Consolidated Research Paper Summaries

Student Engagement and Emotion Recognition

This document contains detailed one-page summaries for ten selected research papers focusing on student engagement analysis and emotion recognition in educational settings.

1. Deep Learning-Based Model for Analyzing Student Engagement in Activities

Title and Authors

- **Title:** Deep learning-based model for analyzing student engagement in activities
- **Author:** Feng Feng
- **Publication:** *Scientific Reports* (Nature), December 2025

Problem Statement Traditional engagement measurement methods, such as surveys and manual observations, are subjective and lack real-time capability. Furthermore, existing automated models often struggle to effectively integrate multimodal signals (video, audio, and digital logs), resulting in incomplete or inaccurate analyses of student behavior in diverse learning activities.

Standard Dataset(s) Used

- **DAiSEE** (Dataset for Affective States in E-Environments) – Used for benchmarking and comparison.
- **Student Privacy Log Dataset** (Kaggle) – Used for log analysis.

Dataset Details

- **Source:** A custom multimodal dataset was collected comprising:
  - **Video:** Classroom recordings capturing facial micro-expressions, action units, and posture.
  - **Audio:** Interaction recordings analyzing speech prosody and tone.
  - **Logs:** Digital activity logs including clickstreams and navigation patterns.
- **Preprocessing:**
  - **Video/Logs:** Min-max normalization to handle scale and illumination variations.
  - **Audio:** Spectral median filtering for noise reduction.
- **Labels:** Engagement states classified as **Active**, **Passive**, or **Disengaged**.

Model / Algorithm Applied

- **Core Architecture: IC-BiSGRU-Net** (Intelligent Chimp-driven Bidirectional LSTM + Scalable GRU).
- **Key Components:**
  - **CNN Encoder:** Extracts visual features (facial micro-expressions).

- **Fusion Layer:** Uses Iterative Attention-based Fusion (IAF²) and Multi-Scale Channel Attention Module (MS-CAM) to weigh and combine multimodal features.
- **Temporal Modeling:** Combines Bi-LSTM (for forward/backward dependencies) with SGRU (Scalable Gated Recurrent Unit) for computational efficiency.
- **Optimization:** "Intelligent Chimp" strategy used to balance accuracy and cost.

Key Findings or Results

- **Accuracy:** The model achieved **96.21%** accuracy on the test data.
- **Precision: 98.31%**.
- **Recall: 97.23%**.
- **F1-Score: 98.13%**.
- **Conclusion:** The IC-BiSGRU-Net outperformed traditional CNN, GRU, and LSTM models in both accuracy and computational efficiency, successfully classifying real-time engagement states.

2. Real-Time Emotion Recognition in Online Learning Using Google Teachable

Title and Authors

- **Title:** Real-Time Emotion Recognition in Online Learning Using Google Teachable
- **Author:** Nazli Rahmeisi
- **Year:** 2026

Problem Statement Online learning environments suffer from a lack of non-verbal cues that are naturally present in face-to-face classrooms. This study addresses the need for an accessible, low-barrier solution to detect learner emotions in real-time without requiring expensive hardware or complex software setups.

Standard Dataset(s) Used

- **FER2013** (Facial Expression Recognition 2013)
- **CK+** (Extended Cohn-Kanade Dataset)

Dataset Details

- **Input:** Live webcam feed capturing facial expressions during online sessions.
- **Training Data:** The model was trained on the standard FER2013 and CK+ datasets to recognize universal emotional expressions.
- **Classes:** Four primary emotional states: **Happy, Sad, Neutral, Angry**.

Model / Algorithm Applied

- **Platform: Google Teachable Machine** (a browser-based, low-code machine learning platform).
- **Architecture:** A standard **CNN-based classifier** (Convolutional Neural Network) optimized for web deployment (likely based on MobileNet or similar lightweight architectures used by Teachable Machine).

Key Findings or Results

- **Overall Accuracy: 91.1%**.
- **Class-Specific Performance:**
    - **Happy:** Highest precision (**95.2%**).
    - **Sad:** Lowest precision (**86.9%**).
- **Conclusion:** The study demonstrates that accessible, browser-based AI tools can effectively bridge the "emotional gap" in online education, providing real-time affective computing capabilities with minimal technical overhead.

3. Understanding the Impact of Emotional Engagement on Learning Outcomes

Title and Authors

- **Title:** Understanding the impact of emotional engagement on learning outcomes in online education: an automated analysis approach
- **Authors:** Guanyu Chen, Guangxin Han, Juan Niu, Juhou He

Problem Statement The physical separation in online education makes it difficult for instructors to gauge student engagement. This study aims to automate the detection of emotional engagement to identify its specific correlation with academic performance and learning outcomes.

Standard Dataset(s) Used

- *Note: This study primarily utilized a self-collected dataset for its core analysis.*

Dataset Details

- **Participants:** 40 undergraduate students engaged in online learning activities.
- **Volume: 71,185 labeled facial images** extracted from video frames.
- **Labels:** Three levels of emotional engagement derived from facial analysis.

Model / Algorithm Applied

- **Core Model:** Optimized **Vision Transformer (ViT)**.
- **Methodology:**
    - **Transfer Learning:** Utilized to adapt the ViT model to the specific domain of student engagement.
    - **Mechanism:** Leveraged ViT's self-attention mechanisms to capture subtle, global dependencies in facial images that traditional CNNs might miss.
    - **Validation:** 5-fold Cross-Validation.

Key Findings or Results

- **Accuracy:** Mean accuracy of **90.62%** (Standard Deviation: 3.09%) across folds. (Abstract notes a peak accuracy of 93.8%).
- **Precision: 89.78%**.

- **Recall: 89.43%**.
- **Pedagogical Insight:** The study established a significant positive correlation between high emotional engagement levels and better learning outcomes, validating the use of automated emotional analytics as a predictor of academic success.

4. Facial Emotion Recognition Based Real-Time Learner Engagement Detection

Title and Authors

- **Title:** Facial emotion recognition based real-time learner engagement detection system in online learning context using deep learning models
- **Authors:** Swadha Gupta, Prateek Kumar, Raj Kumar Tekchandani
- **Year:** 2022

Problem Statement To improve digital education outcomes, there is a critical need to monitor learner engagement in real-time. This study proposes a system to detect engagement levels via facial emotion recognition, enabling immediate feedback.

Standard Dataset(s) Used

- **FER-2013**
- **CK+**
- **RAF-DB** (Real-world Affective Faces Database)

Dataset Details

- **Custom Data:** A proprietary dataset was created to validate the system in real-time online learning scenarios.
- **Input:** Images captured via built-in device cameras.
- **Output:** An calculated **Engagement Index (EI)** classifying the learner as "Engaged" or "Disengaged".

Model / Algorithm Applied

- **Face Detection: Faster R-CNN**.
- **Feature Extraction:** A proposed Modified Face-Point Extractor (**MFACEXTOR**) to identify key facial landmarks.
- **Classification:** Comparative analysis of three deep learning models:
  - **Inception-V3**
  - **VGG19**
  - **ResNet-50**

Key Findings or Results

- **Comparative Accuracy:**
  - **ResNet-50: 92.32%** (Best Performance).
  - VGG19: 90.14%.

    - Inception-V3: 89.11%.
- **Conclusion:** ResNet-50 combined with the proposed feature extractor provided the most robust performance for classifying facial emotions in a real-time learning environment.

5. Real-Time Facial Emotion Recognition for AI-Enhanced Personalized Learning

Title and Authors

- **Title:** Real-Time Facial Emotion Recognition for AI-Enhanced Personalized Learning
- **Authors:** Vijya Tulsani, Prashant Sahatiya, Jignasha Parmar
- **Year:** 2025

Problem Statement Existing systems often lack the speed or precision required for *dynamic* personalization. This study aims to build a system that not only detects emotions but uses them to immediately adapt the pace, difficulty, and instructional method of the content.

Standard Dataset(s) Used

- **FER2013+**
- **AffectNet**

Dataset Details

- **Preprocessing:** Used **MTCNN** (Multi-task Cascaded Convolutional Networks) for precise face detection and alignment.
- **Classes:** Seven universal emotions (Happy, Sad, Neutral, Angry, Fear, Surprise, Disgust).

Model / Algorithm Applied

- **Architecture: Hybrid ViT-BiLSTM**.
- **Components:**
    - **Vision Transformer (ViT):** Extracts high-level spatial facial features using attention mechanisms.
    - **Bi-LSTM (Bidirectional LSTM):** Captures temporal dependencies to track emotional transitions across video frames.

Key Findings or Results

- **Accuracy:** Achieved an overall accuracy of **91.72%**.
- **Latency:** Average response time of **0.8 seconds**, confirming viability for real-time personalization.
- **Conclusion:** The hybrid ViT-BiLSTM model outperformed baseline CNN-LSTM and standalone Transformer models, effectively enabling the system to adapt content based on the learner's emotional state.

6. Revolutionizing Online Education: Advanced Facial Expression Recognition

Title and Authors

- **Title:** Revolutionizing online education: Advanced facial expression recognition for real-time student progress tracking via deep learning model
- **Author:** Mohammed Aly

Problem Statement To empower educators with actionable insights into student well-being and progress by accurately detecting complex emotional states (such as confusion or frustration) in real-time.

Standard Dataset(s) Used

- **RAF-DB**
- **FER2013**
- **CK+**
- **KDEF** (Karolinska Directed Emotional Faces)

Dataset Details

- **Scope:** The study utilized multiple large-scale datasets to ensure the model's robustness across different demographics and image qualities.

Model / Algorithm Applied

- **Architecture: ResNet50 + CBAM + TCNs**.
- **Components:**
  - **ResNet50:** A deep residual network for robust feature extraction.
  - **CBAM (Convolutional Block Attention Module):** An attention mechanism that focuses the network on critical facial regions (eyes, mouth) relevant to expression.
  - **TCN (Temporal Convolutional Networks):** Handles the dynamic, time-variant aspects of facial expressions.

Key Findings or Results

- **Accuracy by Dataset:**
  - **KDEF: 97.08%** (Highest).
  - **CK+: 95.85%**.
  - **RAF-DB: 91.86%**.
  - **FER2013: 91.71%**.
- **Conclusion:** The integration of the CBAM attention module significantly improved the model's ability to detect subtle emotional cues compared to standard baseline models.

7. Detection of Students' Emotions in an Online Learning Environment

Title and Authors

- **Title:** Detection of Students' Emotions in an Online Learning Environment Using a CNN-LSTM Model
- **Authors:** Bilkisu Muhammad Bashir, Hadiza Ali Umar

- **Year:** 2025

Problem Statement General emotion recognition often fails to capture learning-specific states. This research focuses on detecting emotions specifically relevant to learning—**Interest, Boredom, and Confusion**—to help instructors adjust their teaching delivery.

Standard Dataset(s) Used

- **FER2013**
- **CK+**
- **JAFFE** (Japanese Female Facial Expression)

Dataset Details

- **Custom Construction:** A dataset of ~6,000 samples was created by mapping facial action units from the standard datasets (FER2013, CK+, JAFFE) to the specific learning categories (Interest, Boredom, Confusion).
- **Preprocessing:** Images were resized to 48x48 grayscale and normalized.

Model / Algorithm Applied

- **Architecture: CNN-LSTM Hybrid**.
- **Components:**
  - **CNN Layers:** Three layers (32, 64, 128 filters) with ReLU activation for spatial feature extraction.
  - **LSTM Layers:** Two stacked layers (128 hidden units) to model the temporal sequence of expressions.
  - **Optimizer:** Adam optimizer with early stopping to prevent overfitting.

Key Findings or Results

- **Accuracy:** Achieved **98.0%** accuracy on the constructed dataset.
- **Conclusion:** The hybrid approach effectively captured both the static visual features and the dynamic temporal changes required to distinguish between complex states like confusion and boredom.

8. YOLOv5 Based Student Engagement and Emotional States Detection

Title and Authors

- **Title:** YOLOv5 Based Student Engagement and Emotional States Detection in E-Classes
- **Authors:** Shuai Wang, Abdul Samad Shibghatullah, Kay Hooi Keoy, Javid Iqbal

Problem Statement Online classrooms face challenges in monitoring multiple students simultaneously and addressing issues like cyberbullying. The study proposes a detection system for emotions and engagement to improve safety and teaching adaptivity.

Standard Dataset(s) Used

- **FER2013** (Refined/Cleaned version)

Dataset Details

- **Primary Data:** ~11,000 usable samples from a refined FER2013 dataset (removing mislabeled/non-facial images).
- **Secondary Data:** A self-constructed dataset of 668 classroom images and survey responses from 520 students.

Model / Algorithm Applied

- **Core Model: YOLOv5s-6.0** (SER-YOLO).
- **Enhancements:**
  - **Soft-NMS (Non-Maximum Suppression):** Improved handling of overlapping bounding boxes (crucial for crowded classrooms).
  - **Coordinate Attention (CA):** Integrated to strengthen feature map representation.
  - **EIoU Loss:** Used to refine bounding box regression for better localization.

Key Findings or Results

- **Precision: 87.3%** (an improvement of 3.9% over baseline).
- **mAP@0.5: 88.9%** (an improvement of 3.0%).
- **Conclusion:** The enhanced YOLOv5 model proved effective for multi-student detection in single frames, offering a viable solution for monitoring classroom dynamics and safety.

9. Emotion Recognition of College Students' Online Learning Engagement

Title and Authors

- **Title:** Emotion Recognition of College Students' Online Learning Engagement Based on Deep Learning
- **Author:** Chunyan Wang
- **Year:** 2022

Problem Statement Recognizing that positive emotions (curiosity, joy) drive engagement while negative ones (boredom, frustration) hinder it, this study aims to accurately identify these states to predict broader engagement levels in college students.

Standard Dataset(s) Used

- **FER2013** (Used for pre-training)

Dataset Details

- **Testing Data:** Self-collected facial data from college students during online learning.
- **Classes:** 7 standard emotions mapped to three engagement categories (Cognitive, Emotional, Behavioral).

Model / Algorithm Applied

- **Architecture: CNN + Bidirectional LSTM (BLSTM)**.
- **Methodology:**

- **Transfer Learning:** A CNN pre-trained on FER2013 was fine-tuned on the student dataset.
- **BLSTM:** Used to process the sequence of features extracted by the CNN to understand the emotional context over time.

Key Findings or Results

- **Accuracy:** Achieved **>90%** accuracy in emotion recognition tasks.
- **Conclusion:** The study validated that deep learning models could effectively correlate specific emotional states with broader engagement metrics, providing a technological basis for automated student monitoring.

10. Real-Time Emotion Recognition of Students in Online Learning Environments

Title and Authors

- **Title:** Real-Time Emotion Recognition of Students in Online Learning Environments Using Deep Learning Models
- **Authors:** Researchers at the University of Baghdad (Published in *Baghdad Science Journal*)

Problem Statement Post-COVID, monitoring student motivation remains difficult. This study proposes a system to provide instructors with real-time feedback on emotional engagement (Focus/Persistence vs. Boredom/Frustration) to enable immediate pedagogical adjustments.

Standard Dataset(s) Used

- **FER2013**
- **CK+**

Dataset Details

- **Preprocessing:** Applied normalization, resizing, and data augmentation to improve model robustness against variations in lighting and positioning.
- **Classes:** Happiness, Sadness, Anger, Surprise, Fear, Disgust, Neutrality.

Model / Algorithm Applied

- **Architecture: CNN-LSTM Hybrid**.
- **Function:**
    - **CNN:** Extracts spatial features from individual video frames.
    - **LSTM:** Models the temporal dynamics of facial expressions to ensure continuity in emotion detection.
    - **Training:** Implemented using TensorFlow/Keras with the Adam optimizer and cross-validation.

Key Findings or Results

- **Accuracy:** Achieved **>90%** accuracy in emotion recognition tasks.