International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Conference Proceedings: Volume: 05 ACME – 2026 DOI:10.55041/ISJEM.ACME014
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

Predicting Behaviour Change in Students with Special Educational Needs using Multimodal Learning Analytics

Cuddapah Hameed[1]
Department of Computer Science and Engineering
Sri Venkateswara College of Engineering, Karkambadi
Tirupati, India, 517501
mailhameed5@gmail.com

Saritha A[2]
Department of Computer Science and Engineering
Sri Venkateswara College of Engineering, Karkambadi
Tirupati, India, 517501
saritha.a@svcolleges.edu.in

Abstract— One important use of artificial intelligence is facial emotion recognition (FER), which allows machines to recognize and react to human emotions. In order to identify and categorize human emotions from real-time video input, this research shows the design and implementation of a real-time facial emotion recognition system that combines computer vision and deep learning approaches. Angry, Disgust, Fear, Happy, Neutral, Sad, and Surprise are the seven categories into which the system divides facial expressions after identifying faces using Haar Cascade classifiers. Normalized 48x48 pixel facial pictures were used to create and train a Convolutional Neural Network (CNN) architecture. To enhance model generalization and lessen overfitting, data preparation methods like rescaling and picture augmentation like rotation, shifting, and horizontal flipping were used. Multiple convolutional, pooling, and dropout layers are used in the network to extract hierarchical spatial data. Then, fully connected layers are used to classify multiple classes of emotions using a soft-max activation function. To guarantee effective training and precise prediction, the model was optimized using the categorical cross-entropy loss function and the Adam optimizer. The trained model was coupled with Tainter for graphical user interface creation and OpenCV for face identification for real-time deployment. The program records live webcam footage, recognizes facial features, preprocesses the area of interest, and instantly predicts emotions with confidence scores shown on the screen. The suggested system successfully demonstrates the useful application of deep learning in affective computing and provides dependable performance under regulated conditions, according to experimental data. In addition to highlighting the promise of emotion detection systems in fields including behavioural analytics, education, healthcare, and human-computer interaction, this study also identifies areas for future advancements in multimodal integration and robustness

Keywords— Facial Emotion Recognition (FER), Deep Learning and Computer Vision techniques, Convolutional Neural Network (CNN).

1. Introduction

Background and Context: Education systems worldwide are increasingly emphasizing inclusive learning environments that accommodate diverse student needs. Among these learners, students with Special Educational Needs (SEN) require tailored support, adaptive teaching strategies, and continuous monitoring. SEN students may face cognitive, behavioural, emotional, or developmental challenges that influence their learning trajectories and classroom participation. Understanding and predicting behavioural changes in this group is vital for timely interventions that enhance both academic performance and emotional well-being.

Traditional monitoring approaches: such as teacher observations, manual assessments, and periodic reporting, while valuable, are often subjective, time-intensive, and limited in detecting subtle behavioural variations. Recent advances in Artificial Intelligence (AI), Machine Learning (ML), and Learning Analytics have opened new opportunities for data-driven

educational decision-making. Within this domain, Multimodal Learning Analytics (MMLA) has emerged as a powerful approach, capable of analysing diverse data streams simultaneously to generate holistic insights into student behaviour. This research explores the use of MMLA to predict behavioural changes in SEN students by integrating multimodal data, including facial expressions, speech patterns, physiological signals, classroom interaction logs, and academic performance metrics.

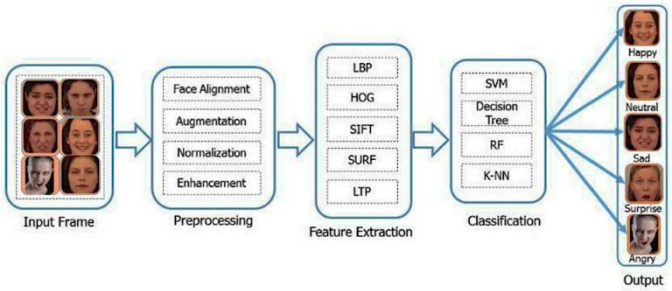

Fig 1: Different Facial Expressions of Students
Source: https://www.mdpi.com/2078-2489/13/6/268

Understanding Special Educational Needs (SEN): SEN encompasses a wide spectrum of conditions, including learning disabilities, ADHD, autism spectrum disorder (ASD), emotional and behavioural disorders, speech and language impairments, and intellectual disabilities. These conditions affect how students process information, regulate emotions, interact socially, and respond to classroom stimuli. Behavioural changes may manifest as anxiety, withdrawal, aggression, reduced attention, emotional distress, disengagement, or sudden academic decline. Early identification of such shifts is critical, as delayed recognition can result in academic setbacks, social isolation, or mental health challenges. However, these changes are often subtle and context-dependent, making them difficult to detect through observation alone. Hence, intelligent systems capable of continuous monitoring and prediction are essential.

Emergence of Learning Analytics: Learning Analytics involves the measurement, collection, analysis, and reporting of learner data to optimize educational processes. Traditional analytics rely on digital traces such as assignment submissions, online activity logs, quiz scores, and attendance records. While useful, these sources provide limited insight into emotional states or behavioural patterns. With advances in sensors, wearable devices, computer vision, and natural language processing, researchers have shifted toward Multimodal Learning Analytics (MMLA). Unlike traditional methods, MMLA integrates multiple modalities—visual, audio, physiological, interaction, and contextual data thereby offering a richer, more comprehensive understanding of learner behaviour.

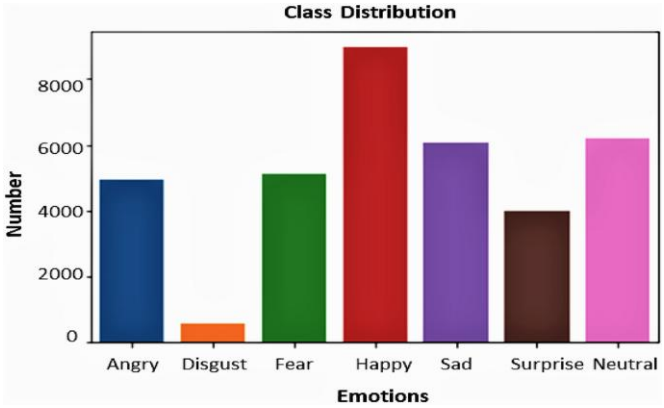

Fig 2: Different emotions in Students
Source: https://www.nature.com/articles/s41598-022-11173-0

Multimodal Learning Analytics in Special Education: Multimodal analytics holds transformative potential in special education. SEN students often express emotional and cognitive states differently from neurotypical peers. For example, a student with autism may avoid eye contact yet remain attentive, while a student with ADHD may appear restless but cognitively engaged. Single-modal analysis often fails to capture these complexities. Multimodal systems, however, can correlate signals across modalities to identify behavioral transitions. For instance, increased fidgeting (visual cue), elevated voice pitch (audio cue), and declining task completion (interaction data) may collectively indicate frustration. Similarly, reduced facial expressiveness, lower participation, and slower response times may signal disengagement. Such cross-modal analysis enables more accurate and earlier detection of behavioural changes.

Technological Foundations: The proposed system leverages advanced ML and deep learning models to process multimodal data. Techniques include Convolutional Neural Networks (CNNs) for facial emotion detection, Recurrent Neural Networks (RNNs) or LSTMs for temporal modelling, audio feature

extraction for speech emotion recognition, and sensor data analysis for physiological monitoring. Data fusion whether early, late, or hybrid integrates these modalities to capture complex behavioural representations and temporal dependencies beyond the reach of single-modal systems.

This research study aims to design and evaluate a multimodal learning analytics framework for predicting behavioural changes among students with SEN. The framework will collect and preprocess multimodal data, extract behavioural features, develop predictive models, evaluate performance, and provide interpretable insights for educators. By bridging advanced analytics and special education practice, this research contributes to inclusive education by demonstrating how multimodal integration can enhance behavioural understanding and enable proactive interventions.

2. Existing System

Manual observation and recurring assessments have been the main methods used in standard educational settings to track and forecast behavioural changes in students with Special Educational Needs (SEN). Teachers usually use qualitative reports to document occurrences, engagement levels, and emotional reactions. These reports are subjective and time-consuming, but they offer useful contextual information. These approaches are not scalable and frequently miss minute behavioural changes that develop gradually over time. Technologically speaking, previous behavioural analysis and emotion detection systems mostly depended on manually developed feature extraction methods as geometric facial landmark analysis, Local Binary Patterns (LBP), and Histogram of Oriented Gradients (HOG). Conventional machine learning classifiers such as Support Vector Machines (SVM), K-Nearest Neighbour's (KNN), and Decision Trees were then used to process these features. These strategies showed some success in controlled laboratory environments, but in real-world classrooms with unpredictable lighting, background noise, and student movement, their efficacy drastically decreased. As a result, current systems have had difficulty offering SEN populations dependable, ongoing, and context-sensitive behavioural monitoring.

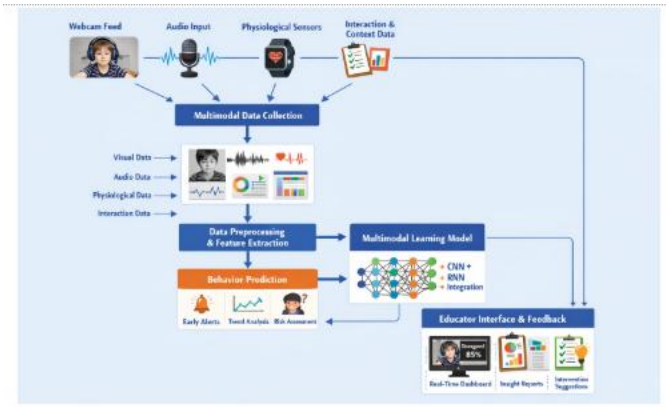

Fig 3: Ethical and Data Security Layer involved in the existing system

Despite their initial promise, traditional systems exhibit several critical limitations that restrict their applicability in special education contexts. First, they are heavily dependent on manual feature engineering, which often fails to capture the nuanced emotional and behavioural cues characteristic of SEN students. For instance, subtle facial expressions or micro-gestures that may indicate anxiety or disengagement are easily overlooked when relying on handcrafted features. Second, most existing systems operate on a single modality, analyzing either visual data or academic performance in isolation. This narrow focus prevents the development of a holistic understanding of student behaviour, as emotional states and engagement levels are often expressed through multiple channels simultaneously. Third, these systems demonstrate poor generalization capability, with performance deteriorating sharply in dynamic classroom environments. Variations in pose, lighting, and background noise frequently compromise accuracy, making them unsuitable for continuous monitoring. Furthermore, traditional models typically perform static classification, failing to account for temporal dynamics and behavioural changes that evolve over time. This lack of temporal modeling means that interventions are often reactive rather than predictive, limiting their effectiveness in preventing escalation.

Another major drawback of existing systems lies in their inability to deliver real-time performance. Computationally expensive feature extraction processes reduce scalability and hinder deployment in live classroom settings where immediate feedback is essential. As a result, behavioural shifts are often

identified only after they have already impacted learning outcomes, delaying intervention and support. This reactive approach undermines the potential for proactive educational strategies that could prevent emotional distress, disengagement, or academic decline. Moreover, the reliance on static models and handcrafted features restricts adaptability, making it difficult to tailor systems to the diverse and individualized needs of SEN students. Taken together, these limitations highlight the pressing need for an intelligent, multimodal, and predictive system designed specifically for special education contexts. Such a system must integrate multiple data modalities, leverage advanced machine learning techniques, and provide interpretable insights to educators

3. Proposed System

The proposed system introduces a Multimodal Learning Analytics (MMLA) framework specifically designed to predict behavioural changes in students with Special Educational Needs (SEN) using advanced deep learning techniques. Unlike traditional approaches that rely on handcrafted features, this system leverages Convolutional Neural Networks (CNNs) to automatically extract hierarchical spatial features from facial expressions, thereby eliminating the limitations of manual feature engineering. Beyond visual analysis, the framework integrates multiple data streams including facial expressions, classroom interaction logs, contextual academic data, and optional physiological signals to construct a comprehensive behavioural profile. This multimodal integration ensures that subtle emotional cues, engagement levels, and contextual influences are captured simultaneously, providing a richer and more accurate understanding of student behaviour.

At the core of the system are several functional components that work together to deliver predictive insights. First, multimodal data acquisition collects diverse inputs such as webcam-based facial expressions, task completion times, engagement metrics, and academic performance indicators. Where available, physiological signals like heart rate or skin conductance can be incorporated to enhance emotional state detection. Second, face detection and emotional feature extraction are performed using OpenCV-based Haar Cascade algorithms for localization, followed by CNN-based emotion classification to identify nuanced facial expressions. Third, behavioural pattern modeling employs temporal analysis of emotional trends combined with machine learning models to predict transitions in behaviour over time. Finally, a real-time behavioural prediction interface presents insights through a graphical user interface (GUI), offering educators live feedback and alerts that support proactive intervention.

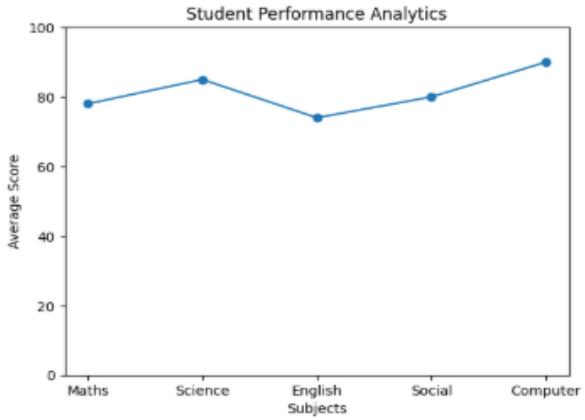

Fig 4: Average Score for the Student Performance

The proposed system offers several distinct advantages over existing methods. Automated feature extraction through CNNs eliminates the need for manual engineering, enabling the detection of complex emotional patterns directly from raw pixel data. Multimodal integration ensures that behavioural predictions are not limited to a single modality but instead reflect a holistic view of student engagement and emotional states. Deep learning models provide improved accuracy and generalization, performing reliably across diverse student populations and varying classroom conditions. Temporal behaviour tracking allows the system to identify early warning signs by analyzing emotional patterns over time, while real-time performance ensures smooth monitoring without delays. Importantly, the educator-friendly GUI makes the system accessible to teachers without requiring technical expertise, and the predictive nature of the framework shifts the paradigm from reactive observation to proactive intervention. Collectively, these advantages position the system as a transformative tool for inclusive education.

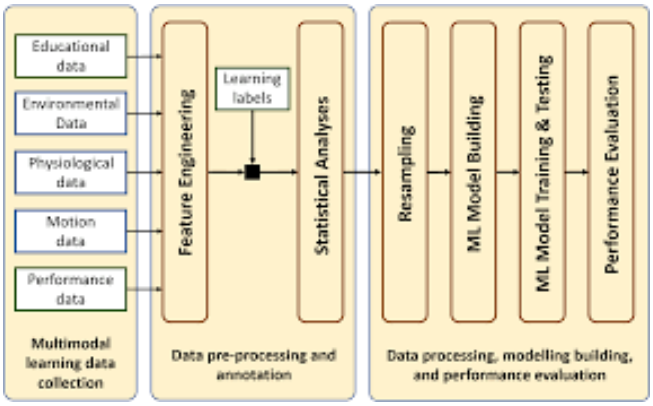

Fig 5: System Architecture for the Model
Source:
https://ieeexplore.ieee.org/iel7/6287639/10005208/10159389.pdf

A feasibility study was conducted to evaluate the practicality of implementing this multimodal predictive framework in educational settings. From a technical perspective, the research is highly feasible due to the availability of mature open-source technologies and AI frameworks. Python serves as the primary programming language, supported by Tensor-Flow and Keras for deep learning, OpenCV for computer vision, NumPy for data processing, and Tkinter for GUI development. The CNN-based emotion recognition architecture, combined with multimodal analytics, is well-supported by existing research, and real-time inference can be achieved using standard CPUs, with GPU acceleration recommended for faster training. Furthermore, multimodal data fusion techniques are increasingly accessible through modern machine learning libraries, ensuring that integration across modalities is both practical and efficient.

Economic and operational feasibility further strengthen the case for implementation. Economically, the system is viable because all required software tools are open-source and free, eliminating the need for proprietary licenses. Hardware requirements are minimal, with a standard computer and webcam sufficient for deployment, while optional GPU usage can enhance performance but is not mandatory. This makes the system accessible to educational institutions with limited budgets, including special education schools. Operational feasibility is equally critical: the system requires minimal technical expertise to operate, provides intuitive visual feedback, and can be seamlessly integrated into classroom setups. Rather than replacing human judgment, it assists teachers by offering clear behavioural insights, supporting informed decision-making while maintaining ethical monitoring guidelines. By combining technical robustness, economic accessibility, and operational simplicity, the proposed system demonstrates strong feasibility and promises to significantly enhance behavioural prediction and intervention in special education contexts.

4. Methodology

The methodology adopted in this study is designed to systematically develop and evaluate a multimodal learning analytics framework capable of predicting behavioral changes in students with Special Educational Needs (SEN). The research follows an applied design, combining experimental modeling with practical classroom integration. The process begins with multimodal data acquisition, where diverse inputs are collected to capture the cognitive, emotional, and contextual dimensions of student behaviour. Visual data is obtained through webcams to analyses facial expressions, while interaction logs record task completion times, participation frequency, and engagement metrics. Academic data such as quiz scores, assignment submissions, and attendance records provide contextual insights into performance trends. Where feasible, physiological signals such as heart rate and skin conductance are incorporated through wearable sensors to enhance emotional state detection. This multimodal acquisition ensures that subtle behavioral cues are not overlooked and that the system can construct a holistic profile of each learner.

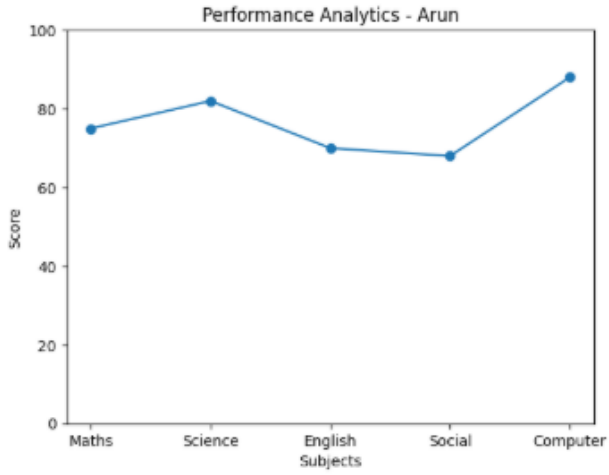

Fig 6: Score for the Performance Analytics for student 1

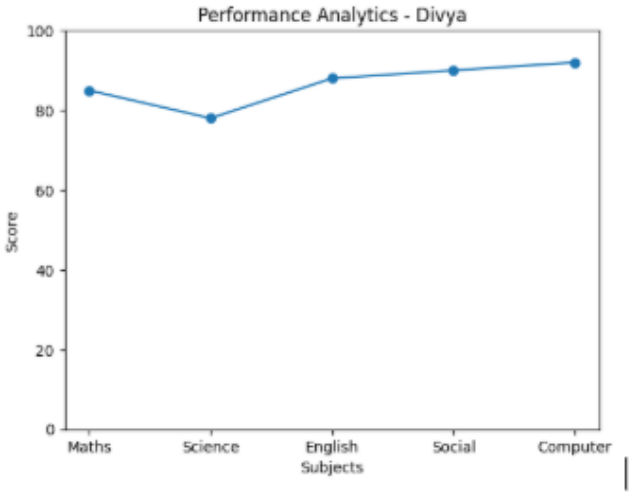

Fig 7: Score for the Performance Analytics for Student 2

Once data is collected, it undergoes preprocessing and feature extraction to ensure quality, consistency, and usability. Visual data is processed using OpenCV-based Haar Cascade algorithms for face localization, followed by Convolutional Neural Networks (CNNs) for emotion classification. Audio and interactional data are analyzed to extract features such as pitch variation, speech rhythm, response latency, and task completion speed. Academic data is normalized to align with behavioural indicators, while physiological signals are filtered to remove noise and synchronized with other modalities. Unlike traditional systems that rely on handcrafted features, the proposed framework leverages deep learning to automate feature extraction. CNNs capture hierarchical spatial features from facial expressions, while sequential models such as Long Short-Term Memory (LSTM) networks analyses temporal dependencies in behavioural patterns. This combination ensures that both static and dynamic aspects of behaviour are effectively modelled.

The next stage involves behavioural pattern modeling and multimodal data fusion. Predictive modeling integrates features across modalities using early, late, and hybrid fusion techniques. Early fusion combines raw features before classification, late fusion integrates predictions from modality-specific models, and hybrid fusion leverages both approaches to maximize accuracy. Temporal analysis is conducted using Recurrent Neural Networks (RNNs) and LSTMs, which are particularly effective in identifying sequential trends and predicting transitions in behaviour over time. For example, the system can detect early warning signs such as disengagement, frustration, or emotional distress by correlating facial cues, interaction logs, and academic performance metrics. This predictive capability shifts the paradigm from reactive observation to proactive intervention, enabling educators to anticipate behavioural changes before they escalate. The modeling process is iterative, with performance continuously refined through training, validation, and testing phases to ensure robustness across diverse classroom environments.

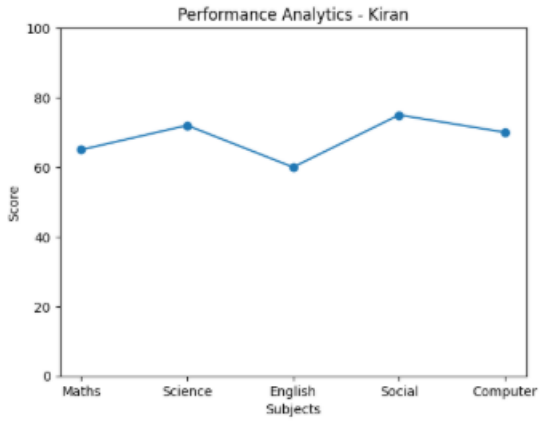

Fig 8: Score for the Performance Analytics for Student 3

5. Results

Finally, the framework is deployed through a real-time behavioural prediction interface designed for educators. A graphical user interface (GUI) presents live feedback, alerts, and visual dashboards that display emotional trends, engagement levels, and predicted behavioural risks. The interface emphasizes interpretability, ensuring that predictions are explained in educator-friendly terms rather than opaque technical outputs. Evaluation of the system is conducted using quantitative metrics such as accuracy, precision, recall, F1-score, and confusion matrix analysis, alongside qualitative feedback from educators regarding usability and practical impact. Real-time responsiveness is also assessed to ensure that the system operates smoothly in classroom settings without latency. Ethical considerations are embedded throughout the methodology: data privacy is maintained through anonymization and secure storage, informed consent is obtained from caregivers and institutions, and monitoring is designed to assist teachers rather than replace human judgment. By integrating multimodal data acquisition, deep learning-based feature extraction, temporal modeling, and educator-friendly interfaces, the

methodology ensures both technical rigor and educational relevance, ultimately supporting inclusive learning environments for SEN students.

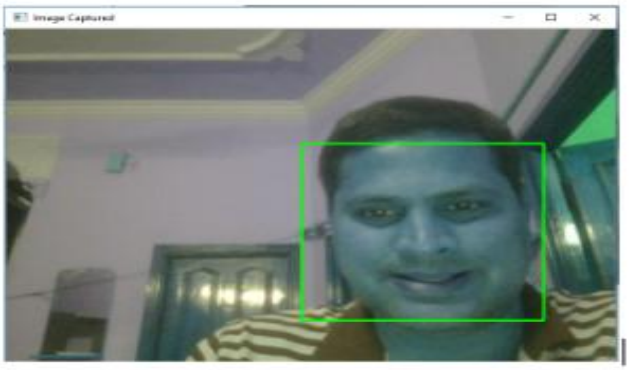

Fig 9: Capturing the image of a Student

7. Conclusion

This research successfully demonstrates the design, development, and deployment of a real-time Facial Emotion Recognition (FER) system that integrates Deep Learning and Computer Vision techniques into a unified and practical application. The system is capable of detecting human faces from a live webcam feed and classifying their emotional expressions into seven predefined categories: Angry, Disgust, Fear, Happy, Neutral, Sad, and Surprise.

At the core of the system lies a Convolutional Neural Network (CNN) developed using TensorFlow and Keras. The model was trained on standardized 48×48pixel facial images, enabling it to learn hierarchical feature representations from low-level edges and textures to high-level abstract patterns associated with emotional expressions. The use of image preprocessing techniques such as normalization significantly improved convergence speed during training, while augmentation strategies including rotation, shifting, and flipping enhanced model robustness and minimized overfitting. Additionally, Dropout regularization layers contributed to improved generalization performance on unseen data.

For real-time face detection, OpenCV's Haar Cascade Classifier provided an efficient and lightweight solution for identifying facial regions within video frames. The detected Region of Interest (ROI) was then pre-processed and fed into the trained CNN model for instant prediction. The seamless integration of video capture, preprocessing, inference, and rendering ensured smooth and low-latency performance. Furthermore, the Tkinter-based Graphical User Interface (GUI) enhanced usability by providing a clear, interactive display of emotion predictions along with confidence scores, making the system accessible even to non-technical users.

The Research meets its primary objectives by delivering a functional, responsive, and reasonably accurate emotion recognition system. It demonstrates the practical applicability of artificial intelligence in interpreting human affective states and highlights the growing potential of FER systems in domains such as education, healthcare, human-computer interaction, and customer experience analytics.

6. Future Work

The proposed multimodal learning analytics framework incorporates a wide range of features designed to capture, process, and interpret behavioural indicators in students with Special Educational Needs (SEN). These features span across multiple modalities, ensuring a holistic and accurate representation of student behaviour. The scope of features can be categorized into the following dimensions:

Visual Features

Facial Expression Analysis: Automatic extraction of emotional states (e.g., happiness, sadness, anxiety, frustration) using CNN-based models.

Micro-Expressions and Gestures: Detection of subtle cues such as fidgeting, avoidance of eye contact, or reduced facial expressiveness.

Posture and Movement Tracking: Identification of restlessness, withdrawal, or physical agitation through webcam-based monitoring.

Audio and Interactional Features

Speech Emotion Recognition: Analysis of tone, pitch, and speech rhythm to detect stress, excitement, or disengagement.

Classroom Interaction Logs: Monitoring task completion time, participation frequency, and response latency.

Engagement Metrics: Tracking clickstreams, keystroke dynamics, and interaction speed in digital learning platforms.

Contextual and Academic Features

Performance Indicators: Integration of assignment scores, quiz results, and attendance records to correlate academic outcomes with behavioural states.

Environmental Context: Consideration of classroom dynamics, peer interactions, and instructional strategies that may influence behaviour.

Temporal Trends: Longitudinal tracking of performance and engagement patterns to identify recurring triggers or gradual behavioural shifts.

Physiological Features

Heart Rate Monitoring: Detection of stress or anxiety through wearable sensors.

Skin Conductance: Measurement of arousal levels to identify emotional intensity.

Multimodal Fusion: Combining physiological signals with visual and interactional data for enhanced predictive accuracy.

Predictive and Interpretive Features

Temporal Behaviour Modeling: Use of RNNs/LSTMs to analyse sequential data and predict transitions in emotional or behavioural states.

Real-Time Alerts: Immediate feedback to educators through a GUI interface, highlighting potential risks such as disengagement or emotional distress.

Interpretable Insights: Presentation of predictions in educator-friendly formats, ensuring transparency and actionable decision-making.

8. References

[1] Ekman, P., & Friesen, W. V. (1978). Facial Action Coding System: A Technique for the Measurement of Facial Movement. Consulting Psychologists Press, Palo Alto, CA.

[2] Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press, Cambridge, MA.

[3] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep Learning. Nature, 521(7553), 436–444.

[4] Krizhevsky, A., Sutskever, I., & Hinton, G. (2012). ImageNet Classification with Deep Convolutional Neural Networks. In Proceedings of Advances in Neural Information Processing Systems (NIPS).

[5] Viola, P., & Jones, M. (2001). Rapid Object Detection using a Boosted Cascade of Simple Features. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR).

[6] Chollet, F. (2017). Deep Learning with Python. Manning Publications.

[7] Zhang, Z., Luo, P., Loy, C. C., & Tang, X. (2016). Learning Deep Representation for Face Alignment with Auxiliary Attributes. IEEE Transactions on Pattern Analysis and Machine Intelligence.

[8] Mollahosseini, A., Hasani, B., & Mahoor, M. H. (2017). AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild. IEEE Transactions on Affective Computing.

[9] OpenCV Contributors. (2023). Open Source Computer Vision Library (OpenCV). OpenCV.

[10] Abadi, M., et al. (2016). TensorFlow: A System for Large-Scale Machine Learning. In Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI). Google.

[11] Kingma, D. P., & Ba, J. (2015). Adam: A Method for Stochastic Optimization. In Proceedings of the International Conference on Learning Representations (ICLR).

[12] Shan, C., Gong, S., & McOwan, P. W. (2009). Facial Expression Recognition Based on Local 7Binary Patterns: A Comprehensive Study. Image and Vision Computing, 27(6), 803–816.