East Journal of Human Science
ISSN: 3079-9384
Volume 1, Issue 5
Publisher: East Publication & Technology
DOI: https://doi.org/10.63496/ejhs.Vol1.Iss5.162

Semantic labeling of Cognitive Visual Streams Based on Learner Behavior Analysis Using Deep-Emotion-AI for Engaging Experience

Muhammad Safyan *[1], Sohail Sarwar [2], Muhammad Kashif [3], Saima Rafiq [4], Muddesar Iqbal [5]

[1]Department of Computer Science, Government College University, Lahore, Pakistan, safyanch@gcu.edu.pk.
[2]Department of Computer Science, Bahria University, Islamabad, Pakistan, sohail.sarwar@seecs.edu.pk.
[3]Department of Advanced Computing, College of Advanced Technology, National University of Science and Technology, Muscat, Oman, muhammadkashif@nu.edu.om
[4]Department of Computer Science, Government College University, Lahore, Pakistan, saima.rafiq@gmail.com
[5]Department of Communication and Networks Engineering, Prince Sultan University, Riyadh, KSA, miqbal@psu.edu.sa
*Corresponding Author.

This article is part of a special issue dedicated to the International Conference on Emerging Technologies in Multidisciplinary Fields (ICETMF25), 8–9 July 2025, organized by Mazoon College, Muscat, Oman.

Received: 19/07/2025, **Revised**: 07/08/2025, **Accepted**: 31/08/2025, **Published**: 03/09/2025

Abstract

The Consumer Electronics (CE) devices (such as smartphones, tablets, laptops, and cameras) equip learners with educational ubiquity, eliminating barriers of cost, time, and space. The features of emotion recognition and analysis in CE devices have the potential to effectively engage learners through personalized and adaptive recommendations of learning content. Engagement activities triggered by the emotions and expressions of the learner rely on the semantic features of the facial coding units. Current research integrates deep emotion AI to analyze learner behavior and label emotions semantically. Hence, the system can dynamically gain deeper insight into learners' emotional states and cognitive responses to maximize the learning engagement based on real-time facial visual streams. Vision-based CE gadgets were used, featuring five major modules and 68 recognized points of interest in each 360x360-pixel image. Deep-Emotion-AI uses a hybrid of Support Vector Machine (SVM) and Convolutional Neural Networks (CNN) to recognize learners, capture their expressions, and semantically label learner's emotions (annotated through Emotions Ontology) on extended Cohn-Kanade (CK+), MMI, and Google datasets. The comparison of the CK+ dataset with Google dataset and MMI highlighted the better performance of the proposed model with CK+. The proposed approach achieved competitive accuracy for labelling emotions between 84 % and 98%. The proposed research has established that the current application of CE with fusion of semantics, deep emotional AI, and a selected set of characteristics can significantly capture the attention of the learner. Potential future directions are deep learning-based gesture analysis, reinforcing content precision in learner feedback, and efficiently catering to the computational requirements of a larger audience.

***Keywords*: Deep Learning, Emotion Recognition, Convolutional Neural Networks, Support Vector Machines, Emotion Ontology, Behavior Analysis**

1. Introduction

The introduction should begin with a broad overview of the topic to provide context and relevance. This sets the Consumer electronics have the potential to play a pivotal role in education by integrating emotion analysis into the feedback mechanisms of academic learners (Nayak, Routray, & Sarma, 2023). These devices may help to gauge learners by implementing features of facial recognition, expression tracking, and emotion analysis to efficiently contextualize learners' sentiments. Facial units serve as valuable indicators of learners' experiences, enabling real time assessment of their engagement and comprehension levels. Fostering of consumer electronics with educational technologies (expected to have a value of USD 605.40 billion by 2027, currently USD 254.80 billion), as mentioned by (Lindner, J. 2023), is attaining widespread attention across various academic domains. Consumer electronics facilitate the seamless integration of semantic emotion analysis into educational activities, ultimately promoting a more supportive and engaging educational environment. Expression or facial actions are the tools that reflect the learner's experience. Facial emotions and expressions have taken diverse attention in recent decades because of immense applications in different academic domains. Facial analysis based on CE enables observers to assess audience engagement and well-being during real-time interactive activities. It has been highlighted that there are eight primary emotions like joy, sadness, acceptance, disgust, fear, anger, surprise, and anticipation (Paikrao, P., Mukherjee, A., & Alnumay, W, 2023). Furthermore, according to research, it is possible to detect multiple faces in video or through real-time consumer electronic gadgets (Acheampong, et al., 2021, and Sai et al., 2024). In real-time videos, images can be taken at a particular time frame and processed for the machine's interpretation. This processing is carried out at the conjecture of expression and resizing of the image to a required size. The system further processes the incoming video streams, while consuming minimum resources in terms of bandwidth and latency in an academic environment (Paikrao, P., Mukherjee, A., & Alnumay, W. 2023). There are constant efforts to achieve the best sensor techniques and algorithms for HCI (Human Computer Interaction) based emotion analysis in prevailing ubiquitous ecosystems such as Apple's Siri (Acheampong et al., 2021), perhaps the first era of emotionally smart and intelligent machines. For example, happiness and excitement may present different levels of intensity for the basic emotion of happiness and may be classified at a finer level of granularity. Current research targets to extend the effectiveness of CE devices on way to provide an engaging learning environment (Sai, S., Goyal, D., Chamola, V., & Sikdar, B. 2024). It proposes a unique approach for semantically recognizing and labeling the facial actions and emotions by using "wheel of emotions" (for seven basic emotions/ expressions, as discussed in section 2) through Emotions-AI (Gursesli et al. 2024, Hwang et al. 2020, Majumder et al. 2016, Clark et al. 2020, Said et al. 2022, Westera et al. 2014). Hence, adaptive aspects can be focused on where learners are fully engaged while using the item or service/content that is offered or created. This analysis has also validated the information by domain experts as captured through CE gadgets (in semantic labeling of emotional states) in real-time and accurately. Once emotions are detected in real time, the learning activities can be tweaked to engage the learners for effective learning outcomes, achieving the following objectives:

- Develop a Consumer Electronics (CE)-based framework for the real-time semantic labelling of cognitive visual streams, specifically focusing on learner behavior analysis using Deep-Emotion-AI.
- Enable dynamic insight into learners' emotional states and cognitive responses from real-time facial visual streams to maximize learning engagement.
- Propose and validate a unique hybrid Deep-Emotion-AI approach, combining Support Vector Machine (SVM) and Convolutional Neural Networks (CNN), for accurate facial action and emotion recognition and semantic labelling, annotated through Emotions Ontology.
- Demonstrate that the integration of semantic emotion analysis with CE devices can significantly enhance educational activities, leading to a more supportive and engaging learning environment and improved learning outcomes.

Based on these objectives, the research posits the following hypotheses:

- H1: Integrating deep emotion AI and semantic features into consumer electronics will significantly enhance learner engagement by providing personalized and adaptive learning content recommendations.
- H2: A hybrid Deep-Emotion-AI model, leveraging Support Vector Machines (SVM) and Convolutional Neural Networks (CNN) for facial expression recognition and semantically labelled using Emotions Ontology, will achieve a high degree of precision and accuracy (e.g., an overall accuracy of 98 % in detecting and classifying learners' emotions from real-time visual streams.
- H3: This proposed framework will demonstrate superior performance in emotion recognition and semantic labelling when evaluated on standard datasets such as Extended Cohn-Kanade (CK+), MMI, and Google datasets, particularly exhibiting better performance with the CK+ dataset.
- H4: The system will be able to efficiently process incoming video streams with minimal resource consumption in terms of bandwidth and latency within an academic environment, thereby providing real-time assessment of learner engagement and comprehension levels.

The hybrid of deep learning approaches produced outputs that identified an optimal hyperplane, identifying new examples to be trained. This enabled us to achieve a relatively higher degree of precision. An average accuracy of 98 % is achieved by the proposed Deep emotions-AI model, utilizing the Emotions Ontology, in semantic emotion labeling. Domain experts have acknowledged it through psychophysical analysis. The rest of the paper is organized as follows. Section II covers the research background and literature review; Section III presents the proposed methodology with associated details. Section IV presents the experimental results, and Section V provides the conclusion and potential future directions.

2. Related Work

In the paper discussed by Hwang et al. (2020), the concept of CE is integrated with education, emotion detection, and multi-action recognition for building an action-based system to recognize facial expressions using a You Only Look Once (YOLO) deep learning architecture. It establishes that analyzing the emotional state of students at the higher education level is of considerable significance for instructors as well as parents to track their progress. There are five universal human emotions commonly available in literature, regardless of where or how they are raised (Majumder et al. 2016, Clark et al. 2020). Some researchers might think that fundamental feelings are not taught but are innate. For example, individuals who are born blind and never saw faces still reveal the quintessential facial expressions of basic sentiments (Said et al. 2022). Previously, Westera et al. (2014) conducted research in e-learning with a focus on machine learning and image effect meters to forecast mental states with detrimental interest and offered a routine aggregation strategy to track substantial negative sentiments and remedial alignment of learning content (Tao et al. 2019 and Corcoran et al. 2020). There are a few fundamental feelings that are valid for every age and cultural distinctions accredited as the universal recognition of six emotions (anger, happiness, fear, surprise, disgust, and sadness) (Prikler et al. 2017 and Pantic et al. 2004). The research by Mijic et al. (2021) and Moore et al. (2011) showcased the implementation and design of an emotional classification system in real time. The range of positive to negative emotions and arousal is described on an active to passive scale. The two-dimensional emotional scheme as proposed by Russell, J. A., Posner, J., & Peterson, B. S. (2005) has been presented as emotions in Figure 1. In a research study, it was concluded that the modern pathway is made up of 4 phases in the contemporary area of recognition, i.e. discovery, alignment, classification, and representation. This research examines orientation phases and the representation stage by using direct 3D face mapping to create a step-by-step facial model from a nine-layer deep neural network. CE and Independent Component Analysis (ICA) have been utilized to obtain facial characteristics and recognition and then attempts to forecast face expressions with the Support Vector Machines (SVM) classification (Greene et al. 20216, Ranzato et al. 2014, Rejeb et al. 2022, Kohonen et al. 2009, Prattico et al. 2024, Prist et al. 2020) also used the same technique for classification and work done by Hassouneh. et al. (2020) and Nikolaidis. et al. (2007) showed how to use rule-based techniques to detect face expression and action units expressed in the Silveira CK + and MMI data set, used by N., Ara´ujo, P., & Beu, L. (2016). A study by Kanade, T., Saragih, J., Ambadar, Z. (2010). interprets the facial expressions in continuous videos using deep emotion AI approaches; they focused on the output of classifiers using head to establish separate temporal parameters of face behavior, which comes with a 3D-face monitor, including the 3D head position and facial movements concurrently.

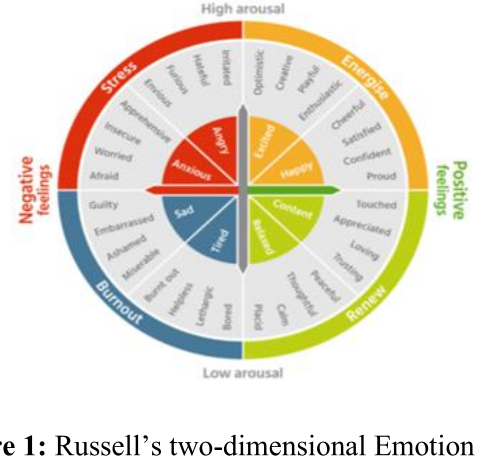

Figure 1: Russell's two-dimensional Emotion Model

In this current research, a CE-based framework is proposed to address the shortcomings in literature. Different ways for face recognition and classification of emotions are highlighted for the assessment of adaptive learning.

3. Proposed Framework

The proposed framework is intelligent, effective in nature, and maintains the semantics of various facial recognition/classification techniques. The modules, as illustrated in Figure 2, present the distinct stages to detect the face, extract features, and classify and label the images.

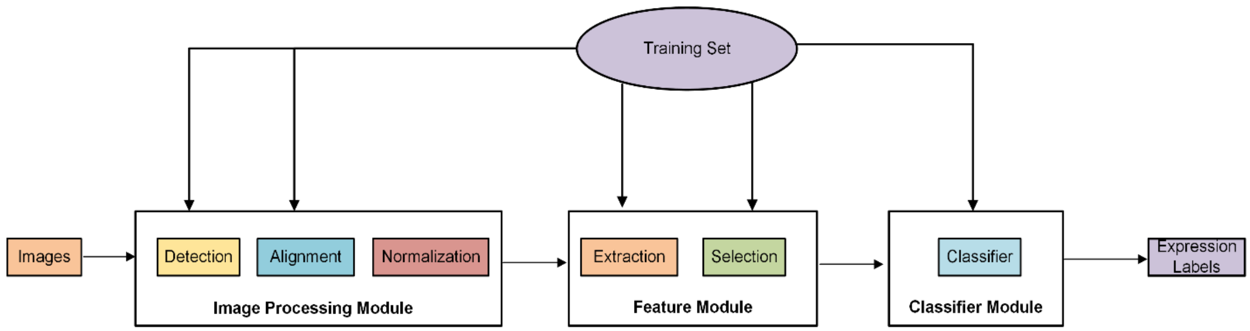

Figure 2: Modular Structure of the Proposed CE-Based Framework for Semantic Emotion Labelling

The identification of facial expression involves both facial movement measurement and image processing. The Deep-Emotion-AI framework employs a hybrid integration strategy where the Convolutional Neural Network (CNN) primarily functions as a robust feature extractor. The CNN processes raw facial image data to learn and generate high-level, discriminative features (learned representations). These extracted features are then passed as input to the Support Vector Machine (SVM) classifier, which performs the final classification of emotions by identifying an optimal hyperplane within this learned feature space. This feature-level fusion leverages the strengths of CNN for automatic and powerful feature learning and the SVM's proven efficacy in classifying complex patterns, leading to the observed high accuracy in emotion labelling.

A three-stage categorization has been performed for the overall strategy of Automatic Facial Expression Analysis (AFEA) as shown in Figure 2 are: Face detection, Extraction and classification of facial characteristics, and Facial expressions recognition. These three components have been implemented through five major modules, namely (1) Detecting and Assessing feelings, (2) Face detection model and identification, (3) Sub Emotions Taxonomy for labeling, (4) Active and Inactive State Analysis, (5) Training-sets. The proposed structure consists of distinct stages to detect the face, extract features, classify, and label images. The details regarding the methodological elements of our Deep-Emotion-AI framework, as depicted in Figure 3, are described as follows:

- **Dataset Usage and Validation Strategy**: The proposed Deep-Emotion-AI model was rigorously evaluated using extended Cohn-Kanade (CK+), MMI, and Google datasets. The CK+ dataset, for instance, comprises 593 video sequences from 123 subjects, including 327 specific annotations. For model evaluation, the datasets were systematically partitioned:" The model evaluation scheme further divides the data set into a training set, trial set, and validation set" . To ensure robustness and generalizability of our findings, a K-fold cross-validation process was employed, specifically," tenfold cross-validation is used" . This robust validation strategy involves running the test set multiple times to select the optimal model configuration based on the validation set. The proposed model demonstrated better performance with CK+, achieving an overall accuracy of 98 % in semantic evaluation and emotion labelling.
- **Pre-processing Steps**: Our framework incorporates several critical pre-processing activities to achieve precise and quicker outcomes. These steps are designed to prepare
the real-time visual streams from Consumer Electronics (CE) gadgets for analysis: 1) Image Conversion: The initial step involves converting the captured images to grayscale. (2) Face Detection and Identification: Facial identification and extraction are performed using the Viola-Jones algorithm (also referred to as the Haar Cascade Principle) in conjunction with the KLT Algorithm. This method is crucial for detecting faces within the images. (3) Facial Landmark Detection: Following face detection, 68 recognized points of interest are

identified and mapped onto each detected face. A facial landmarks shape prediction library is utilized to accurately map these landmarks, which are essential for detailed facial analysis and predicting expressions.(4) Image Resizing and Normalization: Detected faces are then extracted from the original frames, and a new frame is created by resizing the detected face to 360 x 360 dimensions. Additionally, normalization of picture dimensions is performed as part of the pre-processing activities to obtain precise outcomes. (5) Feature Extraction: Feature extraction is a fundamental pre-processing step, implemented as part of the three-stage categorization for Automatic Facial Expression Analysis (AFEA) alongside face detection and classification.

- **Model Training Details**: The Deep-Emotion-AI system employs a hybrid of Support Vector Machine (SVM) and Convolutional Neural Networks (CNN) for regression analysis of facial emotions. The CNN layout specifically includes two coalescent layers, two overall pooling layers, and a 256-neuron invisible layer. For enhancing the CK+ datasets expressibility within the software, specific kernel modifications were applied: the 5x5-recovery kernel was substituted with two 3x3-convolution kernels, and the 7x7- recovery kernel was replaced with one 3x3 kernel and one 5x5-convolution kernel. This fine-tuning resulted in an average accuracy of 98 % for semantic emotion labelling, and a detection accuracy of 97.74 % when a gentle threshold method was applied during testing.
- **Addressing Handling of Class Imbalance and Image Augmentation**: While the provided sources do not explicitly detail the use of image augmentation techniques or specific strategies for handling class imbalance, the robust performance achieved (e.g., 97.74 % detection accuracy on CK+ with gentle thresholding and 98 % overall accuracy) suggests the model's effectiveness with the given datasets and methodology. We acknowledge the importance of these considerations for broader applicability and will consider incorporating detailed strategies for image augmentation and explicit handling of class imbalance in future iterations of the research to further enhance replicability and generalisability.

3.1 Detecting and Assessing Feelings

The primary objective of the proposed framework is to develop a mechanism for detecting and assessing human feelings on facial expression analysis using semantic aspects, so that the resulting applications can evaluate the information accumulated. The proposed multidimensional array-based approach saves the image space and discards images after employing Viola-Jones Face detection method as implemented by Kohonen, et al. (2009) with up to 98 % accuracy of human face detection. Firstly, it converts the image to a gray scale. The rules of face-to-face assessment of mental states are based on the rules of face recognition. The entire method of assessment was split into three successive stages. These stages are the identification and location in the complicated environment of the face in the picture, including normalization, the removal by a classifying agent of suitable characteristics depicting the specified face expression and the associated expression. The two metrics offer a quick and easy assessment of the correspondence, but it is important to take into consideration certain factors as well as robustness against noise, tolerance to deformation, lightweight computationally, the efficiency of a corresponding method with distortions, sizes, and guidance differences.

3.2 Face Detection Model and Identification

Signified facial identification and extraction is done by the Viola-Jones algorithm as implemented by Kohonen, etal. (2009). The landmarks are then detected by 68 recognized points of interest for each rectangle that bounds the image. All 68 landmarks are mapped on their faces for their detailed analysis and handle all necessary transformations of a frame to predict the facial expressions. The facial landmarks shape prediction library has been used to map the landmarks of the detected human face. Then the detected faces from the original frame are extracted, and a new frame, resized from the original frame, is created based on the detected face of 360 x 360 dimensions.

3.3 Sub Emotions Taxonomy

This section examines how various participatory environments enhance the capacity of classification to categorize themselves for new faces. From the analysis and studies of James A. Russell, the fuzziness of boundaries between basic emotions and sub-emotions can further be labeled and assembled into sub-categories of emotions, making basic emotions a tuple.

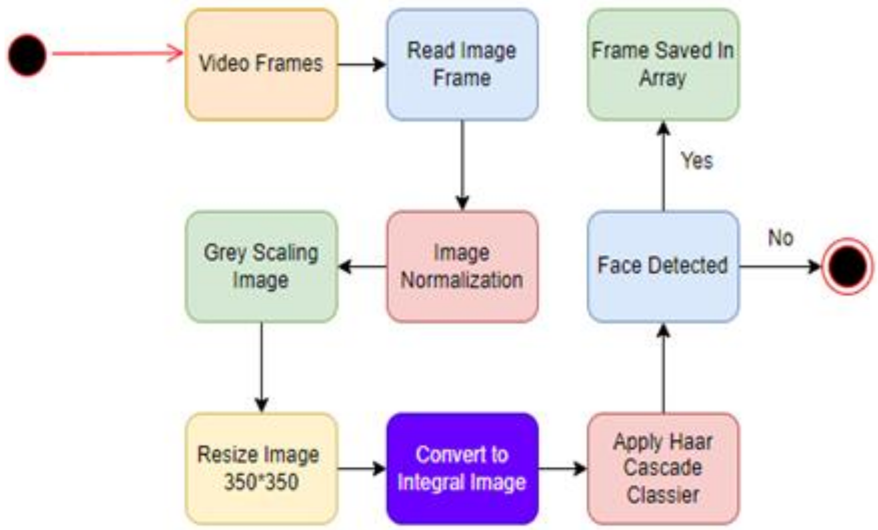

Figure 3: Implementation Flow of the Deep-Emotion-AI Framework for Real-time Learner Emotion Analysis

Our research has taken these steps to categorize an algorithm for its post-processing of basic emotions analysis. The main labels of basic emotions are happy, angry, fear, sad, disgusted, contempt, and neutral. The Algorithm has fully worked on the sub-categories of basic emotions, and the findings are as follows:

- The two maxima of happiness and surprise will give a resultant of delight.
- The two maximum findings of fear and surprise will give you an alert feeling.
- Sleepy is the combination of two basic emotions, sadness and happiness.
- Contempt is the mixture of two emotions, anger and disgust.
- Drowsy is a mix and match of sleepy and sad.
- Unpleasant is made up of sadness and disgust. Relaxed feeling is a combination of sleep and happiness, while trust is the combination of relaxed and happy emotions.

They can be further marked into subcategories, but these emotions are enough to describe the state of being actively or inactively present in the attendees. Supportive proof was acquired by scaling 28 emotion-denoting adjectives in four various respects: Ross' method for the circular order of factors, a complex scaling mechanism centered on presumed resemblance between words, a reductive scaling of hypothetical joy-distress and degrees of arousal, and an assessment of the 343 self-reports by the individuals. It could be said that the community has an inherent concept of an analogy to the intrinsic theory of character or science. The spatial depiction of the emotion by lay people is probably intrinsic in the sense that few, if anyone, could clearly state their full conceptual framework.

Algorithm 1: Emotion Detection Process

```
procedure EmotionDetection
    While classroom session is ongoing do
        Capture Video Frame from CE gadgets
        if Frame captured then
            Save Frame to Disk
            GreyScale Image
            Normalize Image Dimensions
            if Face Detected using Haar Cascade Classifier then
                Extract Facial Features
                Emotion ← Classify Emotion using SVM and CNN hybrid model
                Display Detected Emotion
            else
                Continue capturing
```

```
                    end if
            end if
    end while
end procedure
```

3.4 Training-Sets

The Extended Cohn-Kanade Data set (CK+) [29] is a complete data set for action units and it is specified expression of emotion. The database of Cohn Kanade (CK) was launched in 2000 to promote research on the automatic detection of different facial expressions. CK + data set is a common data set comprising 593 video sequences of 123 items, plus 327 specific annotations, of the Cohn Kanade (CK+) data set. Seven styles of tags are available: dislike, joy, rage, frustration, anxiety, sadness, and shock. The neutral to identified emotions are the source of every series.

4. Experimental results

Machine learning aims to learn patterns that generalize well for unseen data, rather than simply storing the data shown during the training process.

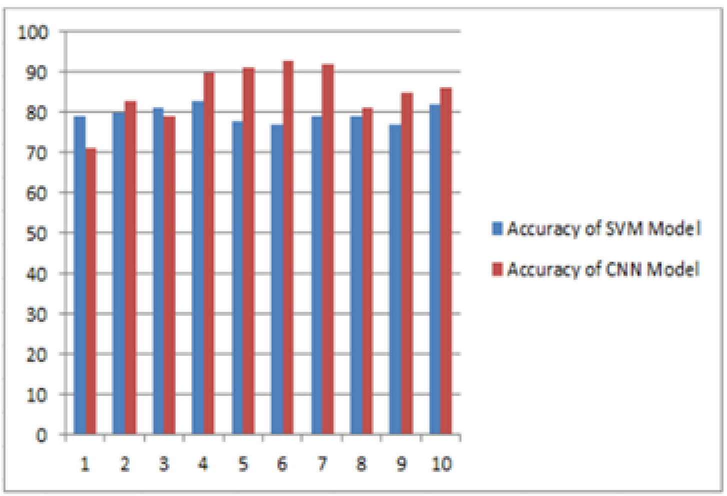

Figure 4: Accuracy of Proposed Framework across 10 folds of Cross Validation

The proposed classification/labeling and outcome optimization show better classification results. Through fine-tuning, the best candidate functionality for classifications was selected. The layout of CNN contains two coalescent layers, 2 overall layers for pooling, and a 256- neuron layer invisible. The model evaluation scheme further divides the data set into a training set, trial set, and validation set, and runs the test set many times to choose the best layout for the validation set during the K-fold cross-validation process. In order to further improve the CK+ expressability of the software, we substitute the 5x5-recovery kernel with two 3x3-convolution kernels in the reference model and remove the 7x7-recovery kernel with one 3x3 kernel and one 5x5-convolution Kernel. Figure 4 visually represents the accuracy of the proposed framework across ten folds of cross-validation. Cross-validation is a method used in machine learning to evaluate the model's performance by dividing the dataset into training, trial, and validation sets, running tests multiple times to select the best layout for the validation set. Specifically, Figure 4 illustrates that the maximum accuracy achieved with Support Vector Machine (SVM) classification is 83.15 % and the maximum accuracy measured by the Convolutional Neural Network (CNN) model is 94.02 %. The study further indicates that the implementation of a gentle threshold significantly improves detection accuracy to 97.74 %, compared to 86.74 % with a hard limit. The Deep-Emotion- AI model, which uses a hybrid of SVM and CNN, achieved an average accuracy of 98 % in semantic emotion labelling. A high accuracy of over 90 % can be obtained using state-of-the-art facial expression recognition approaches on MMI and CK+ data sets. In light of the early samples of pose-invariant facial expressions with limited face occlusion, MMI and CK+, and Google databases have tested the bulk of the approaches to FER. And the FER benchmark is commonly used for these two data sets. FER methods are typically less than 80 % reliable for MMI data sets.

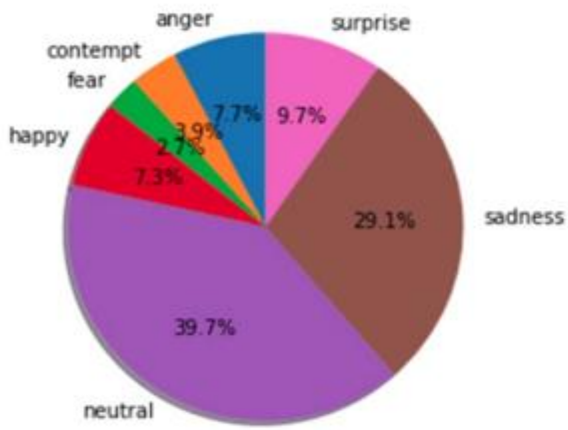

Figure 5: Basic Emotions Calculated in Percentage

The pie chart, as shown in Figure 5, presents the distribution of various emotions, highlighting their relative frequencies. The largest segment, representing Neutral emotions, accounts for 39.7 % of the total. Following closely is Sadness, which comprises 29.1 %. The remaining emotions exhibit smaller proportions, with Anger at 9.7 %, Surprise at 9.0 %, and Contempt at 4.0 %. Additionally, Happy emotions make up 7.3 %, while Fear is the least represented at only 2.7 %. This distribution indicates that Neutral and Sadness are the most prevalent emotions in the dataset, while the other emotions are less frequently observed. The following Table 1 represents the accuracy percentages of different emotion recognition models across various datasets. Each row corresponds to a dataset, and the columns show the accuracy of recognizing specific emotions, such as Happy, Angry, Disgusted, Fear, Surprise, Contempt, and Sad.

Table 1: Emotion Recognition Accuracy by Data Set

Data Sets	Happy	Angry	Disgusted	Fear	Surprise	Contempt	Sad
CK+	96.30%	97.18%	96.00%	100%	91.67%	98.39%	97.41%
MMI	76.67%	79.72%	75.23%	90.92%	82.56%	83.44%	59.23%
Google Dataset	77.65%	72.33%	82.05%	96.32%	79.55%	83.30%	89.80%

5. Conclusion

In current research, a CE based framework has been proposed for evaluating emotions and semantically labeling of emotion sentiments in the learning environment. Emotion AI has been exploited for analysis of Facial Action Units for measuring multiple facial expressions in real time. It improves the amount of facial extraction and leads to elevated precision of identification through a hybrid of deep learning techniques, i.e, SVM classifier and CNN. Proposed approach achieved a competitive accuracy of labeling emotions is between 84% and 98%. While the framework exhibits promising results, certain limitations exist that warrant consideration for future advancements. Although the system is designed to process incoming video streams while consuming minimal resources in terms of bandwidth and latency, efficiently catering to the computational requirements of a larger audience remains a significant area for development and represents a current challenge for widespread real-time scalability. Furthermore, while validated on datasets such as CK+, MMI, and Google, with CK+ showing superior performance, broader dataset diversity and the handling of more varied real-world conditions, including instances of significant face occlusion beyond what is typically found in common benchmark datasets like CK+ and MMI, could be crucial for robust deployment in uncontrolled environments. Despite these considerations, the proposed framework holds substantial educational applications. By dynamically gaining deeper insight into learners' emotional states and cognitive responses through real-time facial visual streams, this system has the potential to effectively engage learners. The real-time detection of emotions enables learning activities to be tweaked to engage learners for effective learning outcomes. This capability can be profoundly leveraged to inform adaptive feedback mechanisms for instructors along with the virtual tutors and facilitate personalized and adaptive recommendations of learning content, thereby maximizing learning engagement and enhancing educational experiences. Potential future endeavors include fine-tuning the developed model to further enhance its performance. The research may also consider the usage of deep learning methods for voice and gesture analysis to provide more precise feedback, aiming to simplify the time complexities involved in the learning environment. Moreover, incorporating gesture analysis for a deeper understanding of both the speaker and audience could further advance the domain of emotion measurement

References:

Nayak, S., Routray, A., & Sarma, M. (2023). GNN-based embedded framework for consumers affect recognition using thermal facial ROIs. IEEE Consumer Electronics Magazine, 12(4), 74-83.

Lindner, J. (2023). Technology in education statistics. Retrieved from https://gitnux. org/technology-in- education-statistics/

Paikrao, P., Mukherjee, A., & Alnumay, W. (2023). Smart emotion recognition framework: A secured IoVT perspective. IEEE Consumer Electronics Magazine, 12(1), 80-86. https://doi.org/10.1109/MCE.2021.3062802

Acheampong, F., Nunoo-Mensah, H., & Chen, W. (2021). Transformer models for textbased emotion detection: A review of BERT-based approaches. Artificial Intelligence Review, 54, 5789–5829.

Sai, S., Goyal, D., Chamola, V., & Sikdar, B. (2024). Consumer electronics technologies for enabling an immersive metaverse experience. IEEE Consumer Electronics Magazine, 13(3), 16-24. https://doi.org/10.1109/MCE.2023.3327530

Abed, E. K. (2019). Electronic learning and its benefits in education. Eurasia Journal of Mathematics, Science and Technology Education, 15(3). https://doi.org/10.29333/ejmste/102668

Gursesli, M., Lombardi, S., Duradoni, M., et al. (2024). Facial emotion recognition (FER) through custom lightweight CNN model: Performance evaluation in public datasets. IEEE Access, 12, 45543-45559. https://doi.org/10.1109/ACCESS.2024.3380847

Hwang, P., Hsu, C., & Wang, W. (2020). Development of a mimic robot−Learning from demonstration incorporating object detection and multi-action recognition. IEEE Consumer Electronics Magazine, 9(3), 79-87. https://doi.org/10.1109/MCE.2019.2956202

Majumder, A., Behera, L., & Venkatesh, S. (2016). Automatic facial expression recognition system using deep network-based data fusion. IEEE Transactions on Cybernetics. https://doi.org/10.1109/TCYB.2016.2625419

Clark, E. A., Kessinger, J., & Duncan, S. E. (2020). The facial action coding system for characterization of human affective response to consumer product-based stimuli: A systematic review. Frontiers in Psychology, 11, 920. https://doi.org/10.3389/fpsyg.2020.00920

Said, A., F´elix-Herr´an, L., Daviz´on, Y., et al. (2022). An active learning didactic proposal with human-computer interaction in engineering education: A direct current motor case study. MDPI Electronics, 11(7), 1059.

Westera, W., Bahreini, K., & Nadolski, R. (2014). Towards multimodal emotion recognition in e-learning environments. Interactive Learning Environments. https://doi.org/10.1080/10494820.2014.908927

Tao, X., Liu, Y., & Jiang, C. (2019). QoE-oriented multimedia assessment: A facial expression recognition approach. IEEE MultiMedia, 26(1), 41-50. https://doi.org/10.1109/MMUL.2018.2879596

Corcoran, P., Costacke, C., Varkarakis, V., et al. (2020). Deep learning for consumer devices and services 3−Getting more from your datasets with data augmentation. IEEE Consumer Electronics Magazine, 9(3), 48-54. https://doi.org/10.1109/MCE.2019.2959062

Prikler, F., & Magdin, M. A. (2017). Real-time facial expression recognition using webcam and SDK. International Journal of Interactive Multimedia and Artificial Intelligence. https://doi.org/10.9781/ijimai.2017.11.002

Pantic, M., & Rothkrantz, L. J. (2004). Facial action recognition for facial expression analysis from static face images. Technical Report, Delft University of Technology, Netherlands.

Miji´c, D., Vranje˘s, M., Grbi´c, R., et al. (2021). Autonomous driving solution based on traffic sign detection. IEEE Consumer Electronics Magazine, 12(5), 39-44. https://doi.org/10.1109/MCE.2021.3090950

Moore, S., & Bowden, R. (2011). Local binary patterns for multi-view facial expression recognition. Computer Vision and Image Understanding, 115(3), 110-112.

Russell, J. A., Posner, J., & Peterson, B. S. (2005). The circumplex model of affect. In An integrative approach to affective neuroscience, cognitive development, and psychopathology (pp. 200-205). *Development and Psychopathology*.

Greene, S., Thapliyal, H., & Caban-Holt, A. (2016). A survey of affective computing for stress detection: Evaluating technologies in stress detection for better health. IEEE Consumer Electronics Magazine, 5(4), 44-56. https://doi.org/10.1109/MCE.2016.2590178

Ranzato, M. A., Taigman, Y., Yang, M., & Wolf, L. (2014). DeepFace: Closing the gap to human-level performance in face verification. In IEEE Conference on Computer Vision and Pattern Recognition (pp. 1701–1708).

Rejeb, S., Duveau, C., & Rebafka, T. (2022). Self-organizing maps for exploration of partially observed data and imputation of missing values. https://doi.org/10.48550/arXiv.2202.07963

Kohonen, T., Tiropanis, H., Davis, D., & Millard, D. (2009). Semantic technologies for learning and teaching in the web 2.0 era. IEEE Intelligent Systems, 24(6), 49-53.

Prattic`o, G., Di Cosmo, D., Piviotti, M., et al. (2024). Effectiveness of computer-based e-learning and virtual reality training system for the preparatory phase of a training on maintenance procedure. In IEEE International Conference on Consumer Electronics (ICCE) (pp. 1-6). https://doi.org/10.1109/ICCE59016.2024.10444354

Prist, M., et al. (2020). Machine learning-as-a-service for consumer electronics fault diagnosis: A comparison between Matlab and Azure ML. In IEEE International Conference on Consumer Electronics (ICCE) (pp. 1-5). https://doi.org/10.1109/ICCE46568.2020.9043014

Hassouneh, A., Mutawa, A. M., & Murugappan, M. (2020). Development of a real-time emotion recognition system using facial expressions and EEG based on machine learning and deep neural network methods. Information in Medicine. https://doi.org/10.1016/j.imu.2020.100372

Nikolaidis, N., Kotsia, P., & Pitas, I. (2007). Facial expression recognition in videos using a novel multi-class support vector machines variant. In IEEE International Conference on Acoustics, Speech and Signal Processing.

Rosales-Perez, A., Garcia, S., & Terashima-Marin, H. (2018). MC2ESVM: Multiclass classification based on cooperative evolution of support vector machines. IEEE Computational Intelligence Magazine, 13(2), 18-29. https://doi.org/10.1109/MCI.2018.2806997

Silveira, N., Ara´ujo, P., & Beu, L. (2016). IMS-based infrastructure for e-learning with a classroom efficiency comparison. In IEEE International Symposium on Consumer Electronics (ISCE) (pp. 109-110). https://doi.org/10.1109/ISCE.2016.7797394

Kanade, T., Saragih, J., & Ambadar, Z. (2010). The extended Cohn-Kanade dataset (CK+): A complete dataset for action unit and emotion specified expression. In IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops.