Indonesian Journal of Electrical Engineering and Computer Science
Vol. 37, No. 2, February 2025, pp. 959~967
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v37.i2.pp959-967

A novel CNN-ANN fusion approach for improved facial emotion detection

Viraj Sawant[1], Husna Shaikh[1], Bhakti Palkar[2], Sanam Kazi[3], Wasim Jasani[1], Lamiya Rampurawala[1], Mohammed Naser Shaikh[1]

[1]Department of Information Technology, M. H. Saboo Siddik College of Engineering, Mumbai, India
[2]Department of Computer Engineering, K. J. Somaiya College of Engineering, Mumbai, India
[3]Department of Engineering and Technology, Bharati Vidyapeeth Deemed to be University, Navi Mumbai, India

Article Info

Article history:

Received Jun 20, 2024
Revised Sep 15, 2024
Accepted Sep 29, 2024

Keywords:

Deep learning
Emotion classification
Facial emotion recognition
Late fusion deep learning
Machine learning

ABSTRACT

In recent years, the field of emotion recognition has witnessed an increased interest due to the rise of deep learning techniques. However, one of the persistent difficulties in this domain, which we have attempted to address, is the variability in image sizes utilized. In this study, we have reviewed the work by different researchers and summarized their key findings. In our research, we introduce a novel technique that integrates the strengths of 1D convolutional neural networks (CNNs) and artificial neural networks (ANNs) through a late fusion model, leveraging CNNs' shared weights and automatic feature learning for spatial and temporal data, alongside ANN's comprehensive feature consideration. Our research findings highlight the effectiveness of this approach, which achieves a remarkable accuracy of 92.42%, along with other evaluation metrics demonstrating notable results. Furthermore, we conduct a comprehensive analysis of the proposed method, comparing it with advanced methods in the field of facial emotion recognition. Through this comparative analysis, we demonstrate the superiority of our proposed approach, addressing challenges that have not yet been addressed till date, thus leading to progress in this field.

Corresponding Author:

Husna Shaikh
Department of Information Technology, M. H. Saboo Siddik College of Engineering
Mumbai – 400008, Maharashtra, India
Email: husnas2933@gmail.com

1. INTRODUCTION

The growth of artificial intelligence makes emotional intelligence an area which is being paid more and more attention to recently. This would aim to obtain an interaction between humans and computers which are similar to those between people. However, this requires machines to understand human emotions, and facial emotion detection serves as a method to reach the goal [1]. Interpreting facial expressions and emotions is seen as a crucial step for creating more natural interactions between humans and machines.

Emotions can be considered as a complex phenomenon of a subject's sensitivity in relation to their environment and specific points of adaptation, being a mixture of behavioural, physiological, and expressive elements, feelings, and nervous system responses appearing automatically and being, partly influenced by cultural factors [2]. Emotions play a pivotal role in various facets of human life, including interpersonal relationships, cognitive processes, perception, and overall well-being [3]. The need to detect an individual's emotions has become increasingly important for many reasons. This stems from the fact that emotions serve as integral components of social interaction. According to research [4], only 7% of human communication relies on language, while 38% is conveyed through paralanguage, and a significant majority, 55%, is communicated through facial expressions.

While emotions can be recognized through EEG, speech or text [5], facial expression serves as a fundamental and distinctive method by which humans communicate their emotional states and intentions, surpassing cultural and linguistic boundaries [6]. With the advent of machine learning and its continuous development, facial expression recognition (FER) systems are being actively researched and built keeping in mind various environmental and sociological factors and the applications of these systems. Researchers use ML and DL algorithms or even a combination of them for FER tasks [7].

Ekman and Friesen [8] recognized six fundamental emotions namely: anger, fear, disgust, happiness, sadness, and surprise through a cross-cultural study [9] which implicated that these emotions are universally recognized regardless of culture [10]. Additionally, a neutral emotion can also be added, summing up the classification of emotions into 7 broad categories [11], [12].

A number of researches have been done in the field of FER [10], [13]. Majority of these works leverage artificial intelligence methods such as machine learning, deep learning and ensemble techniques as well for FER. In the field of computer vision and ML, there is a dedicated drive to generate and study the FER systems that can accurately recognize expressions from face image data.

FER systems usually follow a general sequence which includes face recognition, extraction of features, feature representation and lastly the classification of emotions. The face detection step includes detecting and locating faces within an image or a video frame. Once the faces are detected, the extraction of required features such as facial landmarks, shapes and movements is performed. Then, extracted features are transformed or represented in a suitable format which can be recognized by machine learning systems. Lastly, features which have been transformed are input into a classification algorithm, like neural networks [14] or support vector machines [7], to classify the detected facial expression into predefined emotion categories. The steps mentioned above employ various algorithms to complete each task. The recent review papers on facial emotion recognition were reviewed in accordance to the datasets used, proposed methodology and the outcomes of the method as shown in Table 1.

In our review of FER research, we found a few challenges. Firstly, most datasets used in studies are often small, which limits the diversity of facial expressions captured. Our study aims to tackle these issues by using data augmentation techniques to enlarge the dataset's size so as to handle various environmental factors and facial features. Additionally, most studies utilize images of extremely small size for FER. However, we solve this by resizing the image to a higher suitable size. We also employ the use of histogram of oriented gradient (HOG) for the extraction of relevant features from the data. Moreover, while many FER systems have been developed using state-of-the-art techniques there is still need for further improvements in their performance. Our approach addresses all these gaps.

Table 1. Comparison of existing FER methods

Paper	Dataset	No. of classes	Proposed approach	Accuracy
[15]	FER2013, CK+	7	Hybrid CNN with dense scale invariant feature transform aggregator	FER2013 - 73.4% CK+ - 99.1 ± 0.07
[16]	SEED, CK+	7	Spatial-temporal recurrent neural network (STRNN)	SEED - 89.50% CK+ - 95.4%
[12]	CK+, JAFFE	7	Hierarchical deep learning structure	CK+ - 96.46% JAFFE - 91.27%
[5]	FER2013, CK+, FERG, JAFFE	7	Attentional CNN	FER2013 - 70.02% CK+ - 98.0% FERG - 99.3% JAFFE – 92.8%
	FER2013, CK+, FER+, RAF	7	Multiple branch cross-connected convolutional neural network (MBCC-CNN)	FER2013 – 71.52% CK+ - 98.48% FER+ - 88.10% RAF – 87.34%
[17]	FER2013, CK+	7	Partial transfer learning-based CNN	FER2013 - 82.1% CK+ - 66.7%
[18]	FER2013	7	Ensemble of mini-Xception models	75%
[19]	RAF, AffectNet	7	novel fusion method and ensemble learning	RAF – 91.66% AffectNet- 72.06%
[20]	FER2013	7	Convolutional neural network ensemble	71.27%
[21]	FER2013, AFEW 7.0	7	Temporal ensemble of multi-level CNN	FER2013 - 74.09% AFEW 7.0- 49.3%
[22]	AFEW 7.0, RAF	7	Multi-region ensemble CNN	RAF – 76.73% AFEW – 47.43%
[23]	EmotiW 2017 dataset	7	Supervised scoring ensemble	60.34%

The significant contributions of the authors are as follows:

a) Proposed the fusion model consisting of artificial neural network (ANN) and 1D-CNN for FER which combines the shared weights and feature learning of 1D-CNN with ANN's ability to process all features, enabling automatic learning of spatial and temporal data for effective hierarchical classification of facial emotions.
b) Conducted hyperparameter tuning to enhance the performance and comparative analysis of the model.
c) Comparative result analysis

The subsequent sections of the paper are ordered as: section 2 outlines the implementation procedure; section 3 introduces the proposed methodology; section 4 showcases experimental findings and analysis; and section 5 offers the paper's conclusions.

2. PROCEDURE

2.1. Experiment environment

In this research, we developed and experimented on the proposed system and several other comparative algorithms using the computing resources provided by Google Collaboratory [24]. To construct and prototype the models, we relied on Python [25], including Keras, TensorFlow, Scikit-learn, and Matplotlib. Together, these tools allowed for efficient experimentation and analysis for facial emotion recognition.

2.2. Dataset

In our study, we utilize the Karolinska directed emotional faces (KDEF) [26] for the training and assessment of our methods. The KDEF dataset consists of a wide-ranging collection of 4900 images capturing facial expressions. They feature 70 different individuals, evenly split between 35 females and 35 males, who were selected based on specific criteria, of age 20-30 years and with an absence of facial hair, accessories, or makeup during the photo sessions.

The subjects accurately portray seven distinct emotional expressions namely neutral, happy, angry, afraid, disgusted, sad, and surprised. Each subject's face was photographed from 5 different angles. The images are of standardized dimensions of 562 by 762 pixels.

For the purpose of our study, we have solely included the images that have been captured from a straight angle which sum up to 720 samples in total. Additionally, we have also resized the images to a size of 224 by 224 pixels for our research as compared to other studies that employ the use of images that are smaller in size. Samples from the dataset are shown in Figure 1.

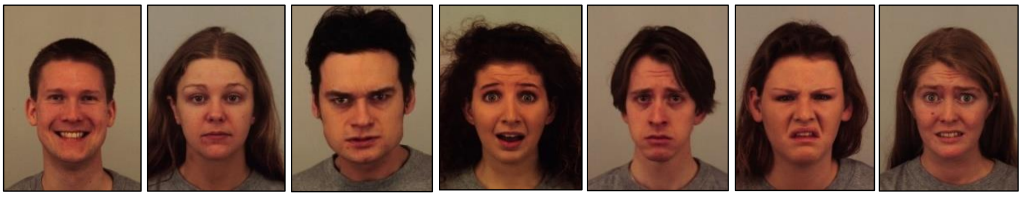


Figure 1. Examples from the KDEF dataset

2.3. Face detection

Face detection stands at the forefront for recognition of facial expressions, being the initial step in understanding and analyzing facial expressions. In our research, we employ a popular and powerful algorithm, Haar Cascade [27] which is based on the Viola Jones algorithm [28]. OpenCV [29] library was utilized to implement Haar Cascade. Furthermore, after the faces are detected, the images are cropped so as to focus on the faces. This model efficiently detects faces in the provided images and crops them. An illustration depicting the detected and cropped images is provided in Figure 2.

Figure 2. Cropped image using haar cascade

2.4. Augmentation

The KDEF dataset contains limited samples due to which under-fitting/over-fitting or limited generalization could take place. The 6137 images were produced from the dataset of 720 original images in batches of varying sizes depending on the different emotion classes. The count of images in each class after augmentation and addition to the original dataset is given in Figure 3. The parameters applied on the images to perform augmentation are mentioned as shown in Table 2. Figure 4 illustrates a sample image after data augmentation was applied.

Table 2. Augmentation parameters

Parameters	Values
Rescale	1./255
Rotation range	10 degrees
Brightness range	0.8, 1.2
Height shift range	10
Width shift range	10
Zoom range	0.1

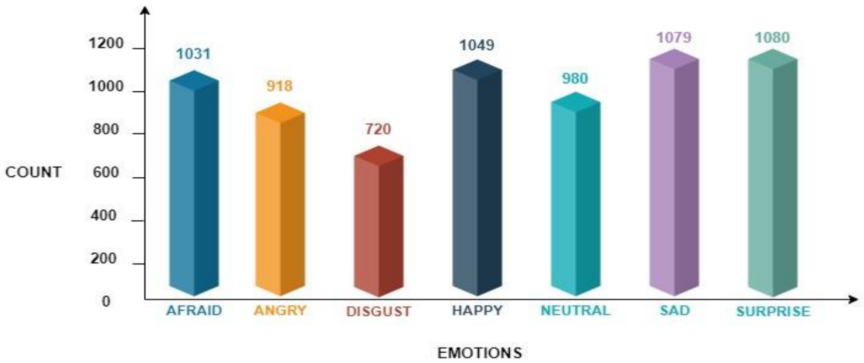


Figure 3. Count of images after augmentation

Figure 4. Original to augmented image

2.5. Feature extraction

To transform raw image data into a more informative representation that captures relevant patterns and characteristics from the image, we have utilized the HOG [30] methodology. This technique dissects each image into smaller regions, capturing the patterns of gradient changes in these regions. By studying these changes in the distribution of gradients, HOG extracts the texture and shape variations from the images. Additionally, HOG was employed separately for each colour channel so as to preserve crucial information about the image's colour composition. These features from different channels are later concatenated into a single array. The parameters used for feature extraction are given in Table 3.

Table 3. HOG parameters for feature extraction

Parameter	Description	Value
Orientations	Number of bins to divide gradient directions in each cell	9
Pixels per cell	Size of each block where gradients are computed	(8,8)
Cells per block	Number of blocks in which gradients are calculated	(2,2)
Visualize	Generate a visual representation of the HOG features	True
Multichannel	Indicate if the image has multiple color channels or grayscale	False

3. METHOD

The proposed architecture consists of 2 independent deep learning models, a 1D convolutional neural network (CNN) and an ANN. These models operate in parallel, extracting high-level features from their respective inputs. Integrating both the models helps in learning complex patterns, with CNNs selecting relevant features and ANNs processing all features. This hierarchical approach enhances the accuracy and robustness of facial emotion recognition. The HOG filter extracts local texture features and shape information from images. Each model is designed to capture different aspects of input data, optimizing emotion recognition classification tasks. 1D-CNN arranges features based on assigned weights whereas the ANN has individual weights for each neuron. CNN1D is used to process one-dimensional data as compared to CNN2D which processes two-dimensional data [31]. The architecture comprises of four convolutional layers, each utilizing the ReLU activation function, followed by a Max Pooling layer and a dropout rate of 0.5 for regularization purposes. Then, a Flatten layer converts the output into a 1D array before concatenating to form the fusion model. The architecture of the model is shown in Figure 5.

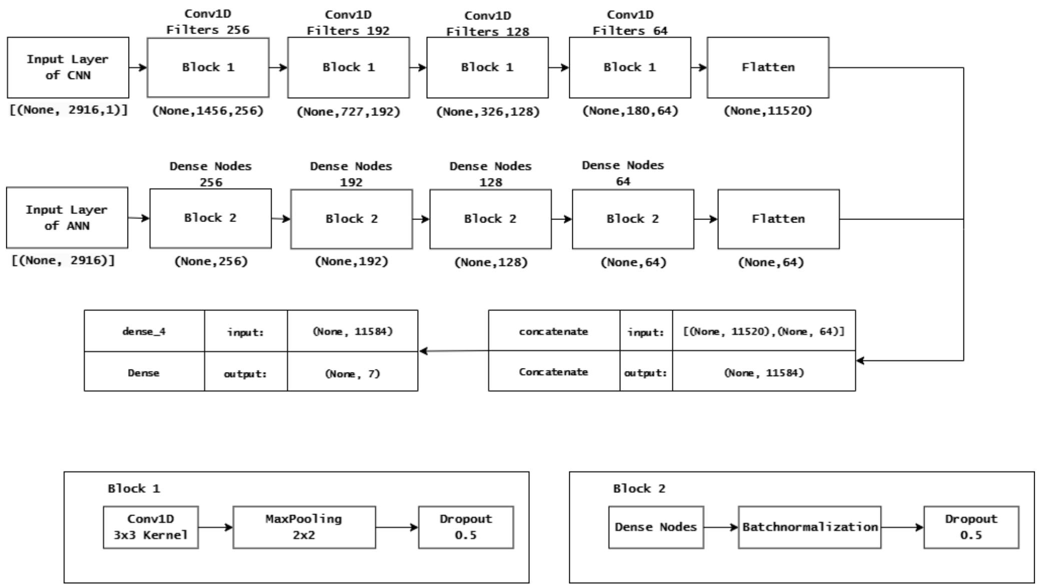


Figure 5. Architecture of ANN-CNN fusion model

The high-level features are fetched from the last fully connected layers. This layer generates feature vectors. These features extracted are concatenated to form a fully combined feature vector which are then arranged according to their weights and passed to the fusion layer where the representations learned by the two networks are combined. The fusion layer allows the models to use their combined strengths while mitigating individual limitations for classification tasks. The model is optimized by using the loss function such as categorical cross-entropy, sparse categorical cross-entropy to minimize the mean squared error. Also, various regularization techniques are used such as dropout and Batch normalization to prevent the model from overfitting. The Adam optimizer [32], which is considered as one of the best for optimization tasks, is used to compile the model. An 80-20 split has been performed for training and validation purposes. The model is trained for a 100 epochs with a batch size of 64. Additionally, early stopping has been implemented with a patience of 10 epochs to avoid overfitting. Facial emotion recognition involves the non-linear relationship between the data points extracted due to difference is pose, lighting condition, structure and size of face for each individual. ANN model proves to be a good model to deal with such non-linear relationships, without knowing the exact function.

This Algorithm 1 describes the process for developing an ensemble model for emotion classification using two neural network types: a convolutional neural network (CNN) and an artificial neural network (ANN). It begins by preparing the data and splitting it into training and testing sets. Both models are trained on the same features, and their predictions are combined to enhance classification accuracy. Finally, the trained model is evaluated and used to classify emotions into seven categories.

Algorithm 1. Proposed model
```
Input: Number of samples (1440, 2916) - 80% for training and 20% for testing.
Output: Trained fusion model classifier for emotion classification
BEGIN
       X_CNN: Data for CNN (2916, 12)
```

```
        X_ANN: Data for ANN (2916, 12)
        y: Training labels (1152, 7)
        Base Learners = [CNN_model, ANN_model]
        Meta- Learner (ensemble model):
               Ensemble_model: CNN_model + ANN_model
        Train Neural Network:
        CNN_pred, ANN_pred
        For i=1: num Epochs
               Ensemble_model
               merge = np.concatenate([CNN_pred, ANN_pred])
               output = Dense (7, activation ='softmax')
        return Output
END
        Evaluate and Test Neural Network
        Classify and Identify Emotions.
```

4. RESULTS AND DISCUSSION

4.1. Experiment results

The fusion model demonstrates a notable accuracy of 92.42%. The training and validation accuracy and loss graphs are depicted in Figures 6 and 7. The model's confusion matrix is presented in Figure 8. Table 4 depicts the evaluation metrics for the model. Additionally, the accuracy, precision, recall and F1-score for each class is given in Table 5.

Table 4. Evaluation metrics per class

Class Labels	Accuracy	Precision	Recall	F1-Score
Afraid	0.93	0.91	0.85	0.88
Angry	0.92	0.91	0.92	0.92
Disgusted	0.88	0.85	0.87	0.86
Happy	0.98	0.93	0.97	0.95
Neutral	0.93	0.98	0.93	0.95
Sad	0.86	0.98	0.97	0.98
Surprised	0.97	0.90	0.93	0.91

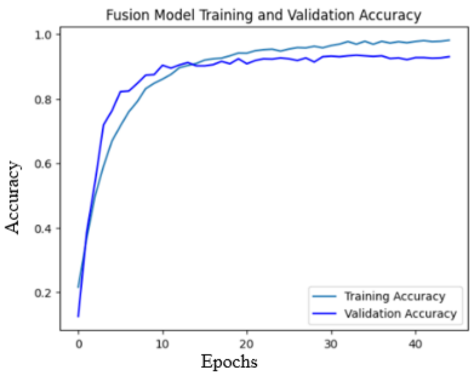


Figure 6. Training and validation accuracy

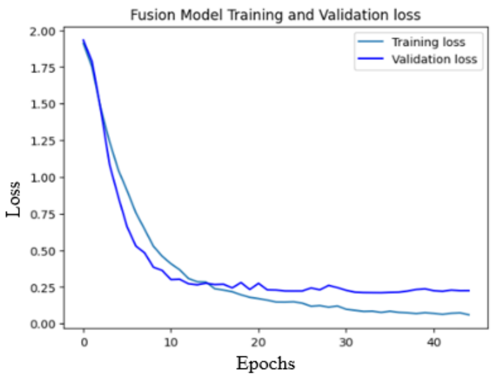


Figure 7. Training and validation loss

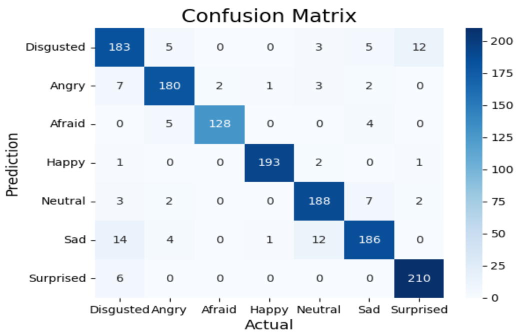


Figure 8. Confusion matrix

Table 5. Evaluation metrics of fusion model

Metric	Value
Accuracy	0.9242
Precision	0.9281
Recall	0.9253
F1-Score	0.9264

4.2. Comparative analysis

A summary of the comparison is given in Table 6. Through our analysis, we have observed that our proposed model demonstrates an accuracy higher than the other models utilized for FER. This validates the usefulness of our fusion model in the field and its effectiveness in discerning the various emotions. In addition, through experimentation, we have noticed that the fusion model has a higher accuracy than the individual ANN and CNN models with 90.01% and 91.47% accuracy respectively. Through the acquired results and comparative analysis, it is observed that our novel ANN-CNN fusion model generates better accuracy and outcomes.

Table 6. Comparison of models with existing methods

Author	Dataset	Method	Accuracy
Zhang *et al.* [16]	SEED	STRNN	89.50%
Shehada *et al.* [17]	FER2013, CK+	Partial transfer learning	82.1%, 66.7%
Ullah *et al.* [19]	RAF-DB, AffectNet	Deep ensemble model	91.66%, 72.06%
Jia *et al.* [20]	FER2013	CNN ensemble	71.27%
Nguyen *et al.* [21]	FER2013, AFEW 7.0	Multi-level CNN ensemble	74.09%, 49.3%
Hu *et al.* [23]	EmotiW 2017	Supervised Scoring ensemble	60.34%
Proposed Method	KDEF	Fusion model	92.42%

5. CONCLUSION

Through our study, we have undertaken a comprehensive investigation into methodologies for FER that have been carried out by researchers in the past. Our research aims to address the challenges of image size, dataset size and variability and extraction of features in this domain, marking a significant step forward in tackling this issue. Our primary contribution lies in the introduction of a novel model combining ANN and CNN through late fusion. This model is fed with an input of the feature extracted using the HOG technique. Through careful architectural considerations and hyperparameter tuning, the Fusion model has demonstrated exceptional performance, achieving an impressive accuracy of 92.42%. Furthermore, it has proven to provide precision, recall, and F1-Scores that are noteworthy, underscoring its efficacy in accurately discerning emotions from facial images. Moreover, by utilizing late fusion technique, we have successfully integrated raw data and feature vectors extracted from different modalities, resulting in a model with enhanced classification accuracy. Moreover, in our comparative analysis against existing methodologies, our model has outperformed state-of-the-art approaches across various datasets. This highlights the efficiency of our suggested methodology. In conclusion, our work contributes valuable insights into the field of facial emotion recognition, providing a solid foundation for future research and applications in this area.

REFERENCES

[1] C. Shi, C. Tan and L. Wang, "A facial expression recognition method based on a multibranch cross-connection convolutional neural network," *in IEEE Access*, vol. 9, pp. 39255-39274, 2021, doi: 10.1109/ACCESS.2021.3063493.

[2] C. Mordka, "What are emotions? structure and function of emotions," *Studia Humana*, vol. 5, no. 3, pp. 29–44, Sep. 2016, doi: 10.1515/sh-2016-0013.

[3] T. Baltrusaitis, C. Ahuja, and L.-P. Morency, "Multimodal machine learning: a survey and taxonomy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 2, pp. 423–443, Feb. 2019, doi: 10.1109/tpami.2018.2798607.

[4] A. Mehrabian, "Communication without words," in *Communication Theory*, Routledge, 2017, pp. 193–200.

[5] S. Minaee, M. Minaei, and A. Abdolrashidi, "Deep-emotion: facial expression recognition using attentional convolutional network," *Sensors*, vol. 21, no. 9, p. 3046, Apr. 2021, doi: 10.3390/s21093046.

[6] C. Darwin, "The expression of the emotions in man and animals," *The Expression Of The Emotions In Man And Animals*. Oxford University PressNew York, NY, pp. iii–vi, Apr. 1998, doi: 10.1093/oso/9780195112719.002.0002.

[7] I. Dagher, E. Dahdah, and M. Al Shakik, "Facial expression recognition using three-stage support vector machines," *Visual Computing for Industry, Biomedicine, and Art*, vol. 2, no. 1, Dec. 2019, doi: 10.1186/s42492-019-0034-5.

[8] P. Ekman and W. V Friesen, "Constants across cultures in the face and emotion," *Journal of Personality and Social Psychology*, vol. 17, no. 2, pp. 124–129, 1971, doi: 10.1037/h0030377.

[9] P. Ekman, "Strong evidence for universals in facial expressions: A reply to Russell's mistaken critique.," *Psychological Bulletin*, vol. 115, no. 2, pp. 268–287, 1994, doi: 10.1037//0033-2909.115.2.268.

[10] S. Li and W. Deng, "Deep facial expression recognition: a survey," *IEEE Transactions on Affective Computing*, vol. 13, no. 3, pp. 1195–1215, Jul. 2022, doi: 10.1109/TAFFC.2020.2981446.

[11] C. A. Corneanu, M. O. Simon, J. F. Cohn, and S. E. Guerrero, "Survey on RGB, 3D, thermal, and multimodal approaches for

facial expression recognition: history, trends, and affect-related applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 8, pp. 1548–1568, Aug. 2016, doi: 10.1109/tpami.2016.2515606.

[12] J.-H. Kim, B.-G. Kim, P. P. Roy, and D.-M. Jeong, "Efficient facial expression recognition algorithm based on hierarchical deep neural network structure," *IEEE Access*, vol. 7, pp. 41273–41285, 2019, doi: 10.1109/access.2019.2907327.

[13] N. U. Khan, "A comparative analysis of facial expression recognition techniques," in *2013 3rd IEEE International Advance Computing Conference (IACC)*, Feb. 2013, pp. 1262–1268, doi: 10.1109/iadcc.2013.6514409.

[14] S. Singh and F. Nasoz, "Facial expression recognition with convolutional neural networks," in *2020 10th Annual Computing and Communication Workshop and Conference (CCWC)*, Jan. 2020, pp. 0324–0328, doi: 10.1109/ccwc47524.2020.9031283.

[15] T. Connie, M. Al-Shabi, W. P. Cheah, and M. Goh, "Facial expression recognition using a hybrid CNN–SIFT aggregator," in *Multi-disciplinary Trends in Artificial Intelligence*, Springer International Publishing, 2017, pp. 139–149.

[16] T. Zhang, W. Zheng, Z. Cui, Y. Zong, and Y. Li, "Spatial–temporal recurrent neural network for emotion recognition," *IEEE Transactions on Cybernetics*, vol. 49, no. 3, pp. 839–847, Mar. 2019, doi: 10.1109/tcyb.2017.2788081.

[17] D. Shehada, A. Turky, W. Khan, B. Khan, and A. Hussain, "A lightweight facial emotion recognition system using partial transfer learning for visually impaired people," *IEEE Access*, vol. 11, pp. 36961–36969, 2023, doi: 10.1109/access.2023.3264268.

[18] G. Xu and W. Q. Yan, "Facial emotion recognition using ensemble learning," in *Deep Learning, Reinforcement Learning, and the Rise of Intelligent Systems*, IGI Global, 2024, pp. 146–158.

[19] Z. Ullah *et al.*, "Emotion recognition from occluded facial images using deep ensemble model," *Computers, Materials & Continua*, vol. 73, no. 3, pp. 4465–4487, 2022, doi: 10.32604/cmc.2022.029101.

[20] C. Jia, C. L. Li, and Z. Ying, "Facial expression recognition based on the ensemble learning of CNNs," in *2020 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*, 2020, pp. 1–5, doi: 10.1109/icspcc50002.2020.9259543.

[21] H.-D. Nguyen, S.-H. Kim, G.-S. Lee, H.-J. Yang, I.-S. Na, and S.-H. Kim, "Facial expression recognition using a temporal ensemble of multi-level convolutional neural networks," *IEEE Transactions on Affective Computing*, vol. 13, no. 1, pp. 226–237, Jan. 2022, doi: 10.1109/taffc.2019.2946540.

[22] Y. Fan, J. C. K. Lam, and V. O. K. Li, "Multi-region ensemble convolutional neural network for facial expression recognition," in *Artificial Neural Networks and Machine Learning – ICANN 2018*, Springer International Publishing, 2018, pp. 84–94.

[23] P. Hu, D. Cai, S. Wang, A. Yao, and Y. Chen, "Learning supervised scoring ensemble for emotion recognition in the wild," Nov. 2017, doi: 10.1145/3136755.3143009.

[24] E. Bisong, "Google Colaboratory," in *Building Machine Learning and Deep Learning Models on Google Cloud Platform*, Apress, 2019, pp. 59–64.

[25] S. Raschka and V. Mirjalili, "Python machine learning: machine learning and deep learning with python, scikit-learn, and tensorflow," *Packt Publishing Ltd*, 2019.

[26] D. Lundqvist, A. Flykt, and A. Öhman, "Karolinska directed emotional faces," *PsycTESTS Dataset*. American Psychological Association (APA), 1998, doi: 10.1037/t27732-000.

[27] S. S. Ali, J. H. Al' Ameri, and T. Abbas, "Face detection using haar cascade algorithm," in *2022 Fifth College of Science International Conference of Recent Trends in Information Technology*, 2022, pp. 198–201, doi: 10.1109/csctit56299.2022.10145680.

[28] P. Viola and M. J. Jones, "Robust real-time face detection," *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, May 2004, doi: 10.1023/b:visi.0000013087.49260.fb.

[29] I. Culjak, D. Abram, T. Pribanic, H. Dzapo, and M. Cifrek, "A brief introduction to OpenCV," *2012 Proceedings of the 35th International Convention MIPRO*, pp. 1725–1730, 2012.

[30] C. Shu, X. Ding, and C. Fang, "Histogram of the oriented gradient for face recognition," *Tsinghua Science and Technology*, vol. 16, no. 2, pp. 216–224, Apr. 2011, doi: 10.1016/s1007-0214(11)70032-3.

[31] S. Kiranyaz, O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj, and D. J. Inman, "1D convolutional neural networks and applications: A survey," *Mechanical Systems and Signal Processing*, vol. 151, p. 107398, Apr. 2021, doi: 10.1016/j.ymssp.2020.107398.

[32] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization*," arXiv preprint arXiv:1412.6980,* 2014.

BIOGRAPHIES OF AUTHORS

Viraj Vilas Sawant has pursued Bachelor of Engineering (B.E.) from University of Mumbai in the field of Information Technology. He has a multifaceted interest in artificial intelligence (AI), dynamic programming, web development, and chess. Additionally, he actively engages in honing his problem-solving skills across various domains. He also participates in multiple hackathons and coding competitions. He also is actively engaged in Kaggle competitions for fun. He can be contacted at email: sawantviraj465@gmail.com.

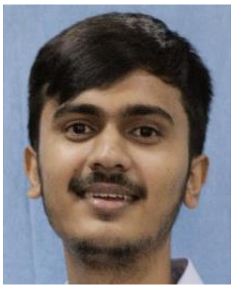
Husna Shehwar Shaikh completed her Bachelor of Engineering (B.E.) in Information Technology from M.H. Saboo Siddik College of Engineering. She is deeply interested in the fields of data science, data analytics, and cloud computing. Husna is dedicated to continuously enhancing her skills and knowledge in these areas, holding several certifications that demonstrate her expertise and commitment to professional growth. She can be contacted at email: husnas2933@gmail.com.

Dr. Bhakti Palkar is a distinguished academic and professional associated with K. J. Somaiya College of Engineering, Somaiya Vidyavihar University, Mumbai. She holds Ph.D. in Computer Engineering. She is working as an Associate Professor at K.J.Somaiya College of Engineering. Her research interests include medical image processing, Geographic Information System and Sentiment Analysis. She can be contacted at email: bhaktiraul@somaiya.edu.

Prof. Sanam Salman Kazi holds a master degree in Engineering (M.E.) in Information Technology. She is currently pursing Ph.D. from Somaiya Vidyavihar university, Mumbai, India. She is working as an Assistant Professor in the Department of Engineering and Technology at Bharati Vidyapeeth Deemed to be University. She has mentored projects like Railway Safety Application for women and bagged a price in Smart India Hackathon 2017. She has reviewed papers in international conferences. Her research interest is artificial Intelligence. She can be contacted at email: sanam.kazi19@gmail.com.

Wasim Jasani completed his B.E. in Information Technology. He is actively interested in the field of Artificial Intelligence and is currently pursuing a career in it. Wasim has participated in multiple competitions and hackathons, showcasing his skills and competitive spirit. He continuously builds various projects to challenge himself and advance his knowledge in Artificial Intelligence. He can be contacted at email: jasaniwasim2002@gmail.com.

Lamiya Rampurawala is a B. Tech student pursuing a degree in Information Technology. She has a strong interest in artificial intelligence (AI) and machine learning (ML). Lamiya has gained practical experience by working on an e-commerce bookselling website using Python and Django, and developing a human detection system with TensorFlow and YOLOv8. She is passionate about advancing in AI and ML and is committed to exploring innovative solutions in these fields. She can be contacted at emai: rampurawallamiya@gmail.com



Mohammed Naser Shaikh recently graduated with a Bachelor of Engineering in Information Technology from Mumbai University. During his academic career, he specialized in Artificial Intelligence (AI) and UX Design. In addition to his academic achievements, he has completed internships at Dr. Balasaheb Konkan Krishi Vidyapeeth and Zentech Systems, where he gained hands-on experience in software engineering and IT consulting. Mohammed Naser is passionate about leveraging technology to solve complex problems and is dedicated to continuous learning and professional development. He can be contacted at email: mohammednaser1572@gmail.com.