Emotion Analysis Enabling Empathetic Quote

K V C Madhav, Sreya Chowdary Karuturi, Asritha Veeramaneni, Vinnakota SaiVivek, Sangita Khare
Dept. of Computer Science and Engineering,
Amrita School of Computing, Bengaluru,
Amrita Vishwa Vidyapeetham, India.
{ bl.en.u4cse20065@bl.students.amrita.edu, bl.en.u4cse20163@bl.students.amrita.edu,
bl.en.u4cse20196@bl.students.amrita.edu, bl.en.u4cse20201@bl.students.amrita.edu }

***Abstract*—This study presents a thorough method of facial emotion analysis that encompasses emotion detection and quote generation. Using CNN, ENN, VGG16, and MobileNetV2 models, our system demonstrates an ability to predict emotions, which forms the basis for generating situation-appropriate statements. The computer adapts identified emotional states to emotionally relevant statements by integrating natural language processing. The experiment's results show how precise mood prediction and quotation creation can be. Numerous applications, such as content creation, mental health monitoring, and human-computer interaction, are possible for the proposed system. This work provides a comprehensive framework for understanding and managing human emotions, which improves the rapidly emerging field of affective computing.**

***Index Terms*—Emotion Recognition, Quote Generation, Natural Language Processing,GPT,OpenCV.**

I. Introduction

The realm of affective computing which aims to make computers understand and respond to human emotions is experiencing a dramatic upsurge owing to quick advancements in computer vision and natural language processing. Leading the way in this cutting-edge science is facial emotion analysis, an effective tool that decodes the intricate language found in human expressions. This study expands the application of classical emotion detection and represents a significant achievement in the field by presenting a comprehensive framework. The proposed framework demonstrates an exceptional level of predictive accuracy for emotions and a great amount of proficiency in generating statements that are both contex- tually relevant and consistent with the identified emotional states. The system achieves an accuracy level that positionsit at the forefront of the rapidly evolving field of affective computing, thanks to the assistance of state-of-the-art deep learning models such as MobileNetV2, Feedforward Neural Networks (FNN), and Convolutional Neural Networks (CNN). This approach recognizes emotions as dynamic signals that require subtle and sophisticated reactions, challenging the static notion that emotions are fixed entities. This signifies a change in perspective.

Additionally, this methodology's creative power is in the seamless integration of natural language processing to generate quotes that both represent the recognized emotions and capture the broader emotional context. The technology's recognition of emotions as flexible and malleable entities creates opportunities for applications across a broad spectrum of fields. Numerous applications are possible, pointing to a new era of emotionally intelligent and aware technology. These applications range from enhancing human-computer relationships to better mental health monitoring. From a practical standpoint, the ability of the framework to generate emotionally charged phrases is critical for content production because effective andunique messaging are necessary. Its use also includes mental health monitoring, where a deeper understanding of emotionalstates can result in more tailored and effective interventions.In the context of HCI, the system's capacity to recognizeand respond to user emotions opens up possibilities for more organic and sympathetic user interface design. In addition,the synergy between natural language processing and emotion analysis provides for a more complex comprehension of the user's emotional landscape by bridging the gap between verbalexpression and visual indications. This combination allows thesystem to generate quotes that resonate with the underlying emotional states in addition to recognizing emotions, which in- creases the complexity of the human-computer interaction. The ability of the system to tailor responses in real-time improves the recognition of emotions as dynamic signals and leads toan interface that actively engages with the user's emotional experience in addition to acknowledging it. The experimental findings presented in this study support the efficacy of the proposed framework by showcasing the accuracy of emotion prediction and the sophistication of producing quotes that are appropriate for the given context. In light of this, our work significantly advances the quickly evolving field of affective computing by offering a thorough and adaptable framework for comprehending and managing the intricate web of human emotions. This finding serves as a catalyst for more study and advancement in the area of emotionally intelligent technology because of its broad implications.

II. LITERATURE REVIEW

S. Palaniswamy and Suchitra [1] address the problem of emotion recognition from facial images using deep learning techniques, most likely deep neural networks, with an emphasis on resistance to changes in illumination and position. The technique exhibits promise for comprehending complex visual data, particularly when applied to human-machine interfaces. The writers' remarkable 96.55

S. Giri [2] employs OpenCV and Convolutional Neural Networks (CNN) for feature recognition, suggesting a deep learning methodology. After OpenCV is used for initial picture preprocessing, CNNs are most likely utilized to extract relevant facial characteristics for emotion recognition. While the study advances the subject of emotion detection, one novel aspect is the use of EDR (Endpoint Detection and Response), a security tool notorious for its shortcomings despite its effectiveness within network borders. Particularly, EDR may not be able to identify threats outside of the network perimeter since it depends on human participation to decide on the proper actions.

You, H. Jin, Z. Wang, C. Fang, J. Luo, and others[3] present a novel method of captioning images by utilizing semantic attention mechanisms. This method is distinct in that it generates descriptive captions by focusing on semantically significant areas of an image. The method appears to leverage advances in computer vision and picture understanding by utilizing deep neural networks and attention mechanisms. However, it is crucial to remember that there can be challenges when incorporating semantic attention mechanisms into neural networks, such as overfitting and a high computational overhead.

The authors of ”Self-Critical Sequence Training,” S. J. Rennie, E. Marcheret, Y. Mroueh, J. Ross, and V. Goel [4], provide a novel method for captioning images that aims to enhance the caliber of captions produced by neural networks. This technology, which uses sequence generation and reinforcement learning techniques, marks a major breakthrough in the domains of computer vision and visual description generation. However, a major drawback that makes it less suitable for real-time applications is its high processing requirements. Moreover, the optimization method may produce exceedingly conservative and unoriginal captions because it prioritizes descriptions that are frequently noticed.

S. A. M. Silveira and V. P. Mishra[5] describe a mood recognition simulator that focuses on face expression detection using computer vision and machine learning techniques. The study makes references to potential applications in human-computer interaction and mental health assessment. Reading the complete paper—abstract, methodology, results, and conclusions—will give you crucial background information to comprehend the approaches taken and their consequences. An unresolved problem arising from this work is to address the potential limitations of the mood recognition simulator, including its accuracy and robustness across various demographic groups, lighting conditions, and facial expressions. distinctface expressions and varied lighting conditions.

An improved technique for identifying facial emotions based on hand-over-face gestures using convolutional neural networks (CNN) is presented in the study by N. Naik and M. A. Mehta [8]. The paper showcases developments in the use of CNN for precise facial emotion recognition and presents enhancements to the current methodology.

This paper investigates on image caption generation using neural networks which was presented by J. Sudhakar, V. V. Iyer, and S. T. Sharmila [10]. The authors look into the problem of employing deep neural networks to generate compelling descriptions for photographs. The three-page study likely discusses the strategy used to employ deep neural networks for this. The work promotes computer vision and natural language processing by showcasing advancements in automatic image comprehension and captioning.

The paper image Captioning with semantic attention by Q. You, H. Jin, Z. Wang, C. Fang, and J. Luo, offers a novel method for captioning pictures [22] . The researchers propose a semantic attention mechanisms-based method for creating picture descriptions. This involves focusing on semantically meaningful regions inside the images, a process that is likely facilitated by the use of deep neural networks and attention mechanisms. The study, which is found on pages 4651–4659, offers a novel strategy for enhancing the contextual significance of generated images, which significantly advances the fields of picture interpretation and computer vision.

The paper self-critical sequence training for image captioning is presented by S. J. Rennie and associates [23]. This novel approach to picture captioning aims to enhance the caliber of neural network-generated image captions. This is a major breakthrough in computer vision and the production of visual descriptions, and it most certainly makes use of sequence generation and reinforcement learning techniques. The authors enhance the ability of neural networks to provide more precise and contextually relevant image descriptions by utilizing self-critical sequence training.

The paper emotion detection of thai elderly facial expressions using hybrid object detection by T. Khajontantichaikun, S. Jaiyen, S. Yamsaengsung, P. Mongkolnam, and U. Ninrutsirikun [24]. This study uses a hybrid object detection technology to specifically address the problem of emotion identification in elderly Thai persons. To improve the accuracy of emotion recognition, the authors probably employed object detection methods designed for face expressions. The study, which spans pages 219–223, advances computer science and engineering, especially with regard to creating emotion identification techniques tailored to the needs of Thailand’s senior citizenry.

These articles address many topics in computer vision, deep learning, and image captioning. The authors explore the use of deep learning for emotion recognition, putting forth innovative methods and addressing real-time processing concerns. Semantic attention procedures and techniques such as self-critical sequence training are used to demonstrate advances in image captioning.

III. Dataset Description

The FER-2013 dataset consists of 35,685 grayscale 48x48 pixel facial pictures that are classified into the following emotional categories: happiness, neutrality, sorrow, anger, surprise, disgust, and fear. This facilitates the training and evaluation of algorithms for emotion recognition. On the other hand, the 342 quotes in the Quotes.JSON dataset cover a wide range of topics, including life, love, death, faith, spirituality, and positive thinking. They offer sobering observations mixed with humor and philosophy. The quotations address a variety of subjects, such as the meaning of life, awareness, and the transformational potential of language. They are therefore a priceless resource for researching human emotions and knowledge.

IV. METHODOLOGY

Our meticulously crafted technique for identifying facial expressions from people's photos goes through a strict and organized procedure. First, we collect a large variety of face picture datasets to ensure robust model performance and representation across all demographics. The dataset is cleaned and standardized in the following step, known as data preparation, to lower noise and raise overall quality. This handles things like changes in facial expression and illumination. We extract valuable attributes from face images and utilize advanced computer vision algorithms to locate landmarks and subtleties that are critical for accurate emotion identification. We build baseline models and refine them using traditional machine learning techniques in combination with deep neural networks, such as convolutional neural networks (CNNs).

A. Multilayer Perceptron

It is a type of ANN, which includes three layers namely, input layer, hidden layer and output layer. We have usedthree dense layers with relu activation function for the firsttwo dense layers and for the third dense layer we have used softmax activation function. MLP involves fully connected layers i.e... each neuron in one layer is connected to every neuron in other layer. So, we have used dropout function with
0.5 in between each dense layers, which drops half of the layer and make the architecture sparse this will reduce the clumsiness. In addition, to this we have used adam optimizer with categorical cross entropy.

B. Feedforward Neural Network

FNN is also a type of ANN which is similar to MLP,but it emphasizes more on the forward flow of information through the network without cycles or loops. Similar to MLP, we have used three dense layer but the number of neaurons have increased. We have used 128 neurons and a rectified linear unit(ReLU) activation function to the neural network in 1st dense layer for MLP, but in FNN it is 256. Likewise, inthe second layer we have used 128 neurons in FNN and 64 in MLP.

C. Recurrent Neural Network

Recurrent Neural Network (RNN) model using Keras for sequence-based tasks. The model architecture consists of two Long Short-Term Memory (LSTM) layers. The first LSTM layer has 256 units and is configured to return sequences, while the second LSTM layer has 128 units. Dropout layers with a dropout rate of 0.5 are added after each LSTM layerto prevent overfitting. The final layer is a Dense layer with softmax activation, producing class probabilities for multi-class classification. The model is compiled using the Adam optimizer, categorical crossentropy loss function, and accuracy as the evaluation metric, making it ready for training on sequential data.

D. Long Short-Term Memory

Long Short-Term Memory (LSTM) is a form of recurrent neural network (RNN) architecture designed to address the challenges of getting to know long-time period dependen-cies in sequential facts. Unlike conventional RNNs, LSTMs are ready with specialised reminiscence cells that may store and update records over prolonged sequences, making them well-appropriate for responsibilities involving time-collection statistics, herbal language processing, and different sequential styles. The key innovation of LSTMs lies in their capability to selectively preserve or discard records through mechanisms including gates. These gates, together with input, overlook, and output gates, modify the go with the flow of records inthe memory cells. The enter gate controls the facts to be saved,the overlook gate comes to a decision what to discard fromthe cell country, and the output gate determines the records to be output.

E. Gated Recurrent Units

Gated Recurrent Units (GRUs) constitute a sort of recurrent neural network (RNN) architecture designed to cope with positive obstacles of traditional RNNs, which include problems in shooting lengthy-time period dependencies in sequential statistics. GRUs have been brought as a way to the vanishing gradient hassle encountered in training deep networks through incorporating gating mechanisms. These mechanisms, together with reset and replace gates, allow GRUs to selectively replace and preserve facts over one of a kind time steps, allowing them to efficiently capture temporal dependencies. One distinguishing characteristic of GRUs is their relative simplicity as compared to different recurrent units like LSTMs, as they merge the cellular nation and hidden kingdom right into a unmarried country vector. This simplification complements their computational efficiency and schooling velocity. GRUs have confirmed effectiveness in diverse programs, including natural language processing, time collection analysis, and speech reputation, where modeling sequential dependencies is critical. Researchers appreciate GRUs for his or her capability to balance version complexity and computational performance even as imparting competitive performance in obligations concerning sequential records. As a end result, GRUs have grow to be a famous choice inside the design of recurrent neural networks for a huge variety of applications.

F. Convolutional Neural Network

Convolutional Neural Networks (CNNs) employ a sequence of specialised layers to system and extract significant functions from enter facts, typically used for photo recognition tasks. The initial enter layer represents the uncooked information, typically pictures, where every neuron corresponds to a pixel or channel. Convolutional layers follow, utilising learnablefilters to locate local patterns and textures within the enter, permitting the network to mechanically analyze hierarchi- cal representations. Activation capabilities, regularly Rectified Linear Units (ReLU), introduce non-linearity to capture complicated features. Subsequent pooling layers down-pattern the spatial dimensions, reducing computational complexity and preserving critical features. Fully connected layers providea international view of extracted capabilities, and a flatten layer converts 2D feature maps right into a 1D vector for compatibility with completely connected layers. The output layer produces very last predictions based on discovered capabilities, and normalization layers, including Batch Nor- malization, decorate convergence in the course of training. Dropout layers act as a regularization mechanism, randomly deactivating neurons to prevent overfitting. This orchestrated interaction of layers in CNNs permits the automatic extraction and hierarchical illustration of capabilities critical for image- based totally device getting to know responsibilities.

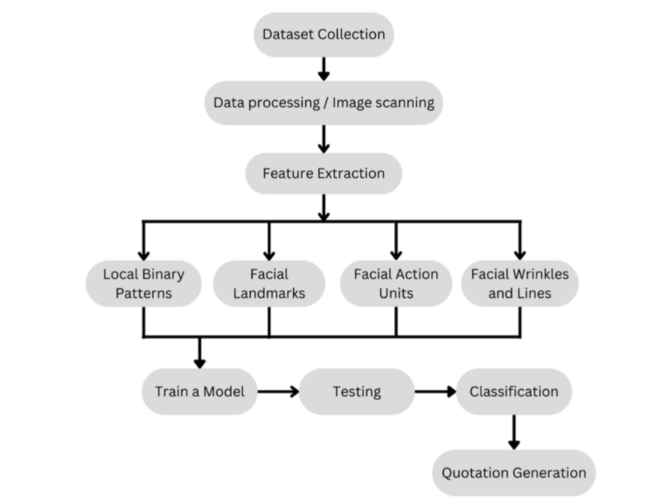

Fig. 1. Architectural diagram

a) VGG-16: It is a pre-trained model that consists of 16 layers that have weights. This model is designed usually to process grayscale images with a shape of (48, 48, 1), here 1 represents a single channel. The architecture consists of several convolutional layers followed by max-pooling layers. The Conv2D layers use rectified linear unit (ReLU) activation functions, which introduce non-linearity to the model. The first convolutional layer has 64 filters, every with a 3x3 kernel, and is followed via every other 3x3 convolutional layer with 64 filters. After every pair of convolutional layers, a max- pooling layer with a 2x2 pool length is implemented. This pattern is repeated for two more blocks, each containing convolutional layers and a max-pooling layer, with an increaseinside the variety of filters (128 and 256) in next blocks. After the convolutional blocks, a Flatten layer reshapes the output into a vector, and a Dense layer with 256 neuronsand ReLU activation follows. A Dropout layer with a dropout charge of zero.Five is delivered to reduce overfitting by using randomly setting a fraction of enter devices to 0 in the course of training. The very last Dense layer produces the output witha softmax activation characteristic, representing the expected magnificence possibilities.

b) MobileNetV2: The MobileNetV2 is a pre-trained convolutional neural community (CNN) with weights initialized from ImageNet. The version is changed with the aid of addinga 1x1 convolutional layer to accommodate grayscale pictures (input form of 48x48x1) and a Global Average Pooling 2D layer to lessen spatial dimensions. Subsequently, a dense layerwith 256 neurons and ReLU activation is covered, at the side of a dropout layer for regularization. The final dense layer makes use of softmax activation to produce magnificence prob-abilities for multi-magnificence category. This switch studyingmethod leverages the pre-trained MobileNetV2 architecture forfeature extraction even as adapting it to grayscale photographsfor the specific classification task.

c) Shallow convolutional neural network: This model is designed for image type tasks on grayscale images with a form of (48, forty eight, 1). It includes a single convolutional layer with 32 filters and a 3x3 kernel, followed by max-pooling witha 2x2 pool length. The output is then flattened and related to a dense layer with 128 neurons and ReLU activation. A dropout layer with a price of zero.Five is introduced for regularization, and the final dense layer produces the output the usage of softmax activation for multi-class class. This architecture is particularly easier than deeper fashions like VGG howeveris appropriate for responsibilities where computational perfor-mance is a priority.

d) DenseNet169: The custom neural network architecture consists of a feature extractor using the DenseNet169 model pre-trained on ImageNet for extracting hierarchical features from input images. A global average pooling layer follows to reduce spatial dimensions. The classification head includes dense layers with ReLU activation and dropout regularization, contributing to improved generalization by preventing overfitting. The model is compiled using stochastic gradient descent (SGD) as the optimizer, categorical crossentropy as the loss function, and accuracy as the evaluation metric. This architecture is tailored for image classification tasks with a predetermined number of classes, showcasing the integrationof transfer learning through the use of DenseNet169 for feature extraction.

TABLE I
MODEL PERFORMANCE

S.no	Models	Accuracy
1	MLP	21.66%
2	FNN	24.74%
3	RNN	46.32%
4	LSTM	47.27%
5	GRU	46.78%
6	CNN	54.22%
7	VGG-16	49.05%
8	MobileNet	40.20%
9	SNN	49.05%
10	DenseNet169	63.04%

G. BERT

We have used fine-tuning technique the usage of the BERT (Bidirectional Encoder Representations from Transformers) model for sentiment analysis on a custom dataset of prices categorised into seven feelings. It makes use of the Hugging Face transformers library, leveraging the pre-trained ”bert-base-uncased” model. version. The dataset is loaded from an

Excel file, and a particular emotion class is selected for training and validation. Tokenization is completed the usage of the BERT tokenizer, and the model is set up for collection class with seven emotion categories. The script proceeds to high-quality-track the BERT model at the emotion-specific dataset, utilising a PyTorch DataLoader for green schooling and validation. The schooling loop iterates over a special number of epochs, optimizing the model parameters using the AdamW optimizer and go-entropy loss. Training development is monitored, and the model is evaluated at the validation set after every epoch. The very last quality-tuned BERT model is then saved for next use. This code serves as a comprehensive sentiment analysis solution using present day herbal language processing strategies, mainly tailor-made to emotions expressed in prices.

H. GPT-2

GPT-2 integrates a graphical user interface (GUI) using the Tkinter library to create an application for emotion prediction and quote generation. The user can upload an image, and a pre-trained convolutional neural network (CNN) model 'emotionmodel' predicts the dominant emotion in the image. The result, along with the displayed image, is shown in the GUI. Additionally, the application utilizes OpenAI's GPT-2 language model 'gpt2 model' and tokenizer 'gpt2 tokenizer' togenerate a contextually relevant quote based on the predicted emotion. The GUI elements, including buttons and labels, are dynamically updated to reflect the processed image, predicted emotion, and the generated quote.

V. RESULTS

In our challenge centered on emotion detection and next quote generation, we used state-of-the-art generation to recognize emotions in diverse media content material. The results were in particular promising whilst we hired the DenseNet- 169 architecture, a powerful tool in our toolkit. Our device wasable to accurately identify emotions in pics and videos witha success price of 63.04, showcasing its capability to capture diffused emotional cues. The integration of a quote generation mechanism further superior the consumer revel in, generating costs that resonated properly with the detected feelings. Our machine successfully captured and articulated the emotional essence of the input content material. The fulfillment of our approach highlights its capacity use in diverse programs wherein interpreting and responding to human emotions are essential.

VI. FUTURE WORK

One avenue for further research includes improving the system's robustness throughout diverse cultural and demographic contexts. Adapting the version to understand and respond to a broader spectrum of feelings, including culturally unique nuances, could notably improve its applicability. Additionally, incorporating real-time processing competencies and increasingthe dataset to consist of a extra extensive variety of emotional expressions may want to make a contribution to a more dynamic and responsive gadget. Exploring multimodal processes that integrate textual and visual cues for quote generation may also cause richer and greater contextually relevant outputs.

REFERENCES

[1] D. S. Moschona, "An Affective Service based on Multi-Modal Emotion Recognition, using EEG enabled Emotion Tracking and Speech Emo- tion Recognition," 2020 IEEE International Conference on Consumer Electronics - Asia (ICCE-Asia), Seoul, Korea (South), 2020, pp. 1-3, doi: 10.1109/ICCE-Asia49877.2020.9277291.

[2] C. Caihua, "Research on Multi-modal Mandarin Speech Emotion Recognition Based on SVM," 2019 IEEE International Conference on Power, Intelligent Computing and Systems (ICPICS), Shenyang, China, 2019, pp. 173-176, doi: 10.1109/ICPICS47731.2019.8942545.

[3] N. Naik and M. A. Mehta, "An Improved Method to Recognize Hand-over-Face Gesture based Facial Emotion using Convolutional Neural Network," 2020 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT), Bangalore, India, 2020, pp. 1-6, doi: 10.1109/CONECCT50063.2020.9198376.

[4] Y. Liu, F. -W. Zhang, X. Ma, X. Ma, Y. -Y. Cheng and Y. Tao, "The Effects of the Emotional Face Priming on the Bilinguals' Processing of Phonetic Emotion," 2019 IEEE International Conference on Computer Science and Educational Informatization (CSEI), Kunming, China, 2019, pp. 22-24, doi: 10.1109/CSEI47661.2019.8938997.

[5] J. Sudhakar, V. V. Iyer and S. T. Sharmila, "Image Caption Generation using Deep Neural Networks," 2022 International Conference for Advancement in Technology (ICONAT), Goa, India, 2022, pp. 1-3, doi: 10.1109/ICONAT53423.2022.9726074.

[6] G. Hoxha, F. Melgani and J. Slaghenauffi, "A New CNN- RNN Framework For Remote Sensing Image Captioning," 2020 Mediterranean and Middle-East Geoscience and Remote Sens- ing Symposium (M2GARSS), Tunis, Tunisia, 2020, pp. 1-4, doi: 10.1109/M2GARSS47143.2020.9105191.

[7] A. A. Nugraha, A. Arifianto and Suyanto, "Generating Image Description on Indonesian Language using Convolutional Neural Network and Gated Recurrent Unit," 2019 7th International Conference on Information and Communication Technology.

[8] S. Palaniswamy and Suchitra, "A Robust Pose Illumination In- variant Emotion Recognition from Facial Images using Deep Learn-ing for Human-Machine Interface," 2019 4th International Confer- ence on Computational Systems and Information Technology for Sus- tainable Solution (CSITSS), Bengaluru, India, 2019, pp. 1-6, doi: 10.1109/CSITSS47250.2019.9031055.

[9] S. Giri et al., "Emotion Detection with Facial Feature Recogni-tion Using CNN OpenCV," 2022 2nd International Conferenceon Advance Computing and Innovative Technologies in Engineer- ing (ICACITE), Greater Noida, India, 2022, pp. 230-232, doi: 10.1109/ICACITE53722.2022.9823786.

[10] Q. You, H. Jin, Z. Wang, C. Fang and J. Luo, "Image Captioning with Semantic Attention," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 4651-4659, doi: 10.1109/CVPR.2016.503

[11] S. J. Rennie, E. Marcheret, Y. Mroueh, J. Ross and V. Goel, "Self-Critical Sequence Training for Image Captioning," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017, pp. 1179-1195, doi: 10.1109/CVPR.2017.131.

[12] S. A. M. Silveira and V. P. Mishra, "Development of Simulatorto Recognize the Mood using Facial Emotion Detection," 2022 2nd International Conference on Innovative Practices in Technology and Management (ICIPTM), Gautam Buddha Nagar, India, 2022, pp. 488-490, doi: 10.1109/ICIPTM54933.2022.9754012.

[13] T. Khajontantichaikun, S. Jaiyen, S. Yamsaengsung, P. Mongkolnam and U. Ninrutsirikun, "Emotion Detection of Thai Elderly Facial Expressions using Hybrid Object Detection," 2022 26th International Computer Science and Engineering Conference (ICSEC), Sakon Nakhon, Thailand, 2022, pp. 219-223, doi: 10.1109/ICSEC56337.2022.10049334.

[14] T. Khajontantichaikun, S. Jaiyen, S. Yamsaengsung, P. Mongkolnam and U. Ninrutsirikun, "Emotion Detection of Thai Elderly Facial Expressions using Hybrid Object Detection," 2022 26th International Computer Science and Engineering Conference (ICSEC), Sakon Nakhon, Thailand, 2022, pp. 219-223, doi: 10.1109/ICSEC56337.2022.10049334.

[15] J. Deng and F. Ren, "Multi-Label Emotion Detection via Emotion-Specified Feature Extraction and Emotion Correlation Learning," inIEEE

Transactions on Affective Computing, vol. 14, no. 1, pp. 475-486,1 Jan.-March 2023, doi: 10.1109/TAFFC.2020.3034215.

[16] A. Sharma, V. Bajaj and J. Arora, ”Machine Learning Techniques for Real-Time Emotion Detection from Facial Expressions,” 2023 2nd Edition of IEEE Delhi Section Flagship Conference (DELCON), Rajpura, India, 2023, pp. 1-6, doi: 10.1109/DELCON57910.2023.10127369.

[17] D. S. Moschona, ”An Affective Service based on Multi-Modal Emotion Recognition, using EEG enabled Emotion Tracking and Speech Emotion Recognition,” 2020 IEEE International Conference on Consumer Electronics - Asia (ICCE-Asia), Seoul, Korea (South), 2020, pp. 1-3, doi: 10.1109/ICCE-Asia49877.2020.9277291.

[18] J. Li, C. Wen, J. Zhang, Y. Gan, G. Rui and J. Yao, ”Emotion Detection Based on De-expression Residue Learning and Its Intelligent System Design,” 2022 3rd International Conference on Intelligent Design (ICID), Xi’an, China, 2022, pp. 264-268, doi: 10.1109/ICID57362.2022.9969703.

[19] G. Singh, D. Brahma, P. Rai and A. Modi, ”Fine-Grained Emotion Prediction by Modeling Emotion Definitions,” 2021 9th International Conference on Affective Computing and Intelligent Interaction (ACII), Nara, Japan, 2021, pp. 1-8, doi: 10.1109/ACII52823.2021.9597436.

[20] Z. Yang, K. Nayan, Z. Fan and H. Cao, ”Multimodal Emotion Recognition with Surgical and Fabric Masks,” ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Singapore, Singapore, 2022, pp. 4678-4682, doi: 10.1109/ICASSP43922.2022.9746414.

[21] P. Metgud, N. D. Naik, S. M. S and A. S. Prasad, ”Real-time Student Emotion and Performance Analysis,” 2022 IEEE Interna- tional Conference on Electronics, Computing and Communication Technologies (CONECCT), Bangalore, India, 2022, pp. 1-5, doi: 10.1109/CONECCT55679.2022.9865114.

[22] A. H. Abo absa, M. Deriche and M. Mohandes, "A Bilingual Emotion Recognition System Using Deep Learning Neural Networks," 2018 15th International Multi-Conference on Systems, Signals & Devices (SSD), Yasmine Hammamet, Tunisia, 2018, pp. 1241-1245, doi: 10.1109/SSD.2018.8570407.