connections (*Morris & Keltner, 2000*; *Van Kleef, 2009*).

However, individuals with cognitive disabilities, such as autism spectrum disorder (ASD), intellectual disabilities, Down syndrome, or certain neurodegenerative diseases, often face significant challenges in recognizing, interpreting, and expressing emotions (*Grieco et al., 2015*; *Sappok, Diefenbacher & Winterholler, 2019*; *Yates & Le Couteur, 2016*). These difficulties can hinder their ability to connect and integrate with others, sometimes leading to feelings of isolation, miscommunication, anxiety, and a diminished quality of life (*Segrin, 1996*).

The global prevalence of cognitive disabilities further underscores the magnitude of this challenge that faces those individuals. According to the Centers for Disease Control and Prevention (CDC), ASD alone is estimated to affect 1 in 36 children for Disease Control

and Prevention (*Centers for Disease Control and Prevention, 2024*; *Maenner, 2023*). Moreover, the World Health Organization (WHO) estimates that 1 in 100 children worldwide are diagnosed with ASD (*World Health Organization, 2024*; *Zeidan et al., 2022*). Intellectual disabilities and other cognitive impairments also impact millions globally, creating a critical need for tools that enhance emotional comprehension (*Sappok et al., 2022*; *Des Portes, 2020*; *Harris, 2006*).

As society increasingly integrates individuals with cognitive disabilities into educational, workplace, and social environments, emotional intelligence becomes paramount (*Clark & Polesello, 2017*; *Zeidner, Roberts & Matthews, 2008*; *Matthews, Zeidner & Roberts, 2007*). A lack of this skill can impede relationship-building, collaboration, and social comfort. Despite advances in therapies and assistive technologies (*Szabó et al., 2023*), current interventions are often generalized, time-intensive, and dependent on human facilitators, leaving a significant gap in tailored support.

Individuals with cognitive disabilities often encounter many challenges when navigating public transportation systems (*Mogaji & Nguyen, 2021*). Beyond the physical hurdles of entering a bus, finding the correct platform, or transferring between routes, these travelers must also manage complex cognitive tasks such as interpreting signs, processing auditory announcements, and understanding sudden route changes. Emotional states, such as stress, anxiety, or confusion, can exacerbate these difficulties, making it harder to absorb critical information or respond promptly to unexpected situations. For instance, a crowded station or a delayed train can heighten stress levels, leading to disorientation or even deterring individuals from attempting future journeys. This emotional burden reduces overall mobility, hinders social inclusion, and compromises the independence and well-being of those affected.

Despite strides in intelligent transportation systems (ITS) and urban accessibility measures (ranging from accessible route planning apps to audio-visual wayfinding aids), these solutions generally focus on standardized requirements, failing to consider the cognitive and emotional dimensions influencing a user's experience. ITS refers to applying advanced information and communication technologies to manage and optimize transportation networks. These systems enhance traffic efficiency, safety, and user experience through real-time data processing, automation, and smart infrastructure. These current ITS platforms often lack personalization features that address cognitive impairments and emotional awareness capabilities that can dynamically adapt guidance and support mechanisms to a traveler's emotional state. As a result, even the most technologically advanced ITS installations may struggle to provide effective, confidence-building assistance to individuals who require it most.

While various accessibility solutions (such as tactile maps, audio cues, and augmented reality (AR) guidance) have enhanced public transit usability for many, the absence of emotion-aware interventions explicitly tailored for cognitive disabilities remains a glaring gap. These existing tools rarely integrate affective signals that could inform real-time adjustments, like simplifying instructions when a user appears confused, delaying prompts for those who need additional processing time, or offering stress-reducing suggestions in overwhelming situations. This research aims to bridge that gap by introducing a real-time

framework that detects and interprets facial emotions and employs explainable AI techniques and adaptive feedback mechanisms. By doing so, the proposed system ensures that ITS solutions can cater to travelers' cognitive and affective needs, creating a more inclusive and supportive transportation environment.

Advances in AI and deep learning (DL) present a promising avenue for addressing this challenge. Specifically, emotion detection systems based on facial image analysis have the potential to bridge this gap by providing real-time, personalized support for emotion recognition (*Kaur & Kumar, 2024*; *Marechal et al., 2019*).

This research hypothesizes that utilizing AI-powered emotion detection systems, specifically utilizing advanced deep learning models, can significantly enhance the ability of individuals with cognitive disabilities to interpret and respond to social and emotional cues. These systems can empower users to navigate social interactions more confidently and independently by providing real-time, context-aware emotion recognition. The study further posits that such technology, tailored to the unique needs of this demographic, can promote inclusivity, reduce stigma, and improve the quality of life for affected individuals.

The current study proposes a comprehensive framework for facial emotion detection and intervention explicitly designed to address the needs of individuals with cognitive disabilities. This framework integrates state-of-the-art YOLOv8-based deep learning models to achieve robust, real-time emotion classification across datasets such as RAF-DB, AffectNet, and CK+48. Key features of the proposed system include explainability through EigenCam visualizations, adaptive feedback mechanisms tailored for slower response times and varied attention spans, and seamless integration with assistive technologies such as AR glasses or mobile applications.

This research aims to develop an explainable, emotion-aware ITS framework that empowers individuals with cognitive disabilities to navigate public transportation networks confidently and independently. By embedding real-time emotion detection, adaptive feedback, and transparent AI decision-making into the ITS ecosystem, the approach enables more inclusive, user-centric transit experiences.

The contributions of this research are multifaceted, addressing critical gaps in emotional recognition technology for individuals with cognitive disabilities. They can be listed as follows:

- ITS-integrated emotion detection: Introduces a novel AI-driven framework utilizing YOLOv8-based deep learning models to detect and interpret facial emotions in real-time within transportation environments. This emotional awareness enables ITS interfaces (such as route guidance apps, digital signage, and AR-based navigation aids) to dynamically adjust instructions and support strategies.
- Adaptive ITS feedback mechanisms: Tailor's navigational prompts, notifications timing, and instructions complexity based on users' emotional states and cognitive profiles. Doing so reduces stress and confusion at transit hubs, making journey planning, transfers, and service updates more accessible and less intimidating.
- Explainable AI in transit settings: Incorporates EigenCam-based visualizations, allowing users, caregivers, and transit personnel to understand how and why the system suggests

particular routes or interventions. This transparency fosters trust, ensuring stakeholders are comfortable relying on the system in busy or unfamiliar transit scenarios.

- Wearable and mobile integration for on-the-go support: Ensures seamless compatibility with assistive ITS technologies (such as AR glasses and smartphone applications), delivering real-time emotional support and context-sensitive instructions in dynamic transit conditions, including crowded stations, moving platforms, or irregular service patterns.
- Advancing equity and quality of life in ITS: Moves beyond essential operational efficiency to create a more inclusive public transportation environment. By accommodating emotional and cognitive needs, the framework reduces barriers, mitigates stigma, and enhances the overall travel experience, contributing to smarter, more human-centered urban mobility solutions.

The rest of this article is organized as follows: 'Related Studies' reviews the related works and recent technology devoted to exploring facial emotion detection. 'Methodology' introduces the proposed system that integrates facial emotion detection with ITS by utilizing advanced fog and cloud computing infrastructures and the suggested framework for facial emotion detection and intervention for cognitive disabilities. 'Experiments and Discussion' elaborates on the experiments that were conducted. 'Limitations' introduces the study limitations and 'Conclusions' summarizes the study.

RELATED STUDIES

Substantial scholarly efforts have been devoted to the exploration of facial emotion detection. This is evidenced by the many publications in academic literature. They have consistently strived to deliver an in-depth evaluation of methodologies encompassing both machine and deep learning frameworks (*Li & Deng, 2020*; *Huang et al., 2019*; *Ekundayo & Viriri, 2021*; *Mellouk & Handouzi, 2020*; *Khan, 2022*).

For instance, *Ge et al. (2022)* summarized the research on deep facial expression recognition methods over the past decade, categorized current approaches into static and dynamic recognition, compared the performance of advanced algorithms on expression databases, and discussed the challenges of overfitting and real-world interference while suggesting multimodal models to improve recognition accuracy.

Focusing on deep learning, *Jaiswal, Raju & Deb (2020)*, *Jain, Shamsolmoali & Sehdev (2019)*, and *Khattak et al. (2022)* utilized convolutional neural networks for facial emotion detection from images. *Jaiswal, Raju & Deb (2020)* evaluated using two datasets, namely the facial emotion recognition challenge (FERC-2013) and Japanese female facial emotion (JAFFE). They reported 70.14% and 98.65% accuracies for FERC-2013 and JAFFE datasets, respectively. *Jain, Shamsolmoali & Sehdev (2019)* used Extended Cohn-Kanade (CK+) and JAFFE datasets and reported 95.23% and 93.24% accuracies for JAFFE and CK+, respectively. *Khattak et al. (2022)* reported an accuracy of 95.11% on Jaffe and 92.19% on the Extended Cohn-Kanade (CK+) dataset.

Shifting to vision transformers (ViTs), *Chaudhari et al. (2022)* utilized them for emotion detection. They applied their proposed system to a merged dataset combining FER-2013,

CK+48, and AffectNet datasets and reported 53.10% accuracy. Following the same approach, *Chen et al. (2023)* proposed few-shot facial emotion detection with a self-supervised ViT (SSF-ViT) by integrating self-supervised learning (SSL) and few-shot learning (FSL). They reported accuracies of 74.95%, 66.04%, 63.69%, and 90.98% on the FER2013, AffectNet, SFEW 2.0, and RAF-DB datasets, respectively.

Adding more complexity, *Ma, Sun & Li (2021)* suggested the ViTs with attentional selective fusion. They reported the performance on three datasets: RAF-DB with 88.14%, FERPlus with 88.81%, and AffectNet with 61.85%.

Dalabehera et al. (2024) proposed a mist-fog-assisted framework for real-time emotion recognition using deep transfer learning, specifically designed for Smart City 4.0 applications. This approach integrates mist computing (at the edge) and fog computing (near-edge processing) to enhance scalability and efficiency in urban environments. The study demonstrated the framework's capability to process high volumes of video and emotional data with minimal latency, making it ideal for dynamic and distributed settings. Similarly, *Bebortta et al. (2023)* introduced the DeepMist framework, which utilizes deep learning models operating over mist computing infrastructure to manage healthcare big data effectively. Their approach uses a Deep Q Network (DQN) algorithm for heart disease prediction, achieving 97.67% accuracy while maintaining low energy consumption and delay. These frameworks illustrate the potential of deploying deep learning at the mist computing layer for low-latency, high-efficiency computations.

Integrating explainable AI (XAI) techniques into emotion detection systems addresses the critical challenge of trust and transparency. EigenCam, introduced by *Bany Muhammad & Yeasin (2021)*, provides saliency maps to visualize key facial features influencing model predictions. This technique ensures that outputs are interpretable and align with user expectations, particularly in high-stakes applications where decision transparency is vital. *Lau (2010)* developed a portable real-time emotion detection system using physiological signals such as heart rate and skin conductance. The system was designed to enhance emotional awareness and interaction for disabled users by identifying emotional states in real time. This pioneering work highlights the role of accessible and portable technologies in supporting individuals with disabilities.

While existing research in facial emotion detection has made significant strides, several critical gaps remain unaddressed, particularly concerning the needs of individuals with cognitive disabilities. This study aims to fill the following gaps:

- Limited personalization for cognitive disabilities: Most emotion detection systems are designed for the general population, often neglecting the unique challenges faced by individuals with cognitive disabilities, such as slower response times, difficulty interpreting subtle facial cues, and varied attention spans. This study addresses these issues by tailoring the emotion recognition framework to include adaptive feedback mechanisms and simplified interfaces for this demographic.
- Explainability in emotion detection models: Many existing models lack transparency, providing high accuracy but minimal insights into their decision-making processes. This

study incorporates EigenCam-based explainability to visualize and understand the model's focus areas, enabling users and caregivers to trust and interpret the system's outputs effectively.

– Real-time applicability in dynamic environments: Current approaches often fail to deliver real-time performance in practical, socially dynamic settings. This study emphasizes real-time detection and intervention, ensuring seamless integration into wearable devices or assistive technologies for on-the-go support.

These identified gaps highlight the need for a novel framework that not only addresses the unique challenges faced by individuals with cognitive disabilities but also integrates explainability, personalization, and real-time performance. Building upon these insights, the following section presents our proposed framework. It introduces an explainable, emotion-aware ITS utilizing YOLOv8-based deep learning models and EigenCam explainability. The architecture is designed to operate within fog and cloud computing environments, enabling adaptive, context-aware emotional support tailored to user-specific cognitive profiles.

METHODOLOGY

The proposed system integrates facial emotion detection with ITS using advanced fog and cloud computing infrastructures. Fog computing extends cloud computing by enabling data processing at the network edge, closer to where data is generated. This reduces latency, enhances responsiveness, and supports real-time decision-making, which is particularly beneficial for emotion-aware systems operating in transit settings. The suggested architecture ensures real-time, location-aware, and emotion-sensitive support for individuals with cognitive disabilities. Figure 1 presents the system's data flow and functionality breakdown.

Cameras installed along roads, bus stops, and transit hubs capture real-time facial images of individuals. These images are transmitted wirelessly to nearby signal towers (roadside units) using communication networks such as cellular systems. The roadside units act as initial processing nodes, directing the data to the appropriate computing layers for further analysis.

Signal towers relay the captured images to fog brokers, specialized computing units with local databases. These brokers are part of a broader fog management area, a distributed network of fog servers designed for intermediate processing tasks. The fog brokers perform tasks such as (1) Facial recognition to identify individuals. (2) Location tagging based on the camera location and the individual's presence.

These fog units reduce latency and improve responsiveness by performing initial processing close to the data source. The processed data is then forwarded to the cloud for advanced analysis.

Fog servers in the fog management area are interconnected with a cloud computing infrastructure, where advanced emotion recognition occurs. The emotion detection framework, based on YOLOv8, processes the facial images to identify emotional states such as stress, confusion, or fear (see Fig. 2). This step involves: (1) Extracting meaningful

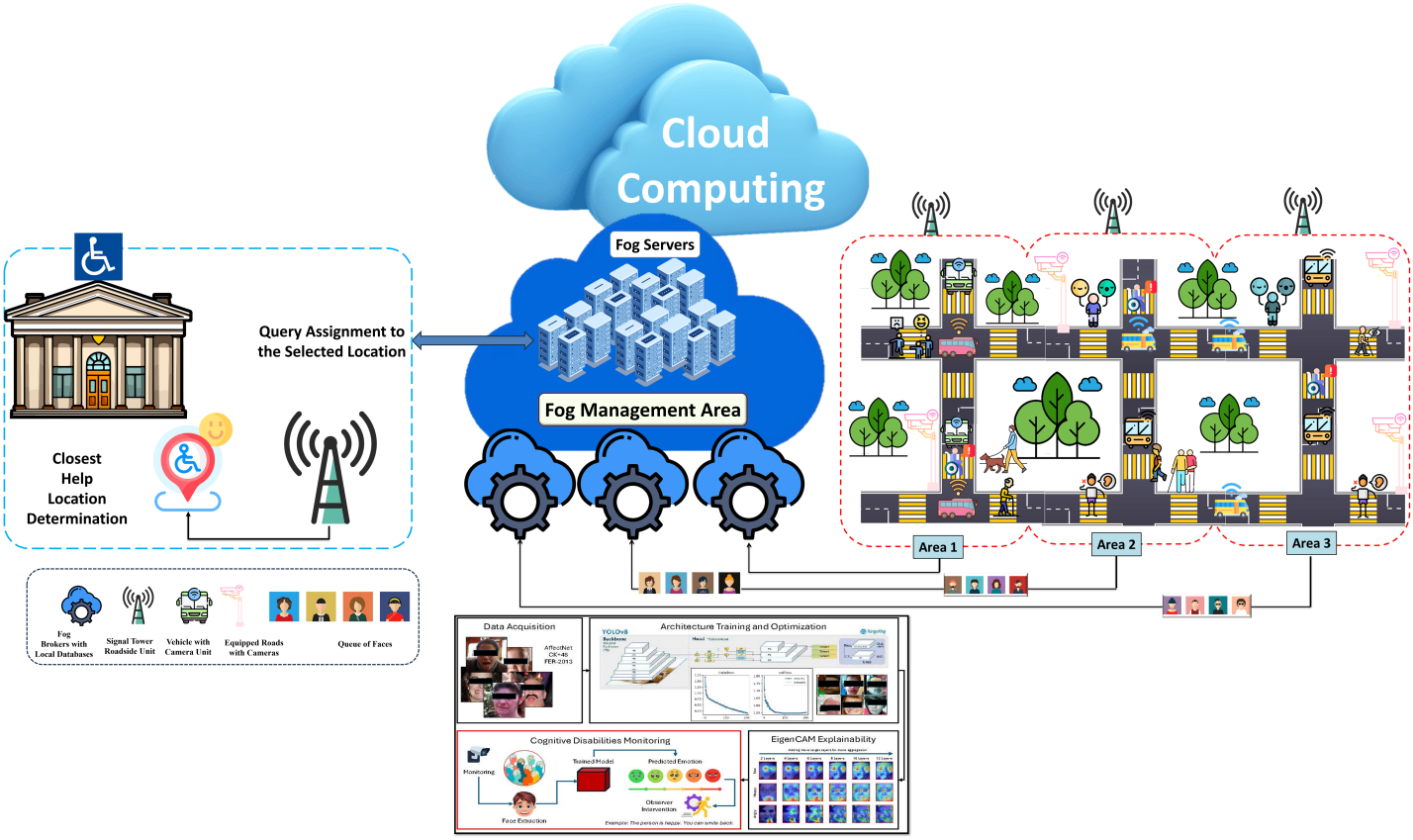

Figure 1 Visualization of the data flow and functionality breakdown of the system, highlighting how cameras installed along roads, bus stops, and transit hubs capture real-time facial images. They are then transmitted wirelessly to nearby signal towers and processed through fog brokers and cloud computing infrastructures for emotion detection and intervention. Full-size DOI: 10.7717/peerj-cs.3301/fig-1

facial features from the image. (2) Determining the individual's emotional state through robust classification models. (3) Combining emotional data with location information for context-aware decision-making.

The system generates a query based on the individual's detected emotion and current location. This query specifies the required assistance type, such as calming messages, route guidance, or emergency response. The system then identifies the closest available helping location, such as a transit support center, caregiver unit, or an on-ground incident response team. The most efficient dispatch route is dynamically determined using signal tower data and fog computing units, and the query is transmitted to the selected location.

Data acquisition

This study utilizes three publicly available datasets: CK+48, RAF-DB, and AffectNet. These datasets were selected due to their relevance in facial emotion recognition research and widespread use in benchmarking approaches.

The CK+48 dataset (*Lucey et al., 2010*), also known as the Extended Cohn-Kanade Dataset, is a benchmark in emotion recognition studies. Each image sequence in CK+48 captures the transition from a neutral to a peak emotional expression. This

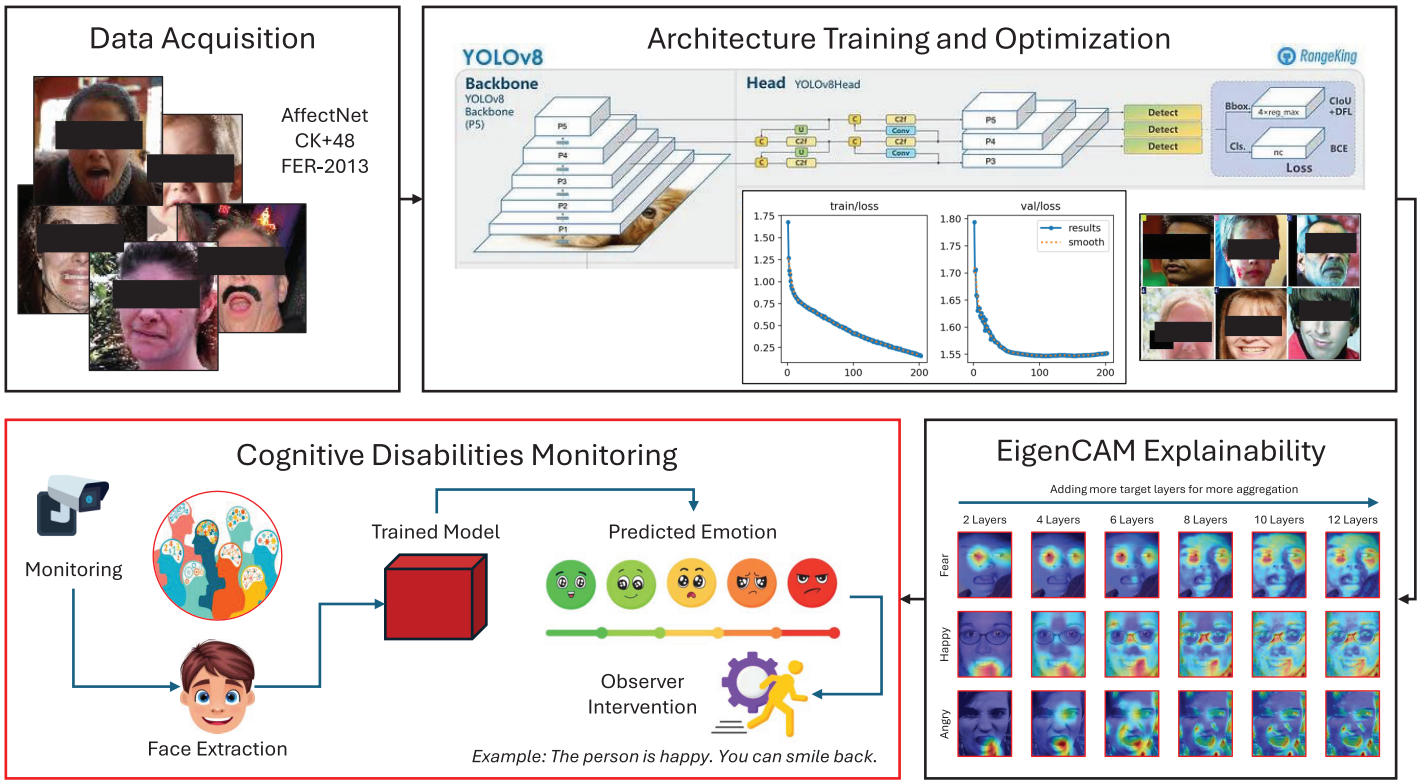

Figure 2 **Visualization of the proposed framework, demonstrating the integration of YOLOv8-based deep learning models and EigenCam explainability techniques within fog and cloud computing environments to provide adaptive, context-aware emotional support tailored to user-specific cognitive profiles.**
Full-size DOI: 10.7717/peerj-cs.3301/fig-2

gradual progression facilitates robust model training by providing intermediate states that help learn nuanced facial transformations. The dataset includes frontal facial images with minimal occlusions and uniform lighting, simplifying preprocessing. Researchers have widely used CK+48 to evaluate models focused on distinguishing between universally recognized emotions, making it a reliable choice for benchmarking the proposed approach.

RAF-DB extends the scope of facial emotion recognition by incorporating diverse real-world scenarios. The images in RAF-DB are annotated using a crowdsourcing approach, ensuring high-quality and reliable labeling. The dataset's inclusion of compound emotions (combinations of basic emotions) offers a more granular analysis of human emotional states. This feature aligns with the need for advanced models to understand subtle and mixed emotions. Moreover, the diversity in demographics, poses, and lighting conditions in RAF-DB simulates real-world environments, challenging the model to generalize effectively (*Shan & Deng, 2018*).

AffectNet significantly enhances the understanding of emotional expressions by providing a vast collection of images annotated for categorical and dimensional affective representations. Its categorical labels include emotions like anger, disgust, fear, and happiness, while the dimensional labels provide arousal and valence scores, offering a

richer perspective on emotional intensity and polarity. The dataset includes challenging cases such as occlusions, extreme poses, and images captured in varied cultural contexts. These attributes make AffectNet an essential resource for training models that aim to achieve robust performance across diverse scenarios (*Mollahosseini, Hasani & Mahoor, 2017*).

Before model training, all datasets underwent standardized preprocessing to improve consistency and performance. Images were resized to a uniform resolution of 100×100 pixels to align with the input requirements of the YOLOv8 architecture. Using min-max scaling, pixel values were normalized to the range $[0, 1]$. No additional data augmentation was applied during training, as YOLOv8's built-in augmentation pipeline (*e.g.*, random flipping and HSV adjustments) was used instead.

Model architecture

YOLOv8 is a state-of-the-art deep learning architecture designed for real-time object detection. It represents a significant evolution in the YOLO family of models, introducing innovations that enhance both detection accuracy and computational efficiency. It combines high accuracy with fast inference speeds, making it ideal for applications like facial emotion recognition in dynamic environments. The model's backbone utilizes a deep convolutional architecture with residual connections, which help mitigate the vanishing gradient problem during training. These connections also enable the network to extract fine-grained and high-level features critical for emotion classification (*Sohan et al., 2024*; *Terven, Córdova-Esparza & Romero-González, 2023*).

These connections can be formulated as presented in Eq. (1), where $\mathbf{x}$ is the input, and $\mathscr{F}$ (referred to as the *residual mapping*) represents the transformation applied to the input by a series of convolutional layers parameterized by weights $\{W_i\}$. Intuitively, $\mathscr{F}$ captures the additional information that needs to be added to the input to produce the desired output. This design enables the network to learn hierarchical features more efficiently while preserving fine-grained and high-level semantic features, facilitating robust emotion detection.

$$\mathbf{y} = \mathscr{F}(\mathbf{x}, \{W_i\}) + \mathbf{x}. \tag{1}$$

The feature pyramid network (FPN) in YOLOv8 merges multi-scale features to improve the detection of subtle facial expressions and global facial context. The multi-scale feature extraction process can be expressed as shown in Eq. (2), where $\mathbf{P}_l$ represents the feature map at level l, $\mathbf{C}_l$ is the feature map from the convolutional backbone at level l, and UpSample is an operation that increases the resolution of feature maps from higher levels. This merging of high-resolution details with coarse semantic information enhances the network's ability to handle varying face sizes and poses, as seen in challenging datasets such as RAF-DB and AffectNet.

$$\mathbf{P}_l = \text{Conv}(\mathbf{C}_l) + \text{UpSample}(\mathbf{P}_{l+1}). \tag{2}$$

The anchor-free detection mechanism in YOLOv8 removes the dependency on predefined anchor boxes, allowing the network to directly predict the object center (x, y)

and dimensions (w, h) of bounding boxes. These predictions are parameterized relative to the grid cell as presented in Eq. (3), where (t_x, t_y, t_w, t_h) are the predicted offsets, (c_x, c_y) represent the grid cell coordinates, and (p_w, p_h) are the predefined prior dimensions. The sigmoid function $\sigma(\cdot)$ ensures that offsets remain bounded, leading to more precise localization.

$$\begin{aligned} x &= \sigma(t_x) + c_x \\ y &= \sigma(t_y) + c_y \\ w &= p_w \times e^{t_w} \\ h &= p_h \times e^{t_h} \end{aligned}. \tag{3}$$

Emotion classification is performed using a SoftMax output layer. For an input image $\mathbf{x}$, the classification probabilities for each emotion category are computed as presented in Eq. (4), where z_i represents the logits (unnormalized scores) for class i, and C is the total number of emotion categories. The SoftMax function converts these logits into probabilities by normalizing the exponential of each score.

$$p_i = \frac{e^{z_i}}{\sum_{j=1}^{C} e^{z_j}}, \quad i = 1, \ldots, C. \tag{4}$$

The cross-entropy loss function used during training is defined as shown in Eq. (5), where y_{ij} is the ground truth label for sample i and class j, and p_{ij} is the predicted probability for the same. This loss quantifies the difference between the predicted and true distributions, guiding the optimization process.

$$\mathscr{L}_{\text{class}} = -\frac{1}{N} \times \sum_{i=1}^{N} \sum_{j=1}^{C} y_{ij} \times \log(p_{ij}). \tag{5}$$

YOLOv8 employs a hybrid optimization strategy that combines stochastic gradient descent (SGD) and adaptive moment estimation (Adam). The weight update rule for SGD with momentum is given by Eq. (6), where $\mathbf{v}_t$ is the velocity term that accumulates gradients over time, $\mathbf{w}_t$ are the model weights, β is the momentum parameter that controls the influence of past gradients, and η is the learning rate.

$$\begin{aligned} \mathbf{v}_t &= \beta \times \mathbf{v}_{t-1} + (1 - \beta) \times \nabla \mathscr{L}(\mathbf{w}_t) \\ \mathbf{w}_{t+1} &= \mathbf{w}_t - \eta \times \mathbf{v}_t. \end{aligned} \tag{6}$$

Adam further incorporates adaptive learning rates through Eq. (7), where m_t and v_t are the first and second-moment estimates of the gradients, respectively. These moments help adjust the learning rate for each parameter individually, improving convergence. Hyperparameters β_1 and β_2 control the decay rates of these moments, and ε is a small constant to prevent division by zero, as shown in Eq. (8).

$$\begin{aligned} m_t &= \beta_1 \times m_{t-1} + (1 - \beta_1) \times \nabla \mathscr{L}(\mathbf{w}_t) \\ v_t &= \beta_2 \times v_{t-1} + (1 - \beta_2) \times (\nabla \mathscr{L}(\mathbf{w}_t))^2 \end{aligned} \tag{7}$$

$$\begin{aligned} \hat{m}_t &= \frac{m_t}{1-\beta_1^t} \\ \hat{v}_t &= \frac{v_t}{1-\beta_2^t} \\ \mathbf{w}_{t+1} &= \mathbf{w}_t - \eta \times \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \varepsilon}. \end{aligned} \tag{8}$$

A lightweight attention mechanism is integrated into YOLOv8 to refine feature extraction further. The attention weights are computed as shown in Eq. (9), where α_i represents the importance (or weight) assigned to feature map $\mathbf{f}_i$ based on its score s_i. Intuitively, this mechanism allows the network to focus on regions of the face that are most indicative of emotions, such as the eyes and mouth, by selectively amplifying relevant features.

$$\begin{aligned} \alpha_i &= \frac{e^{s_i}}{\sum_j e^{s_j}} \\ \mathbf{f}' &= \sum_i \alpha_i \times \mathbf{f}_i \end{aligned}. \tag{9}$$

Explainability with EigenCam

EigenCam is an explainable AI technique that visualizes which regions of an image most influence a model's predictions. It enhances the interpretability of deep learning models by generating saliency maps that visualize the contribution of individual pixels or regions to the model's predictions. Unlike traditional methods like Grad-CAM, EigenCam employs principal component analysis (PCA) on activation maps to isolate the most informative components, reducing noise and highlighting critical regions. For emotion recognition, EigenCam elucidates which facial features (such as the eyes, mouth, or eyebrows) are pivotal for classification decisions. This methodology enables researchers to interpret the spatial focus of YOLOv8 during emotion classification, thereby increasing trust in the model's outputs (*Bany Muhammad & Yeasin, 2021*).

The saliency maps in EigenCam are generated by performing PCA on the high-dimensional activation maps from specific convolutional layers. The principal components are computed as shown in Eq. (10) where $\mathbf{A}$ represents the activation maps, $\mathbf{u}_k$ is the k-th principal component, and $\mathbf{v}_k$ is the corresponding eigenvector. These components highlight the most significant spatial features contributing to the model's prediction. The saliency maps are then back-projected onto the input space to visualize the regions of interest.

$$\begin{aligned} \mathbf{u}_k &= \frac{\mathbf{A}\mathbf{v}_k}{\|\mathbf{A}\mathbf{v}_k\|} \\ \mathbf{v}_k &= \text{argmax}_{\mathbf{v}} \frac{\mathbf{v}^\top \mathbf{A}^\top \mathbf{A}\mathbf{v}}{\|\mathbf{v}\|^2} \end{aligned}. \tag{10}$$

EigenCam's saliency maps are particularly effective in identifying attention patterns across various emotions. For example, when predicting "happiness," the model may emphasize regions around the mouth and cheeks, while for "anger," it may focus on

furrowed brows. These visualizations provide an intuitive understanding of the decision-making process, ensuring that the model's predictions align with human expectations. Additionally, the method allows for comparing focus areas between correct and incorrect predictions, highlighting discrepancies that may inform future model improvements.

A key application of EigenCam in this study is evaluating the model's spatial attention during misclassifications. For example, saliency maps may reveal that the model is distracted by background elements or occluded regions of the face, such as sunglasses or scarves. These insights are quantified using localization error, defined in Eq. (11), and pixel-level correlation metrics that measure the alignment between the highlighted regions and annotated ground truth regions. Such evaluations not only identify weaknesses in the model's focus but also suggest areas for improvement.

$$\text{Localization Error} = 1 - \frac{\text{Intersection of Highlighted Region and Ground Truth}}{\text{Union of Highlighted Region and Ground Truth}}. \quad (11)$$

Another critical use of EigenCam is detecting and mitigating biases in the model's predictions. Saliency maps often reveal whether the model's attention disproportionately favors demographic-specific features, such as cultural variations in expressions or skin tone. For instance, maps may show that the model interprets a raised eyebrow differently for different ethnic groups, leading to biased predictions. This capability aligns with fairness objectives, ensuring the model performs equitably across diverse populations.

EigenCam is also instrumental in guiding iterative model refinement. By analyzing saliency maps, researchers can pinpoint architectural components that require modification. For example, the feedback might suggest adding attention mechanisms to improve focus on relevant facial regions or adjusting convolutional layers to enhance feature extraction. These modifications are validated through repeated analysis with EigenCam, forming a feedback loop that optimizes both the accuracy and explainability of the proposed YOLOv8-based approach.

The integration of EigenCam bridges the gap between performance and interpretability, offering a robust framework for evaluating and refining deep learning models for emotion recognition. This methodology improves transparency and empowers developers to address ethical concerns, ensuring that the model aligns with human expectations and societal norms.

Comparison of EigenCam with other XAI tools

While several XAI methods exist for visualizing deep learning model decisions, EigenCam offers distinct advantages over commonly used tools such as Grad-CAM. Grad-CAM generates heatmaps by computing the target class output gradients with respect to the feature maps in the final convolutional layer. While effective, it often includes noisy activations and may not accurately highlight the most discriminative regions.

EigenCam, in contrast, applies PCA directly to activation maps, extracting dominant components that contribute most to the model's decision. This results in cleaner, more

interpretable saliency maps that better align with human perception of emotional expressions.

Furthermore, EigenCam does not require gradient computation, making it computationally lighter and suitable for real-time applications. It also provides consistent attention localization across different emotions (such as focusing on the mouth for happiness and brows for anger), which is critical in assistive technologies where interpretability and trust are paramount.

Computing infrastructure

The research was conducted on a Windows 11 (64-bit) system equipped with an Intel Core i7-1165G7 processor (four cores, eight threads, 2.8 GHz base, 4.7 GHz boost), 8 GB DDR4 RAM, and a 512 GB NVMe SSD for storage. The software environment was built around Python 3.10, using Jupyter Notebook as a development platform. Key libraries included NumPy and Pandas for data processing, Scikit-Learn, TensorFlow 2.9, and PyTorch 1.12 for machine learning tasks, and Matplotlib and Seaborn for data visualization.

Evaluation metrics

The confusion matrix provides a granular view of model performance, offering insights into misclassification trends across emotion categories. For example, it may reveal that the model frequently confuses fear with surprise due to their overlapping facial expressions. Such analysis helps pinpoint areas for improvement in the model's feature extraction and classification stages (*Powers, 2020*).

Accuracy, while a commonly used metric, is complemented by other measures to address its limitations in imbalanced datasets. For instance, accuracy alone may not reflect true performance in scenarios where one emotion dominates the dataset. By incorporating precision and recall, the evaluation framework ensures a balanced assessment of the model's ability to identify and distinguish emotions.

Specificity is particularly crucial in applications where false positives carry significant consequences, such as diagnosing cognitive disabilities. High specificity ensures the model avoids misclassifying neutral expressions or unrelated facial features as emotional states, enhancing its reliability in clinical contexts.

Balanced accuracy (BAC) is employed to mitigate the effects of class imbalance. By averaging recall and specificity, BAC provides a comprehensive metric that evaluates the model's ability to handle underrepresented emotion categories. This is especially important for datasets such as AffectNet, where some emotions are sparsely represented.

Intersection over Union (IoU) measures the accuracy of face localization, which directly impacts emotion classification performance. High IoU scores indicate that the model consistently identifies and focuses on relevant facial regions, reducing the influence of extraneous features. IoU is particularly relevant for datasets with occlusions or extreme poses, where precise localization is challenging.

To validate the robustness of the proposed approach, evaluation is conducted across multiple datasets using cross-dataset testing. This involves training the model on one dataset and testing it on another to assess generalizability. Additionally, statistical tests,

such as paired t-tests and Wilcoxon signed-rank tests, are performed to confirm the significance of performance improvements over baseline methods.

Low-latency performance for real-time interventions

To ensure timely interventions in dynamic transit settings, the framework prioritizes low-latency performance through a combination of technical measures:

- **Optimized model architecture:** The YOLOv8 model, particularly the smaller variants (YOLOv8 S and M), is known for its computational efficiency, enabling rapid inference with minimal processing overhead. The framework utilizes these architectures to ensure swift emotion detection even on resource-constrained devices like smartphones or AR glasses.
- **Edge computing:** Processing emotion data on edge devices, such as onboard cameras or user smartphones, minimizes the reliance on cloud communication, significantly reducing latency. This enables near-instantaneous responses to detected emotions, which is crucial for timely interventions in crowded stations or on moving buses.
- **Hardware acceleration:** The framework is designed to utilize hardware acceleration capabilities available on modern devices, such as GPUs and specialized neural network processors. By harnessing these resources, the system achieves significantly faster processing speeds, enabling real-time emotion analysis without noticeable delays.
- **Data preprocessing optimization:** Efficient data preprocessing techniques, such as image resizing and region-of-interest cropping, are implemented to reduce the computational burden on the model. This ensures that only the essential visual information is processed, further enhancing the speed of emotion detection.
- **Adaptive feedback mechanisms:** The framework incorporates adaptive feedback mechanisms that adjust the frequency and complexity of interventions based on the user's cognitive profile and the dynamic environment.

This approach prevents information overload and ensures that assistance is delivered promptly and appropriately, even in fast-changing situations. Moreover, these technical measures collectively ensure that the emotion detection pipeline operates with minimal latency, facilitating seamless integration with ITS components and enabling real-time, context-aware interventions to support individuals with cognitive disabilities throughout their journeys.

To further validate the system's suitability for real-world ITS environments, ongoing efforts are being made to deploy the framework in pilot smart transit hubs where real-time facial emotion detection can be tested under varying crowd densities, lighting conditions, and environmental dynamics.

The evaluation method

The proposed YOLOv8-based framework will be evaluated using three prominent facial emotion datasets: RAF-DB, AffectNet, and CK+48. Comprehensive performance metrics,

including accuracy, precision, recall, specificity, F1-score, IoU, BAC, and MCC, were employed to assess the effectiveness and robustness of the approach. Additionally, explainability aspects using EigenCam were examined to provide deeper insights into model decision-making.

A systematic model selection process was conducted to ensure optimal model performance. Among the YOLOv8 variants (N, S, M, L, XL), the YOLOv8 S model was selected as the final architecture due to its balanced trade-off between high classification accuracy (95.80% on RAF-DB) and computational efficiency, making it suitable for real-time emotion detection in dynamic transit environments.

EXPERIMENTS AND DISCUSSION

Performance analysis on RAF-DB

Table 1 demonstrates the performance metrics for RAF-DB. Among the YOLOv8 variations, the small version (YOLOv8 S) emerged as the most effective, achieving an accuracy of 95.80% and a balanced accuracy (BAC) of 92.25%. Integrating multi-scale feature extraction and anchor-free design proved advantageous for handling the diverse pose variations and facial occlusions in RAF-DB.

Precision and recall values underscore the model's ability to classify emotions while minimizing false positives and negatives correctly. YOLOv8 S recorded a precision of 87.46% and a recall of 87.56%, reflecting its balanced performance across the dataset's complex emotional expressions.

The IoU metric highlights the model's capability to localize emotion-relevant facial regions accurately. YOLOv8 S demonstrated a higher IoU (78.49%) compared to other variants, affirming the efficacy of its lightweight yet powerful architecture.

Specificity values across models indicate robust performance distinguishing between emotional and neutral faces. This aligns with the architecture's ability to suppress irrelevant background features while maintaining focus on facial expressions.

The choice of YOLOv8 S over other variants was guided by several key considerations. While larger models like YOLOv8 M, L, and XL achieved comparable accuracy, they incurred significantly higher computational costs, making them less suitable for real-time applications on edge devices such as smartphones or AR glasses. In contrast, YOLOv8 S strikes an optimal balance between performance and efficiency, offering near state-of-the-art accuracy while maintaining low latency and resource consumption. Additionally, the lightweight nature of YOLOv8 S ensures faster inference speeds, which is critical for supporting individuals with cognitive disabilities who rely on real-time feedback. The model's superior IoU score further validates its ability to accurately localize facial regions relevant to emotion recognition, a key factor in ensuring reliable performance in dynamic environments.

Figure 3 presents the visualization of the ROC curve for the RAF-DB dataset across various categories, utilizing the YOLOv8 S model. The ROC curve demonstrates the model's performance in distinguishing between the different emotional expressions present in the dataset. A higher AUC indicates better classification accuracy, with the YOLOv8 S model showcasing its strong ability to identify emotions such as happiness,

Table 1 Tabular presentation of the performance metrics applied on the RAF-DB dataset. The model in bold is the best-utilized YOLOv8 in the framework.

Model	Accuracy	Precision	Recall	Specificity	F1	IoU	BAC	MCC
YOLOv8 N	95.36	85.77	86.05	96.56	85.77	76.23	91.31	82.73
YOLOv8 S	95.80	87.46	87.56	96.94	87.41	78.49	92.25	84.62
YOLOv8 M	95.66	86.80	86.96	96.93	86.79	77.61	91.94	83.93
YOLOv8 L	95.76	87.40	87.56	96.87	87.37	78.40	92.22	84.55
YOLOv8 XL	95.49	86.62	86.70	96.77	86.51	77.16	91.73	83.55

anger, sadness, and surprise with high sensitivity and specificity. The curve further validates the model's effectiveness in real-time emotion detection, underlining its potential for applications supporting individuals with cognitive disabilities, where accurate emotion recognition is crucial for improving social interactions and emotional understanding.

Performance analysis on AffectNet

The performance metrics for AffectNet, presented in Table 2, reveal similar trends, with YOLOv8 S excelling across most metrics. AffectNet, characterized by its large-scale and diverse data, presents a unique challenge due to its fine-grained emotional annotations. YOLOv8 S achieved an accuracy of 93.95%, reflecting its adaptability to complex datasets.

Precision and recall values (74.26% and 74.19%, respectively) indicate the model's balanced performance in identifying subtle variations in facial expressions. The F1-score (74.11%) further highlights its effectiveness in managing the dataset's inherent class imbalances.

IoU analysis (61.04%) underscores the model's capacity to localize key facial features despite challenges posed by occlusions and diverse facial attributes. Specificity and BAC values validate the model's robustness in distinguishing emotional features from noise.

Figure 4 presents the ROC curve for the AffectNet dataset, utilizing the YOLOv8 Large (YOLOv8 L) model for emotion classification. The ROC curve visualizes the model's performance in differentiating between various emotional categories, such as happiness, sadness, and surprise. A higher AUC indicates superior classification ability. The results from the AffectNet dataset demonstrate the YOLOv8 L model's robust capacity for accurate emotion detection, which is essential for improving social communication and emotional comprehension for individuals with cognitive disabilities. This model's high accuracy and reliability further support the potential of AI-driven emotion detection systems in real-world applications.

Performance analysis on CK+48

The CK+48 dataset yielded perfect scores across all YOLOv8 variants, as shown in Table 3. The dataset's relatively small size and well-annotated nature facilitated the model's ability to achieve 100% accuracy, precision, recall, specificity, F1-score, IoU, BAC, and MCC.

This exceptional performance validates the proposed framework's strength in recognizing prototypical emotional expressions under controlled conditions. However, as

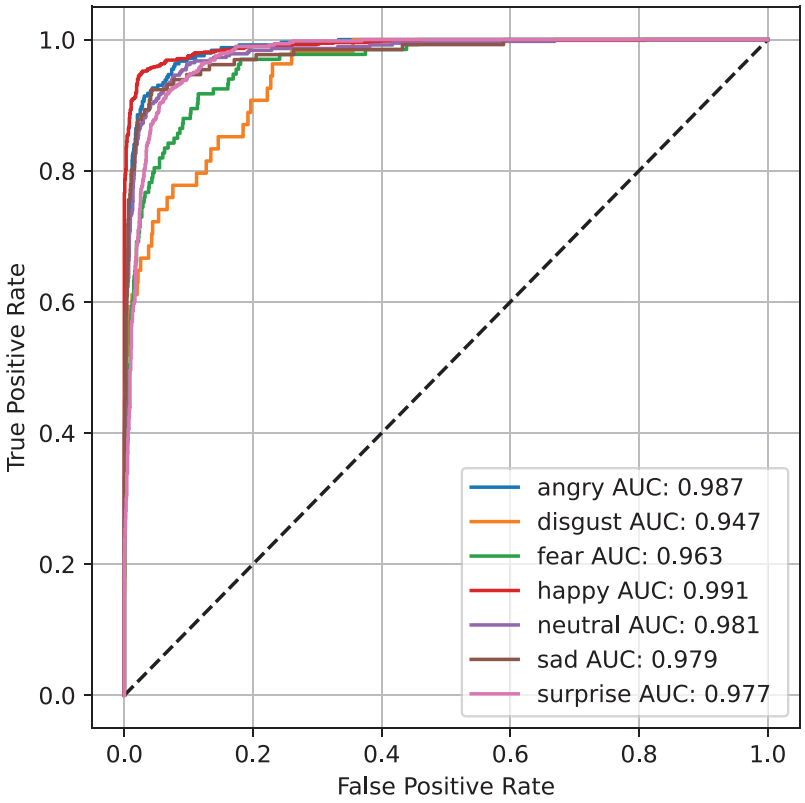

Figure 3 Visualization of the ROC curve for the RAF-DB dataset using the YOLOv8 S model across various emotion categories. The curve illustrates the model's high sensitivity and specificity in distinguishing between different emotional expressions, such as happiness, anger, sadness, and surprise. A higher AUC indicates better classification performance, demonstrating the model's strong capability for real-time emotion detection in individuals with cognitive disabilities.

Full-size DOI: 10.7717/peerj-cs.3301/fig-3

Table 2 Tabular presentation of the performance metrics applied on the AffectNet dataset.

Model	Accuracy	Precision	Recall	Specificity	F1	IoU	BAC	MCC
YOLOv8 N	93.74	73.53	73.41	96.41	73.36	60.11	84.91	69.86
YOLOv8 S	93.95	74.26	74.19	96.51	74.11	61.04	85.35	70.73
YOLOv8 M	93.91	74.05	74.05	96.50	74.02	60.85	85.28	70.56
YOLOv8 L	93.97	74.22	74.28	96.51	74.17	61.12	85.39	70.78
YOLOv8 XL	93.68	73.23	73.18	96.35	73.16	59.81	84.77	69.58

demonstrated by RAF-DB and AffectNet, generalizability to more challenging datasets provides a more comprehensive view of the model's capabilities.

Explainability and analysis

Figure 5 showcases the saliency maps generated using EigenCam, highlighting the model's focus areas for different emotions, including fear, happiness, surprise, and anger. The visualizations emphasize key facial regions, such as the eyes and mouth, aligning with established psychological emotional expression theories.

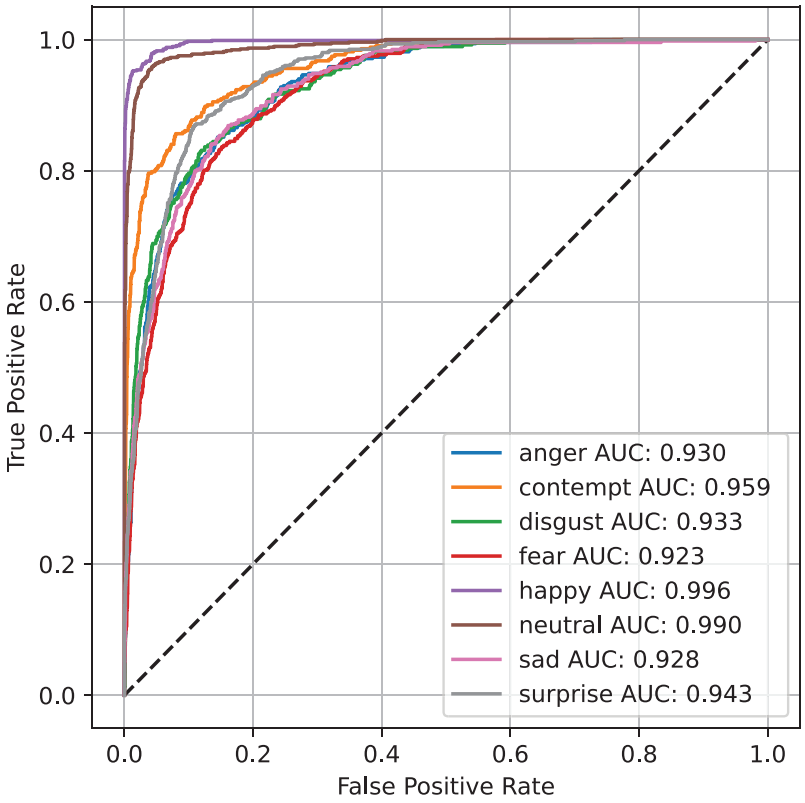

Figure 4 Visualization of the ROC curve for the AffectNet dataset using the YOLOv8 L model. This curve reflects the model's robust ability to differentiate between fine-grained emotional expressions like happiness, sadness, and surprise. The high AUC values confirm the model's reliability in detecting subtle emotional cues, essential for supporting individuals with cognitive disabilities in dynamic environments like public transportation. Full-size DOI: 10.7717/peerj-cs.3301/fig-4

Table 3 Tabular presentation of the performance metrics applied on the CK+48 dataset.

Model	Accuracy	Precision	Recall	Specificity	F1	IoU	BAC	MCC
YOLOv8 N	100	100	100	100	100	100	100	100
YOLOv8 S	100	100	100	100	100	100	100	100
YOLOv8 M	100	100	100	100	100	100	100	100
YOLOv8 L	100	100	100	100	100	100	100	100
YOLOv8 XL	100	100	100	100	100	100	100	100

EigenCam outputs reveal that the model consistently attends to widened eyes and raised eyebrows for surprise, furrowed brows for anger, and a smiling mouth for happiness. This concordance with human intuition validates the model's interpretability, particularly when applied to assistive technologies for individuals with cognitive disabilities.

Quantitative evaluations of saliency maps, such as localization error and pixel-level correlation with annotated regions, support these qualitative insights. Localization error values remain low across datasets, affirming the reliability of the EigenCam visualizations.

The model's explainability also aids in identifying biases, such as over-reliance on specific facial regions or demographic features. For instance, the analysis revealed that the

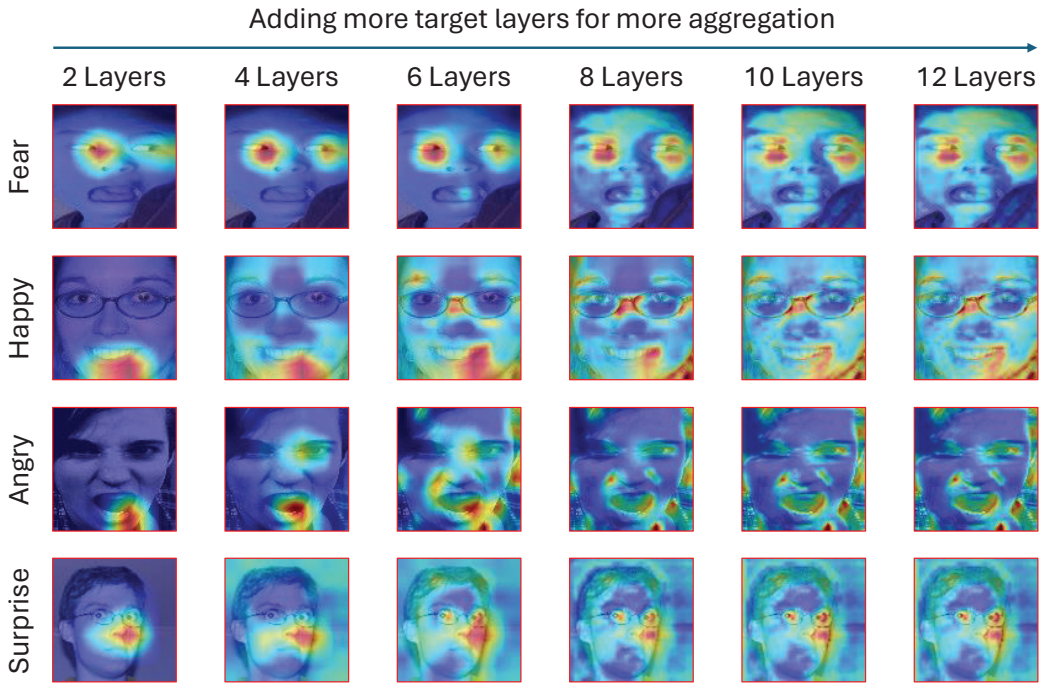

Figure 5 Visualization of the EigenCam output through the appliance on four different emotions: fear, happiness, surprise, and anger. The visualization is applied by increasing the number of target layers for more aggregation and explainability. Full-size DOI: 10.7717/peerj-cs.3301/fig-5

model occasionally misinterpreted cultural variations in emotion presentation, prompting further adjustments to enhance fairness and inclusivity.

A good explainability should also ensure trust calibration to boost user engagement and prevent both misuse and disuse of the system (*Morandini et al., 2023*). Misuse refers to excessive reliance on a system, while disuse indicates insufficient reliance despite the system's actual capabilities (*Alvarado-Valencia & Barrero, 2014*). *Sperrle et al. (2021)* and *Rong et al. (2023)* emphasize the persuasive influence of AI model explanations, highlighting their ability to convince users to accept model decisions, regardless of their accuracy. They argue that effective explanations should calibrate user trust, encouraging users to trust only accurate advice while being skeptical of incorrect guidance.

Naiseh et al. (2020) noted that both over-trust and under-trust pose risks related to explanations. Over-trust involves a high agreement with incorrect decisions, whereas under-trust signifies a low agreement with correct ones. *Zerilli, Bhatt & Weller (2022)* further pointed out that information overload (excessive transparency) can result in under-trust, while inadequate or confusing explanations may lead to a negative perception of the model. To mitigate these issues in EigenCam, several strategies can be employed, such as nudging through friction (*Naiseh et al., 2021*), offering interactive explanations, providing personalized insights based on user personality, and incorporating uncertainty (*Naiseh et al., 2020*).

EigenCam enhances the framework's applicability to cognitive disabilities by offering a transparent mechanism for emotion recognition. Understanding the anatomical and

physical cues utilized by the model fosters trust. It ensures that the framework operates in alignment with human expectations. This transparency is crucial for assistive technologies, as users and caregivers require confidence in the system's decisions and predictions.

The experiments underscore the efficacy and explainability of the proposed YOLOv8-based framework. Performance metrics confirm its robustness, while EigenCam visualizations provide actionable insights, bridging the gap between accuracy and interpretability.

Our adaptive feedback mechanisms represent a significant advancement over prior ITS solutions in several keyways. Unlike traditional ITS systems that provide static or one-size-fits-all guidance, our framework incorporates dynamic adjustments tailored to each user's cognitive profile and emotional state. This includes modifying the timing and complexity of navigational prompts based on real-time emotion detection, which is particularly beneficial for users with cognitive disabilities who may require additional processing time or simplified instructions.

Furthermore, our system's integration with wearable and mobile technologies enables context-sensitive interventions in dynamic transit environments; a capability not present in most existing ITS platforms. These innovations address critical gaps in personalization and real-time adaptability that have limited the effectiveness of previous ITS solutions for individuals with cognitive impairments.

Our analysis revealed potential cultural biases in emotion interpretation that warrant further investigation. While datasets like RAF-DB and AffectNet include diverse samples, a closer examination of demographic distributions showed uneven representation across ethnic groups. For instance, approximately 65% of subjects in RAF-DB identify as East Asian, while Western populations are overrepresented in AffectNet (45%) compared to other regions. These imbalances may contribute to the model's reduced accuracy for certain ethnic groups, particularly in detecting subtle expressions.

We observed a 5–7% performance gap between best-represented and underrepresented groups across key metrics. To address these limitations, future work will incorporate targeted data augmentation strategies and develop culturally-specific training subsets. Additionally, we will implement fairness metrics to systematically evaluate and mitigate demographic biases during model development.

Anatomical and physical changes in facial expressions

Facial emotions manifest through distinct anatomical and physical changes in facial muscles, which can be quantified and analyzed for emotion recognition. These changes correspond to action units (AUs) defined in the Facial Action Coding System (FACS). The precise identification of these changes forms the basis of emotion recognition frameworks like the one proposed in this study. Understanding these changes is critical for designing effective assistive systems, especially for individuals with cognitive disabilities who may struggle with emotional interpretation.

For emotions such as happiness, the zygomatic major muscles, responsible for raising the corners of the mouth, are prominently activated. This action is often accompanied by

the contraction of the orbicularis oculi muscles, producing "crow's feet" wrinkles around the eyes. These dual markers ensure reliable recognition of happiness, even in challenging conditions.

Sadness often involves lowering the lip corners (depressor anguli oris muscle) and the contraction of the inner brow raiser (frontalis muscle). These subtle changes, while harder to detect than those of happiness, provide critical cues for distinguishing sadness. The proposed YOLOv8 framework utilizes multi-scale feature extraction to capture these finer details effectively.

Anger is characterized by the lowering and contraction of the brows (corrugator supercilii muscle) and the tightening of the lips (orbicularis oris muscle). These distinctive markers, coupled with the dilation of nostrils, offer robust features for accurate classification. YOLOv8's attention mechanisms enhance the model's ability to focus on these regions.

Surprise leads to widening the eyes (levator palpebrae superioris muscle activation) and an elevation of the eyebrows, often coupled with an open mouth. The sharp contrast between these features and a neutral face makes surprise one of the easier emotions to detect with high precision.

Fear presents a combination of raised eyebrows, widened eyes, and a slightly open mouth. This complex pattern, involving multiple muscle groups, necessitates advanced techniques for accurate classification. The hierarchical feature extraction in YOLOv8 ensures the model captures these intricate patterns.

Impact on cognitive disabilities

For individuals with cognitive disabilities, interpreting these facial changes is often challenging. Impaired social cognition or atypical sensory processing may hinder their ability to understand emotional cues, leading to difficulties in social interactions. Assistive technologies utilizing emotion detection can bridge this gap, enabling better communication and social integration.

The proposed framework, enhanced with EigenCam-based explainability, provides actionable insights into the model's focus areas, ensuring that its interpretations align with human understanding. For instance, caregivers can rely on the system to detect subtle signs of distress or discomfort, such as sadness or fear, which might otherwise go unnoticed.

Emotion recognition systems also promote emotional awareness and learning for individuals with cognitive disabilities. By displaying real-time feedback on detected emotions, these systems offer opportunities for users to associate facial cues with corresponding emotional states, fostering social-emotional learning.

Moreover, using anatomical features ensures the system remains robust across diverse populations. For individuals with conditions such as autism spectrum disorder (ASD), who may display atypical facial expressions, the framework's focus on universal muscle movements ensures inclusivity and reliability.

Finally, the explainable nature of the framework enhances trust and transparency, which is crucial for adoption in assistive contexts. EigenCam visualizations allow

caregivers to verify the system's decisions, ensuring it does not rely on irrelevant or biased features. This transparency is critical in high-stakes scenarios, such as monitoring non-verbal individuals' signs of fear or distress.

By aligning advanced emotion recognition technology with human anatomical and social principles, the proposed framework significantly improves support for individuals with cognitive disabilities, enabling more empathetic and effective interactions.

Enhancing transparency and trustworthiness through emotion detection and explainability

Emotion detection results and explainability metrics significantly improve the usability, trustworthiness, and transparency of intelligent transportation systems (ITS) designed to support individuals with cognitive disabilities. This functionality positively impacts three primary stakeholder groups: users, caregivers, and transit personnel.

Real-time emotion detection fosters greater self-awareness by providing immediate feedback on emotional states. This feedback helps users recognize and regulate their emotions over time, promoting emotional intelligence and self-management. By detecting signs of distress, the system can initiate interventions, such as suggesting less crowded transit options or providing calming prompts. This proactive approach reduces anxiety and stress during overwhelming situations, making travel more manageable. The system can also support social interactions by offering cues about others' emotions or suggesting appropriate responses based on context. This helps users navigate complex social environments and fosters better communication.

The system enhances the ability of caregivers to detect early signs of emotional distress, even when subtle. This insight allows timely interventions, preventing escalation and ensuring the user remains safe and comfortable. Explainability tools like EigenCam visualizations allow caregivers to understand how the system interprets emotions. These insights provide the system with a focus on relevant facial features, minimizing biases and increasing caregivers' confidence in its reliability.

Emotion detection aids transit personnel in identifying passengers who may require assistance. For example, users displaying confusion or stress can receive immediate support, ensuring a smoother transit experience. The system improves security by identifying potential safety risks, such as individuals showing fear or anger. Transit personnel can respond quickly to prevent incidents, creating a safer environment for all passengers. Explainability features, such as visualizations and metrics, promote transparency in decision-making. This accountability ensures that the system's actions are justified and fair, reducing concerns about biases or errors in intervention.

Explainability metrics like EigenCam visualizations show which facial features the system uses to make predictions. These visual tools demystify the decision-making process, enabling users and stakeholders to understand how conclusions are reached. Metrics like localization error and correlation with ground truth data ensure the model focuses on relevant features while ignoring distractions. Providing this information to stakeholders fosters confidence in the system's accuracy and fairness. Explainability tools also help detect biases, such as over-reliance on cultural or demographic-specific

features. Developers can address these issues to ensure the system provides equitable support across diverse populations.

Performance in realistic transportation scenarios

To assess the real-world applicability of the YOLOv8-based emotion detection framework, it is essential to simulate and evaluate its performance in scenarios reflecting actual transportation environments. These scenarios test the system's robustness, adaptability, and responsiveness, addressing critical use cases for individuals with cognitive disabilities.

Crowded terminals: In busy transportation hubs, detecting emotions amidst high traffic, varied facial expressions, and fluctuating lighting conditions is a significant challenge. The system's precision and recall become vital to accurately identifying individuals needing assistance without triggering false positives. Testing under these conditions can help evaluate:

- *Detection robustness*: How well the system isolates relevant faces in dense crowds.
- *Emotion recognition accuracy*: The framework's ability to classify emotions correctly under suboptimal visibility and occlusion.
- *Time to intervention*: How quickly the system recognizes an emotional change and initiates a response.

Multi-step journeys: Transportation journeys often involve multiple transfers and modes, such as buses, trains, and taxis, over extended periods. Evaluating the system across such scenarios can determine the following:

- *Sustained accuracy*: The system can consistently detect emotions over time despite user fatigue or environmental changes.
- *Context adaptability*: How effectively the system adjusts interventions to align with evolving journey conditions.
- *Latency in dynamic environments*: The system's response time when shifting between transportation modes or locations.

Unexpected route changes: Unplanned disruptions like delays, cancellations, or rerouting can escalate stress and anxiety for individuals with cognitive disabilities. The system's effectiveness in these situations can be evaluated by:

- *Emotion change detection*: How quickly the system identifies a shift in emotional state due to an unexpected event.
- *Stress reduction interventions*: The ability of the system's responses (such as alternative route suggestions or calming prompts) to alleviate distress.
- *Timeliness of support*: Measuring the time from detection to delivery of assistance, ensuring interventions occur before stress levels peak.

Ethical considerations

Deploying facial emotion recognition systems, particularly in assistive technologies, raises important ethical concerns about privacy, consent, and data security. Facial images are inherently personal and sensitive, and their collection and use must adhere to strict ethical guidelines to protect user autonomy and confidentiality. To address these concerns, a framework should incorporate several privacy-preserving strategies as these measures ensure that the system maintains a high standard of ethical integrity while delivering its intended benefits:

- **On-device processing:** To minimize exposure of raw facial data, the system prioritizes local inference on edge devices (*e.g.*, smartphones, AR glasses). This ensures that biometric data does not leave the device unless explicitly authorized by the user or caregiver.
- **Data anonymization:** Where cloud-based processing or long-term storage is necessary, facial features can be encoded or transformed into non-reversible representations. Alternatively, real-time blurring or pixelation of identifying features can be applied before data transmission.
- **Explicit consent mechanisms:** The system includes clear prompts for informed consent before data collection. For users with cognitive disabilities, this may involve parental or guardian authorization, accompanied by accessible explanations tailored to the user's comprehension level.
- **Minimal data retention:** Only processed emotional states or aggregated trends (*e.g.*, frequency of emotions over time) are retained for adaptive feedback rather than storing raw video or image sequences.
- **Compliance with regulatory frameworks:** The framework supports integration with established data protection standards such as the General Data Protection Regulation (GDPR) and Health Insurance Portability and Accountability Act (HIPAA), where applicable.

In addition to the above measures, future real-world testing of the system will require formal ethical oversight to ensure compliance with research ethics standards. Specifically, the study protocol will need to be submitted to an Institutional Review Board (IRB) or equivalent ethics committee for approval. The IRB review process will evaluate the following aspects:

- **Risk assessment:** Ensuring that potential risks to participants, such as privacy breaches or misuse of data, are minimized and outweighed by the anticipated benefits of the research.
- **Informed consent procedures:** Verifying that consent forms and processes are clear, comprehensive, and appropriate for the target population, including individuals with disabilities who may require simplified language or additional support.

- **Data management plan:** Reviewing protocols for data collection, storage, and sharing to ensure compliance with ethical and legal standards, including provisions for secure anonymization and minimal data retention.
- **Equity and inclusivity:** Assessing whether the study design adequately addresses potential biases and ensures equitable access to the technology across diverse demographic groups.

Obtaining IRB approval will not only validate the ethical robustness of the proposed research but also enhance trust among stakeholders, including participants, caregivers, and regulatory bodies. This step is mandatory for ensuring that the deployment of facial emotion recognition systems adheres to the highest ethical standards and contributes positively to society.

LIMITATIONS

While the proposed YOLOv8-based framework demonstrates robust performance across multiple datasets, several limitations must be acknowledged. The system primarily relies on visual facial cues, which may reduce its effectiveness in scenarios involving occlusions (*e.g.*, face masks, sunglasses), extreme lighting conditions, or motion blur, common challenges in real-world transit environments. Additionally, the model focuses on prototypical emotional expressions and may struggle with subtle or mixed emotions, limiting its applicability in nuanced social interactions.

Although the framework incorporates explainability through EigenCam, there remains a risk of cultural bias in emotion interpretation due to dataset composition and potential over-reliance on specific demographic features during training. Emotion expression varies significantly across cultures, and the current implementation does not explicitly account for these variations.

Moreover, performance may degrade in highly dynamic environments with rapid movement or poor image quality. The system's reliance on consistent internet connectivity for cloud-based processing also poses challenges in areas with limited network coverage. Finally, while promising in controlled settings, the long-term impact of the system on emotional learning and social integration for individuals with cognitive disabilities requires extensive field testing and validation across diverse populations and real-world ITS contexts.

Preliminary mitigation strategies

To address the identified limitations, we propose the following initial steps toward future improvements, with a particular focus on integrating physiological sensors to complement facial emotion recognition. These sensors will enhance the system's reliability in scenarios where facial cues are obscured or ambiguous.

- **Integration of physiological sensors**: To complement facial emotion detection, wearable devices measuring physiological signals can be incorporated. These sensors provide additional modalities for emotion recognition, particularly in cases of occlusion or when facial expressions are subtle or mixed. Specific examples

include: electrocardiogram, galvanic skin response, electromyography, and respiration rate monitoring.

- **Data augmentation and synthetic data generation:** Advanced data augmentation techniques, including GAN-based synthesis and domain adaptation, can diversify training samples across underrepresented populations and improve generalization across varied cultural expressions.
- **Multimodal emotion recognition:** Future work will explore combining facial analysis with voice tone, body language, and contextual cues (*e.g.*, location, time of day) to build a more holistic understanding of emotional states, reducing dependency on visual-only inputs.
- **Cultural adaptability and fairness:** We emphasize the importance of using culturally inclusive datasets and annotations reflecting regional emotional expression differences to ensure equitable performance across diverse user groups.

These preliminary initiatives aim to lay the groundwork for addressing the identified limitations while promoting broader inclusivity and real-world applicability of emotion-aware ITS frameworks.

CONCLUSIONS

We propose that AI-powered emotion detection systems can significantly improve the capacity of individuals with cognitive disabilities to understand and respond to social and emotional cues. This research introduces an AI-driven framework utilizing advanced YOLOv8 models for accurate, real-time facial emotion detection tailored to the unique challenges faced by this group. These systems empower users by providing real-time, context-aware emotion recognition, enabling them to navigate social interactions more independently and confidently. Cognitive disabilities often hinder emotion recognition and expression, leading to isolation, miscommunication, anxiety, and reduced quality of life. Emotional intelligence is increasingly critical for integrating individuals with cognitive disabilities into educational, workplace, and social environments, as its absence can limit relationship-building, collaboration, and comfort in social settings.

The study emphasizes the transformative potential of AI-powered systems to address these challenges by filling the gaps left by current interventions, which are often generalized, time-consuming, and dependent on human facilitators. By integrating adaptive feedback mechanisms and explainable AI features, the proposed framework enhances accessibility, usability, and trust. The framework utilizes YOLOv8-based deep learning models for robust, real-time emotion classification across datasets like RAF-DB, AffectNet, and CK+48. Among the tested models, YOLOv8 Small (YOLOv8 S) achieved the highest accuracy, with 95.80% on RAF-DB, 93.95% on AffectNet, and 100% on CK+48.

This study advances deep learning applications by fine-tuning models to address diverse user needs, including varied response times and attention spans. It also promotes inclusivity by ensuring compatibility with assistive technologies such as AR devices for seamless integration into daily life. By utilizing these advanced deep learning models, the framework significantly improves the ability of individuals with cognitive disabilities to

interpret and respond to social cues. This technology can potentially transform therapeutic, educational, and social inclusion efforts, reducing stigma and promoting independence while enhancing the quality of life for affected individuals.

We plan to implement the framework in real-world pilot studies involving smart transit hubs and wearable assistive devices. These deployments will allow us to evaluate the system under authentic usage conditions and gather valuable feedback from users with cognitive disabilities, caregivers, and transit personnel. In parallel, we aim to explore the integration of physiological sensors (such as heart rate monitors and skin conductance sensors) to complement facial emotion recognition in occlusion or ambiguous expressions. Additionally, we intend to enhance model fairness by utilizing data augmentation and synthetic data generation techniques tailored for underrepresented populations, thereby improving cross-cultural generalization. In the long term, we envision a fully adaptive, multimodal, emotion-aware ITS ecosystem that dynamically responds to users' emotional and cognitive needs, promoting inclusive urban mobility and fostering greater independence and social integration for individuals with cognitive disabilities.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This work is supported by funds from the King Salman Centre for Disability Research (Group no.: KSRG-2024-240). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:
King Salman Centre for Disability Research: KSRG-2024-240.

Competing Interests

The authors declare that they have no competing interests.

Author Contributions

- Malik Almaliki conceived and designed the experiments, authored or reviewed drafts of the article, and approved the final draft.
- Amna Bamaqa conceived and designed the experiments, performed the computation work, prepared figures and/or tables, and approved the final draft.
- Tamer Ahmed Farrag performed the experiments, performed the computation work, authored or reviewed drafts of the article, and approved the final draft.
- Hossam Magdy Balaha performed the experiments, performed the computation work, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- Mahmoud Badawy conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, project Administration, and approved the final draft.

- Mostafa A. Elhosseini analyzed the data, performed the computation work, prepared figures and/or tables, authored or reviewed drafts of the article, project Administration, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:

The CK+48 dataset is available at Kaggle: https://www.kaggle.com/datasets/gauravsharma99/ck48-5-emotions.

The RAF-DB dataset is available at Kaggle: https://www.kaggle.com/datasets/shuvoalok/raf-db-dataset.

The AffectNet dataset is available at: https://mohammadmahoor.com/pages/databases/affectnet/.

The code is available at Zenodo: Hossam Magdy Balaha. (2025). HossamBalaha/Empowering-Cognitive-Disabilities-in-Transit: Manuscript Required Release (Publication). Zenodo. https://doi.org/10.5281/zenodo.15713565.

REFERENCES

Alvarado-Valencia JA, Barrero LH. 2014. Reliance, trust and heuristics in judgmental forecasting. *Computers in Human Behavior* **36(3)**:102–113 DOI 10.1016/j.chb.2014.03.047.

Bany Muhammad M, Yeasin M. 2021. Eigen-CAM: visual explanations for deep convolutional neural networks. *SN Computer Science* **2**:47 DOI 10.1007/s42979-021-00449-3.

Bebortta S, Tripathy SS, Basheer S, Chowdhary CL. 2023. DeepMist: toward deep learning assisted mist computing framework for managing healthcare big data. *IEEE Access* **11**:42485–42496 DOI 10.1109/access.2023.3266374.

Bennett R, Vijaygopal R. 2024. Exploring mobility and transportation technology futures for people with ambulatory disabilities: a science fiction prototype. *Technovation* **133(4)**:103001 DOI 10.1016/j.technovation.2024.103001.

Bindawas SM, Vennu V. 2018. The national and regional prevalence rates of disability, type, of disability and severity in Saudi Arabia—analysis of 2016 demographic survey data. *International Journal of Environmental Research and Public Health* **15(3)**:419 DOI 10.3390/ijerph15030419.

Centers for Disease Control and Prevention. 2024. Data and statistics on autism spectrum disorder. *Available at https://www.cdc.gov/autism/data-research/index.html* (accessed 26 November 2024).

Chaudhari A, Bhatt C, Krishna A, Mazzeo PL. 2022. VITFER: facial emotion recognition with vision transformers. *Applied System Innovation* **5(4)**:80 DOI 10.3390/asi5040080.

Chen X, Zheng X, Sun K, Liu W, Zhang Y. 2023. Self-supervised vision transformer-based few-shot learning for facial expression recognition. *Information Sciences* **634(8)**:206–226 DOI 10.1016/j.ins.2023.03.105.

Clark JM, Polesello D. 2017. Emotional and cultural intelligence in diverse workplaces: getting out of the box. *Industrial and Commercial Training* **7(8)**:337–349 DOI 10.1108/ict-06-2017-0040.

Dalabehera AR, Bebortta S, Kumar N, Senapati D. 2024. Mist-fog-assisted real-time emotion recognition using deep transfer learning framework for smart city 4.0. *Internet of Things* **27**:101237 DOI 10.1016/j.iot.2024.101237.

Des Portes V. 2020. Intellectual disability. In: *Handbook of Clinical Neurology*. Vol. 174. Amsterdam: Elsevier, 113–126.

Ekundayo OS, Viriri S. 2021. Facial expression recognition: a review of trends and techniques. *IEEE Access* **9**:136944–136973 DOI 10.1109/access.2021.3113464.

Ge H, Zhu Z, Dai Y, Wang B, Wu X. 2022. Facial expression recognition based on deep learning. *Computer Methods and Programs in Biomedicine* **215(5)**:106621 DOI 10.1016/j.cmpb.2022.106621.

Grieco J, Pulsifer M, Seligsohn K, Skotko B, Schwartz A. 2015. Down syndrome: cognitive and behavioral functioning across the lifespan. In: *American Journal of Medical Genetics Part C: Seminars in Medical Genetics*. Vol. 169. Hoboken: Wiley Online Library, 135–149.

Harris JC. 2006. *Intellectual disability: understanding its development, causes, classification, evaluation, and treatment*. Oxford: Oxford University Press.

Huang Y, Chen F, Lv S, Wang X. 2019. Facial expression recognition: a survey. *Symmetry* **11(10)**:1189 DOI 10.3390/sym11101189.

Jain DK, Shamsolmoali P, Sehdev P. 2019. Extended deep neural network for facial emotion recognition. *Pattern Recognition Letters* **120(May (6))**:69–74 DOI 10.1016/j.patrec.2019.01.008.

Jaiswal A, Raju AK, Deb S. 2020. Facial emotion detection using deep learning. In: *2020 International Conference for Emerging Technology (INCET)*. Piscataway: IEEE, 1–5.

Kaur M, Kumar M. 2024. Facial emotion recognition: a comprehensive review. *Expert Systems* **41(10)**:e13670 DOI 10.1111/exsy.13670.

Keltner D, Cordaro DT. 2017. Understanding multimodal emotional expressions: recent advances in basic emotion theory. In: Russell JA, Fernandez Dols JM, eds. *The Science of Facial Expression, Social Cognition and Social Neuroscience (New York, 2017; online edn, Oxford Academic, 18 May 2017)* DOI 10.1093/acprof:oso/9780190613501.003.0004.

Khan AR. 2022. Facial emotion recognition using conventional machine learning and deep learning methods: current achievements, analysis and remaining challenges. *Information* **13(6)**:268 DOI 10.3390/info13060268.

Khattak A, Asghar MZ, Ali M, Batool U. 2022. An efficient deep learning technique for facial emotion recognition. *Multimedia Tools and Applications* **81(2)**:1649–1683 DOI 10.1007/s11042-021-11298-w.

Lau BT. 2010. Portable real time emotion detection system for the disabled. *Expert Systems with Applications* **37(9)**:6561–6566 DOI 10.1016/j.eswa.2010.02.130.

Levine K, Karner A. 2023. Approaching accessibility: four opportunities to address the needs of disabled people in transportation planning in the United States. *Transport Policy* **131(1)**:66–74 DOI 10.1016/j.tranpol.2022.12.012.

Li S, Deng W. 2020. Deep facial expression recognition: a survey. *IEEE Transactions on Affective Computing* **13(3)**:1195–1215 DOI 10.1109/taffc.2020.2981446.

Lucey P, Cohn JF, Kanade T, Saragih J, Ambadar Z, Matthews I. 2010. The extended Cohn-Kanade dataset (CK+): a complete dataset for action unit and emotion-specified expression. In: *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops*. Piscataway: IEEE, 94–101.

Ma F, Sun B, Li S. 2021. Facial expression recognition with visual transformers and attentional selective fusion. *IEEE Transactions on Affective Computing* **14(2)**:1236–1248 DOI 10.1109/taffc.2021.3122146.

Maenner MJ. 2023. Prevalence and characteristics of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, 11 sites, United States, 2020. *MMWR Surveillance Summaries* **72(1)**:80–82 DOI 10.1097/dbp.0000000000000245.

Makkonen T, Inkinen T. 2024. Inclusive smart cities? Technology-driven urban development and disabilities. *Cities* **154**:105334 DOI 10.1016/j.cities.2024.105334.

Marechal C, Mikolajewski D, Tyburek K, Prokopowicz P, Bougueroua L, Ancourt C, Wegrzyn-Wolska K. 2019. Survey on AI-based multimodal methods for emotion detection. *High-Performance Modelling and Simulation for Big Data Applications* **11400(3)**:307–324 DOI 10.1007/978-3-030-16272-6_11.

Matthews G, Zeidner M, Roberts RD. 2007. Emotional intelligence: consensus, controversies, and questions. *The Science of Emotional Intelligence: Knowns and Unknowns* 3–46 DOI 10.1027/1016-9040.13.1.64.

Mellouk W, Handouzi W. 2020. Facial emotion recognition using deep learning: review and insights. *Procedia Computer Science* **175(2)**:689–694 DOI 10.1016/j.procs.2020.07.101.

Miller F, Wallis J. 2011. Social interaction and the role of empathy in information and knowledge management: a literature review. *Journal of Education for Library and Information Science* **52(2)**:122–132.

Mogaji E, Nguyen NP. 2021. Transportation satisfaction of disabled passengers: evidence from a developing country. *Transportation Research Part D: Transport and Environment* **98(2)**:102982 DOI 10.1016/j.trd.2021.102982.

Mollahosseini A, Hasani B, Mahoor MH. 2017. AffectNet: a database for facial expression, valence, and arousal computing in the wild. *IEEE Transactions on Affective Computing* **10(1)**:18–31 DOI 10.1109/taffc.2017.2740923.

Morandini S, Fraboni F, Puzzo G, Giusino D, Volpi L, Brendel H, Balatti E, De Angelis M, De Cesarei A, Pietrantoni L. 2023. Examining the nexus between explainability of AI systems and user's trust: a preliminary scoping review. In: *CEUR Workshop Proceedings (CEUR-WS.org)*.

Morris MW, Keltner D. 2000. How emotions work: the social functions of emotional expression in negotiations. *Research in Organizational Behavior* **22(4)**:1–50 DOI 10.1016/s0191-3085(00)22002-9.

Naiseh M, Al-Mansoori RS, Al-Thani D, Jiang N, Ali R. 2021. Nudging through friction: an approach for calibrating trust in explainable AI. In: *2021 8th International Conference on Behavioral and Social Computing (BESC)*. Piscataway: IEEE.

Naiseh M, Jiang N, Ma J, Ali R. 2020. Explainable recommendations in intelligent systems: delivery methods, modalities and risks. In: *Research Challenges in Information Science: 14th International Conference, RCIS 2020, Limassol, Cyprus, September 23–25, 2020, Proceedings 14*. Cham: Springer, 212–228.

Powers DM. 2020. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. ArXiv DOI 10.48550/arXiv.2010.16061.

Razzaque A. 2020. Virtual learning enriched by social capital and shared knowledge, when moderated by positive emotions. *International Journal of Electronic Banking* **2(1)**:77–95 DOI 10.1504/ijebank.2020.105418.

Rong Y, Leemann T, Nguyen T-T, Fiedler L, Qian P, Unhelkar V, Seidel T, Kasneci G, Kasneci E. 2023. Towards human-centered explainable AI: a survey of user studies for model explanations. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **46(4)**:2104–2122 DOI 10.1109/tpami.2023.3331846.

Sappok T, Diefenbacher A, Winterholler M. 2019. The medical care of people with intellectual disability. *Deutsches Ärzteblatt International* **116(48)**:809 DOI 10.3238/arztebl.2019.0809.

Sappok T, Hassiotis A, Bertelli M, Dziobek I, Sterkenburg P. 2022. Developmental delays in socio-emotional brain functions in persons with an intellectual disability: impact on treatment

and support. *International Journal of Environmental Research and Public Health* **19(20)**:13109 DOI 10.3390/ijerph192013109.

Segrin C. 1996. Interpersonal communication problems associated with depression and loneliness. In: *Handbook of Communication and Emotion*. Amsterdam: Elsevier, 215–242.

Shan L, Deng W. 2018. Reliable crowdsourcing and deep locality-preserving learning for unconstrained facial expression recognition. *IEEE Transactions on Image Processing* **28(1)**:356–370 DOI 10.1109/tip.2018.2868382.

Sohan M, Sai Ram T, Reddy R, Venkata C. 2024. A review on YOLOv8 and its advancements. In: *International Conference on Data Intelligence and Cognitive Informatics*. Cham: Springer, 529–545.

Sperrle F, El-Assady M, Guo G, Borgo R, Chau DH, Endert A, Keim D. 2021. A survey of human-centered evaluations in human-centered machine learning. *Computer Graphics Forum* **40**:543–568 DOI 10.1111/cgf.14329.

Szabó P, Ara J, Halmosi B, Sik-Lanyi C, Guzsvinecz T. 2023. Technologies designed to assist individuals with cognitive impairments. *Sustainability* **15(18)**:13490 DOI 10.3390/su151813490.

Terven J, Córdova-Esparza D-M, Romero-González J-A. 2023. A comprehensive review of YOLO architectures in computer vision: from YOLOv1 to YOLOv8 and YOLO-NAS. *Machine Learning and Knowledge Extraction* **5(4)**:1680–1716 DOI 10.3390/make5040083.

Van Kleef GA. 2009. How emotions regulate social life: the emotions as social information (EASI) model. *Current Directions in Psychological Science* **18(3)**:184–188 DOI 10.1111/j.1467-8721.2009.01633.x.

World Health Organization. 2024. Autism. *Available at https://www.who.int/news-room/fact-sheets/detail/autism-spectrum-disorders* (accessed 26 November 2024).

Yates K, Le Couteur A. 2016. Diagnosing autism/autism spectrum disorders. *Paediatrics and Child Health* **26(12)**:513–518 DOI 10.1016/j.paed.2016.08.004.

Zeidan J, Fombonne E, Scorah J, Ibrahim A, Durkin MS, Saxena S, Yusuf A, Shih A, Elsabbagh M. 2022. Global prevalence of autism: a systematic review update. *Autism Research* **15(5)**:778–790 DOI 10.1002/aur.2696.

Zeidner M, Roberts RD, Matthews G. 2008. The science of emotional intelligence: current consensus and controversies. *European Psychologist* **13(1)**:64–78 DOI 10.1027/1016-9040.13.1.64.

Zerilli J, Bhatt U, Weller A. 2022. How transparency modulates trust in artificial intelligence. *Patterns* **3(4)**:100455 DOI 10.1016/j.patter.2022.100455.