

An emotion recognition system: bridging the gap between humans-machines interaction

Ahmed Nouman^{1,3}, Hashim Iqbal², Muhammad Khurram Saleem³

¹Centre for Applied Autonomous Sensor Systems, Örebro University, Örebro, Sweden

²Department of Mechanical, Mechatronics and Manufacturing Engineering, UET Lahore, Lahore, Pakistan

³Department of Mechanical Engineering, University of Engineering and Technology Taxila, Taxila, Pakistan

Article Info

Article history:

Received Feb 26, 2023

Revised Apr 24, 2023

Accepted May 28, 2023

Keywords:

Deep learning

Emotion recognition systems

Facial expressions

Human-machine interaction

Machine learning

Manual image descriptors

ABSTRACT

Human emotion recognition has emerged as a vital research area in recent years due to its widespread applications in psychology, healthcare, education, entertainment, and human-robot interaction. This research article comprehensively analyzes a machine learning-based six-emotion classification algorithm, focusing on its development, evaluation, and potential applications. The study aims to assess the algorithm's performance, identify its limitations, and discuss the importance of selecting appropriate image descriptors for accurate emotion classification. The algorithm achieved an overall accuracy of 92.23%, showcasing its potential in various fields. However, the classification of specific emotions, particularly "excited" and "afraid", demonstrated some limitations, suggesting further refinement. The study also highlights the significance of choosing suitable image descriptors, with the manual distance calculation used in the framework proving effective. This article offers insights into developing and evaluating a six-emotion classification algorithm using a machine learning framework, emphasizing its strengths, limitations, and possible applications in multiple domains. The findings contribute to ongoing efforts in creating robust, accurate, and versatile emotion recognition systems that cater to the diverse needs of various applications across psychology, healthcare, robotics, education, and entertainment.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Ahmed Nouman

Machine Perception and Interaction Laboratory, Centre for Applied Autonomous Sensor Systems,

Örebro University

70281, Örebro, Sweden

Email: ahmed.nouman@oru.se

1. INTRODUCTION

Human emotion recognition is a rapidly growing field that has gained significant attention in recent years due to its potential applications in various areas, including psychology, healthcare, education, and entertainment. Emotions are complex and subjective experiences crucial to communication, decision-making, and well-being. Understanding emotions is essential for effective human-robot interaction, personalized mental health interventions, and many other applications.

Recent technological advances have made it possible to detect and analyze human emotions using various techniques. Researchers have explored different approaches to recognizing and classifying emotions accurately, including machine learning algorithms, such as support vector machines, artificial neural

networks, and deep learning models. Other techniques involve analyzing physiological signals, such as electroencephalography (EEG), electrocardiography (ECG), and galvanic skin response (GSR), to extract features related to emotional responses.

Emotion recognition systems have numerous potential applications in various fields. For example, we can use emotion recognition to monitor and improve mental health conditions like depression, anxiety, and post-traumatic stress disorder (PTSD). In entertainment, emotion recognition can personalize video and audio content to the viewer's emotional state. In security, emotion recognition can detect and prevent crimes by identifying suspicious behavior and emotional states.

Emotion recognition is an important research field that has gained considerable attention in recent years due to its applications in various areas such as psychology, healthcare, robotics, education, and entertainment. Emotion recognition refers to identifying and analyzing human emotions from multiple sources, such as facial expressions, physiological signals, speech, and gestures. In this literature review, we will discuss the recent advances in emotion recognition from different modalities and the methods proposed by various researchers.

2. LITERATURE REVIEW

This section covers various state-of-the-art approaches to emotion recognition. We discuss four areas of emotion recognition: facial, speech, hybrid, and physiological signal-based methods. Numerous studies and techniques emphasize deep learning approaches and multimodal systems for improved accuracy. The research spans multiple applications, for example, human-computer interaction, real-world deficits, and physiological signals.

2.1. Facial emotion recognition

Jain *et al.* [1] contributed in classifying each image into one of six facial emotion classes. Balasubramanian *et al.* [2] covered the datasets and algorithms used for facial emotion recognition (FER). The algorithms used were Gabor filters [3], a histogram of oriented gradients (HoG) [4], and local binary pattern (LBP) [5] for feature extraction [6]. Hassouneh *et al.* [7] aimed to classify physically disabled people (deaf, dumb, and bedridden) and autism children's emotional expressions based on facial landmarks, and electroencephalograph (EEG) signals using convolutional neural network (CNN), and long short-term memory (LSTM) classifiers by establishing an algorithm to recognize real-time emotion using virtual markers through an optical flow algorithm that works effectively in uneven lighting and subject head rotation (up to 25°), multiple backgrounds, and various skin tones. Mellouk *et al.* [8] aimed to study recent works on automatic facial emotion recognition (FER) via deep learning. Deep learning techniques in human-computer interactions were employed in [9] on the advancement of artificial intelligence as an efficient system application procedure. Hayes *et al.* [10] studied about how understanding real-world deficiencies and task selection in upcoming emotion recognition studies are affected by the variability in the age effects of various facial emotion recognition task designs. Ulusoy *et al.* [11] examined patients with BD, their parents, and healthy controls' capacity to recognize and distinguish between facial emotions.

2.2. Speech emotion recognition

Zhang *et al.* [12] presented a novel attention-based fully convolutional network for speech emotion recognition. Albanie *et al.* [13] considered learning embeddings for speech classification without access to labeled audio. Deep learning techniques are utilized as an alternative to traditional approaches in speech emotion recognition. Khalil *et al.* [14] discussed some recent literature where these methods are used for speech-based emotion recognition. At the emotion classification stage, an algorithm was proposed to determine the structure of the decision tree [15]. By utilizing the characteristics of CNN in modeling contextual information, Latif *et al.* [16] proved that there is still potential to enhance the performance of emotion recognition from raw speech. Both verbal and nonverbal sounds within an utterance were thus considered for emotional recognition of real-life conversations [17]. To create more accurate multimodal feature representations, Xu *et al.* [18] suggested using an attention mechanism to understand the alignment between speech frames and text words.. Koduru *et al.* [19] mainly contributed in improving a system's speech emotion recognition rate using the different feature extraction algorithms. Siriwardhana *et al.* [20] explore the use of modality-specific "BERT-like" pre-trained self-supervised learning (SSL) architectures to represent both speech and text modalities for the task of multimodal speech emotion recognition. Pepino *et al.* [21] proposed a transfer learning method for

speech emotion recognition where features extracted from pre-trained wav2vec 2.0 models are modeled using simple neural networks.

2.3. Hybrid approaches for emotion recognition

Albraikan *et al.* [22] presented a hybrid sensor fusion approach based on a stacking model allowing information from various sensors and emotion models to be jointly incorporated within a user-independent model. Pane *et al.* [23] proposed strategies incorporating emotion lateralization and ensemble learning approach to enhance the accuracy of EEG-based emotion recognition. Alswaidan *et al.* [24] critically surveyed the state-of-the-art research for explicit and implicit emotion recognition in text. He discussed the different approaches in the literature, detailed their main features, advantages, and limitations, and compared them with tables. During human-computer interaction, it might be tricky to automatically recognize facial emotions. Sandhu *et al.* [25] used the hybrid CNN approach to recognize human emotions and categorize them into sub-categories based on their features. A hybrid system consisting of three feature extraction stages, dimensionality reduction, and feature classification was proposed for speech emotion recognition (SER) [26], [27].

Moreover, a novel emotion recognition system, based on a number of modalities, including electroencephalogram (EEG), galvanic skin response (GSR), and facial expressions, was introduced [28]. Siddiqui *et al.* [29] presented a multimodal automatic emotion recognition (AER) framework capable of accurately differentiating expressed emotions. To predict emotions by examining facial expressions in an image, a convolution neural network (CNN)-based deep learning method has been proposed [30].

2.4. Physiological signal-based emotion recognition

Shu *et al.* [31] gave an in-depth analysis of physiological signal-based emotion recognition that covered emotion models, emotion elicitation techniques, published emotional and physiological datasets, features, classifiers, and the entire framework for emotion recognition based on physiological signals. Li *et al.* [32] presented an extensive and organized taxonomy for recognizing emotions based on physiological signals. The emotion recognition methods based on multi-channel EEG and multimodal physiological signals are reviewed [33]. Kim *et al.* [34] presented a robust physiological model called a deep physiological affect network (DPAN) for recognizing human emotions. Li *et al.* [35], [36] proposed a multimodal attention-based BLSTM network framework for efficient emotion recognition. The work attempts to fuse the subject individual EDA features and the external evoked music features [37]. Yin *et al.* [38] proposed an end-to-end multimodal framework, the one-dimensional residual temporal and channel attention network (RTCAN-1D). Chen *et al.* [39] proposed a single SP-signal-based method for emotion recognition. Stappen *et al.* [40] offered four different sub-challenges: i) MuSe-Wilder and ii) MuSe-Stress, which concentrate on continuous emotion (valence and arousal) prediction; iii) MuSe-Sent, which requires participants to identify five classes for valence and arousal; and iv) MuSe-Physio, which asks participants to predict a novel aspect of "physiological-emotion." For this year's challenge, the Ulm-TSST dataset, which displays people in stressful depositions, is introduced [40]. To fill a gap in the present literature, Ahmad *et al.* [41] reviewed the impact of inter-subject data variance on emotion recognition, essential data annotation techniques for emotion recognition and their comparison, data pre-processing methods for each physiological signal, data splitting techniques to enhance the generalization of emotion recognition models and multiple multimodal fusion methods and their comparison.

3. RESEARCH METHOD

The complete methodology of the emotion detection framework is shown in Figure 1. First, we capture the image and crop it to the processing size. Next, we perform RGB to grayscale conversion and apply histogram equalization. Next, Canny edge detection and Hough circle transform are performed to locate a person's eyes. Next, we identify the critical points in the image as image descriptors for classification. We explain the process in detail in this section, and the working of the algorithm shown in Figure 2 is elaborated in this section.

3.1. Image capturing

We use a high-end webcam, the Logitech Brio 4K, designed for professional use. It can capture video in 4K Ultra HD at 30 frames per second or 1080p or 720p at up to 60 frames per second. It has advanced features such as autofocus and 5× digital zoom and supports high dynamic range (HDR) imaging for improved color and contrast in difficult lighting conditions.

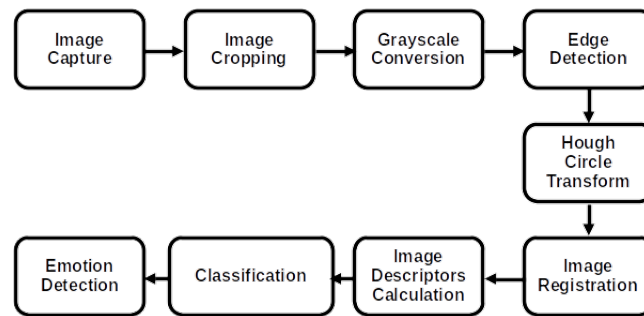


Figure 1. Research methodology of our emotion detection framework

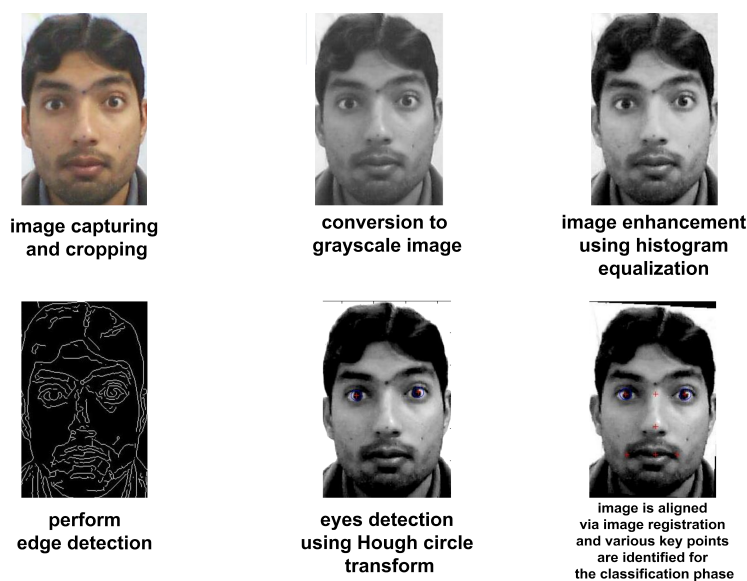


Figure 2. Snapshots showing the step-by-step implementation of our human emotion detection framework

3.2. Conversion to grayscale image

We convert the captured image in the red, green, and blue (RGB) color space to a grayscale image in a single channel. Grayscale images are often used in computer vision tasks, such as object recognition, image segmentation, and edge detection, as they reduce the complexity of the image by eliminating color information while preserving the overall structure and contrast of the image. In a grayscale image, each pixel is represented by a single channel, with values ranging from 0 to 255, where 0 corresponds to black and 255 to white.

3.3. Histogram equalization

Next, we apply histogram equalization to the grayscale image. Histogram equalization is a method of contrast enhancement in digital image processing. It is a technique that redistributes the pixel intensities in an image to make the overall image contrast better. In other words, it adjusts the dynamic range of an image by spreading the intensity levels over the whole range. It is done by calculating a histogram of pixel intensities in the image and then modifying the pixel values so that the histogram becomes more evenly distributed. It can benefit images with a very narrow or compressed range of pixel intensities, making them appear flat or low in contrast. The result of histogram equalization is an image with higher contrast and better visibility of details.

3.4. Edge detection

Edge detection is a standard image processing technique used to identify and highlight the edges or boundaries within an image. The edges in an image represent areas of rapid changes in brightness or intensity, such as the boundaries between objects or the contours of shapes. The edge detection algorithm works by

analyzing the intensity differences between adjacent pixels in an image and identifying areas where there is a sharp change in intensity.

There are several algorithms for edge detection, but some of the most common ones include Sobel [42], Canny [43], and Roberts operators [44]. The Sobel operator calculates the image intensity gradient in horizontal and vertical directions. In contrast, the Canny operator uses a multi-stage algorithm that includes smoothing, edge detection, and hysteresis thresholding to produce high-quality edge detection results. The Roberts operator is a simple but effective operator that calculates the gradient using a pair of 2×2 kernels. We apply the Canny edge detection technique to extract features from our images.

3.5. Hough circle transform

The Hough circle transform [45], [46] is a feature extraction technique used in digital image processing to detect circular shapes in images. It is an extension of the Hough transform algorithm to detect straight image lines. The Hough circle transform converts the image from Cartesian coordinates to polar coordinates, representing circles as points in a two-dimensional parameter space. Each point in this parameter space corresponds to a circle in the original image, with the radius and center coordinates of the circle encoded in the coordinates of the point. We use the Hough circle transform algorithm to detect the iris of the human in our binary images provided by the Canny edge detection techniques.

3.6. Image registration

Image alignment, also known as image registration, is the process of aligning multiple images of the same scene or object. The goal is to find a transformation that maps one image onto another so that they are in the same coordinate system. Image alignment is important in various applications, including computer vision, remote sensing, medical imaging, and astronomy. Next, we align the eyes in the images via the image registration method.

3.7. Keypoints detection and image descriptors

Keypoint detection, also known as interest point detection, is a technique in computer vision that identifies and localizes distinctive features or points in an image. These key points are regions in an image with certain properties, such as high contrast, sharp edges, or corners, making them easily distinguishable from the surrounding areas. The process of keypoint detection typically involves analyzing an image using a series of algorithms to identify areas likely to be key points. Some popular algorithms for keypoint detection include Harris corner detection [47], scale-invariant feature transform (SIFT) [48], speeded-up robust features (SURF) [49], and ORB (Oriented FAST and Rotated BRIEF) [50].

We use a manual keypoint detection technique for the classification of emotions in humans for our methodology. We detect the bottom of the forehead by joining a line between the eyes in the images and locating its center. Next, we calculate the distances from the bottom of the forehead to key point locations like lips, cheeks, ears, chin, and forehead. We calculate 28 key points that serve as image descriptors for the classification methods.

3.8. Classification

Image classification is a task in computer vision that assigns a label or category to an input image. Image classification aims to teach a computer to recognize visual patterns in images and to classify them into one of several pre-defined categories or classes. This task is usually accomplished by training a machine learning model on a large dataset of labeled images, where the labels represent the correct category or class of the image.

There are several techniques used for image classification, including traditional machine learning algorithms such as support vector machines (SVM) [51] and k-nearest neighbors (KNN) [52], [53], Mixture of Experts (MoE) [54], AdaBoost [55], as well as deep learning methods such as convolutional neural networks (CNN) [56]. Deep learning has become the dominant approach for image classification in recent years due to its ability to learn complex features directly from raw image data. We constructed the dataset manually with annotated 1800 images of persons in six different emotions. We took fifty candidates and captured six images of each person for every six emotions. Next, we divided the dataset into 75% (1350) images for training while the rest, 25% (450), were utilized to evaluate or generalize the ANN. The datasets are equally divided for each emotion to remove any inherent dataset bias.

We use an artificial neural network (ANN) with 28 input neurons, 56 second-layer hidden neurons, and six neurons in the output layer. The six output neurons classify human emotions in images from fear,

anger, sadness, happiness, excitement, and normal. The learning algorithm used is the Levenberg-Marquardt backpropagation algorithm [57]–[60]. The network converges to the goal at around 252 epochs as shown in Figure 3 and the output of the ANN, in this case, normal is demonstrated in Figure 4.

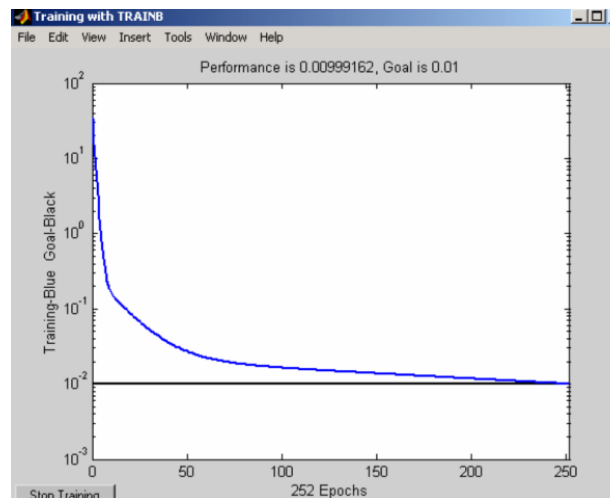


Figure 3. Training curve for the artificial neural network model shows convergence to goal around 250 epochs

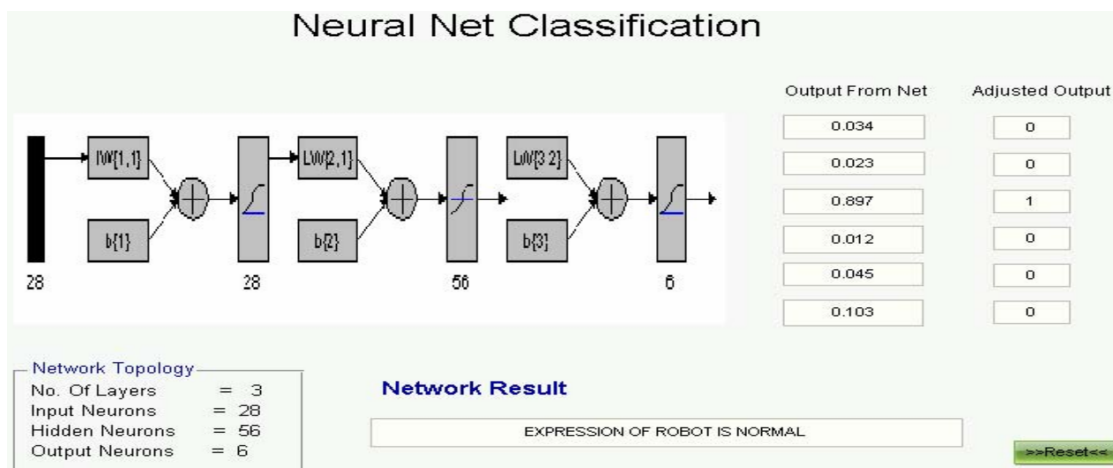


Figure 4. Output for the trained artificial neural network model shows detection of “normal” mode

4. RESULT AND DISCUSSION

The study’s results reveal the effectiveness of the six-emotion classification algorithm in accurately classifying emotions using a machine learning framework. The algorithm’s overall accuracy of 92.23% demonstrates its potential for practical applications in healthcare, marketing, education, and human-computer interaction. The successful classification of emotions using machine learning algorithms can enhance human-machine interactions, personalized user experiences, and informed decision-making processes in various fields.

The confusion matrix generated in the study offers a deeper understanding of the algorithm’s performance as shown in Figure 5. It highlights that the algorithm had difficulties classifying “excited” and “afraid” emotions, with 14 and 8 false classifications, respectively. A closer analysis of the confusion matrix indicates that most false classifications occurred between “excited” and “happy” and between “excited” and “afraid”. The algorithm also exhibited confusion between “excited” and “happy” and between “normal” and “happy”. These misclassifications suggest that the algorithm might require further refinement to improve its performance in distinguishing emotions with similar expressions or characteristics.

		Predicted Classes					
		Normal	Happy	Sad	Afraid	Angry	Excited
Actual Classes	Normal	70	3	0	0	0	2
	Happy	2	69	0	0	0	4
	Sad	1	0	71	1	2	0
	Afraid	1	0	0	67	0	7
	Angry	1	0	0	0	71	1
	Excited	4	9	1	0	0	61

Figure 5. Confusion matrix for emotions classification from 450 test images of random people dataset

Another key finding of the study is the effectiveness of the manual distance calculation used as the image descriptor in the machine learning framework. It highlights the importance of selecting appropriate image descriptors for accurate emotion classification. Choosing the right image descriptor can significantly impact the algorithm's performance and the overall accuracy of emotion classification.

While the results demonstrate the potential of machine learning algorithms for accurately classifying emotions, it is essential to acknowledge the limitations in classifying specific emotions, particularly "excited" and "afraid". Further research and development of the algorithm could focus on addressing these limitations and enhancing its performance in classifying these challenging emotions. It might involve exploring alternative or additional image descriptors, refining the feature extraction process, or incorporating other machine-learning techniques to improve classification accuracy.

5. CONCLUSION

The research article's main conclusions highlight the potential and effectiveness of machine learning algorithms in accurately classifying emotions, an area of growing research interest with significant implications for fields like healthcare, marketing, education, and human-computer interaction. The evaluated six-emotion classification algorithm achieved an overall accuracy of 92.23%, showcasing its potential for practical applications in these fields. However, the confusion matrix generated during the study revealed some limitations in classifying specific emotions, particularly "excited" and "afraid". The majority of false classifications occurred between "excited" and "happy" and between "excited" and "afraid". It indicates the need for further algorithm refinement to improve its performance in distinguishing emotions with similar expressions or characteristics. The article also emphasizes the importance of selecting appropriate image descriptors for accurate emotion classification. The manual distance calculation used as the image descriptor in the machine learning framework proved effective, suggesting that the choice of image descriptor plays a significant role in determining the algorithm's performance and overall accuracy. The article's conclusions underscore the effectiveness of a six-emotion classification algorithm using a machine learning framework in emotion classification. However, addressing the limitations in classifying specific emotions will enhance the algorithm's performance and expand its practical applications in various fields that rely on accurate emotion recognition.

REFERENCES

- [1] D. K. Jain, P. Shamsolmoali, and P. Sehdev, "Extended deep neural network for facial emotion recognition," *Pattern Recognition Letters*, vol. 125, pp. 407–413, 2019.
- [2] B. Balasubramanian, P. Diwan, R. Nadar, and A. Bhatia, "Analysis of facial emotion recognition," in *3rd International Conference on Trends in Electronics and Informatics (ICOEI)*, 2019, pp. 248–252, doi: 10.1109/ICOEI.2019.8862908.
- [3] A. C. Tsoi, J. Mammen, and D. Mukherjee, "Texture analysis using Gabor filters," in *Proceedings of the 1998 IEEE Signal Processing Society Workshop*, 1998, pp. 333–342.





- [4] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 2005, vol. 1, pp. 886–893, doi: 10.1109/CVPR.2005.177.
- [5] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, Jul. 2002, doi: 10.1109/TPAMI.2002.1017623.
- [6] D. Mehta, M. F. H. Siddiqui, and A. Y. Javaid, "Recognition of emotion intensities using machine learning algorithms: a comparative study," *Sensors*, vol. 19, no. 8, Apr. 2019, doi: 10.3390/s19081897.
- [7] A. Hassouneh, A. M. Mutawa, and M. Murugappan, "Development of a real-time emotion recognition system using facial expressions and EEG based on machine learning and deep neural network methods," *Informatics in Medicine Unlocked*, vol. 19, 2020, doi: 10.1016/j.imu.2020.100360.
- [8] W. Mellouk and W. Handouzi, "Facial emotion recognition using deep learning: review and insights," *Applied Sciences*, vol. 10, no. 21, 2020.
- [9] L. Zahara, P. Musa, E. P. Wibowo, I. Karim, and S. B. Musa, "The facial emotion recognition (FER-2013) Dataset for prediction system of micro-expressions face using the convolutional neural network (CNN) algorithm based Raspberry Pi," in *2020 Fifth International Conference on Informatics and Computing (ICIC)*, 2020, pp. 1–5.
- [10] G. S. Hayes et al., "Task characteristics influence facial emotion recognition age-effects: a meta-analytic review," *Psychology and Aging*, vol. 35, no. 7, 2020, doi: 10.1037/pag0000476.
- [11] S. Işık Ulusoy, Ş. A. Gülsiren, N. Özkan, and C. Bilen, "Facial emotion recognition deficits in patients with bipolar disorder and their healthy parents," *General Hospital Psychiatry*, vol. 65, pp. 9–14, Jul. 2020, doi: 10.1016/j.genhosppsych.2020.04.008.
- [12] Y. Zhang, J. Du, Z. Wang, J. Zhang, and Y. Tu, "Attention based fully convolutional network for speech emotion recognition," in *2018 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, Nov. 2018, pp. 1771–1775, doi: 10.23919/APSIPA.2018.8659587.
- [13] S. Albanie, A. Nagrani, A. Vedaldi, and A. Zisserman, "Emotion recognition in speech using cross-modal transfer in the wild," in *Proceedings of the 26th ACM International Conference on Multimedia*, Oct. 2018, pp. 292–301, doi: 10.1145/3240508.3240578.
- [14] R. A. Khalil, E. Jones, M. I. Babar, T. Jan, M. H. Zafar, and T. Alhussain, "Speech emotion recognition using deep learning techniques: a review," *IEEE Access*, vol. 7, pp. 172885–172913, 2019.
- [15] L. Sun, S. Fu, and F. Wang, "Decision tree SVM model with fisher feature selection for speech emotion recognition," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2019, no. 1, pp. 1–10, 2019, doi: 10.1186/s13636-019-0143-x.
- [16] S. Latif, R. Rana, S. Khalifa, R. Jurdak, and J. Epps, "Direct modelling of speech emotion from raw speech," in *Interspeech 2019*, Sep. 2019, pp. 3920–3924, doi: 10.21437/Interspeech.2019-3252.
- [17] K.-Y. Huang, C.-H. Wu, Q.-B. Hong, M.-H. Su, and Y.-H. Chen, "Speech emotion recognition using deep neural network considering verbal and nonverbal speech sounds," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2019, pp. 5866–5870, doi: 10.1109/ICASSP.2019.8682283.
- [18] H. Xu, H. Zhang, K. Han, Y. Wang, Y. Peng, and X. Li, "Learning alignment for multimodal emotion recognition from speech," in *Interspeech 2019*, Sep. 2019, pp. 3569–3573, doi: 10.21437/Interspeech.2019-3247.
- [19] A. Koduru, H. B. Valiveti, and A. K. Budati, "Feature extraction algorithms to improve the speech emotion recognition rate," *International Journal of Speech Technology*, vol. 23, no. 1, pp. 45–55, Mar. 2020, doi: 10.1007/s10772-020-09672-4.
- [20] S. Siriwardhana, A. Reis, R. Weerasekera, and S. Nanayakkara, "Jointly fine-tuning 'BERT-like' self supervised models to improve multimodal speech emotion recognition," in *Proceedings of the Annual Conference of the International Speech Communication Association*, 2020, pp. 3755–3759, doi: 10.21437/Interspeech.2020-1212.
- [21] L. Pepino, P. Riera, and L. Ferrer, "Emotion recognition from speech using Wav2Vec 2.0 embeddings," in *Interspeech 2021*, Aug. 2021, vol. 1, pp. 3400–3404, doi: 10.21437/Interspeech.2021-703.
- [22] A. Albraikan, D. P. Tobón and A. El Saddik, "Toward user-independent emotion recognition using physiological signals," in *IEEE Sensors Journal*, vol. 19, no. 19, pp. 8402–8412, 1 Oct. 1, 2019, doi: 10.1109/JSEN.2018.2867221.
- [23] E. S. Pane, A. D. Wibawa, and M. H. Purnomo, "Improving the accuracy Of EEG emotion recognition by combining valence lateralization and ensemble learning with tuning parameters," *Cognitive Processing*, 2019, doi: 10.1007/s10339-019-00907-w.
- [24] N. Alswaidan and M. E. B. Menai, "A survey of state-of-the-art approaches for emotion recognition in text," *Knowledge and Information Systems*, 2020, doi: 10.1007/s10115-019-01372-1.
- [25] N. Sandhu, A. Malhotra, and M. K. Bedi, "Human emotions detection using hybrid CNN approach," in *2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC)*, 2020, doi: 10.1109/ICCMC48073.2020.9198344.
- [26] F. Daneshfar and S. J. Kabudian, "Speech emotion recognition using discriminative dimension reduction by em-

- ploying a modified quantum-behaved particle swarm optimization algorithm,” *Multimedia Tools and Applications*, vol. 79, no. 1–2, pp. 1261–1289, Jan. 2029, doi: 10.1007/s11042-019-08222-8.
- [27] F. Daneshfar and S. J. Kabudian, “Speech emotion recognition using hybrid spectral-prosodic features of speech signal/glottal waveform, metaheuristic-based dimensionality reduction, and Gaussian elliptical basis function network classifier,” *Applied Acoustics*, vol. 166, Sep. 2020, doi: 10.1016/j.apacoust.2020.107360.
- [28] Y. Cimtay, E. Ekmekcioglu, and S. Caglar-Ozhan, “Cross-subject multimodal emotion recognition based on hybrid fusion,” *IEEE Access*, vol. 8, pp. 168865–168878, 2020, doi: 10.1109/ACCESS.2020.3023871.
- [29] M. F. H. Siddiqui and A. Y. Javaid, “A multimodal facial emotion recognition framework through the fusion of speech with visible and infrared images,” *Multimodal Technologies and Interaction*, vol. 4, no. 3, Aug. 2020, doi: 10.3390/mti4030046.
- [30] G. Verma and H. Verma, “Hybrid-deep learning model for emotion recognition using facial expressions,” *The Review of Socionetwork Strategies*, vol. 14, no. 2, pp. 171–180, Oct. 2020, doi: 10.1007/s12626-020-00061-6.
- [31] L. Shu et al., “A review of emotion recognition using physiological signals,” *Sensors*, vol. 18, no. 7, Jun. 2018, doi: 10.3390/s18072074.
- [32] W. Li, Z. Zhang, and A. Song, “Physiological-signal-based emotion recognition: An odyssey from methodology to philosophy,” *Measurement*, vol. 172, Feb. 2021, doi: 10.1016/j.measurement.2020.108747.
- [33] J. Zhang, Z. Yin, P. Chen, and S. Nichele, “Emotion recognition using multi-modal data and machine learning techniques: A tutorial and review,” *Information Fusion*, vol. 59, pp. 103–126, Jul. 2020, doi: 10.1016/j.inffus.2020.01.011.
- [34] B. H. Kim and S. Jo, “Deep physiological affect network for the recognition of human emotions,” *IEEE Transactions on Affective Computing*, vol. 11, no. 2, pp. 230–243, 2020, doi: 10.1109/TAFFC.2018.2790939.
- [35] C. Li, Z. Bao, L. Li, and Z. Zhao, “Exploring temporal representations by leveraging attention-based bidirectional LSTM-RNNs for multi-modal emotion recognition,” *Information Processing and Management*, vol. 57, no. 3, May 2020, doi: 10.1016/j.ipm.2019.102185.
- [36] Z. Li, G. Zhang, L. Wang, J. Wei, and J. Dang, “Emotion recognition using spatial-temporal EEG features through convolutional graph attention network,” *Journal of Neural Engineering*, vol. 20, no. 1, 2023, doi: 10.1088/1741-2552/acb79e.
- [37] G. Yin, S. Sun, D. Yu, D. Li, and K. Zhang, “A efficient multimodal framework for large scale emotion recognition by fusing music and electrodermal activity signals,” *arXiv*, vol. 14, no. 8, pp. 1–15, Aug. 2020, doi: 10.1145/3490686.
- [38] G. Yin, S. Sun, D. Yu, D. Li, and K. Zhang, “A multimodal framework for large-scale emotion recognition by fusing music and electrodermal activity signals,” *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 18, no. 3, pp. 1–23, Aug. 2022, doi: 10.1145/3490686.
- [39] S. Chen et al., “Emotion recognition based on skin potential signals with a portable wireless device,” *Sensors*, vol. 21, no. 3, Feb. 2021, doi: 10.3390/s21031018.
- [40] L. Stappen et al., “The MuSe 2021 multimodal sentiment analysis challenge,” in *Proceedings of the 2nd on Multimodal Sentiment Analysis Challenge*, Oct. 2021, pp. 5–14, doi: 10.1145/3475957.3484450.
- [41] Z. Ahmad and N. Khan, “A survey on physiological signal-based emotion recognition,” *Bioengineering*, vol. 9, no. 11, Nov. 2022, doi: 10.3390/bioengineering9110688.
- [42] I. E. Sobel, “Isotropic 33 image gradient operator,” US patent, 1983.
- [43] J. Canny, “A computational approach to edge detection,” in *Readings in Computer Vision*, Elsevier, 1987, pp. 184–203.
- [44] L. G. Roberts, “Machine perception of three-dimensional solids,” *MIT Lincoln Laboratory*, vol. 25, no. 2, pp. 1–15, 1963.
- [45] H.-M. Yuen, J. Princen, J. Illingworth, and J. Kittler, “Hough circle transform,” *Computer Vision, Graphics, and Image Processing*, vol. 46, no. 1, pp. 39–59, 1989.
- [46] P. V. C. Hough, “A method and means for recognizing complex patterns,” *US Patent*, vol. 2, 1959.
- [47] C. Harris and M. Stephens, “A combined corner and edge detector,” in *Alvey Vision Conference*, 1988, doi: 10.5244/C.2.23.
- [48] D. G. Lowe, “Object recognition from local scale-invariant features,” in *Proceedings of the Seventh IEEE International Conference on Computer Vision*, 1999, vol. 2, pp. 1150–1157, doi: 10.1109/ICCV.1999.790410.
- [49] H. Bay, T. Tuytelaars, and L. Van Gool, “SURF: speeded up robust features,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 3951, 2006, pp. 404–417.
- [50] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, “ORB: an efficient alternative to SIFT or SURF,” in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2564–2571, 2011, doi: 10.1109/ICCV.2011.6126544.
- [51] C. Cortes and V. Vapnik, “Support-vector networks,” *Machine Learning*, vol. 20, no. 3, pp. 273–297, Sep. 1995, doi: 10.1007/BF00994018.
- [52] E. Fix and J. L. Hodges Jr., “Discriminatory analysis-nonparametric discrimination: consistency properties,” *International Statistical Review*, vol. 57, no. 3, pp. 238–247, 1951.





- [53] T. Cover and P. Hart, "Nearest neighbor pattern classification," in *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21–27, Jan. 1967, doi: 10.1109/TIT.1967.1053964.
- [54] R. A. Jacobs, M. I. Jordan, S. J. Nowlan, and G. E. Hinton, "Adaptive mixtures of local experts," *Neural Computation*, vol. 3, no. 1, pp. 79–87, 1991, doi: 10.1162/neco.1991.3.1.79.
- [55] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of Computer and System Sciences*, vol. 55, no. 1, pp. 119–139, Aug. 1997, doi: 10.1006/jcss.1997.1504.
- [56] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," in *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998, doi: 10.1109/5.726791.
- [57] K. Levenberg, "A method for the solution of certain non-linear problems in least squares," *Quarterly of Applied Mathematics*, vol. 2, no. 2, pp. 164–168, 1944, doi: 10.1090/qam/10666.
- [58] D. W. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters," *Journal of the Society for Industrial and Applied Mathematics*, vol. 11, no. 2, pp. 431–441, Jun. 1963, doi: 10.1137/0111030.
- [59] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, Oct. 1986, doi: 10.1038/323533a0.
- [60] T. H. M. and B. M. M., "Training feedforward networks with the Marquardt algorithm," *IEEE Transactions on Neural Networks*, vol. 5, no. 6, pp. 989–993, 1994.

BIOGRAPHIES OF AUTHORS







Ahmed Nouman     is a distinguished academic and researcher in artificial intelligence and robotics, having served as an assistant professor at the University of Engineering and Technology in Taxila, Pakistan, for 14 years. He is a postdoctoral researcher at the Centre for Applied Autonomous Sensor Systems (AASS) in Sweden, contributing to various research projects. With a Ph.D. in mechatronics engineering from Sabanci University, Istanbul, his research interests encompass cognitive robotics, machine perception, planning under uncertainty, service robotics, and multiagent coordination. Ahmed Nouman's work is pivotal in advancing intelligent machines and robotics, leading to innovative technologies that can revolutionize industries and improve lives. He can be contacted at ahmed.nouman@oru.se.



Hashim Iqbal     is an assistant professor at the University of Engineering and Technology Lahore, specializing in medical robotics and haptic devices, focusing on developing innovative robotic devices for medical applications and enhancing human-machine interactions. By teaching various robotics and mechatronics engineering courses, Dr. Iqbal is shaping the next generation of engineers. At the same time, his research has the potential to transform healthcare and other industries by improving precision, efficiency, and safety in surgical procedures and beyond. He can be contacted at hashim.iqbal@uet.edu.pk.



Muhammad Khurram Saleem     is an assistant professor in the Mechanical Engineering department at UET Taxila, focusing on haptics and mechatronics systems. He holds a Ph.D. in Mechanical Engineering from KOC University, Turkey, and an M.Sc. and B.Sc. in mechatronics engineering from the University of Engineering and Technology Lahore. Dr. Saleem's research interests have led to several publications in reputable journals and presentations at international conferences. In addition to supervising postgraduate theses, he actively participates in workshops and courses to stay current in his field and share his expertise with students and colleagues. He can be contacted at khurram.saleem@uettaxila.edu.pk.