AI Pioneering Ethical, Analytical and Real time Emotional Recognition in Dynamic Human Expressions

Akhilesh Venkiteswaran
Department of Artificial Intelligence and Data Science, Chennai Institute of Technology,
Chennai 600 069, India
akhileshvenkiteswaran.aids2023@citchennai.net

Dr. D. Jagadiswary
Professor, Department of Artificial Intelligence and Data Science, Chennai Institute of Technology,
Chennai 600 069, India
jagadiswaryd@citchennai.net

Abstract – Facial Emotional Recognition (FER) has emerged an important field in AI, allowing machines to interpret and analyze human emotions through facial expressions. This study dives deep into FER using deep learning, machine learning and real-time AI-based facial recognition methodologies to improvise accuracy, adaptability and adhere to ethical practices. It signifies the limitations of existing FER models, especially in privacy and ethical considerations, real-time analyses. Leveraging CNNs, FER-Former and Graph-based approach, this paper approaches sophisticated FER systems. Furthermore, this study investigates ethical implications of facial data, privacy, encryption and federated learning. The findings also demonstrate the need to include more efficient methods like geospatial information systems (GIS) and extended reality (XR) which helps enhancing applications, which include especially for proctoring systems, human-computer interactions. This paper provides a comprehensive study of state-of-the-art FER models, focusing on key challenges and moving forward for future advancements in privacy concerns, ethical considerations and real-time facial recognition tech.

keywords- Neural Networks, Artificial Intelligence, Facial Expression Recognition, proctoring systems, real-time analysis, Privacy, Encryption

I. INTRODUCTION

Various methods of expression and emotions, from one human to another human and with animals are the major force of highlighting their feelings. The FER is a matter of great importance in computer vision and machine learning since it examines ways to study and recognize microscopic muscle movements in facial structure. It is important to understand and categorize the fundamental human emotions into distinct segments of classification. They are divided into two types - [1]

- Arousal - Is defined how inclined an individual is to behave based on their emotional state.
- Valence - determines how pleasant or terrible a sensation of an emotion.

The exponential growth of the facial expression recognition (FER) methods performed using computer vision, deep learning, and AI has been observed. The scope of this review is to dive deep into the prospects of FER in various aspects, the required techniques and timelines used in various applications and to draw out a conclusion about the ethical considerations that are being followed and finding ways to follow them. Additionally, it seeks to provide understanding of the latest findings in this field of technology and considerations that should be taken for further research and applications.

The below Fig. 1, give us a clear classification of variety of emotions.

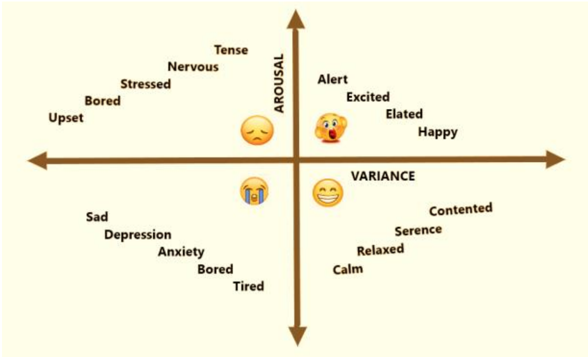

Fig. 1. Arousal-Variance Relation with different emotional expressions.

II. LITERATURE REVIEW

A. Data Collection & Preprocessing

Certain databases were collected, and their study was conducted with respect to each database. Each database had its own properties in showcasing their collection and usage of their data in methodologies which is discussed below.

- **AffectNet** offers more than one million images containing faces with extracted facial landmark points. Approximately half of them were manually labelled, which may lead to potential subjectivity issues, but can identify 6-7 basic human expressions [2]. It requires high computational cost.
- **Extended Cohn-Kanade Dataset (CK+)** image sequences can be analysed for both action units and prototypic emotions. It provides protocols and baseline results for facial feature tracking, AUs, and emotion recognition, with image resolutions of 640x480 and 640x490 at 8-bit precision for grayscale values [3]. Although clear labelling, it has limited expressions covered, and its large dataset causes operational slow processing and has underrepresented groups mostly including only black, Asian and female subjects.
- **FER2013** FER2013 is a large-scale, unconstrained database automatically collected using the Google image search API. FER2013 includes 28,709 training images, 3,589 validation images, and 3,589 test images labelled

with seven expressions anger, disgust, fear, happiness, sadness, surprise, and neutral [4]. However, bias can occur due to lack of age variety.

B. Linear Discriminant Analysis

Linear Discriminant Analysis (LDA) has been used in Facial Expression Recognition (FER) to improve dimensionality reduction by identifying latent topics or features in facial data. LDA is used [5] for labels with a face recognition dataset. LDA make sures that variance of the same category of the data groups after projection is as small as possible and the variance amongst the groups to be as great as possible. The dimensionality reduction of LDA is supervised, and it should use the label data to divide the data of different categories as far as possible. The below Fig. 2 shows the before and after LDA categorize the data and help in dimensionality reduction.

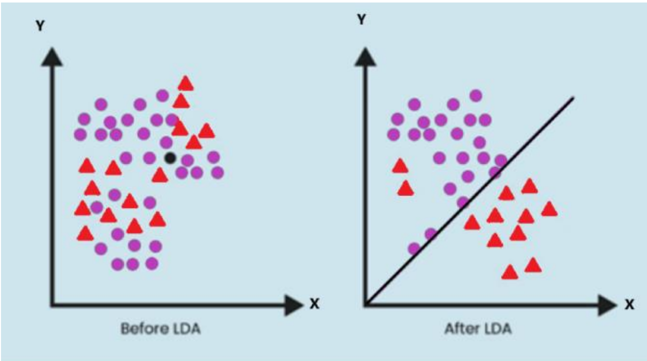

Fig. 2. LDA Graph - Before vs After

III. OVERVIEW OF FACIAL EXPRESSION RECOGNITION AND MACHINE LEARNING DEEP LEARNING AND AI TECHNIQUES

A. Methodologies in FER

There are various methodologies that are used in FER, various architecture of neural networks, that are used in FER.

- Convolutional Neural Networks (CNN)

CNNs are a subclass of Artificial Neural network that specializes in processing grid-like data. Kahou et al. [6] presented an expanding dataset to develop high-capacity models that enhance FER's performance without overfitting. Common preprocessing methods, such as data augmentation, cropping, down sampling, and normalization, typically make CNNs have a robust neural network, with Lopes et al. [7] demonstrating that combining these techniques significantly boosts model accuracy. Konda et al., [8] introduced a zero-bias model for CNN's fully connected layers which helps in the additional optimization of the FER. Some notable research was also conducted in many other methods and datasets, is given In Table 1.

TABLE 1. TECHNIQUES USED IN CNN FOR FER

AUTHOR	DATA SETS	UTILISED TECH	OUTCOME
Y. Khaireddin et al [9]	FER2013	Tuned hyperparameters in optimized model in VGGNet	Accuracy - 73.28%.
Jain Deepak et al [10]	JAFFE & CK+	Combining FCN and residual block cloud improved the overall result and efficiency.	Accuracy Rate - JAFFE - 95.23% CK+ - 93.24%

Fig. 3 shows the structure of CNN. The input layer takes in the input from the user, process the image with the help of algorithms to identify the person and gives the output as shown below.

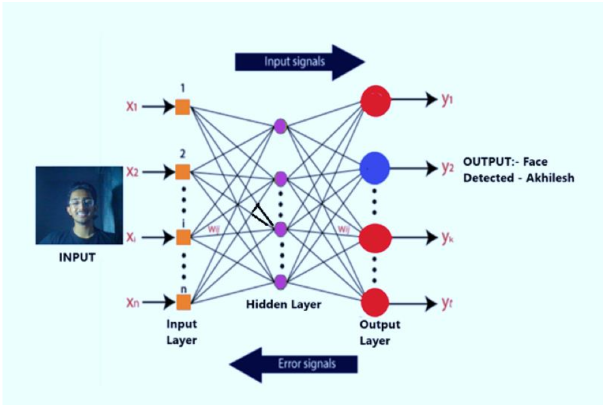

Fig. 3. Architecture of CNN in Deep learning

- Graphical Representation & Characteristics for Facial Expression Recognition

Graphical representation, commonly called as graphs, serve as modelling tools that captures complex relationships between facial features and expressions, allowing the machine to have a fine-grained analysis of facial affect. A Review was conducted by Y. Liu et al [11] where they had researched the various ways by which graphs are used. The following Table 2 helps in understanding the graphs as given below.

TABLE 2. VARIOUS TYPES OF GRAPHICAL REPRESENTATIONS

GRAPH NAME	CHARACTERSTICS & FINDINGS
Graph Based Representation	• Region-Level Graph - Focussed on Region of Interest (ROI), which primarily took eyes and mouth. Some graph models built NPI (Non-Prior Information), graphs which uses without landmark reliance allowing flexible distinction properties [12]. • Landmark-Level Graph - Nodes uses 68-point marking system for creating a detailed analysis of the expression with advanced technologies [13].
AU Label Graphs	• It uses data-driven, symmetrical adjacency matrices for high-dimensional AU node attributes, correcting existing labels and generating unknown labels. [14].
AU-Map Graphs	• Represent multiple-AUs at the same Region of Interest (ROI) by leveraging the local

	feature maps. Employ deep features using ResNet, integrate structured learning for complex FAA with AU intensity estimation using CNN and Bayesian Network. [15].

The FER using graphs are shown in Fig. 4 as A) Happy, B) Sad, C) Neutral face, D) Angry.

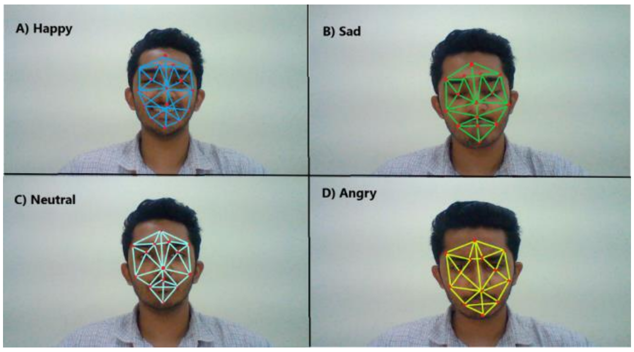

Fig. 4. Graphical Representation of Facial Structure

- FER-Former

Li, Y., Wang et al, proposed a FER-former architecture which is one of the latest forms of the facial expression recognition [16]. The FER-former is a cutting-edge latest transformer-based model designed to enhance facial expression recognition in real-world environments. It uses a hybrid self-attention scheme in its FER-specific encoder to deal with the typical ambiguity in facial expression data (images and text annotations) using stepwise linear discriminant analysis, processing tokens that include both one-hot and text semantics for richer understanding, thereby giving a much better cutting-edge advantage compared to traditional LSTM-CNN Hybrid model. This model was referred to as the multi-model supervision. Their approach consists, hybrid self-attention scheme, multi-granularity embedding integration, and heterogeneous domain steering supervision. Results confirm that FER-former outpaced other state-of-the-art methods in FER tasks. The model utilised pretrained IR-50 CNN as a feature extractor, by using a custom encoder to process images and text features together. The MGEI (Multi-granularity Embedding Integration) model was used to produce spatial feature patches so it could handle diverse scale. The HDSS Module (Heterogeneous Domains-Steering Supervision Module) calculated the similarities between texts and images so the machine can have less ambiguous classification problems, the authors proposed the inclusion of this technique referencing from. The datasets used for the research on FER-Former is shown in Table 3.

TABLE 3. DATASETS USED FOR FER-FORMER

DATASET	ACCURACY ACHIEVED
RAF DB Dataset	91.3%
FERPlus Dataset	90.96%
SFEW 2.0 Dataset	62.18%

Fig. 5 shows the step-by-step process followed in FER-Former to convert an input into an output.

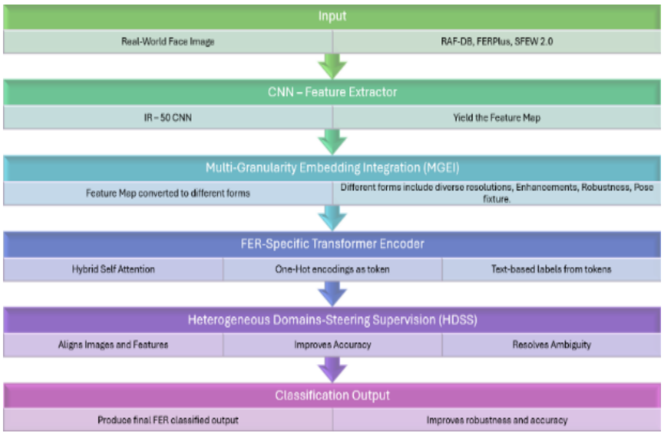

Fig. 5. FER-Former Process

- Geospatial Information Systems, Artificial Intelligence & Extended Reality

A study was conducted by Somaiieh Rokhsaritalemi et. al., [17] on creating an FER, that utilizes AI, Geospatial information (GIS) and Extended reality (XR). They tried and tested to create a model by combining the three factors together, where previous studies had only focussed on each aspect.

Fig. 6 shows the next-gen technology is being widely used to create intelligent emotional analysis and facial expressions systems, leading to a very powerful, robust, complex and widely available machines to capture and analyse human emotions.

Fig. 6. Multi-Model Architecture of AI-GIS-XR FER machines.

The diagram Fig. 7 explains the exact structure and process of emotional analysis in FER. The privacy is protected by anonymizing the data or encrypting them, finally generating an analysis of the emotion. Rule based methods, videos and audio are sent to machine learning and deep learning models where the processing of data takes place. Privacy of data is ensuring by anonymizing it, or by encrypting the data to avoid misuse of data. Generation of output is an important step after the model is able to predict the emotional state.

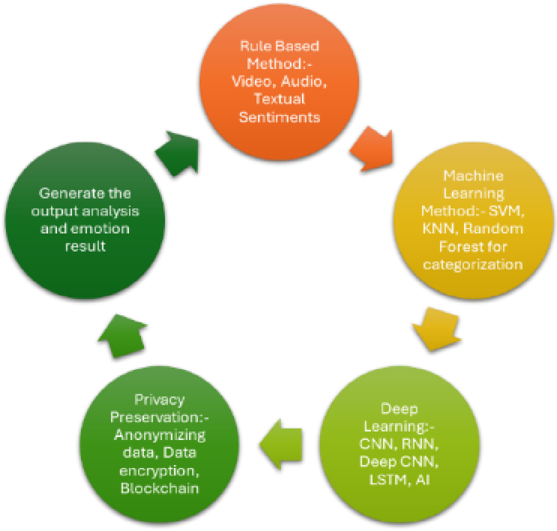

Fig. 7. Structure of Emotional Analysis in FER.

- Support Vector Machine

Support Vector Machines (SVM) is a supervised machine learning model that uses the classification method to classify the objects and data. Research conducted by M. Rashad et.al., [18] had found that by using hybrid models with SVM and other deep learning tech like CNN, these models incorporated feature optimization techniques demonstrated a better balance of speed and performance in real time applications.

Fig. 8 is where the machine was able to categorize the expression Anger and Happy, all the blue dots consisted of the happy expression and all the red dots were categorized as Anger.

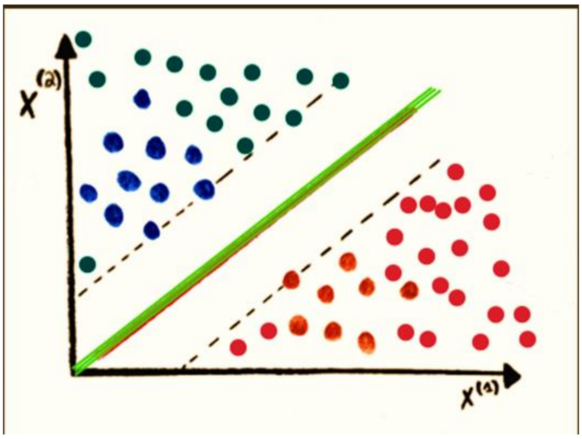

Fig. 8. Support Vector Machine Categorization.

IV. *GAP IN REVIEW ANALYSIS*

From the research studies and reviews there is a gap in real time analysis of facial expressions. A real time analysis should assist in detecting at the point of time, an individual was undergoing through the type of emotions at various intervals.

Proctoring Systems are used to detect the student's emotional status and to check if they are involved in any sort of malpractices. The real time analysis can help in finding out at what point of time did the student had performed any suspicious activity in the middle of the examination.

Manichaeus et. al., [19], proposed the idea of using a communicative method that used the emotional status of patients during a video call. M.Misuraca et. al., [20] proposed the implementation of library packages like R language, for creating analysis of live video.

From the above review and understandings, an image with a proposed idea is shown in Fig. 9. It explains the facial expression systems to be used for real time analysis. The person is trying to cheat by using an unidentified object) that is a mobile phone or a table or any other object) indicated by red line. The red line when click will be helpful for the proctor to understand and decide on when and if the candidate taking the examination has done any malpractice. The red line will be marked in real-time making sure that the machine is able to detect whatever suspicious activity it finds and proctors need not take a lot of time of reviewing the footage.

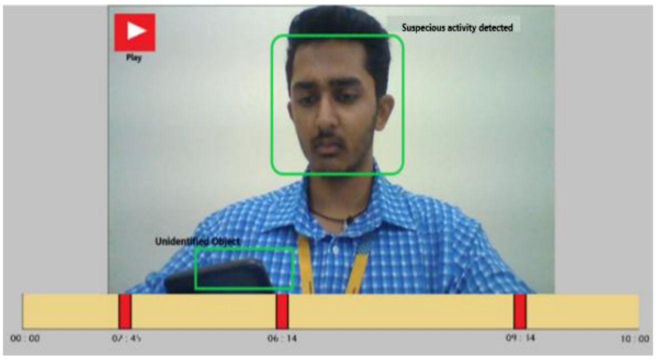

Fig. 9. Real time Analysis

V. *ETHICAL CONCERNS OF LIVE VIDEO USING FER.*

A. *Privacy Concerns*

There are a lot of privacy concerns, related to facial data extraction, such as voices, facial structures, saving information related to user ID on third-party servers has always been a major concern. Anwar et. al., [21] projected an improved system using EEG data for privacy-preserving emotion detection. Their system intricated the usage of (ANN) that is, Artificial neural network, to organise emotional expressions based on the 3D classification of arousal, dominance, and valence graph. S. Latif et.al., [22] proposed on the idea of using a federated learning model to address privacy concerns for detecting the emotional status of a person via speech recognition. Yuta Nakashima et. al., [23] proposed an image melding-based approach that transforms facial regions that is visually not intrusive while preserving the emotional expressions. Peter Lewinski et. al., [24] has cut through the system of AFRS (Automated Facial Recognition System), which standardizes and follows the EU data protection laws by the means of anonymizing of information and ensuring consent. X.Wang et. al., [25] suggested the use of multi-model discriminate analysis on dataset and using normalization technique with the help of

AAM algorithm. Honggu Liu et. al., [26] proposed the method to use bidirectional protection approach, where two ways, Face-out-detection (FOD) ensuring detectability and Face-in-forensics (FIF) ensuring traceability when protected facial data is trying to be replaced and swapped respectively. The authors also discussed a watermarking method – Fragile Watermarking (for FOD) and Robust Watermarking (for FIF) as a defensive mechanism, to avoid potential use of that data into deepfakes. It had also achieved an accuracy of 95% success rate when it came to avoid usage of facial data in deepfakes.

VI. CHALLENGES & FUTURE DIRECTION

From the analysis of the above case studies, several key challenges in the development and the implementation of FER systems were identified. Addressing them is essential to ensure the continued advancement and widespread adoption in these technologies. Firstly, the computational cost, FER systems require robust, efficient and a lot of computational resources. Simplification and optimization of system needs to be done. Secondly, user-friendly reporting needs to be done to ensure analysis reports are made easy to understand and professional. Additionally, efficiency also plays a significant role to provide faster, better and smarter systems. Lastly, overfitting is a major drawback, as noisy data can lead to machines while taking wrong decisions. Addressing these challenges, the computational cost, user-friendly reporting, data privacy, error management, efficiency, and overfitting will pave the way for the future improvements of more robust, ethical and effective FER systems capable of meeting crucial demands.

VII. ABBREVIATIONS

VGGNet – Visual Geometry Group Network.
KDEF - Karolinska Directed Emotional Faces database.
AAM – Active Appearance Model
FERM – Facial Expression Recognition and Modality.
EDA – Electrodermal Activity
EGG – Electrogastrograph,

REFERENCES

[1] Kopalidis, T., Solachidis, V., Vretos, N., & Daras, P. (2024). Advances in Facial Expression Recognition: A Survey of Methods, Benchmarks, Models, and Datasets. Information, 15(3), 135. https://doi.org/10.3390/info15030135

[2] A. Molla Hosseini, B. Hasani, and M. H. Mahoor, “Affectnet A database for facial expression, valence, and arousal computing in the wild,” IEEE Transactions on Affective Computing, vol. PP, no. 99, pp. 1–1, 2017.

[3] Lucey, P.; Cohn, J.F.; Kanade, T.; Saragih, J.; Ambadar, Z.; Matthews, I. The extended Cohn-Kanade Dataset (CK+) A complete dataset for action unit and emotion-specified expression. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, San Francisco, CA, USA, 13–18 June 2010; pp. 94–101.

[4] I. J. Goodfellow, D. Erhan, P. L. Carrier, A. Courville, M. Mirza, B. Hamner, W. Cukierski, Y. Tang, D. Thaler, D.-H. Lee et al., “Challenges in representation learning A report on three machine learning contests,” in International Conference on Neural Information Processing. Springer, 2013, pp. 117–124.

[5] Turk, M.; Pentland, A. Eigenfaces for recognition. J. Cogn. Neurosci. 1991, 3, 71–86.

[6] Kahou, S.E.; Pal, C.; Bouthillier, X.; Froumenty, P.; Gülçehre, Ç.; Memisevic, R.; Vincent, P.; Courville, A.; Bengio, Y.; Ferrari, R.C.; et al. Combining modality specific deep neural networks for emotion recognition in video. In Proceedings of the 15th ACM on International Conference on Multimodal Interaction, Sydney, Australia, 9–13 December 2013; pp. 543–550.

[7] Lopes, A.T.; De Aguiar, E.; De Souza, A.F.; Oliveira-Santos, T. Facial expression recognition with convolutional neural networks Coping with few data and the training sample order. Pattern Recognit. 2017, 61, 610–628.

[8] Konda, K.; Memisevic, R.; Krueger, D. Zero-bias autoencoders andthe benefits ofco-adapting features. arXiv 2014, arXiv 1402.3337.

[9] EEE, 2017. [9] Khaireddin, Yousif, and Zhuofa Chen. "Facial emotion recognition State of the art performance on FER2013." arXiv preprint arXiv 2105.03588 (2021).

[10] Jain, Deepak Kumar, Pourya Shamsolmoali, and Paramjit Sehdev. "Extended deep neural network for facial emotion recognition." Pattern Recognition Letters 120 (2019) 69-74

[11] Y. Liu, X. Zhang, Y. Li, J. Zhou, X. Li, and G. Zhao, "Graph-Based Facial Analysis A Review," in IEEE Transactions on Affective Computing, vol. 14, no. 4, pp. 2657-2677, 1 Oct.-Dec. 2023, doi 10.1109/TAFFC.2022.3215918.

[12] Z. Zhang, T. Wang, and L. Yin, “Region of interest-based graph convolution A heatmap regression approach for action unit detection,” in Proc. ACMInt.Conf. Multimedia,2020, pp.2890–2898

[13] R. Zhao, T. Liu, Z. Huang, D. P.-K. Lun, and K. K. Lam, “Geometry-aware facial expression recognition via attentive graph convolutional networks,” IEEE Trans. Affect. Comput., tobe published, doi 10.1109/TAFFC.2021.3088895.

[14] Z. Cui, Y. Zhang, and Q. Ji, “Label error correction and generation through label relationships,” in Proc. AAAI Conf. Artif. Intell., 2020, pp. 3693–3700.

[15] R. Walecki, O. Rudovic, V. Pavlovic, B. Schuller, and M. Pantic, “Deep structured learning for facial expression intensity estimation,” Image Vis. Comput., vol. 259, pp. 143–154, 2017.

[16] Li, Y., Wang, M., Gong, M., Lu, Y., & Liu, L. (2023). Fer-former Multi-modal transformer for facial expression recognition. arXiv preprint arXiv 2303.12997.

[17] Rokhsaritalemi, S., Sadeghi-Niaraki, A., & Soo-Mi Choi, S. M. (2023). Exploring emotion analysis using artificial intelligence, geospatial information systems, and extended reality for urban services. IEEE Access.

[18] Rashad, M., Alebiary, D. M., Aldawsari, M., El-Sawy, A. A., & AbuEl-Atta, A. H. (2024). CCNN-SVM automated model for emotion recognition based on custom convolutional neural networks with SVM. Information, 15(7), 384.

[19] Menychtas, M. Galliakis, P. Tsanakas, and I. Maglogiannis, “Real time integration of emotion analysis into homecare platforms,” in Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC), Jul. 2019, pp. 3468–3471.

[20] M. Misuraca, A. Forciniti, G. Scepi, and M. Spano, “Sentiment analysis for education with R Packages, methods and practical applications,” 2020, arXiv 2005.12840.

[21] M. A. Anwar, M. Agrawal, N. Gahlan, D. Sethia, G. K. Singh, and R. Chaurasia, “FedEmo A privacy-preserving framework for emotion recognition using EEG physiological data,” in Proc. 15th Int. Conf. Commun. Syst. Netw. (COMSNETS), Jan. 2023, pp. 119–124.

[22] S. Latif, S. Khalifa, R. Rana, and R. Jurdak, “Federated learning for speech emotion recognition applications,” in Proc. 19th ACM/IEEE Int. Conf. Inf. Process. Sensor Netw. (IPSN), Apr. 2020, pp. 341–342.

[23] Y. Nakashima, T. Koyama, N. Yokoya and N. Babaguchi, "Facial expression preserving privacy protection using image melding," 2015 IEEE International Conference on Multimedia and Expo (ICME), Turin, Italy, 2015, pp. 1-6, doi: 10.1109/ICME.2015.7177394.

[24] Lewinski, P., Trzaskowski, J., & Luzak, J. (2016). Face and emotion recognition on commercial property under EU data protection law. Psychology & Marketing, 33(9), 729-746.

[25] Wang, X; Xiong, C; Pei, Q; Qu, Y (2018). Expression preserved face privacy protection based on multi-mode discriminant analysis. Deakin University. Journal contribution. https://hdl.handle.net/10536/DRO/DU:30115189

[26] Honggu Liu, Xiaodan Li, Wenbo Zhou, Han Fang, Paolo Bestagini, Weiming Zhang, Yuefeng Chen, Stefano Tubaro, Nenghai Yu, Yuan He, and Hui Xue. 2023. BiFPro: A Bidirectional Facial-data Protection Framework against DeepFake. In Proceedings of the 31st ACM International Conference on Multimedia (MM '23). Association for Computing Machinery, New York, NY, USA,

Akhilesh Venkiteswaran is an undergraduate student at Chennai Institute of Technology, specializing in Artificial Intelligence and Data Science. With a strong academic foundation and practical experience in AI and machine learning, he has developed expertise in areas such as data analytics, natural language processing, and software development. As an enthusiastic researcher, he is currently exploring advanced topics like facial expression recognition in real time video streams and the application of AI in Human Resource Management. Akhilesh is also a top 10 finalist in the Green FinTech Hackathon conducted by RBI-H IIT Jodhpur TISC. Through various projects, internships, and research initiatives, he is building a solid foundation in AI while contributing to cutting-edge technologies and methodologies.

E-Mail Id akhileshvenkiteswaran.aids2023@citchennai.net, akhileshvenki05@gmail.com

Dr. D. Jagadiswary M.Tech., Ph.D., Professor in the Department of Artificial Intelligence and Data Science, Chenai Institute of Technology, Chenai. She received her Ph.D through QIP, AICTE from Pondicherry university. She received the bachelor's degree in Electronics and Communication and master's degree in information security from Pondicherry University. She is a life member of ISTE. She published 12 International Journals, 2 National Journals, 24 International Conferences and 4 Patents. Her research interests include Machine Learning, Security Analytics, Logic Circuit Design, Electromagnetics, Cryptography and Network Security, Formal Languages and Automata Theory.

E-Mail Id jagadiswaryd@citchennai.net, jagadiswary@ptuniv.edu.in