Transfer Learning for Face Emotions Recognition in Different Crowd Density Situations

Amirah Alharbi[1†],

University of Umm Alqura, College of Computing, KSA

Abstract
Most human emotions are conveyed through facial expressions, which represent the predominant source of emotional data. This research investigates the impact of crowds on human emotions by analysing facial expressions. It examines how crowd behaviour, face recognition technology, and deep learning algorithms contribute to understanding the emotional change according to different level of crowd. The study identifies common emotions expressed during congestion, differences between crowded and less crowded areas, changes in facial expressions over time. The findings can inform urban planning and crowd event management by providing insights for developing coping mechanisms for affected individuals. However, limitations and challenges in using reliable facial expression analysis are also discussed, including age and context-related differences.

Keywords:
facial expressions; crowd; crowd management; emotions; bias.

1. Introduction

Modern cities often characterized with high congestion and densely populated areas, posing various challenges for urban planning which can impact residents' quality of life. Human feelings can be affected when personal space is crossed. To understand how humans interpret emotions, a renowned psychologist in 1976 [1], discovered through his research that emotional information, described by humans as emotions, is segmented. His findings indicate that a mere 7% of the entire emotional data is conveyed through language, while 38% is conveyed through language adjuncts, which vary across cultures, encompassing elements like speech rhythm, tone, pitch, etc. Notably, the facial expression exhibits the highest percentage of emotional data, reaching 55%.

Several studies tried to explore monitoring the mood of crowd generally [2][3][4]. The study in [33] focused on understanding how crowd emotions impact observers. They indicated that observers respond emotionally to both positive and negative crowd emotions. Another interesting study [34] aimed to investigate the duration of fixation on emotional faces in a crowd, comparing seven primary emotions and embarrassment to a neutral face expression. The study supported the happiness superiority effect, indicating that observers tend to fixate longer on faces expressing happiness compared to other emotions.

Our research presented here sheds the light on different perspective of crowd emotions research. It aims to explore the impact of different congestion levels on human emotions through facial expression analysis (FEA), which is not investigated yet. This research aims to investigate the effect of crowd density on human feelings using the techniques of face recognition and facial expressions analysis. Facial expression analysis involves examining the various movements and positions of muscles on the face to determine an individual's emotional state.
Although facial expression analysis alone may not be capable of offering a holistic comprehension of an individual's emotions, particularly those pertaining to congestion, it can offer valuable insights into their overall emotional well-being. By analysing specific facial muscle movements and patterns such as furrowed brows, squinting, or nose scrunching, researchers and observers can infer the discomfort or frustration experienced by individuals with congestion. However, it is important to consider other factors that may influence facial expressions, such as individual cultural differences, personal expression styles and contextual factors affecting facial expressions for example, social norms [5].

Facial expression analysis is crucial in several fields for mass gathering event organisers. Hence, by accurately interpreting facial expressions, researchers can assess the impact of certain stimuli or situations on individuals' emotions and consequently taking the best decision for their target.

Manuscript received April 5, 2024
Manuscript revised April 20, 2024
https://doi.org/10.22937/IJCSNS.2024.24.4.3

This paper is structured as follows: In the introduction section, a review of current literature on expression recognition and analysis is presented, and the contextual foundation for the study is established. The primary research objective is stated to address an identified gap in understanding the impact of crowds on human emotions. We emphasise that the overarching aim is to explore emotional dynamics across varied crowd densities. Also in this section, we hypothesize that crowd behavior, coupled with face recognition technology and deep learning algorithms, contributes significantly to understanding emotional changes.

In method section, the research design is explained by integrating crowd behaviour analysis, advanced face recognition technology, and state-of-the-art deep learning algorithms. We show a systematic analysis of facial expressions during congestion which covered distinctions of emotional expressions between crowded and less crowded areas, in addition to analysing the temporal changes in facial expressions.

The conclusion of the study reveals valuable insights into prevalent emotions during congestion, distinctions in emotional expressions between crowded and less crowded areas, and the dynamic nature of facial expressions over time. These results have implications for urban planning and crowd event management. We conclude by discussing the potential future implications of our findings, emphasizing their relevance to inform the development of coping mechanisms for individuals affected by crowd dynamics.

2. Related Work

This research has two perspectives of related work. On one hand, facial expression analysis related work and on another hand, the research of facial expression analysis that are conducted in crowd context.

1) Facial expression analysis

Facial expression analysis is a computer vision objective focused on recognizing and classifying emotional expressions displayed on a human face. The main aspects of facial expression analysis involve the recognition and interpretation of facial expressions to understand the emotional states and communicative signals conveyed by an individual. Key aspects include facial expression recognition (FER) which identify and categorize facial expressions into emotional states such as happiness, sadness, anger, surprise, fear, and disgust [6]. Analysing facial muscle movements and configurations to decode the underlying emotional expressions known as facial action coding system (FACS). FACS provides a comprehensive framework for describing facial expressions based on individual muscle actions. It categorizes different facial movements into specific action units (AUs). These AUs are then combined to understand the overall expression displayed by an individual [7].

Other studies in FEA focus on studying micro-expressions which aims at detecting subtle and brief facial movements that occur within a fraction of a second. This research hypothesize that micro-expressions often revealing concealed emotions or intentions [8]. Besides micro-expressions research, some researchers believe that quantifying the intensity or strength of expressed emotions, allowing for a more accurate understanding of the emotional state [9].

Integrating information from multiple modalities is an interesting area of FEA. Several papers tried to combine facial expressions with voice tone or body language, to enhance the accuracy of emotion recognition what they called it multimodal analysis [10][35][36]. Real-Time analysis is a key topic of FEA which considers implementing systems capable of analysing facial expressions in real-time. Hence, enabling applications like emotion-aware human-computer interaction or effective computing. FEA can be applied in various fields, including human-computer interaction, virtual reality, mental health assessment, user experience design, and security for instance, emotion-based security systems [11]. Most of above-mentioned studies use machine learning algorithms. Unlike the research that rely on Electromyography (EMG) measurements as discussed in [12]. In this study, EMG used to measure muscle activity using wearable electrodes placed on the face. It helps identify subtle changes in muscle contractions that occur during different emotional states.

2) Facial expression analysis and crowd management

The paper [13] examines the impact of high-density urban environments on individual emotional responses and the potential restorative effects of parks and open spaces. The study uses a within-subject and between-subject measurement design, with 30 students walking a selected route in Hong Kong, capturing data on crowd encounters, stress perception, and psychological emotion responses using wearable devices. It focuses only on two negative emotions like stress and aggression. It should be noted that, the paper highlights the need for further research in different cultural contexts to understand the perception of crowding and personal space invasions. The research in [14] outlines the implementation of a solution for the emotion recognition in the Wild 2017 Challenge, specifically addressing group-level emotion recognition. The task involves classifying a group's emotion as positive, neutral, or negative. The proposed approach integrates both image context and facial information extracted from images for classification.

Convolutional Neural Networks (CNNs) are employed to predict facial emotions from detected faces, and these predictions are combined with scene-context information obtained by another CNN using fully connected neural network layers. The paper explores various techniques for combining and training these two Deep Neural Network models to achieve group-level emotion recognition. Evaluation on the Group Affective Database 2.0, provided with the challenge, demonstrates promising performance improvements. The proposed approach shows an approximately 37% enhancement over the competition's baseline model on the validation dataset. Another similar research [4] has developed a novel crowd monitoring algorithm centered on estimating crowd emotion through Facial Expression Recognition (FER). The goal is to isolate various emotions within a crowd, enabling the prediction of the overall mood even in non-panic scenarios. The proposed algorithm's efficacy is validated through cross-validation tests, utilizing a unique Crowd Emotion dataset with known ground-truth emotions. The results demonstrate the algorithm's accuracy and efficiency in predicting multiple crowd emotion classes, even in situations where movement and density information may be incomplete.

The study of [2] introduced EmoNet, a high-performance, computationally efficient network designed for robust facial expression feature extraction. It proposes a fusion mechanism called Non-Volume Preserving-based Fusion (NVPF), to handle group-level emotion in crowded scenes, addressing scenarios with unclearly identified faces, such as large sports events. The NVPF is extended to TNVPF to incorporate temporal information in crowd videos. Comprehensive experiments on emotion recognition at individual, group, and video levels, using datasets including AffectNet and EmotiW2018, demonstrate the robustness and effectiveness of EmoNet, NVPF, and TNVPF. The paper also introduces a new dataset, GEVC, aiming to stimulate further research in predicting group-level emotion in videos. The limitations of face detectors are acknowledged, suggesting potential enhancements for crowd scene understanding.

Interesting study [15] has explores the cognitive processes involved in assessing the intensity of facial expressions in individual faces within a more naturalistic crowd context. The research reveals a bias in judgments of the intensity of happy and angry expressions towards the group mean expression intensity, even when the faces belong to different individuals. The researchers in [16] [17] have proposed a novel method in this task, participants are tasked with determining the predominant expression displayed by most faces in a crowd. Also, one aspect of gender differences in how congestion affects facial expressions is related to the societal expectations and norms surrounding gender roles.

3. Methodology

This research consists of two main steps which are data preparation and facial expression analysis as illustrated in Figure 1. Both steps are described elaborately in this section.

Step 1: Data preparation and metadata extraction
Dataset

This project has allocated funding for the purpose of expediting research endeavours aimed at advancing the development of a smart social city and facilitating the transformation of Makkah into a prominent city within the sphere of digital transformation. The dataset that is used here represents a religious event in Makkah city in Saudi Arabia. It should be noted that the proposed method can be applied to any crowded events. This data is in the form of video recorded on various days at the Grand Mosque, where the prayer attendance fluctuates between peak and normal periods. The videos can be accessed from the official channel of KSA Quran TV. In data preparation step, two sub-tasks are conducted that are crowd video acquiring and crowd density estimation to extract meta-data of crowd level.

Crowded scene acquiring: This task aims to identify the source of crowd events that are recorded on videos or captured on still images. The video stream is split into multiple frames that have faces of crowds. The frames are used as input for the next step. For meta-data extraction, crowd estimation techniques are used. There are several methods can be used for crowd counting from images, thus in our experiment dense-based estimation model [18] and [19] was used because of its credibility in this area. The frames are fed to crowd density estimation tool to assess the degree of crowding. In our experiment, the crowd density is categorised into five levels: very high, high, medium, low, and very low as conducted previously in this research [20]. Other crowd estimation methods can be adopted are mentioned in [21][20][22].

Step 2: Facial Expression Analysis (FEA)

Face emotion recognition is basically a multi-model classification task. The task aims to categorize or classify the observed facial features into predefined emotion or expression categories. It is a supervised learning problem where the model is trained on a dataset with labelled examples of facial expressions, and it learns to generalize its understanding to correctly classify new, unseen examples. It includes:

a) *Face detection and extraction:* There are numerous techniques for face detection however, in this research the RetinaFace model [23] is used for different reasons. The most important reason is that RetinaFace model is preliminary designed for detecting faces in crowd and has shown significant results for detecting tiny faces and occluded faces [24].

b) *Face expression recognition:* Our method for face expression recognition is based on transfer learning which enables the utilization of knowledge acquired from a large dataset through a deep convolutional neural network. This allows for the application of the network's feature extraction capabilities in our domain scenario. This approach is suggested by considerable number of research [25][26]. The adopted model is Resnet50 trained on Face2[27]. In order to fine tune the pre-trained model, the adaptation class mode is followed here with small amount of our labelled data.

c) *Transfer learning of face emotion recognition using Resnet50 deep model*: The model Resnet50, which stands for Residual Network, is a specific type of convolutional neural network (CNN) that was introduced by Microsoft Research in 2015 [28]. It is used in this research as it showed a state-of-art results of face expression recognition according to [29], [28] and [27]. ResNet50 is a pretrained model with the Imagenet data [30]. It is highly efficient in training deep networks by adding extra layers and creating a more profound network. By incorporating skip connections, it effectively addresses the vanishing gradient problem. The powerful advantage of pretrained model is in making it publicly accessible and enable researchers to evaluate even smaller detest. This involves utilizing a previously trained network capable of extracting valuable features. weights of the Resnet50 model can be downloaded from Deepface python package. The general approach for transfer learning in the context of facial expression recognition requires importing a pre-trained Convolutional Neural Network (CNN) model, freezing parameters (weights) within the lower convolutional layers of the model, integrating a classifier, incorporating multiple layers endowed with trainable parameters, into the model and then training on the classifier layers using the provided training data in our task.

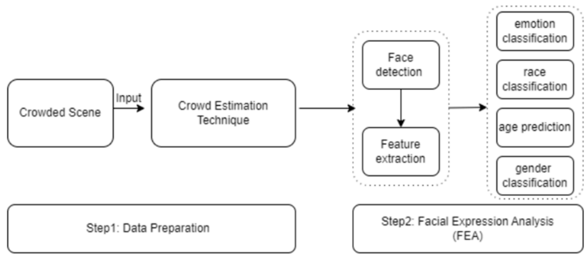


Figure 1:Research Design

4. Results

Crowd density estimation technique has counted the number of people in a sense. Interestingly, the estimated number of people in crowd sense is different than the number of analysable faces as shown in Table 1. This difference is because of some detected faces are blurred and analysing its expression is not applicable.

Table 1:: Difference between crowd estimation counting based on map density and the number of analyzable faces.

Crowd size	Very low	Low	Medium	High	Very high
crowd counting	106	229	650	830	1282

number of analysable faces	100	220	350	411	627

The results of crowd-density estimation are demonstrated in Figure2 showing five different crowd scenes at various densities. The first column shows actual photos of the Grand Mosque area with crowds at these five levels. The second column provides an estimated count of individuals present for each level. The third column presents corresponding density maps created using colours to represent the concentration of people; darker colours indicate higher densities. The crowd levels range from very low to very high, with corresponding increases in the number of people present and the colour intensity on the density maps. The crowd counts increase progressively from top to bottom: 106 for very low, 229 for low, 650 for medium, 830 for high, and 1282 for very high. The density maps use colour gradients to visually represent crowd concentrations; they become more intense as the crowd size increases.

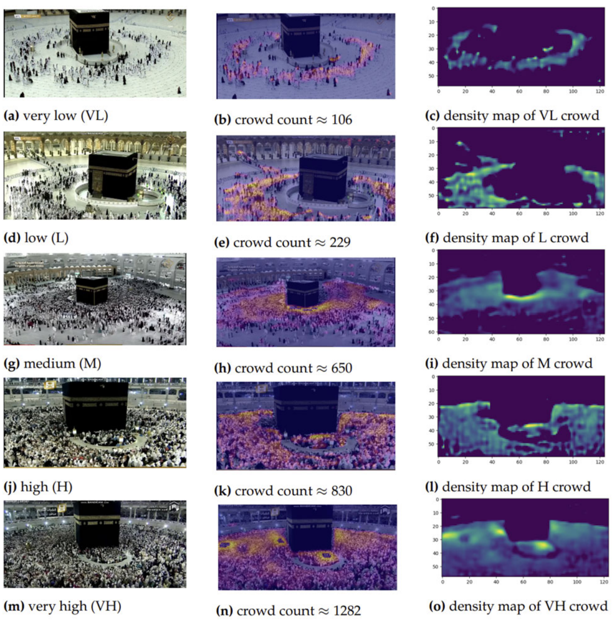


Figure 2: Different crowd scenes

4.1 Pre-trained model performance

The dataset sample used for training is 400 faces selected from different crowd density level and 240 faces were annotated manually to be used for validation. The model overall performance was excellent with accuracy of 93% without any tailored pre-processing steps. The recognition accuracy of ”Angry” and ”Sad” is achieved at 87%, while other expressions such as "Surprise","Neutral" and "Fear" are better recognized with 97%. However, the emotions "Disgust" and "Happy" have a moderate recognition accuracy with 90% and 93% respectively. Figure 3 shows the confusion matrix of model performance on a validation dataset.

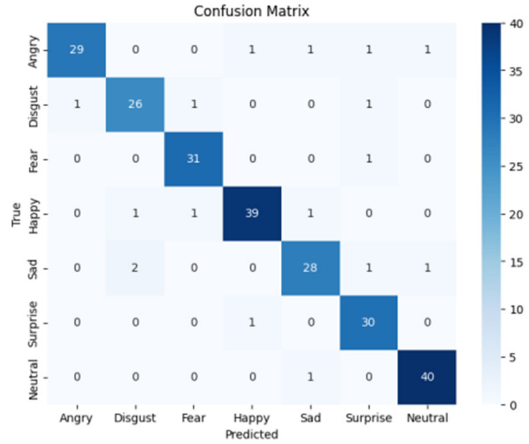


Figure 3:Confusion matrix of face expression model performance

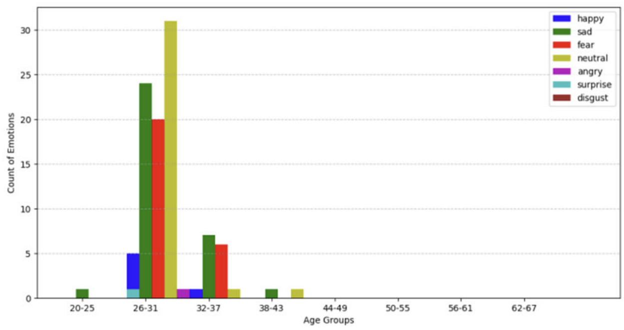
4.2 Facial expression analysis

The model has recognized 7 general emotions which are: angry, sad, happy, fear, surprise, disgust and neutral as demonstrated in Figure 4. The bar graphs representing the count of emotions across different age groups. The x-axis represents age groups: 20- 25, 26-31, 32-37, 38-43, 44-49, 50-55, 56-61 and 62-67. The y-axis represents the count of emotions ranging from 0 to over 80. There are seven different emotions represented by distinct colors: happy, sad, fear, neutral, angry, surprise and disgust for all crowd levels. In (a) ’very low’ crowd, age group “26–31” has the highest emotional counts overall. Emotion “neutral” is most prominent in age group “26–31”. Emotion “sad” is second most common in this same age group. Age groups “50–55”, “56–61”, and “62–67” have very low counts of all emotions.

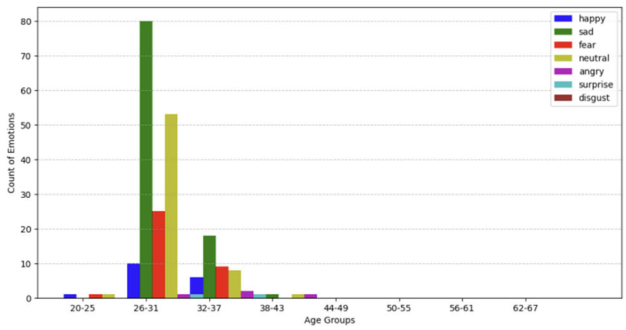
In (b) ’low’ crowd density images, ’sad’ was the major feeling of the faces with 100 out of 220. furthermore, ’neutral’ and ’fear’ have 60 and 35 respectively.

Face expression analysis of (c) the ’medium’ crowd density event has revealed that age group 26–31 has the highest count of fear emotion. Sad emotion is most prominent in age group 32–37. Other emotions are distributed at lower counts across all age groups. Interestingly, the crowd of medium density has shown for the first time ’disgust’ feeling. Whereas ’sad’, ’fear’ and ’neutral’ were observed in all groups of age with 158, 70 and 62 respectively.

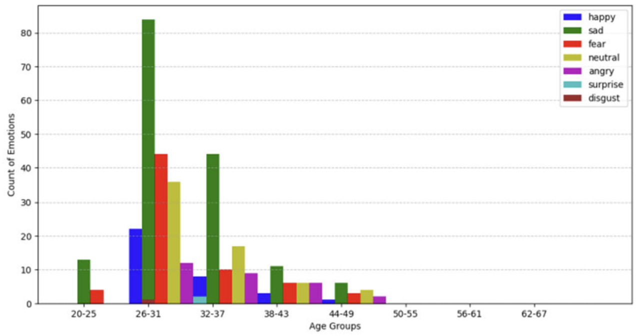
In (d) and (e) that show ’high’ and ’very high’ crowd scale, all age groups have high counts of ’sad’,’neutral’ and ’fear’ emotions. Sadness peaks in the age group of (e). Anger appears to be consistent across all age groups but peaks at ages between 44–49 of ’very high’ crowd. Surprise and disgust have relatively low counts across all age groups in both (d) and (e).



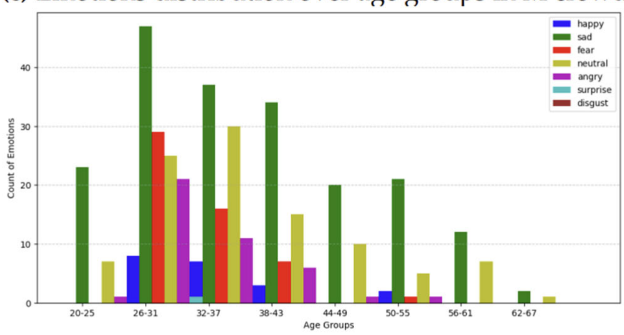
(a) Emotions distribution over age groups in VL crowd



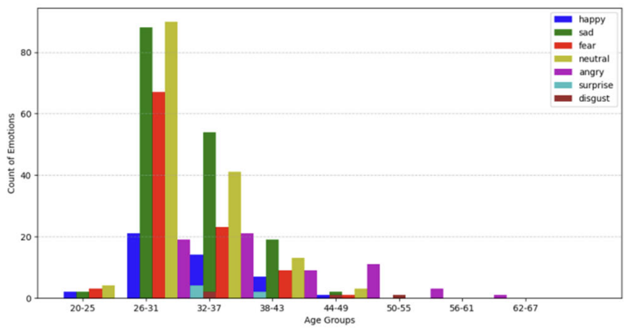
(b) Emotions distribution over age groups in L crowd



(c) Emotions distribution over age groups in M crowd



(d) Emotions distribution over age groups in H crowd



(e) Emotions distribution over age groups in VH crowd

Figure 4:Emotions distribution over different crowd densities

Factors influencing gender differences in facial expressions was not included in this experiment. Firstly, hormonal variations between males and females can influence emotional responsiveness and subsequently affect facial expression patterns. Additionally, cultural and societal expectations play a significant role in shaping how individuals express their emotions. For instance, certain cultures may encourage men to hide any signs of vulnerability or

discomfort, leading them to exhibit less varied facial expressions compared to women.
Furthermore, individual personality traits also contribute to the diversity of facial expressions exhibited during congestion. Some people may naturally have a more expressive face regardless of their gender, while others might be more reserved in displaying their emotions publicly.

5. Discussion and Conclusion

Addressing the discrepancies in face expression recognition in religious contexts is the most critical finding of this research. Emotion recognition systems can play a pivotal role in understanding human behaviour and responses in various scenarios. However, our experiment focusing on face expression recognition in the context of a religious event revealed a notable discrepancy between the labelled emotion ('sad') and the observed expression ('tranquillity'). The noticeable count of sadness feeling in the experiment has derived us to dive deeper with the data used. Examples from our data cannot be shown here for protecting faces privacy. However, this incongruity prompts a critical examination of the contextual and cultural precise inherent in emotion recognition technologies. It should be noted that, human emotions related to tranquillity can be ambiguous in terms of its interpretation as mentioned in [31]. This challenge we encountered lies in the cultural and contextual sensitivity of emotion recognition algorithms. Emotions, especially in the context of religious events, may manifest in ways that diverge from conventional labels. In our case, faces labelled as 'sad' predominantly reflected a sense of tranquillity and reflection synonymous with the religious setting. The issue is further compounded by the bias inherent in training datasets which was highlighted also in [27]. Often, these datasets lack the diversity needed to capture the full spectrum of human expressions across various contexts. As a result, models may struggle to generalize to novel situations, such as religious events, where expressions might deviate from typical dataset norms. Moreover, emotion recognition can benefit significantly from a multimodal approach. Beyond facial expressions, integrating contextual information, voice tone, and body language can enhance the model's capacity to capture the intricacies of emotions in diverse and nuanced situations. Hence, these challenges urge the need for reconsidering facial expression analysis as a standalone assessment method. It is essential to acknowledge the limitations of emotion recognition technology and communicate these limitations transparently to users. Recognizing the potential biases and contextual constraints of the model ensures responsible and ethical deployment in sensitive settings.

This study conducted to explore the influence of crowd size on human emotions by leveraging facial expression analysis, examining the interplay between crowd behaviour, face recognition technology, and deep learning algorithms. The objective was to discover if varying crowd levels induce notable changes in emotional states as reflected in facial expressions. However, the results of our investigation reveal a subtle and unexpected outcome. Contrary to initial hypotheses, the analysis indicates that varying crowd sizes do not yield a significant impact on the distribution of emotional states as inferred from facial expressions. Despite leveraging advanced face recognition technology and employing sophisticated deep learning algorithms, the observed emotional changes across different crowd levels appear to be minimal. These findings prompt a reassessment of the presumed correlation between crowd dynamics and alterations in emotional states as expressed through facial expressions [32]. The lack of substantial impact on feeling distribution across diverse crowd sizes suggests a complexity in the relationship between crowd behaviour and individual emotional responses, which may extend beyond the current scope of understanding or necessitate a more refined methodological approach. While the study provides valuable insights into the interplay of crowd size and human emotions, the unexpected nature of the results underscores the complicated and multifaceted nature of emotional responses within a crowd setting. Further research and exploration are justified to unravel the underlying factors contributing to emotional dynamics in crowded environments and to refine models that may better capture the complexity of these interactions. Moreover, it is important to consider individual differences and cultural variations when interpreting these expressions. Facial expression analysis provides valuable insights into the emotional state of congestion but should be used alongside other assessment methods for a comprehensive understanding.

Acknowledgment

The author would like to express her gratitude to Umm Alqura University for supporting this research (ID:4401095348).

References

[1] Mehrabian, A.; Ferris, S.R. Inference of attitudes from nonverbal communication in two channels. Journal of consulting psychology 1967, 31, 248.

[2] Quach, K.G.; Le, N.; Duong, C.N.; Jalata, I.; Roy, K.; Luu, K. Non-volume preserving-based fusion to group-level emotion recognition on crowd videos. Pattern Recognition 2022, 128, 108646.

[3] Veltmeijer, E.A.; Gerritsen, C.; Hindriks, K.V. Automatic emotion recognition for groups: a review. IEEE Transactions on Affective Computing 2021, 14, 89–107.

[4] Holder, R.P.; Tapamo, J.R. Using facial expression recognition for crowd monitoring. In Proceedings of the Image and Video Technology: 8th Pacific-Rim Symposium, PSIVT 2017, Wuhan, China, November 20-24, 2017, Revised Selected Papers 8. Springer, 2018, pp. 463–476.

[5] Dildine, T.C.; Atlas, L.Y. The need for diversity in research on facial expressions of pain. Pain 2019, 160, 1901.

[6] Ghosh, R.; Sinha, D. Human emotion recognition by analyzing facial expressions, heart rate and blogs using deep learning method. Innovations in Systems and Software Engineering 2022, pp. 1–9.

[7] Rosenberg, E.L.; Ekman, P. What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS); Oxford University Press, 2020.

[8] Li, S.; Deng, W. Deep Facial Expression Recognition: A Survey. IEEE Transactions on Affective Computing 2022, 13, 1195–1215.

[9] Turan, B.; Algedik Demirayak, P.; Yildirim Demirdogen, E.; Gulsen, M.; Cubukcu, H.C.; Guler, M.; Alarslan, H.; Yilmaz, A.E.; Dursun, O.B. Toward the detection of reduced emotion expression intensity: an autism sibling study. Journal of Clinical and Experimental Neuropsychology 2023, pp. 1–11.

[10] Döllinger, L.; Laukka, P.; Högman, L.B.; Bänziger, T.; Makower, I.; Fischer, H.; Hau, S. Training emotion recognition accuracy: Results for multimodal expressions and facial micro expressions. Frontiers in Psychology 2021, 12, 708867.

[11] Rokhsaritalemi, S.; Sadeghi-Niaraki, A.; Choi, S.M. Exploring Emotion Analysis using Artificial Intelligence, Geospatial Information Systems, and Extended Reality for Urban Services. IEEE Access 2023.

[12] Gjoreski, M.; Kiprijanovska, I.; Stankoski, S.; Mavridou, I.; Broulidakis, M.J.; Gjoreski, H.; Nduka, C. Facial EMG sensing for monitoring affect using a wearable device. Scientific Reports 2022, 12, 16876.

[13] Engelniederhammer, A.; Papastefanou, G.; Xiang, L. Crowding density in urban environment and its effects on emotional responding of pedestrians: Using wearable device technology with sensors capturing proximity and psychophysiological emotion responses while walking in the street. Journal of Human Behavior in the Social Environment 2019, 29, 630–646.

[14] Abbas, A.; Chalup, S.K. Group emotion recognition in the wild by combining deep neural networks for facial expression classification and scene-context analysis. In Proceedings of the Proceedings of the 19th ACM international conference on multimodal interaction, 2017, pp. 561–568.

[15] Griffiths, S.; Rhodes, G.; Jeffery, L.; Palermo, R.; Neumann, M.F. The average facial expression of a crowd influences impressions of individual expressions. Journal of Experimental Psychology: Human Perception and Performance 2018, 44, 311.

[16] Bucher, A.; Voss, A. Judging the mood of the crowd: Attention is focused on happy faces. Emotion 2019, 19, 1044.

[17] Halamová, J.; Strnádelová, B.; Kanovsky, M.; Moró, R.; Bieliková, M. Anger or happiness ` superiority effect: A face in the crowd study involving nine emotions expressed by nine people. Current Psychology 2023, 42, 15381–15387.

[18] Teršek, M.; Kljun, M.; Peer, P.; Emerši˘c, Ž. Re-evaluation of the CNN-based state-of-the-art crowd-counting methods with enhancements. Computer Science and Information Systems 2022, 19, 1177–1198.

[19] Shami, M.B.; Maqbool, S.; Sajid, H.; Ayaz, Y.; Cheung, S.C.S. People counting in dense crowd images using sparse head detections. IEEE Transactions on Circuits and Systems for Video Technology 2018, 29, 2627–2636.

[20] Fu, M.; Xu, P.; Li, X.; Liu, Q.; Ye, M.; Zhu, C. Fast crowd density estimation with convolutional neural networks. Engineering Applications of Artificial Intelligence 2015, 43, 81–88.

[21] Still, G.K. Applied crowd science; CRC Press, 2021.

[22] Fan, Z.; Zhang, H.; Zhang, Z.; Lu, G.; Zhang, Y.; Wang, Y. A survey of crowd counting and density estimation based on convolutional neural network. Neurocomputing 2022, 472, 224–251.

[23] Deng, J.; Guo, J.; Ververas, E.; Kotsia, I.; Zafeiriou, S. Retinaface: Single-shot multi-level face localisation in the wild. In Proceedings of the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020, pp. 5203–5212.
[24] International journal of computer vision 2023.
[25] Akhand, M.; Roy, S.; Siddique, N.; Kamal, M.A.S.; Shimamura, T. Facial emotion recognition using transfer learning in the deep CNN. Electronics 2021, 10, 1036.
[26] Rescigno, M.; Spezialetti, M.; Rossi, S. Personalized models for facial emotion recognition through transfer learning. Multimedia Tools and Applications 2020, 79, 35811–35828.
[27] Kong, Y.S.; Suresh, V.; Soh, J.; Ong, D.C. A systematic evaluation of domain adaptation in facial expression recognition. arXiv preprint arXiv:2106.15453 2023.
[28] Atanassov, A.; Pilev, D. Pre-trained Deep Learning Models for Facial Emotions Recognition. In Proceedings of the 2020 International Conference Automatics and Informatics (ICAI). IEEE, 2020, pp. 6.
[29] Li, S.; Deng, W. A deeper look at facial expression dataset bias. IEEE Transactions on Affective Computing 2020, 13, 881–893.
[30] Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. Imagenet large scale visual recognition challenge. International journal of computer vision 2015, 115, 211–252.
[31] Summa, M. Are emotions “recollected in tranquility”? Phenomenological reflections on emotions, memory, and the temporal dynamics of experience. In Feeling and value, willing and action: Essays in the context of a phenomenological psychology; Springer, 2014; pp. 163–181.
[32] Reicher, S. The psychology of crowd dynamics. Blackwell handbook of social psychology: Group processes 2001, pp. 182–208.
[33] Qureshi, A. W., Monk, R. L., Quinn, S., Gannon, B., McNally, K., & Heim, D. Catching a smile from individuals and crowds: Evidence for distinct emotional contagion processes. Journal of Personality and Social Psychology 2023. Advance online publication. https://doi.org/10.1037/pspi0000445
[34] Halamová, J., Strnádelová, B., Kanovský, M., Moró, R. and Bieliková, M. Anger or happiness superiority effect: A face in the crowd study involving nine emotions expressed by nine people. Current Psychology 2023, 42(18), pp.15381-15387.
[35] Siddiqui MF, Javaid AY. A multimodal facial emotion recognition framework through the fusion of speech with visible and infrared images. Multimodal Technologies and Interaction 2020, 4(3):46.
[36] Ezzameli K, Mahersia H. Emotion recognition from unimodal to multimodal analysis: A review. Information Fusion 2023, 101847.

Funding: “The author would like to thank the Deanship of Scientific Research at Umm Al-Qura University for supporting this work by Grant Code: (23UQU43101400DSR008).”