Emotion Recognition Systems: An Analysis of Emerging Trends and Technologies

Shreeya S. Halwasia
Computer Engineering
Thadomal Shahani Engineering College Bandra Mumbai

Vansh B. Wadhwa
Computer Engineering
Thadomal Shahani Engineering College Bandra Mumbai

Sonali B. Jadhav
Computer Engineering
Thadomal Shahani Engineering College Bandra Mumbai

ABSTRACT
Human-Computer Interaction (HCI) encompasses the study of how humans interact with technology, including computer software, hardware, mobile devices, websites, and other digital interfaces. HCI traces back to the early stages when human-computer interactions were limited to command-line interfaces, but with the advancement in the digital world, HCI has become a significant area of research and development. As we journey towards a more emotionally intelligent technological era, emotion recognition in HCI stands as a critical enabler of richer, more meaningful human-computer interactions. This review paper aims to shed light on the progress made, challenges faced, and potential for growth in this fascinating domain. Additionally, it aims to analyze the various methodologies used to implement human-computer interaction using facial features and voice. The paper reviews various pre-existing systems that implement sentiment analysis on multiple datasets, their technology, the output, the accuracy they achieved, and their conclusion. Most of the systems reviewed here were implemented using Deep Learning algorithms like Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN), Deep Neural Networks (DNN), and various other concepts like Feature Extraction, Data Preprocessing, etc.

General Terms
Emotion Recognition, Artificial Intelligence, Deep Learning, etc.

Keywords
Facial Expression Recognition, Convolutional Neural Network, Preprocessing. Statistical Approach, Electrooculography, etc

1. INTRODUCTION
In the rapidly evolving environment of technology, Human-Computer Interaction (HCI) continues to redefine the way humans interact with machines and digital environments. As computing systems become increasingly integrated into our daily lives, the need for seamless, intuitive, and emotionally aware interfaces becomes more apparent. Emotions play a crucial role in human cognition and behavior, they significantly influence decision-making, learning, understanding, and social interactions. Emotion recognition, a subfield of HCI, emerges as a promising avenue to bridge the gap between humans and machines, leading us to a new era of emotionally intelligent interactions. Emotion recognition technologies can be used by systems to perceive, analyze, and respond appropriately to human emotions. This can be done by deciphering emotional gestures and tailoring responses to them. By examining numerous applications, this paper intends to provide a complete review of the current level of emotion recognition in HCI. This paper provides an in-depth analysis of emotion recognition systems, exploring three distinct methodologies: Speech Emotion Recognition, Eye Emotion Recognition, and Facial Emotion Recognition using Deep Learning. The Speech Emotion Recognition System utilizes statistical approaches and lexicon-based classification. The Eye Emotion Recognition System employs electrooculography and noise reduction algorithms. Facial Emotion Recognition using Deep Learning utilizes Convolutional Neural Networks without preprocessing. The final section introduces DeepFEVER, a feature extractor for facial expression recognition that emphasizes the importance of generalization techniques.

2. IMPORTANT TERMINOLOGIES
In order to effectively study certain systems, it is necessary to establish a thorough understanding of key concepts.

2.1 Convolutional Neural Network
A Convolutional Neural Network (CNN) is a type of Deep Learning neural network architecture commonly used in Computer Vision [10].The Convolutional Neural Network (CNN) architecture comprises multiple layers, including the input layer, Convolutional layer, Pooling layer, and fully connected layers. The Convolutional layer utilizes filters to perform feature extraction on the input image, while the Pooling layer serves to downsample the image, enhancing computational efficiency. The ultimate prediction is generated by the Fully Connected layer using the processed data. To optimize network performance, gradient descent, and backpropagation techniques are employed to iteratively refine these filters [10].

2.2 Preprocessing
Preprocessing is the initial step conducted meticulously on any raw data to get it ready for another subsequent data processing method. Data preprocessing changes the data into a streamlined format that can be effectively processed in machine learning, data mining, and other data science tasks more quickly and efficiently. The first stage of data preprocessing is sampling [11]. In this step, a subset of data from a sizable comprehensive population dataset is chosen. Following this, the transformation stage manipulates the raw data to create a single output. Denoising is the following stage, which eliminates noise from the data. Additionally, the missing values are filled in, and inaccurate values are corrected or removed. Normalization, the final phase, organizes data for quicker access.

2.3 Feature Extraction
It is the process in which relevant features significant in the required context are pulled out [11]. Feature extraction is a fundamental process that streamlines raw data by identifying and isolating key attributes, thereby enhancing data management. This method is imperative in the context of

handling extensive datasets, as it allows for resource conservation without compromising essential data. Large datasets frequently encompass a multitude of variables, necessitating substantial computational power. Through the process of feature extraction, these variables are amalgamated and selectively retained, ultimately yielding concise features that reduce data volume while retaining critical information from the initial dataset.

3. DETAILED STUDY OF EMOTION RECOGNITION SYSTEMS

In order to effectively study certain In this section, we conduct an extensive examination of emotion recognition systems, encompassing three distinct methodologies. The first approach centers on Speech Emotion Recognition, employing statistical techniques for identifying emotions in speech, encompassing text preprocessing and lexicon-based classification. We evaluate similarity through Jaccard, Cosine, and Correlation measures. The Speech Emotion Recognition System adopts a statistical framework, utilizing databases and preprocessing to classify emotions from spoken words [1].

The second system follows an Eye Emotion Recognition approach, leveraging electrooculography (EOG) for monitoring eye movements and detecting emotions via changes in EOG values. This process involves noise reduction, feature extraction, and classification using a linear Multiclass Support Vector Machine (SVM). The Eye Emotion Recognition System analyzes emotions through electrooculogram and eye motion [14] [15] [16].

The third system utilizes Deep Learning for Facial Emotion Recognition, employing Convolutional Neural Networks (CNNs) for facial expression recognition without preprocessing. It addresses challenges such as lighting, poses, and landmark detection. Facial Emotion Recognition using Deep Learning employs Convolutional Neural Networks to categorize emotions based on facial expressions [18] [19].

Additionally, we studied DeepFEVER, an autonomous feature extractor for Facial Expression Recognition, which involves training teacher and student networks through knowledge distillation to achieve generalization across diverse datasets. The Facial Feature Extraction system deploys a Deep Facial Expression Vector Extractor and employs Knowledge Distillation to underscore the importance of intricate facial feature extraction for accurate emotion recognition and the pivotal role of generalization techniques in this field [4].

3.1 Speech Emotion Recognition System Using Statistical Approach

This paper suggests an approach for identifying emotions in speech. The gathering and preparation of the speech database is the initial phase. Preprocessing involves refining the dataset by cleaning the data, removing outliers, and eliminating words that are often used. Furthermore, for the purpose of improving textual accuracy, the Snowballc tool is used to eliminate suffixes from the dataset. The Porter word stemming algorithm, which reduces words to a common root to facilitate vocabulary comparison, is implemented by Snowballc, an R interface to the C 'libstemmer' package. This algorithm stands out as one of the most advanced and efficient techniques in its domain [12]. In the second stage, the emotions are classified using lexicons.

These lexicons comprise a collection of terms, each possessing two polarities, that is, two opposites that can be used as a scale to classify emotions is known as a lexicon. The process of categorizing emotions involves the junction of vocabulary and term documents. The document loads each sentence, and each word is separated into a character vector [1]. A character vector, in this context, consists of elements composed solely of alphabetic characters (either in lowercase or uppercase), irrespective of their length [13]. Following this, the character vector undergoes evaluation through a function that compares it to the emotion lexicon, yielding a result. The documents are converted from text to vector to look for similarities. Semantically comparable texts are those that are similar to one another.

In this study, three alternative approaches are used to determine how similar the two documents are.

3.1.1 Jaccard Similarity

It assesses how similar two datasets are.

3.1.2 Cosine similarity

This metric assesses how closely two documents' contents coincide.

3.1.3 Correlation

It is a statistical tool for determining the linear relationship between two variables [1].

On various datasets with different numbers of samples under each category, additional statistical analysis is conducted to determine the effectiveness of the Jaccard similarity coefficient, Cosine similarity coefficient, and correlation similarity value.

3.2 Eye Emotion Recognition System

Instead of discussing the many classification techniques used for predicting emotions, this work focuses on categorizing emotions. based on the results of a comparative analysis of similarity measurements. Within this study, emotions are identified by analyzing eye movements, a process facilitated by the use of specialized equipment like eye trackers and cameras. The proposed methodology adopts a dataset acquired from Twente Medical Systems International (TMSI), encompassing both horizontal and vertical eye movement signals captured through the use of electrooculography (EOG). EOG, or electrooculogram, constitutes an electrophysiological test that measures the inherent resting electrical potential between the cornea and Bruch's membrane. It essentially records the dynamic changes in the corneoretinal potential difference present within the eye [14].

Basically, an EOG system considers the human eye as a dipole where the retina of the eye acts as a negative pole and the cornea of the eye acts as the positive pole. During our eye movements, the position of the cornea and retina considerably change, and the potential charge varies accordingly. For example - when we cry our eyeballs contract, therefore the position of the cornea and retina change due to which the potential of the eye changes i.e., the EOG value of the eye changes. Hence a predefined set of EOG values for different emotions (sad, happy, fear, etc) can be used to recognize emotions.

The first stage in the process suggested is the entry of EOG values from the dataset. Preprocessing the EOG signals is the next stage. The EOG signals could be affected by noise due to electrical supply or wires in the surrounding. In order to achieve accurate results in this step the noise from the EOG signals is reduced. In order to remove noise from EOG signals two different algorithms were used.

3.2.1 Low Pass Filter

Low pass filters are used to keep the signal's low-frequency components while removing high-frequency noise from the signal. Low-pass filters attenuate (lower) the high-frequency components of an input signal while allowing the low-frequency components to flow through.

3.2.2 Median Filter

A non-linear digital filtering method known as the median filter is frequently employed to eliminate noise from signals. A typical pre-processing procedure to enhance the outcomes of subsequent processing is noise reduction. The essential principle behind the median filter is to iteratively replace each element in the signal with the median of its nearby entries [15].

The median filter algorithm was employed for noise reduction because it was found that it performed efficiently better than the low pass filter. The third step entails the extraction and selection of features. Within this phase, various statistical parameters, including mean, kurtosis, variance, and others, are computed for both the horizontal and vertical EOG signals. A redundant method for feature selection is Independent Component Analysis (ICA). The individual mean value is calculated for each component of the de-noised signal by ICA. After extracting the features required to generate the output, classification is carried out.

As a classification process, a linear Multiclass Support Vector Machine (SVM) is employed. A Multiclass SVM is a variation on conventional Support Vector machines. In multiclass SVM, the data is classified into two classes, a positive class, and a negative class or we can say that binary classification is performed. Each of these classes are further subdivided into another binary class. The valence arousal model is employed to differentiate these classes. The degree of autonomic activation that an experience induces, known as arousal (or intensity), can range from calm (or low) to excited (or high) [16]. In addition, while both melancholy and fury share a negative valence, rage has a higher arousal level than sadness does.

The final step encompasses the evaluation of results, revealing that this system has the potential to be employed within a Personal Assistance framework. This integration allows the system to automate actions or respond in accordance with the user's emotions. It is further determined that this system is superior to previous systems with complex structures since it uses an optimum way to capture a person's sentiment. It is therefore not ideal, but it is very accurate [2].

3.3 Facial Emotion Recognition using Deep Learning

Another existing application implements Facial Expression Recognition (FER) using CNN. The program was executed without any preprocessing or feature extraction.

The problems faced by machines in terms of FER:

i. Detection of image section as face due to hidden parts.

ii. Lightening

iii. Poses

iv. Landmark Detection

The Facial Action Coding System (FACS) is a comprehensive tool for analyzing facial expressions and identifying emotions, it is universally accepted. The paper states the seven emotions of the system: happy, sad, surprise, fear, anger, disgust, and neutral.

The paper has reviewed various systems that were built using Deep Learning algorithms such as CNNs and Siamese Networks. These were tested with certain datasets and their methodology and accuracy levels have been specified in the paper.

The proposed system employs the Dlib facial landmark detector, a computer vision tool that accurately identifies 68 key points on a face, including eyes, nose, and mouth, enabling facial analysis, expression recognition, and feature manipulation in various applications [18]. In FER (Facial Expression Recognition), landmarks are key facial points used to analyze and interpret emotions from an image. This model can accurately locate the key points on a face, even under varying lighting conditions and different facial poses. It has been trained using the iBUG 300-W dataset. A Convolutional Neural Network (CNN) comprising of 6 convolutional layers is implemented, utilizing the Rectified Linear Unit (ReLU) activation function. ReLU (Rectified Linear Unit) is an activation function in neural networks that outputs the input if it is positive, or zero otherwise. The architecture includes 3 max-pooling layers, followed by 2 convolution layers, 2 dropouts with a value of 0.2, 1 flatten layer, and 2 dense layers.

Pooling layers are also referred to as down sampling layers. It decreases the amount of input parameters by performing dimensionality reduction.

Pooling is classified into two types:

3.3.1 Max Pooling

As the filter traverses the input, it chooses the pixel with the highest value to transfer to the output array.

3.3.2 Average pooling

This As the filter traverses the input, it computes the average value within the receptive field and sends it to the output array [19].

Convolutional Layer is the fundamental building component of a CNN and precisely where the majority of computation takes place. It necessitates a few components, including input data, a filter, and a feature map [19].

Flattening Layer is used to combine all the 2-Dimensional arrays produced by pooling feature maps into a single long continuous linear vector [20].

The system achieved 61.7% accuracy on the FER2013 Dataset without preprocessing or feature extraction. The highest accuracy was attained with a batch size of 512 and 10 epochs. Out of 467 angry faces, the system accurately predicted 276 images. For happy faces, out of 895 instances, the system made accurate predictions for 699 images. The emotions most prone to misclassification were fear and sadness, with accuracies of 43.95% and 49.77%, respectively.

The system was developed using Keras, which is a Python-based open-source deep-learning library. Keras is a high-level API that allows users to quickly and efficiently design and train neural networks. Google's TensorFlow, on the other hand, is a sophisticated deep-learning framework. It provides a low-level interface for building and training complex machine-learning models.

In this system, the architecture and design of the neural network, along with its layers and activation functions, were specified using Keras. TensorFlow, serving as the backend, handled the low-level computations and optimization, efficiently utilizing the available hardware resources such as CPUs and GPUs to accelerate the training process.

The system encountered an overfitting issue, with a training dataset accuracy of 99.64%, which prompted them to explore the importance of data augmentation in deep FER. Additionally, the use of preprocessing and feature extraction in the future was also proposed [3].

3.4 Facial Feature Extraction Using Knowledge Distillation and Generalization

Facial Expression Recognition (FER) remains a challenging problem in computer vision, and existing methods often lack the ability to generalize across diverse datasets and tasks. The proposed system, known as the Deep Facial Expression Vector Extractor (DeepFEVER), offers a promising solution to this issue. DeepFEVER is an independent feature extractor specifically designed for facial expressions, capable of generalized performance on any FER task or dataset.

3.4.1 Teacher Network Architecture and Training

The teacher network is a crucial component of DeepFEVER, responsible for recognizing facial expressions in images. To ensure generalization, the teacher model undergoes a two-step training process: pre-training and further training. The pre-training phase starts with a FaceNet model, originally designed for person re-identification, which is adapted for FER using labeled datasets, namely "AffectNet" and "Google Facial Expression Comparison (FEC)." The teacher model's architecture consists of a modified FaceNet model, extended with DenseNet blocks to learn facial expression features effectively. A 1x1 convolution operation followed by global average pooling is used to create a Dface-dimensional vector representation, named "Dface." To enhance robustness, two teacher networks are trained independently with different random seeds and penultimate layer sizes (Dface = 256 and Dface = 128). The outputs of both teachers are combined to create reliable targets for the self-distillation process.

3.4.2 Student Network Architecture and Training

The student network shares similarities with the teacher network, as it also aims to recognize facial expressions in images. However, the student network is trained differently and benefits from the knowledge distilled from the teacher network. The pre-training phase starts with DenseNet201, which has been trained on the general ImageNet dataset to learn essential features. The student network training follows a specific algorithm similar to the teacher network but with an additional step. In addition to labeled data from "Google FEC" and "AffectNet" datasets, the student network uses batches of unlabelled data from the "PowderFaces" dataset, created specifically for this project.

3.4.3 Knowledge Distillation for Generalization

The key to achieving generalization lies in the knowledge distillation process, where the student network learns from the outputs of the two trained teacher networks. The teacher networks produce two types of predictions for the Google FEC and AffectNet tasks. These predictions are normalized to ensure consistent scaling and concatenated into an 80-dimensional target vector. During training, the student network's predictions are compared to the target vector using the "Relational Knowledge Distillation" loss function. This additional loss is combined with the standard losses for the AffectNet and Google FEC tasks, ensuring that the student network learns the expressive power of the teacher networks and generalizes across diverse datasets.

Table 1. Accuracy Results

Dataset	Without Distillation	With Distillation (with Unlabeled Data)
AffectNet (eight-class)	58.8%	61.1%
Google FEC	85%	86.4%

Comparatively, the influence of the unlabelled data on improving performance was minimal, indicating that distillation played a more crucial role in enhancing the student model's capabilities. The absence of distillation led to a substantial decrease in performance. Additionally, the study reveals that while the use of unlabelled data (PowderFaces dataset) had some positive effect on performance, its impact was not as significant as distillation [4].

4. LIMITATIONS OF THE SYSTEMS

4.1 Limited dataset diversity

Many of the reviewed systems used limited datasets for emotion recognition, which may lead to biased results and reduced generalizability to real-world scenarios.

4.2 Hardware requirements

Some of the deep learning models used in the reviewed systems are computationally intensive and require high-end hardware, making them less accessible to researchers with limited resources.

4.3 Sensitivity to noise

Emotion recognition systems based on facial expressions and voice may be sensitive to environmental noise, variations in lighting conditions, and different facial poses, impacting their accuracy.

4.4 Ethical challenges

Emotion recognition technologies raise concerns about privacy, data security, and the possibility for commercial or manipulative use of user emotions.

5. FUTURE WORK

5.1 Dataset diversity and size

Future research should focus on using diverse and larger datasets to improve the generalizability and reliability of emotion recognition models.

5.2 Multimodal emotion recognition

Integrating multiple modalities such as facial expressions, voice, and physiological signals can lead to more robust and accurate emotion recognition systems.

5.3 Real-time emotion recognition

Developing real-time emotion recognition systems that can adapt to dynamic emotional states will enable more seamless and interactive human-computer interactions.

5.4 Explainable AI

Enhancing the interpretability and competence of emotion recognition models can build trust in these systems and facilitate their adoption in various applications.

6. CONCLUSION

The review paper highlights the significance of emotion recognition in Human-Computer Interaction (HCI) and its potential to enhance the quality of human-computer interactions. Various methodologies based on deep learning algorithms were reviewed, including those utilizing facial features and voice for emotion recognition. While progress has been made in this domain, there are still challenges to address, such as dataset limitations, hardware requirements, generalization, sensitivity to noise, etc. The future of emotion recognition in HCI lies in further advancements in dataset diversity, larger data collection, and the integration of multiple modalities for more accurate and robust recognition. Additionally, the development of real-time emotion recognition systems and the consideration of ethical implications will be imperative in shaping the future of emotionally intelligent technological interactions. As research continues to evolve, emotion recognition will likely find applications in diverse fields, leading us toward a more emotionally aware and empathetic technological era.

7. REFERENCES

[1] C. Manasa, D. Dheeraj and V. S. Deepthi, "Statistical Analysis of Voice Based Emotion Recognition using Similarity Measures," 2019 1st International Conference on Advanced Technologies in Intelligent Control, Environment, Computing & Communication Engineering (ICATIECE), Bangalore, India, 2019, pp. 46-50, doi: 10.1109/ICATIECE45860.2019.9063783.

[2] R. S. Soundariya and R. Renuga, "Eye movement based emotion recognition using electrooculography," 2017 Innovations in Power and Advanced Computing Technologies (i-PACT), Vellore, India, 2017, pp. 1-5, doi: 10.1109/IPACT.2017.8245212.

[3] S. Singh and F. Nasoz, "Facial Expression Recognition with Convolutional Neural Networks," 2020 10th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA, 2020, pp. 0324-0328, doi: 10.1109/CCWC47524.2020.9031283.

[4] L. Schoneveld and A. Othmani, "Towards a General Deep Feature Extractor for Facial Expression Recognition," 2021 IEEE International Conference on Image Processing (ICIP), Anchorage, AK, USA, 2021, pp. 2339-2342, doi: 10.1109/ICIP42928.2021.9506025.

[5] H. L. Sharma and M. Sharma, "Using CNN and Open CV, Mood Identification with Face Feature Learning," 2022 11th International Conference on System Modeling & Advancement in Research Trends (SMART), Moradabad, India, 2022, pp. 1109-1112, doi: 10.1109/SMART55829.2022.10047674.

[6] T. Sînziana, T. Tudor, M. Cristian and D. Laura, "An initial study of feature extraction's methods in facial expression recognition," 2019 IEEE 15th International Conference on Intelligent Computer Communication and Processing (ICCP), Cluj-Napoca, Romania, 2019, pp. 251-257, doi: 10.1109/ICCP48234.2019.8959538.

[7] P. Sudhakaran, P. K. Nair and A. Suraj, "Music Recommendation using Emotion Recognition," 2022 IEEE 2nd Mysore Sub Section International Conference (MysuruCon), Mysuru, India, 2022, pp. 1-7, doi:10.1109/MysuruCon55714.2022.9972635.

[8] K. Patel et al., "Facial Sentiment Analysis Using AI Techniques: State-of-the-Art, Taxonomies, and Challenges," in IEEE Access, vol. 8, pp. 90495-90519, 2020, doi: 10.1109/ACCESS.2020.2993803.

[9] "Human Computer Interaction," Interaction Design Foundation, https://www.interaction-design.org/literature/topics/human-computer-interaction. Accessed on July 26,2023.

[10] "Introduction to Convolution Neural Network," geeksforgeeks, March 2023, https://www.geeksforgeeks.org/introduction-convolution-neural-network/. Accessed on July 26, 2023

[11] George Lawton, "data preprocessing," Tech Target, January 2022, https://www.techtarget.com/searchdatamanagement/definition/data-preprocessing, Accessed on July 26, 2023

[12] Milan Bouchet-Valat, "Snowballc," cran.r-project, April 2023,https://cran.r project.org/web/packages/SnowballC/index.html#:~:text=SnowballC%3A%20Snowball%20Stemmers%20Based%20on,to%20aid%20comparison%20of%20vocabulary. Accessed on July 26, 2023

[13] Naveen, "Explain Character Vector in R?", SparkBy{Examples}, July 2022,https://sparkbyexamples.com/r-programming/explain-character- vector-in-r/#:~:text=A%20vector%20where%20each%20element,( )%20or%20character()%20functions.Accessed on July 26, 2023

[14] Hema L. Ramkumar, MD, Vinay A. Shah M.D., Koushik Tripathy, MD (AIIMS), FRCS (Glasgow), Zachary Koretz, MD, MPH, Robert A Hyde, MD, PhD, Jennifer I Lim MD, Peter A.Karth, MD, "Electrooculogram," EyeWiki, May 2023, https://eyewiki.aao.org/Electrooculogram. Accessed on July 26, 2023

[15] "Median Filter," Wikipedia, April 2023, https://en.wikipedia.org/wiki/Median_filter. Accessed on July 26, 2023

[16] Bestelmeyer PEG, Kotz SA, Belin P. Effects of emotional valence and arousal on the voice perception network. Soc Cogn Affect Neurosci. 2017 Aug 1;12(8):1351-1358. doi: 10.1093/scan/nsx059. PMID: 28449127; PMCID: PMC5597854.

[17] "About Keras," Keras, https://keras.io/about/. Accessed on July 26, 2023.

[18] Italo José, "Facial mapping (landmarks) with Dlib + python," Medium, June 2018, https://towardsdatascience.com/facial-mapping-landmarks-with-dlib-python-160abcf7d672. Accessed on July 25, 2023.

[19] "What are convolutional neural networks?," IBM ,https://www.ibm.com/topics/convolutional-neural-networks. Accessed on July 26, 2023.

[20] Avijeet Biswal, "Convolutional Neural Network Tutorial," Simplilearn, April 2023, https://www.simplilearn.com/tutorials/deep-learning-tutorial/convolutional-neural-network#:~:text=Flattening%20is%20used%20to%20convert,layer%20to%20classify%20the%20image. Accessed on July 26, 2023.

[21] Priti Rai Jain, S. M. K. Quadri, and Muskan Lalit. 2021. Recent Trends in Artificial Intelligence for Emotion Detection using Facial Image Analysis. In Proceedings of

the 2021 Thirteenth International Conference on Contemporary Computing (IC3-2021). Association for Computing Machinery, New York, NY, USA, 18–36. https://doi.org/10.1145/3474124.3474205

[22] M, Maithri & U, Raghavendra & Gudigar, Anjan & Samanth, Jyothi & Barua, Prabal Datta & Murugappan, Prof & Chakole, Yashas & Acharya, U Rajendra. (2022). Automated Emotion Recognition: Current Trends and Future Perspectives. Computer Methods and Programs in Biomedicine. 215. 106646. 10.1016/j.cmpb.2022.106646.