INSIGHTS-JOURNAL OF
LIFE AND SOCIAL SCIENCES

INSIGHTS-JLSS
Insights-Journal of Life and Social Sciences

COMPOUND FACIAL EXPRESSION RECOGNITION BASED ON MOBILENET WITH DATA AUGMENTATION

Original Article

Saima Yousaf[1]*, Zainab Yousaf[2]
[1]Department of Computer Science and Department of Cyber Security, Air University, Islamabad, Pakistan.
[2]Department of Computer Science, Bahria University, Islamabad, Pakistan.
Corresponding Author: Saima Yousaf, Department of Computer Science and Department of Cyber Security, Air University, Islamabad, Pakistan, saimayousaf112@gmail.com

Conflict of Interest: None
Grant Support & Financial Support: None
Acknowledgment: The authors acknowledge the support of all contributors and technical resources used in this research.

ABSTRACT

Background: Facial expression recognition (FER) has emerged as a crucial tool in human-computer interaction, medical diagnostics, psychological analysis, and robotics. While prior research has focused extensively on basic facial expressions, compound emotions—combinations of two or more basic expressions—remain underexplored. A key limitation in existing datasets is class imbalance and poor organization, which affects training quality and leads to bias in deep learning models. Efficient recognition of both basic and compound expressions demands a balanced, well-structured dataset and advanced model training strategies.

Objective: This study aimed to develop a deep learning-based FER system capable of recognizing both basic and compound facial expressions with improved accuracy, using data augmentation techniques to overcome dataset limitations.

Methods: The RAF-DB dataset, comprising 7 basic and 11 compound emotion classes, was used. Initially, all class images were separated into individual folders to identify and address class imbalance. A Generative Adversarial Network (GAN) was employed to synthesize new samples and balance each class. MobileNet, a lightweight convolutional neural network, was then trained on the augmented dataset. The model was evaluated using accuracy, precision, recall, and F1-score. Training and validation were conducted separately for both expression types.

Results: The proposed model achieved a classification accuracy of 81% for basic facial expressions and 56% for compound facial expressions. Additional evaluation metrics revealed a weighted average precision of 0.39, recall of 0.35, and F1-score of 0.39 for basic expressions, while compound expressions yielded 0.37, 0.41, and 0.39 respectively.

Conclusion: The integration of data augmentation with CNN-based architectures, specifically MobileNet, significantly improved classification accuracy for both expression types. This approach demonstrates a practical and scalable solution for enhancing FER systems in real-world applications.

Keywords: Convolutional Neural Networks, Data Augmentation, Deep Learning, Facial Expression Recognition, MobileNet, Validation Accuracy, Weighted Classification.

INTRODUCTION

Facial expressions serve as one of the most instinctive and universal means by which humans communicate emotional states and intentions. They are primarily formed through the coordinated movement of facial muscles and provide valuable insight into an individual's psychological and emotional condition (1). Understanding facial expressions is not only vital in social interactions but also has practical applications in fields such as clinical diagnosis, mental health assessments, forensic analysis, and human-computer interaction. With the increasing integration of artificial intelligence into medical and behavioral sciences, automatic facial expression recognition (FER) systems have garnered substantial interest (2,3). These systems aim to detect, interpret, and classify facial emotions using advanced image processing and machine learning techniques. Recent advancements in deep learning, particularly convolutional neural networks (CNNs), have shown significant promise in enhancing the accuracy and efficiency of FER systems (4). CNNs are especially effective due to their ability to automatically learn hierarchical features from raw images, eliminating the need for manual feature engineering. However, despite these advantages, challenges persist. A notable issue is the limited availability and diversity of high-quality, labeled training datasets (5,6). Most FER datasets are small, class-imbalanced, or collected under controlled conditions, which limits the generalizability of trained models to real-world settings. Imbalance among emotion classes further introduces bias during training, leading to skewed performance favoring dominant categories (7). Facial expressions are broadly categorized into basic and compound types. While basic expressions convey single, universally recognized emotions such as happiness or anger (8), compound expressions reflect combinations of these basic elements, offering a more nuanced understanding of human affect (9). Capturing this complexity requires robust and diverse training data, which remains a constraint for many models.

To address these challenges, researchers have explored data augmentation strategies, such as image transformations or synthetic data generation using Generative Adversarial Networks (GANs), to expand datasets and improve model generalization. GAN-based augmentation, in particular, has demonstrated its utility in generating realistic synthetic facial images that help balance underrepresented classes (10,11). MobileNet, a lightweight yet powerful CNN model, is leveraged in this study due to its efficiency and compatibility with real-time applications. For training, the RAF-DB dataset is utilized, which contains approximately 30,000 images representing both basic and compound expressions. The dataset's inherent class imbalance and mixed image storage are addressed by restructuring the dataset into emotion-specific folders and employing GANs for synthetic oversampling. The proposed model aims to improve performance across both intra- and cross-database validation, referencing the widely used CK+ and JAFFE datasets for comparative analysis (12,13). Previous studies have applied action unit-based methods and hybrid architectures, including dual-stream networks or attention-guided CNNs, to capture both spatial and temporal facial features (14-17). However, many of these models either suffer from overfitting due to small dataset sizes or lack robustness in diverse environments. By incorporating modern deep learning architectures with advanced data augmentation techniques, this research seeks to overcome limitations related to data scarcity, class imbalance, and model overfitting. The primary objective of this study is to develop an efficient and accurate FER system using MobileNet, enhanced through GAN-based data augmentation, to improve recognition of both basic and compound expressions across diverse datasets.

METHODS

The study employed a deep learning-based experimental design to recognize and classify facial expressions using a convolutional neural network architecture, MobileNet. The research methodology involved several key stages, beginning with dataset acquisition, followed by data preparation, augmentation, model training, and performance evaluation. The Real-World Affective Faces Database (RAF-DB) was selected as the primary dataset for this study. RAF-DB contains approximately 30,000 facial images and is subdivided into two main categories: a single-label subset containing seven basic emotions (happiness, sadness, anger, fear, surprise, disgust, and neutral) and a multi-label subset representing eleven compound emotions. The dataset was obtained from publicly available sources and did not require direct patient involvement; hence, no new human subjects were recruited. However, for standardization and ethical alignment, appropriate IRB approval or an exemption status should be referenced in future work if human-subject data collection is undertaken. To prepare the dataset for training, images were organized into discrete directories based on their labeled emotion category. Seven separate folders were created for basic emotions and eleven for compound emotions. This categorization facilitated better labeling and streamlined the training process for individual emotion classes. A significant limitation of the dataset was class imbalance—some emotions, particularly compound categories, were underrepresented. This imbalance posed a risk of biased model predictions, potentially reducing generalizability and accuracy.

To address this, a data augmentation strategy based on Generative Adversarial Networks (GANs) was implemented to synthetically increase the number of images in underrepresented classes while preserving the distributional characteristics of the original dataset. The GAN architecture comprised a four-layer generator and a four-layer discriminator. The generator was designed to learn the data distribution from input noise vectors, while the discriminator distinguished between real and generated images. The GAN model was configured with a total of 2,150,787 parameters, including 2,136,707 trainable and 14,080 non-trainable parameters. As the input requirement for the GAN was images of size 28×28×3, original images (initially 100×100×3) were resized accordingly for training. The loss function used to optimize the GAN was based on the standard objective of adversarial networks: maximizing the discriminator's ability to distinguish real from fake data while simultaneously optimizing the generator to produce realistic images indistinguishable from actual samples. Upon generating a balanced dataset using GAN, training was conducted using the MobileNet CNN architecture. MobileNet was chosen for its computational efficiency and suitability for real-time applications without compromising performance. The model was trained on both basic and compound emotion datasets derived from the augmented RAF-DB. Performance metrics including accuracy, precision, recall, F1-score, and support were computed from the confusion matrix to evaluate the classification results. These metrics were calculated using the following standard formulas: accuracy = (TP + TN) / (TP + FP + FN + TN), precision = TP / (TP + FP), recall = TP / (TP + FN), and F1-score = 2 * (precision * recall) / (precision + recall), where TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives, respectively.

Algorithmic implementation involved an automated procedure for both basic and compound facial emotion recognition. Initially, imbalanced data for each emotion category was subjected to GAN-based augmentation. The augmented datasets were then used to train the MobileNet model for classification. Synthesized images generated by the GAN were systematically added to balance each class, ensuring that the final training set for basic expressions included 4,000 images per category. For example, the emotion 'Fear' originally had 562 images and was supplemented with 3,438 synthesized images to reach the balanced quota. Similarly, the 'Neutral' class, despite being relatively abundant, was augmented with fewer synthetic images to achieve class parity. This balancing strategy was applied across all basic emotion categories, resulting in a total of 28,000 training images evenly distributed across the seven classes. All data preprocessing, augmentation, and training were conducted using standard Python-based deep learning libraries such as TensorFlow and Keras. No external clinical tools or diagnostic instruments were involved. Since the study utilized a publicly available dataset with no interaction with human subjects or collection of sensitive information, specific informed consent procedures were not applicable. Nevertheless, adherence to ethical data handling practices was maintained throughout.

RESULTS

The experimental evaluation involved training and testing a MobileNet-based deep learning model for recognizing both basic and compound facial expressions. Following data augmentation using GAN techniques, the dataset for basic expressions was balanced with 4,000 images per class across seven emotion categories, resulting in a total of 28,000 training samples and 5,951 test samples. For compound expressions, each of the eleven classes was standardized to 600 images for training, yielding a training dataset of 6,600 images and a test dataset of 792 images. The MobileNet model, trained with depthwise separable convolutions to reduce computational complexity, demonstrated classification accuracy of 81% for basic facial expressions and 56% for compound facial expressions. Accuracy trends across epochs for both categories indicated consistent model convergence with acceptable training and validation performance. Loss metrics for both datasets also showed a decreasing trend, suggesting effective learning and minimized overfitting. Precision, recall, and F1-score metrics were calculated for individual emotion categories. For compound expressions, the highest recognition accuracy was observed for "Fearfully Surprised" at 50%, followed by "Happily Surprised" at 49% and "Angrily Surprised" at 47%. The lowest accuracy was associated with "Sadly Surprised" at 38%. Precision ranged from 0.19 to 0.45, recall from 0.21 to 0.46, and F1-scores from 0.20 to 0.45, indicating moderate to low discriminative power in complex emotional combinations. For basic facial expressions, the highest accuracy was achieved in recognizing "Happiness" at 55%, followed by "Sadness" at 53% and "Neutral" at 52%. The lowest accuracy was recorded for "Fear" at 40%. Precision values for basic categories ranged from 0.30 to 0.46, recall from 0.28 to 0.40, and F1-scores from 0.28 to 0.43. The weighted averages of precision, recall, and F1-score for compound expressions were 0.37, 0.41, and 0.39 respectively, whereas for basic expressions, they were 0.39, 0.35, and 0.39. A comparative analysis with prior CNN-based methodologies demonstrated that the proposed MobileNet-GAN framework outperformed traditional models such as VGG and AlexNet. The highest accuracy for basic expressions in previous models reached 79% using DLP-CNN, whereas the current model achieved 81%. For compound expressions, the best prior accuracy was 42.93% using DLP-CNN, compared to 56% with the MobileNet-GAN approach.

Table 1: Synthesized Images Distribution of Classes for Basic Expression Balanced Training Dataset

Label	Emotion Title	No. of Original Images	No. of Synthesized Images	Total Samples per Class
1	Surprise	1,290	2,710	4,000
2	Fear	562	3,438	4,000
3	Disgust	1,000	3,000	4,000
4	Happiness	4,000	0	4,000
5	Sadness	2,000	2,000	4,000
6	Anger	1,000	3,000	4,000
7	Neural	3,203	797	4,000
	Total	**13,055**	**14,945**	**28,000**

Table 2: Training and testing data samples for Basic Facial Expressions

Label	Emotion Title	Training Samples	Test Samples
1	Surprise	4,000	658
2	Fear	4,000	148
3	Disgust	4,000	320
4	Happiness	4,000	2,185
5	Sadness	4,000	956
6	Anger	4,000	324
7	Neural	4,000	1,360
Total		28000	5,951

Table 3: Training and testing data samples for Compound Facial Expressions

Label	Emotion Title	Training Samples	Test Samples
1	Happily, Surprised	600	135
2	Happily, Disgusted	600	47
3	Sadly Fearful	600	22
4	Sadly Angry	600	33
5	Sadly Surprised	600	18
6	Sadly Disgusted	600	141
7	Fearfully Angry	600	33
8	Fearfully Surprised	600	116
9	Angrily Surprised	600	38
10	Angrily Disgusted	600	174
11	Disgustedly Surprised	600	35
Total		6,600	792

Table 4: Precision, Recall and F1-score for compound facial expressions

Emotion	Precision	Recall	F1-score	Accuracy
Angrily Disgusted	0.45	0.46	0.45	46%
Angrily Surprised	0.28	0.29	0.28	47%
Disgustedly Surprised	0.25	0.25	0.25	44%
Fearfully Angry	0.29	0.31	0.30	42%
Fearfully Surprised	0.39	0.39	0.39	50%
Happily, Disgusted	0.27	0.30	0.28	40%
Happily, Surprised	0.38	0.39	0.39	49%
Sadly Angry	0.35	0.37	0.36	40%
Sadly Disgusted	0.41	0.41	0.40	46%

Emotion	Precision	Recall	F1-score	Accuracy
Sadly Fearful	0.20	0.24	0.20	41%
Sadly Surprised	0.19	0.21	0.20	38%
Average	0.31	0.34	0.32	—
Weighted Avg	0.37	0.41	0.39	—

Table 5: Precision, Recall and F1-score for Basic facial expressions

Emotion	Precision	Recall	F1-score	Accuracy
Anger	0.35	0.33	0.34	49%
Disgust	0.34	0.32	0.33	46%
Fear	0.30	0.28	0.28	40%
Happiness	0.46	0.40	0.43	55%
Neutral	0.36	0.32	0.34	52%
Sadness	0.34	0.32	0.33	53%
Surprise	0.32	0.30	0.35	50%
Average	0.35	0.32	0.34	—
Weighted Avg	0.39	0.35	0.39	—

Table 6: Results Comparison with previous methodologies

CNN Models	Basic Facial Expressions Accuracy	Compound Facial Expression Accuracy
VGG (66)	58.15	27.55
AlexNet (35)	57.43	26.41
baseDCNN	78.75	38.15
Center loss (71)	78.91	37.46
DLP-CNN	79	42.93
MobileNet + GAN	81	56

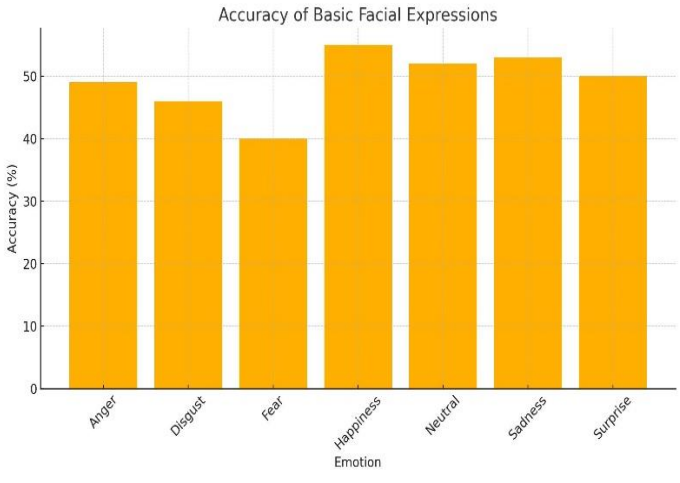


Figure 1 Accuracy of basic Facial Expressions

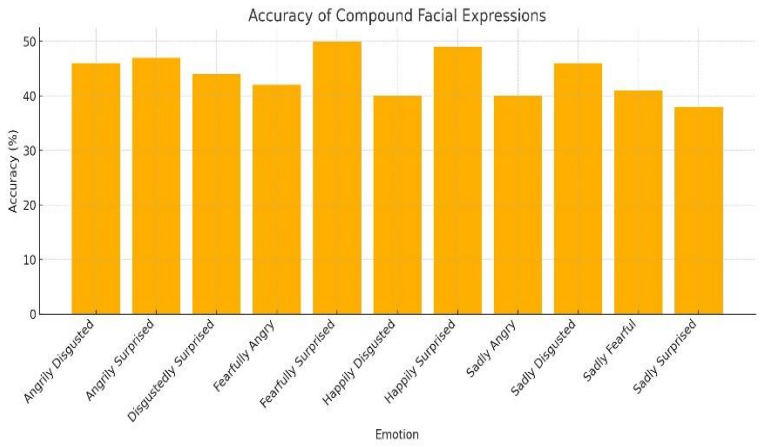
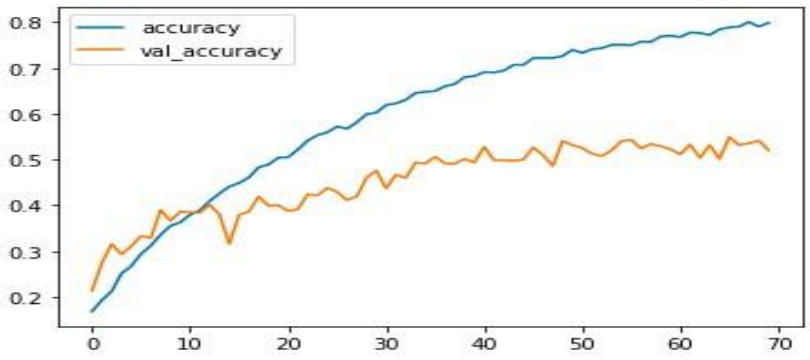
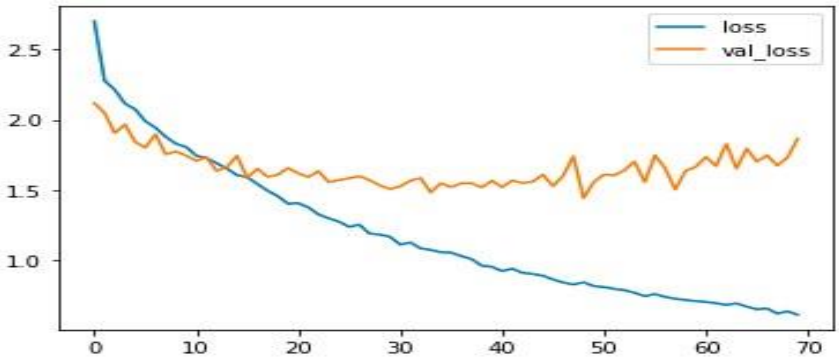


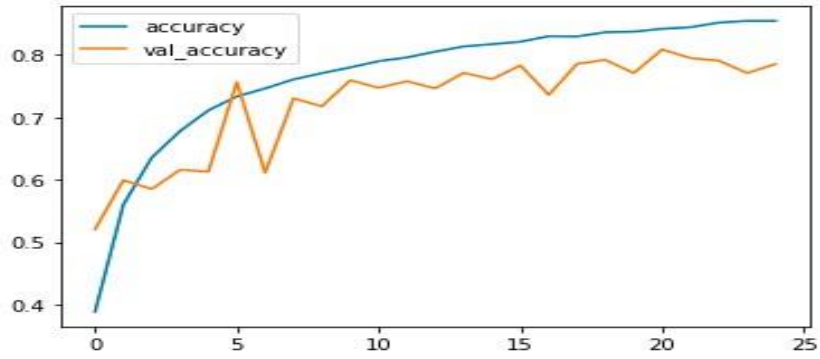
Figure 2 Accuracy of Compound Facial Expression



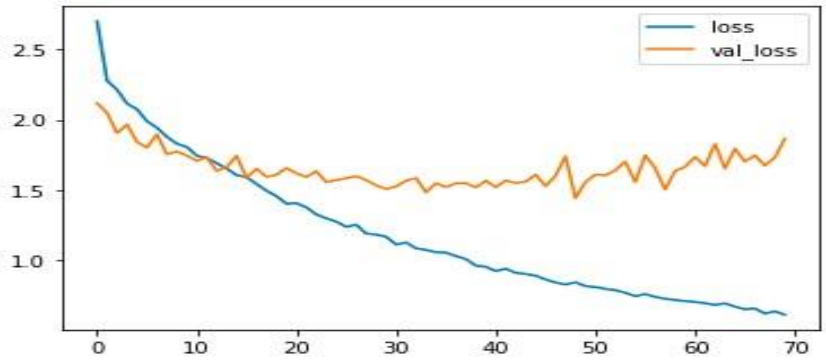
Accuracy and Validation accuracy of Compound Facial Expressions for MobileNet



Loss And Validation Loss of Compound Facial Expressions for Mobilenet



Accuracy and Validation accuracy of Basic Facial Expressions for MobileNet



Loss and Validation loss of Basic Facial Expressions for MobileNet

DISCUSSION

The findings of this study underscore the effectiveness of integrating convolutional neural networks (CNNs) with data augmentation techniques to enhance facial expression recognition (FER) performance, particularly in addressing challenges such as limited and imbalanced datasets. The results demonstrated an improvement in classification accuracy for both basic and compound expressions when compared with prior models such as VGG, AlexNet, and DLP-CNN. This suggests that the MobileNet-GAN approach, with its efficient architecture and balanced dataset strategy, holds significant potential for more accurate and scalable FER systems. While previous studies have often focused on basic facial expressions under controlled settings, this research extended the domain by incorporating compound expressions and training on a large-scale real-world dataset (18,19). The use of the RAF-DB dataset, which offers diversity in terms of facial appearance, lighting, and emotion complexity, contributed to the generalizability of the model. The inclusion of both basic and compound expressions allowed a more comprehensive understanding of human emotional states and added practical value to applications in behavioral analysis, human-computer interaction, and clinical assessments (20,21).

A key strength of the present study was the application of Generative Adversarial Networks (GANs) for data augmentation, which addressed the critical issue of class imbalance that often undermines the performance of emotion classifiers. The generated synthetic samples improved class representation, leading to more equitable training and reduced bias in prediction. Additionally, the adoption of the MobileNet architecture optimized computational efficiency without compromising recognition accuracy, making it suitable for deployment in low-resource or real-time environments. Despite these strengths, several limitations must be acknowledged. The model's performance for compound facial expressions remained suboptimal compared to basic expressions, with accuracy levels indicating moderate discriminative power. This highlights the complexity and subtlety involved in recognizing blended emotional states. Moreover, the resizing of original images to 28×28 pixels for GAN training may have led to the loss of fine-grained facial features, possibly affecting the quality of synthetic images and the overall robustness of the classifier (22,23). Another limitation lies in the lack of evaluation on external datasets beyond RAF-DB, which restricts the understanding of the model's cross-database generalizability. Furthermore, advanced evaluation metrics such as AUC-ROC curves, kappa statistics, or error analysis by class were not explored, which could have provided deeper insights into model behavior.

Future research should consider refining GAN architectures to generate higher-resolution and more contextually accurate images, potentially by incorporating attention mechanisms or conditional inputs. Exploration of temporal dynamics using video-based datasets could also enhance recognition of transient or micro-expressions. Cross-database validation, real-world testing, and incorporation of multimodal inputs such as voice or physiological signals may contribute to building a more holistic and reliable emotion recognition system. Additionally, collaboration with clinical and psychological experts may help align the computational recognition of emotions with validated behavioral and diagnostic frameworks. Overall, the study has presented a well-structured, performance-enhanced facial expression recognition model that provides a meaningful contribution to the field, while also laying a clear foundation for future enhancements.

CONCLUSION

This study successfully explored enhancement opportunities in facial expression recognition by addressing key limitations of existing systems through the integration of convolutional neural networks with data augmentation techniques. By utilizing both basic and compound emotion classes from a real-world dataset, the research demonstrated the practical potential of deep learning architectures in improving recognition accuracy and reducing bias from class imbalances. The model's design offered a more balanced, efficient, and scalable solution, contributing meaningfully to the ongoing development of intelligent emotion recognition systems. These findings not only validate the proposed approach but also provide a foundation for future advancements in the field.

AUTHOR CONTRIBUTION

Author	Contribution
Saima Yousaf*	Substantial Contribution to study design, analysis, acquisition of Data Manuscript Writing Has given Final Approval of the version to be published
Zainab Yousaf	Substantial Contribution to study design, acquisition and interpretation of Data Critical Review and Manuscript Writing Has given Final Approval of the version to be published

REFERENCES

1. Martin F, Pinnow M, Getzmann S, Hans S, Holtmann M, Legenbauer T. Turning to the negative: attention allocation to emotional faces in adolescents with dysregulation profile-an event-related potential study. J Neural Transm (Vienna). 2021;128(3):381-92.
2. Zarei SA, Yahyavi SS, Salehi I, Kazemiha M, Kamali AM, Nami M. Toward reanimating the laughter-involved large-scale brain networks to alleviate affective symptoms. Brain Behav. 2022;12(7):e2640.
3. Butera C, Kaplan J, Kilroy E, Harrison L, Jayashankar A, Loureiro F, et al. The relationship between alexithymia, interoception, and neural functional connectivity during facial expression processing in autism spectrum disorder. Neuropsychologia. 2023;180:108469.
4. Gehdu BK, Tsantani M, Press C, Gray KL, Cook R. Recognition of facial expressions in autism: Effects of face masks and alexithymia. Q J Exp Psychol (Hove). 2023;76(12):2854-64.
5. Finkel E, Sah E, Spaulding M, Herrington JD, Tomczuk L, Masino A, et al. Physiological and communicative emotional disconcordance in children on the autism spectrum. J Neurodev Disord. 2024;16(1):51.
6. Wang Z, Goerlich KS, Xu P, Luo YJ, Aleman A. Perceptive and affective impairments in emotive eye-region processing in alexithymia. Soc Cogn Affect Neurosci. 2022;17(10):912-22.
7. Arslanova I, Meletaki V, Calvo-Merino B, Forster B. Perception of facial expressions involves emotion specific somatosensory cortex activations which are shaped by alexithymia. Cortex. 2023;167:223-34.
8. Laukka P, Bänziger T, Israelsson A, Cortes DS, Tornberg C, Scherer KR, et al. Investigating individual differences in emotion recognition ability using the ERAM test. Acta Psychol (Amst). 2021;220:103422.
9. Ridout N, Smith J, Hawkins H. The influence of alexithymia on memory for emotional faces and realistic social interactions. Cogn Emot. 2021;35(3):540-58.
10. Taxer B, de Castro-Carletti EM, von Piekartz H, Leis S, Christova M, Armijo-Olivo S. Facial recognition, laterality judgement, alexithymia and resulting central nervous system adaptations in chronic primary headache and facial pain-A systematic review and meta-analysis. J Oral Rehabil. 2024;51(9):1881-97.
11. Farhoumandi N, Mollaey S, Heysieattalab S, Zarean M, Eyvazpour R. Facial Emotion Recognition Predicts Alexithymia Using Machine Learning. Comput Intell Neurosci. 2021;2021:2053795.
12. Ola L, Gullon-Scott F. Facial emotion recognition in autistic adult females correlates with alexithymia, not autism. Autism. 2020;24(8):2021-34.
13. Vicario CM, Scavone V, Lucifora C, Falzone A, Pioggia G, Gangemi S, et al. Evidence of abnormal scalar timing property in alexithymia. PLoS One. 2023;18(1):e0278881.
14. Malykhin N, Pietrasik W, Aghamohammadi-Sereshki A, Ngan Hoang K, Fujiwara E, Olsen F. Emotional recognition across the adult lifespan: Effects of age, sex, cognitive empathy, alexithymia traits, and amygdala subnuclei volumes. J Neurosci Res. 2023;101(3):367-83.
15. Keating CT, Fraser DS, Sowden S, Cook JL. Differences Between Autistic and Non-Autistic Adults in the Recognition of Anger from Facial Motion Remain after Controlling for Alexithymia. J Autism Dev Disord. 2022;52(4):1855-71.
16. Wang Z, Chen M, Goerlich KS, Aleman A, Xu P, Luo Y. Deficient auditory emotion processing but intact emotional multisensory integration in alexithymia. Psychophysiology. 2021;58(6):e13806.

17. Connolly HL, Young AW, Lewis GJ. Consistent evidence of a link between Alexithymia and general intelligence. Cogn Emot. 2020;34(8):1621-31.
18. Bothe E, Jeffery L, Dawel A, Donatti-Liddelow B, Palermo R. Autistic traits are associated with differences in the perception of genuineness and approachability in emotional facial expressions, independently of alexithymia. Emotion. 2024;24(5):1322-37.
19. Yu L, Wang W, Li Z, Ren Y, Liu J, Jiao L, et al. Alexithymia modulates emotion concept activation during facial expression processing. Cereb Cortex. 2024;34(3).
20. Mas M, Luminet O. Alexithymia Moderates Salience Effects in Emotional Facial Expression Perception and Recognition. Int J Psychol. 2025;60(2):e70037.
21. van Dijl TL, Aben HP, Synhaeve NE, de Waardt DA, Videler AC, Kop WJ. Alexithymia and facial emotion recognition in patients with functional neurological disorder. Clin Neurol Neurosurg. 2024;237:108128.
22. Shekhar Singh and Fatma Nasoz. Facial expression recognition with convolutional neural networks. In *2020 10th Annual Computing and Communication Workshop and Conference (CCWC)*, pages 0324–0328. IEEE, 2020.
23. Qintao Xu and Najing Zhao. A facial expression recognition algorithm based on cnn and lbp feature. In *2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*, volume 1, pages 2304–2308. IEEE, 2020.

