Management Science and Information Technology

MSIT

2024, Volume 1, Issue 1, pp. 20-30
ISSN-Print: 3079-1006
ISSN-Online: 3079-1014
https://msit.refconf.com

Human-Computer Interaction for Affective State Detection Using Facial Expression

Ebenezer Koukoyi [a*], Li Xiaoxia [a], Andrews Dodzi Kobla Dzikunu [b], Dablu Etse Bobobee [c,d] and Benjamin Odoi-Lartey [e]

[a] *School of Information Engineering, Southwest University of Science and Technology, Mianyang 621010, China*
[b] *Zhejiang Normal University, P. R. China*
[c] *Electric Power College, Inner Mongolia University of Technology, Inner Mongolia 010000, China*
[d] *Smart Energy Storage Institute, Inner Mongolia 010000, China*
[e] *Kwame Nkrumah University of Science and Technology, Kumasi, Ghana*

Abstract

If computers could understand and respond to clients' nonverbal communication, human-computer communication (HCI) would improve in a friendly and non-intrusive way. Looks and feelings are usually seen to have a link, according to research. Inactive improvements, like watching videos, are often used to acknowledge emotional states. This paper examines emotional state acknowledgement using dynamic improvements, such as client looks when they attempt electronic assignments, especially across common computer frameworks according to programmatic experience, proposing a structure joining face effective recognition (FER) method with various stages. An information collection investigation is introduced to collect data from regular users while they complete a set of tasks based on programming. A cutting-edge AI approach for look-based full-of-feeling state recognition with Euclidean distance-based component portrayal and a modified encoding for clients' self-detailed emotional states is proposed and implemented for the accomplishments of this work. The results confirm the suggested method's effectiveness, efficiency, and robustness.

Keywords: Facial Expression; Human-Computer Interaction; Hierarchical Machine Learning; Face Effective Recognition; Man-Machine Communication; Articulation.

1. Introduction

The most effective method of nonverbal communication is facial expression. The way someone expresses themselves may provide insight into their level of excitement, anguish, sharpness, character, social connection, and physiological signals. It is challenging to do a programmed gaze inspection since the articulations are predominantly mixed and subject-dependent(Spaniol, Wehrle, Janz, Vogeley, & Grice, 2024; Yamashita, Takimoto, Oishi, & Kumada, 2024). People pay attention to not only what each other says but also how they seem, how they conduct themselves, and how they move their heads while they are working together or communicating. Expression recognition in programming is difficult, but it has a wide range of uses and is attracting increasing interest(Kessous, Castellano, & Caridakis, 2009; Yan et al., 2024). Looks are the primary mode of communication among humans. According to Vinciarelli, Pantic, and Bourlard (2009), shortly, man-machine communication and emotional figuring will include the capacity to identify human emotions(Govindaraju & Thangam, 2024). According to (Jarosz et al., 2021), recognizing appearances is an interdisciplinary challenge that can lead to touch design in HCI. The field of brain research meets the administration of images and videos(Mancuso, Borghesi, Bruni, Pedroli, & Cipresso, 2024). A connection exists between feelings and influence(Key & Brown, 2024), but they are not the same thing. Since these two concepts refer

*Corresponding author email address: ebenkouk@mails.swust.edu.cn
DOI: 10.22034/ISS.2024.7561.1015

to circumstances and the outcomes of logical processes, they are typically interchangeable. Because of its closeness, this artwork can impact feelings in a two-way fashion.

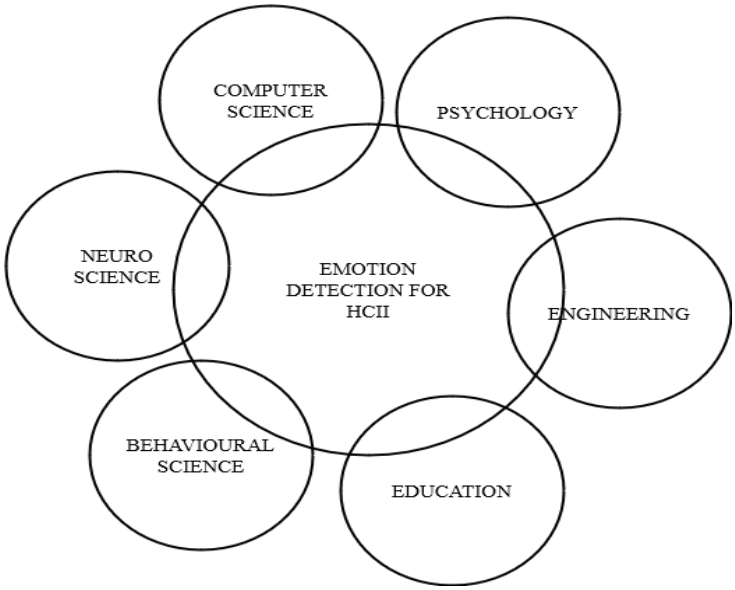

Figure 1. Emotion Detection of the HCI

The user interface (UI) is the primary concentration of HCI. The Human-computer interaction interface (HCII) is an active subfield in software engineering(Abdat, Maaoui, & Pruski, 2011; Annela, 2023). The UI should be created with a clear grasp of its users and the demands that they have (Nayak et al., (2021). This is because humans and computers do not speak the same language. The planning cycle of a UI decreases the amount of input required from customers to communicate and comprehend the outcomes. This upgrade is seen in both the UI and the client/user experience (UX) during collaborative work. HCI interfaces need to be able to recognize unpretentious client behavior, especially intense emotions(Altameem & Altameem, 2020; Chen, Dey, Wang, Bi, & Liu, 2024), and make connections based on this information rather than just reacting to requests from clients. For instance, a sensitive learning environment that identifies and responds to the discontent of students is anticipated to both inspire and increase learning (Giroux et al., 2021). Human-to-human collaboration, which established the framework for the field(Rathore & Gautam, 2024), enabled the programming of computers to understand passionate articulations for HCII Altameem and Altameem (2020). In another way, this contributes to the body of research on computerized look recognition. Several FER methods have been proposed to manage appearance data in facial portrayals(Khan, 2022). These methods make use of computational reasoning and deep learning, which results in improved accuracy and performance in comparison to conventional approaches. This paper is structured as follows; Section 1 is the introduction to the topic, Human-Computer Interaction for Affective State Detection Using Facial Expression. Section 2 discusses previous research and related work. Section 3 presents the proposed methodology and its implementation. Section 4 discusses the results and the final part is the conclusion which gives a summary of the research work.

2. Related Work

Facial expressions have long been recognized as a vital channel for conveying human emotions, intentions, and physical states(Achour-Benallegue, Pelletier, Kaminski, & Kawabata, 2024; Tipirneni & Leal, 2023). Automated facial expression analysis has consequently become a focal point of research in the field of affective computing(Dubey, 2020; Giroux et al., 2021), which explores the detection, synthesis, and understanding of emotional states through computational means(Saganowski et al., 2022). Inferring an individual's affective state by leveraging their facial expressions holds significant potential for enhancing human-computer interaction, as it would enable systems to adapt their behaviour and better accommodate the user's needs and preferences(Praneesh, 2024). Indeed, a growing body of research has explored the feasibility of utilizing computer vision techniques to automatically recognize human emotions based on facial cues. One key challenge in this domain, however, is the limited availability of standardized

datasets that capture authentic, real-world emotional expressions, as the majority of existing datasets tend to feature posed or exaggerate facial displays(Chu, 2023; Han et al., 2023).

The term "HCI" refers to the exchange of data between humans and machines. Alongside the development of deep learning technology, HCI technology has evolved. To collect facial image data, authors B. Li and Lima (2021) utilized real-time face images in addition to video. These models are ineffective for detecting faces in real time because they have such a large number of features(Hussain et al., 2022). For face identification, the author (McDonnell et al. 2021) proposed a grammatical evolution (GE) region-based fully conventional neural network (R-FCNN) that adapts hyper-parameters utilizing GE heuristics. The authors suggest a face identification and tracking system that is powered by deep learning and that makes use of the Regression Network-based Face Tracking model. According to Xiao and Hu (2021), the first layer of CNN's feature-based tuning chain helps to prevent facial and textural variations that might lead to incorrect detection. The authors (Yu et al. 2020) proposed a method for feature extraction that is based on the deep residual network ResNet-50. This method mixes convolutional neural networks with deep residual networks to recognize facial emotions. The author ranked the different blend shape expressions in the first experiment that was published on the perceptibility of facial action units at various activation levels Kumar and Shafi, (2021). In (J. Li et al. 2020), the authors proposed a feature similarity-based FEC network. The model accounts for both the ambiguity and occlusion of labeling.

According to (Giroux et al., 2021), understanding facial expressions and overall facial appearance can be aided by local informative dynamics amid ambiguous expressions. It was proposed that fundamental emotion intensities may be predicted using emotion distribution learning (EDL) that was based on surface electromyography (sEMG). Depressor supercilii, zygomaticus major, frontalis medial, and depressor angulioris muscles exhibited sEMG signals (Satyanarayana et al., 2021). Children can transmit their feelings in social settings through the use of facial expressions (McDonnell et al., 2021). The author investigated the facial expressions of emotion in children who were born with hearing impairment. To improve network speed, LBP features first extract the texture of the image before looking for very tiny face motions. The neural network may be trained to pay attention to aspects that are helpful thanks to an attention mechanism. According to (Cowie et al., 2008), the attention model can be improved by combining LBP properties with an attention mechanism.

Facial expression recognition for affective state detection has already found applications in various HCI domains. E-learning systems that adapt to a learner's emotional state can enhance engagement and retention(Gumelar, Wulandari, Lestari, & Ruswandi, 2024). Similarly, in healthcare, affective computing systems can assist in diagnosing and treating mental health conditions by tracking emotional states over time(Sneha & Raza, 2024). In gaming and virtual reality environments, facial expression analysis allows for more immersive and interactive experiences by adapting the game based on the player's emotions(Sumi & Sato, 2022).

3. Proposed Methodology

Initially, the input face image is taken and undergoes preprocessing followed by a normalization process to diminish the impedance of face-like objects behind the scenes. After preprocessing, the region of interest is selected for extracting the relevant features extracted for the detection of the face. Subsequently, the CNN model operates over the extracted features to detect the output image.

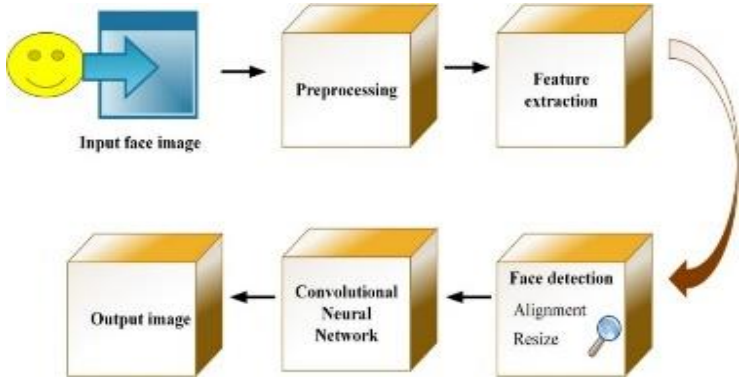

Figure 2. Proposed framework

3.1 Image Preprocessing Based on Interface

An interface is utilized, which incorporates calculations for facial component following, head present assessment, facial trait location, and so on, as the apparatus of preprocessing. Interfaces can likewise be utilized to recognize various faces simultaneously(Hossain, E-Shan, & Kabir, 2021; Juliandy, Ng, & Darwin, 2024). The critical provisions of every face can be perceived adequately, and the look can be identified by rectangular layouts likewise; The rectangle layouts are enlarged by 1.15 times to cover more facial substance, preventing facial data overlooking while reducing foundation turmoil. The detailed calculations are shown in Equations 1-9.

$$I_{preprocessed}(x) = \frac{1}{w_p}\sum_{x_k \epsilon \Omega} I(x_k) F_r(\|I(x_k) - I(x)\|)_{G_s}(\|x_k - x\|) \tag{1}$$

The independent component is separated to facilitate processing.

$$\begin{aligned} z(a,b) &= \ln f(a,b) = \ln(i(a,b)r(a,b)) \\ z(a,b) &= \ln i(a,b) + \ln r(a,b) \end{aligned} \tag{2}$$

Processing $z(a,b)$ using a homomorphic filter function *h*,

$$s(a,b) = h * z(a,b) = h * \ln i(a,b) + h * \ln r(a,b) \tag{3}$$

According to the convolution theorem,

$$S(c,d) = H(c,d)Z(c,d) = H(c,d)F_i(c,d) + H(c,d)F_r(c,d) \tag{4}$$

where $S(c,d), H(c,d), Z(c,d)$ are the transform of $s(a,b), h(a,b), z(a,b)$

Inverse transform of $S(c,d)$ to compute $s(a,b)$ is given as

$$s(c,d) = \Im^{-1}\{S(c,d)\} \tag{5}$$

$$s(c,d) = \Im^{-1}\{H(c,d)F_i(c,d)\} + \Im^{-1}\{H(c,d)F_r(c,d) \tag{6}$$

where $\Im^{-1}$ is an inverse operator of the homomorphic filter

$$\begin{aligned} i'(a,b) &= \Im^{-1}\{H(c,d)F_I(c,d)\} \\ r'(a,b) &= \Im^{-1}\{H(c,d)F_r(c,d) \end{aligned} \tag{7}$$

Then the inverse transform of the homomorphic filter is

$$s(a,b) = i'(a,b) + r'(a,b) \tag{8}$$

The enhanced image after filtering using a homomorphic filter is $g(a,b)$ which is represented as

$$g(a,b) = e^{s(a,b)} = e^{i'(a,b)}e^{r'(a,b)} = i_0(a,b)r_0(a,b) \tag{9}$$

Where $i_0(a,b) = e^{i'(a,b)}$ and $r_0 = e^{r'(a,b)}$ are the illumination and reflectance components of the fundus image after filtering?

3.2 Feature Extraction and Distance-based Representation

Mathematically related methods for investigation of the expression depend upon finding the facial focuses and deciding the area and state of related facial parts. The investigation introduced here utilized the "Chehra" instrument to remove the area of facial focus (Ramakrishnan et al. 2020). In any case, this methodology is not sufficiently thorough to allow acknowledgment of what was not given in the information (Cr and Mp 2014). This is because of

the way that the heavenly body of these focuses shifts among the horde of facial shapes that include diverse facial morphologies Thusly, resultant provisions are addressed by tracking down the Euclidean distances among all facial milestone focuses. Thus, the look is at last addressed as an 1176-measurement including vector, coming about because of 49 cartesian organized blends.

3.3 The Learning Model Based on CNN

Convolutional neural networks, often known as CNNs, are one of the most popular types of deep learning algorithms that are used for voice and image recognition. According to Cowie et al. (2008), this suggests that the stimulus is only capable of activating the neurons in a specific location. According to Qu et al. (2018), the fundamental processes of pooling consist of first dividing the feature images into several cells that range in size from 2*2 to 4*4 and then computing the output of each cell. In this investigation, the photos that need to be identified are initially converted into a grayscale format with a dimension of 28 by 28. Following this step, the convolution kernel is constructed, and then the proper activation function is selected. This solution makes use of a 10-layer convolutional neural network, as can be seen in Figure 3. Within that network are five convolution layers and two pooling layers. Convolution kernels with a size of 16 by 3 are employed to acquire the features of the input images. After that, 16 feature images with a resolution of 28 x 28 are obtained. These images are then convolved using 32 convolution kernels of 3 x 32, which leads to the acquisition of 32 features that have been extracted across the two layers (Minaee, Minaei, and Abdolrashidi 2021).

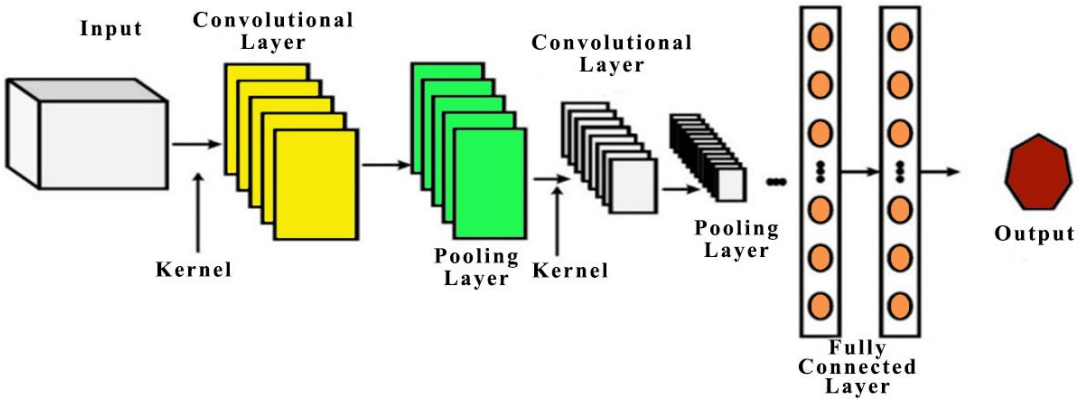

Figure 3. Structure of the CNN model

The third layer uses 64 convolution kernels and 64 local features of 28*28 size. The fourth layer uses max pooling to reduce computational complexity, resulting in 64 feature maps with a size of 14*14 from the first pooling layer and 64 convolution kernels. The output of the fourth convolution layer is max pooled to produce 64 feature maps with a 7*7 size, and a subsequent convolution layer continues to produce 128 feature maps with the same size(Nayak, Nagesh, Routray, & Sarma, 2021). An applied CNN-based massive information model and its earlier demonstration in FER over the competition have already been proven(Gursesli et al., 2024; Mozaffari, Brekke, Gajaruban, Purba, & Zhang, 2023). After a 32-element convolutional layer, there are 3 squares: 2 convolutional layers and 1 max-pooling layer with 62 component maps each. The information image is compressed to a quarter as a result of the piece size of the first convolutional layer being set to 2 2, the second to 4 4, and the maximum pooling layers both having a bit of size 1 and step 2. In three fully interconnected layers of 2148 and 1124 cells, Rectified Linear Units (ReLUs) actuate. The network learns residually. H(y) is a mapping to fit utilizing stacked network layers, where y represents layer inputs. H's residual function is

$$F(y) \coloneqq H(y) - y \tag{10}$$

To reduce the dimensionality, a dropout is introduced after each of the two entirely associated layers, which will supply a chunk of neurons according to the presetting drop-likelihood; the two attributes are both set to 0.5 in this article. Softmax is used to organize the expressions examined as far as indignation, scorn, terror, pleasure, compassion, surprise, hate, and neutrality in the following yield layer, which is made up of 8 units.

The algorithm for training the data in CNN is shown below.

Algorithm 1: Training data in Convolutional Neural Network

```
for each data m = 0, ...M − 1 do
Initialize ΔW = 0
Get the m^th mini-batch, Dm.
    for each layer l = 0, ...L − 1
      do
        {
          Calculate activations Al based on Dm
        }
   for each layer l = L − 1, ...0
     do
       {
        Calculate errors El
       }
   for each layer l = 0, ...L − 1
     do
     {
      Calculate weight gradients ΔWl
     }
 Update parameters, Wl and Bl
```

3.4 Performance Evaluation of the proposed method

The performance of the proposed method is evaluated based on accuracy, precision, recall, and F1-score.

Accuracy

Accuracy measures how accurately the system model operates. Generally, it is the proportion of correctly predicted observations to all observations. Accuracy is uttered in Eqn. (11),

$$Accuracy = \frac{T_{Pos} + T_{Neg}}{T_{Pos} + T_{Neg} + F_{Pos} + F_{Neg}} \tag{11}$$

Precision

Precision is estimated as the number of correct positive estimates alienated by the overall positive estimates. It is the fraction of precise diagnosis of the affected region to be cancer that is computed utilizing Eqn. (12),

$$P = \frac{T_{Pos}}{T_{Pos} + F_{Pos}} \tag{12}$$

Recall

Recall is defined as the ratio of the entire true positives and false negatives to the right positive forecasting accuracy. It states what percentage of forecasts were accurate in their diagnosis of cancer that is represented in Eqn. (13)

$$R = \frac{T_{Pos}}{T_{Pos} + F_{Neg}} \tag{13}$$

F1-Score

The F1-score measurement combines precision and recall. Precision and recall are used to calculate the F1-score measure that is symbolized in Eqn. (14),

$$F1-score=\frac{2\times precision\times recall}{precision+recall} \tag{14}$$

4. Results and Discussion

The CNN model is fed images and information to assess the system's viability. The difference is that a face may have components of different appearances, and the demeanor introduced on this face will be named by the most probable demeanor chosen by these elements, but the general demeanor of an image with multiple countenances is chosen by the number of different demeanor highlights in each face. Using the approved datasets CK-8 and KDEF, HPBCNN with a distance-based component was used to make two models. Each future model will be able to group the video edges of the accounts gathered during computer-based tasks at a rate of one case per second. Table 1 shows six fundamental emotion recognition confusion matrices. The bigger the difference between specificity and generality, the more accurate the results. Figure 4 shows the ROC curves for all six moods.

Table 1. Confusion matrix

Factor	Happy	Sad	Fear	Disgust	Surprise	Anger
Happy	**0.94**	0.11	0.21	0.11	0.11	0.21
Sad	0.11	**0.95**	0.11	0.21	0.11	0.11
Fear	0.31	0.11	**0.93**	0.11	0.21	0.11
Disgust	0.11	0.11	0.21	**0.91**	0.21	0.21
Surprise	0.21	0.11	0.21	0.11	**0.95**	0.11
Anger	0.21	0.11	0.21	0.21	0.11	**0.93**

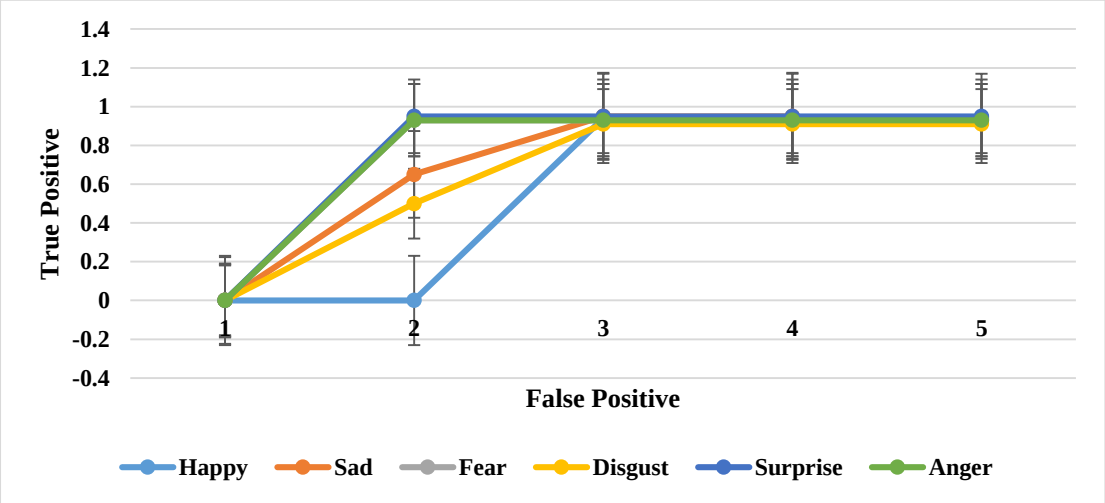

Figure 4. ROC Curve

Figure 5 presents the articulation rates utilizing an arrangement model that is prepared on the CK-8 dataset. Additionally, Figure 6 presents the articulation rates utilizing a prepared classifier utilizing the KDEF dataset. Various rates for every articulation have been found for the two models. Notwithstanding, one can see these rates distinctively

by considering the way that a few articulations are substantially more precisely perceived than others. For the most part, distinguishing states, for example, gladness and shock, is equivalently more prevalent than recognizing different states like disdain, impartiality, dread, fury, misery, and nausea, which is perhaps because of the similitude in the mathematical state of those articulations.

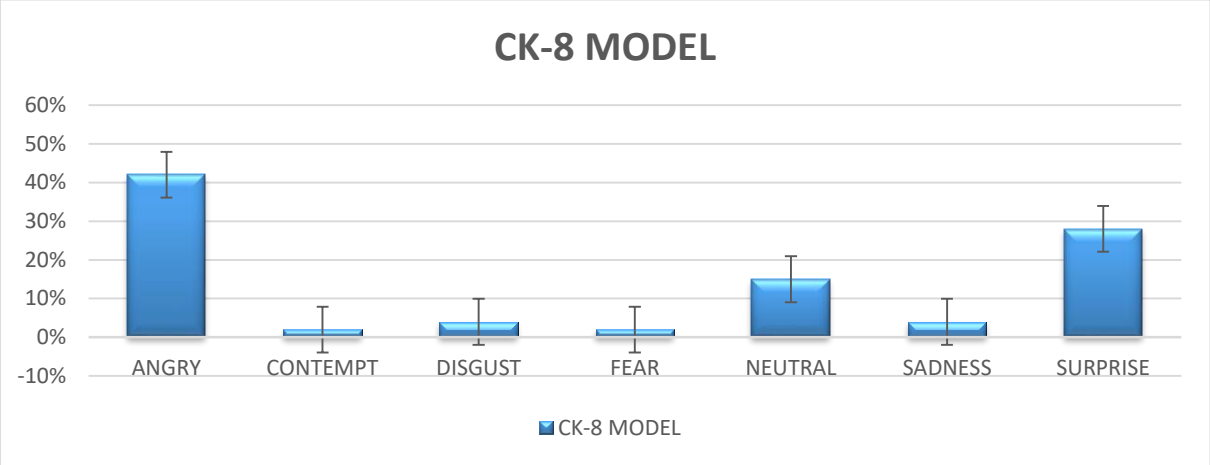

Figure 5. Articulation rates utilizing the proposed model in the CK-8 dataset

The work also highlighted this type of collection by establishing the level of connected articulations and indicating where they are low. Thus, the negative state can be addressed by combining the CK-8 dataset's articulation marks of anger, hatred, disdain, dread, and bitterness. The KDEF dataset names anxious, furious, scorn, and tragedy can be combined to address the negative state. From the circumplex model, these articulations tend to place bad names on the negative side of the charming-horrible continuum, such as the valence hub, as seen in Fig 5. CK-8 outcomes were also affected by the negative state gathering. The resultant rates were obtained by averaging the yields of the two prepared models from then on. The outcomes also provide rates of each articulation from recordings made during each assignment, using the normal rates from both the CK-8 and KDEF-prepared models.

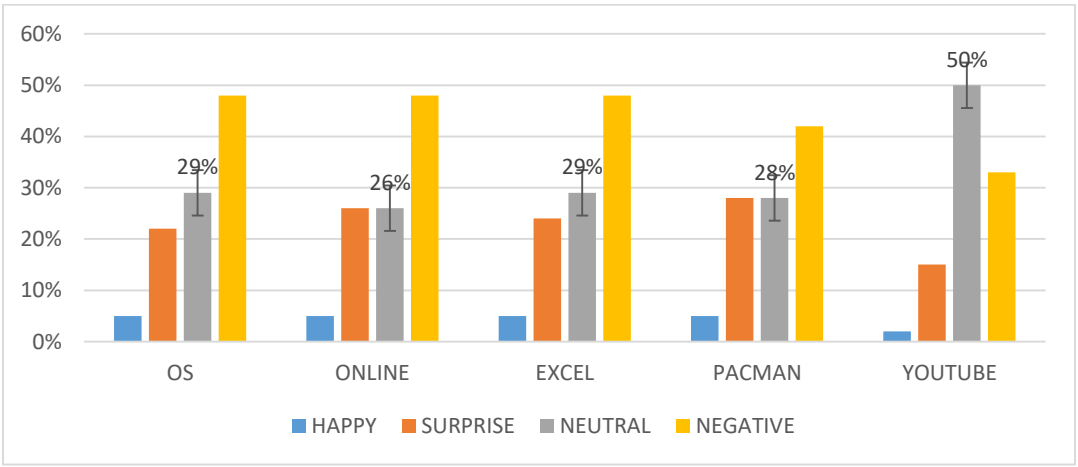

Figure 6. Articulation rates utilizing the proposed model in KDEF dataset

It is seen that impartial and negative articulations possess the most elevated rates across the various duties, including dynamic dialogue by members. Paradoxically, the latent communication YouTube task shows more impartial articulation. Although all assignments have some happy articulation, the YouTube work has the least. Thus, dynamic collaboration projects may have higher participant diversity than uninvolved communication projects. As we'll see later in this paper, these assertions may not reflect members' feelings. Averaging the results from each database yields precision, recall, and F-measure statistics. Comparing past results requires comparable methods. These equations

explain precision, recall, and F-measure values. Table 2 shows that face images have an eventual accuracy of 92, 22.57% higher than original photos, and a maximum accuracy of 95%. Table 2 compares our research's precision, recall, and F-measure values to past techniques.

Table 2. Maximum Accuracy on Face Image and Original Image

Input	Face (%)	Original (%)
Maximum	95	70
Eventual	92	69

The suggested method had the highest accuracy of 98% for recognizing facial emotions when contrasted to all the previous works.

Conclusion

This study creates a model to categorize emotions based on how they seem from the perspective of a virtual encounter. The model is developed using stages and a smaller learning model reliant on the design of CNN. The feelings were classified in the proposed system as follows: wrath, scorn, dread, bliss, difficulty, shock, hatred, and impartiality. An authentic image that includes the faces of all of the photographs taken at the same time is obtained. This is done so that the suitability of this structure may be evaluated in a real-world environment. By providing this image to the CNN model that is being used, the labels of enthusiasm for each significant face are obtained, and the overall sentiment that is associated with them is captured. It has been established that the structure has a great deal of materiality and ineffective activities and that it plays a positive role in assuming responsibility for addressing the problems. Calculations that have better execution and more limited activity time, including preprocessing and learning models, will be consistently developed after some time, according to the point of view of developments. This improvement will come about as a result of the evolution of the virtual experience, which will enhance For example, image preprocessing includes face identification, arrangement, revolution, and resizing; however, when dealing with issues such as backdrop illumination, shadows, and facial inadequacy brought about by complex conditions, these current techniques are consistently weak, and these deficiencies may be addressed at a later time. Even if the proposed model performs better than the CNN model, models with higher-order accuracy and greater learning capacity will eventually replace it. To ensure that the system will continue to be effective over a longer period, it has to be modified and kept updated regularly, and it also needs to be updated with newly created calculations and technological advancements.

Looking ahead, integrating facial expression recognition into wearable technology and Internet of Things (IoT) systems opens new possibilities for ubiquitous emotion-aware computing. Furthermore, the ethical implications of emotion detection—such as privacy concerns, data security, and the potential for emotional manipulation—are critical areas for future research and regulation

References

Abdat, F., Maaoui, C., & Pruski, A. (2011, 16-18 Nov. 2011). *Human-Computer Interaction Using Emotion Recognition from Facial Expression.* Paper presented at the 2011 UKSim 5th European Symposium on Computer Modeling and Simulation.

Achour-Benallegue, A., Pelletier, J., Kaminski, G., & Kawabata, H. (2024). Facial icons as indexes of emotions and intentions. *Frontiers in Psychology, 15*, 1-13. doi:https://doi.org/10.3389/fpsyg.2024.1356237

Altameem, T., & Altameem, A. (2020). Facial expression recognition using human machine interaction and multi-modal visualization analysis for healthcare applications. *Image and Vision Computing, 103*, 1-19. doi:https://doi.org/10.1016/j.imavis.2020.104044

Annela, M. (2023). Human Computer Interaction. 1-4. doi:https://www.researchgate.net/publication/379693822

Chen, J., Dey, S., Wang, L., Bi, N., & Liu, P. (2024). Attention-Based Multi-Modal Multi-View Fusion Approach for Driver Facial Expression Recognition. *IEEE Access, PP*, 1-20. doi:https://doi.org/10.1109/ACCESS.2024.3462352

Chu, Z. (2023). Facial expression recognition for a seven-class small and medium-sized dataset based on transfer learning CNNs. *Applied and Computational Engineering, 4*, 1-6. doi:https://doi.org/10.54254/2755-2721/4/2023394

Dubey, A. K., and Vanita Jain. (2020). Automatic Facial Recognition Using VGG16 Based Transfer Learning Model. *Journal of Information and Optimization Sciences*, 1-9. doi:https://doi.org/10.1080/02522667.2020.1809126

Giroux, F., Léger, P.-M., Brieugne, D., Courtemanche, F., Bouvier, F., Chen, S. L., . . . Senecal, S. (2021). Guidelines for Collecting Automatic Facial Expression Detection Data Synchronized with a Dynamic Stimulus in Remote Moderated User Tests. In (pp. 1-13).

Govindaraju, D., & Thangam, D. (2024). Emotion Recognition in Human-Machine Interaction and a Review in Interpersonal Communication Perspective. In (pp. 1-16).

Gumelar, W. S., Wulandari, S. F., Lestari, T. S., & Ruswandi, R. (2024). The Correlation Between Teachers' Emotional Intelligence and Students' Learning Engagement in EFL Class. *JEELS (Journal of English Education and Linguistics Studies), 11*, 1-25. doi:https://doi.org/10.30762/jeels.v11i2.3377

Gursesli, M., Lombardi, S., Duradoni, M., Bocchi, L., Guazzini, A., & lanatà, A. (2024). Facial Emotion Recognition (FER) Through Custom Lightweight CNN Model: Performance Evaluation in Public Datasets. *IEEE Access, PP*, 1-17. doi:https://doi.org/10.1109/ACCESS.2024.3380847

Han, S., Guo, Y., Zhou, X., Huang, J., Shen, L., & Luo, Y. (2023). A Chinese Face Dataset with Dynamic Expressions and Diverse Ages Synthesized by Deep Learning. *Scientific Data, 10*, 1-9. doi:https://doi.org/10.1038/s41597-023-02701-2

Hossain, M., E-Shan, S., & Kabir, H. (2021). *An Efficient Way to Recognize Faces Using Mean Embeddings*.

Hussain, T., Hussain, D., Hussain, I., Alsalman, H., Hussain, S., Sajid, S., & Al-Hadhrami, S. (2022). Internet of Things with Deep Learning-Based Face Recognition Approach for Authentication in Control Medical Systems. *Computational and Mathematical Methods in Medicine, 2022*, 1-17. doi:https://doi.org/10.1155/2022/5137513

Juliandy, C., Ng, P. W., & Darwin. (2024). Modeling Face Detection Application Using Convolutional Neural Network and Face-API for Effective and Efficient Online Attendance Tracking. *Jurnal Online Informatika, 9*, 1-8. doi:https://doi.org/10.15575/join.v9i1.1203

Kessous, L., Castellano, G., & Caridakis, G. (2009). Multimodal Emotion Recognition in Speech-based Interaction Using Facial Expression, Body Gesture and Acoustic Analysis. *Journal on Multimodal User Interfaces, 3*, 1-16. doi:https://doi.org/10.1007/s12193-009-0025-5

Key, B., & Brown, D. (2024). Making sense of feelings. *Neuroscience of Consciousness, 2024*, 1-9. doi:https://doi.org/10.1093/nc/niae034

Khan, A. (2022). Facial Emotion Recognition Using Conventional Machine Learning and Deep Learning Methods: Current Achievements, Analysis and Remaining Challenges. *Information, 13*, 1-17. doi:https://doi.org/10.3390/info13060268

Mancuso, V., Borghesi, F., Bruni, F., Pedroli, E., & Cipresso, P. (2024). Mapping the landscape of research on 360-degree videos and images: a network and cluster analysis. *Virtual Reality, 28*, 1-19. doi:https://doi.org/10.1007/s10055-024-01002-2

Mozaffari, L., Brekke, M., Gajaruban, B., Purba, D., & Zhang, J. (2023). *Facial Expression Recognition Using Deep Neural Network.*

Nayak, S., Nagesh, B., Routray, A., & Sarma, M. (2021). A Human–Computer Interaction framework for emotion recognition through time-series thermal video sequences. *Computers & Electrical Engineering, 93*, 107280. doi:https://doi.org/10.1016/j.compeleceng.2021.107280

Praneesh, M. (2024). Visual Emotion Recognition Through Affective Computing. In (pp. 147-162): SpringerLink.

Rathore, D., & Gautam, P. (2024). UTILIZING MACHINE LEARNING TECHNIQUES TO IDENTIFY EMOTIONAL CORRELATES IN PHRASE ARTICULATION. *ShodhKosh: Journal of Visual and Performing Arts, 5*, 1-8. doi:https://doi.org/10.29121/shodhkosh.v5.i5.2024.2107

Saganowski, S., Komoszyńska, J., Behnke, M., Perz, B., Kunc, D., Klich, B., . . . Kazienko, P. (2022). Emognition dataset: emotion recognition with self-reports, facial expressions, and physiology using wearables. *Scientific Data, 9*, 1-12. doi:https://doi.org/10.1038/s41597-022-01262-0

Sneha, & Raza, S. (2024). Affective Computing for Health Management via Recommender Systems: Exploring Challenges and Opportunities. In (pp. 163-182).

Spaniol, M., Wehrle, S., Janz, A., Vogeley, K., & Grice, M. (2024). *The influence of conversational context on lexical and prosodic aspects of backchannels and gaze behaviour*. Paper presented at the Speech Prosody.

Sumi, K., & Sato, S. (2022). Experiences of Game-Based Learning and Reviewing History of the Experience Using Player's Emotions. *Frontiers in Artificial Intelligence, 5*, 1-10. doi:https://doi.org/10.3389/frai.2022.874106

Tipirneni, S., & Leal, S. (2023). Deciphering Facial Expressions: factors that affect emotion recognition. *Journal of Student Research, 11*, 1-10. doi:https://doi.org/10.47611/jsrhs.v11i1.2460

Yamashita, J., Takimoto, Y., Oishi, H., & Kumada, T. (2024). How do personality traits modulate real-world gaze behavior? Generated gaze data shows situation-dependent modulations. *Frontiers in Psychology, 14*, 1-19. doi:https://doi.org/10.3389/fpsyg.2023.1144048

Yan, J., Li, P., Du, C., Zhu, K., Zhou, X., Liu, Y., & Wei, J. (2024). Multimodal Emotion Recognition Based on Facial Expressions, Speech, and Body Gestures. *Electronics, 13*, 1-22. doi:https://doi.org/10.3390/electronics13183756