Decoding Emotional Complexity: Challenges in Emotions Classification Using the CMU Dataset

Zineb Bougriche[1,*] and Rafał Gasz[1]

1 Opole University of Technology, Prószkowska Street 76, 45-758 Opole, Poland

Abstract

The aim of the paper was to show that the problem of emotion recognition is complex for both artificial intelligence computer methods and people. Emotion recognition is still a difficult problem because of the complexity and overlap between emotional expressions. In response to the challenge, we investigate the challenges of classifying emotions with the Carnegie Mellon University (CMU) dataset. Using a basic level of the wheel of emotions, 220 students were asked to annotate images in a survey. Though they were simple choices, a large number of responders had trouble choosing the emotion that best fit and indicated this by answering unsure. Such uncertainty has a negative impact on the proper preparation of training data for the machine learning process. For the algorithm to work well, it is crucial to properly train such a model. By categorizing the classification errors with confusion matrices and analyzing recognition rates in detail, this paper demonstrates why emotion perception is such a hard problem facing fierce challenges for new machine learning algorithms development.

Keywords

emotion classification, emotion detection, emotion recognition problems

1.Introduction

Emotion recognition is a vital area of study that bridges the disciplines of psychology and artificial intelligence. It entails identifying and classifying emotions expressed by individuals, commonly through facial expressions, body language, or vocal tones. Accurately recognizing and understanding emotions is essential for various applications, ranging from enhancing human-computer interactions to advancing mental health diagnostics. However, this task poses significant challenges, especially when emotions are intricate or composite rather than basic and easily discernible.

The Carnegie Mellon University (CMU) dataset [3] is a widely utilized resource in emotion recognition research, offering a diverse array of images designed to elicit specific emotional responses. This dataset has been instrumental in furthering our understanding of how emotions can be recognized and categorized. Nevertheless, numerous images in this

ITTAP'2024: 4th International Workshop on Information Technologies: Theoretical and Applied Problems, November 20–22, 2024, Ternopil, Ukraine, Opole, Poland *

Corresponding author.
† These authors contributed equally.
z.bougriche@student.po.edu.pl (Z. Bougriche); r.gasz@po.edu.pl (R. Gasz); 0009-0002-8415-2441 (Z. Bougriche); 0000-0002-3629-0578 (R. Gasz);

dataset exhibit complex or composite emotions that do not align neatly with fundamental categories, such as joy, sadness, anger, and fear.

In our study, we aimed to delve into the intricacies associated with categorizing these complex emotions. We carried out a survey involving 220 students who were asked to annotate images from the CMU dataset. Participants were provided with a predetermined set of basic emotions based on the wheel of emotions, a model that classifies emotions into primary and secondary levels. Despite these simplified choices, we found that numerous participants struggled to accurately identify the appropriate emotions, often marking their responses as uncertain.

The central objective of this article is to draw attention to the challenges associated with categorizing emotions using existing models. This study aims to explore the difficulties faced by human annotators in accurately recognizing emotions and how these challenges impact the reliability of emotion recognition systems. Through an analysis of the patterns of misclassification in the survey responses, we hope to shed light on the intricacies and inherent problems of emotion categorization. Our findings highlight the limitations of current annotation methods and emphasize the need for improved datasets and techniques that can capture the nuances of human emotional expression, ultimately leading to more accurate and reliable emotion recognition systems in the future.

2.Background on Emotion Recognition

Emotion recognition is the process of identifying and categorizing emotions expressed by individuals, which is typically done through facial expressions, body language, or vocal tones. This area of study is essential for applications in psychology, human-computer interaction, and artificial intelligence. Basic emotions, such as joy, sadness, anger, and fear, are universally recognized and have been extensively studied since the pioneering work of Paul Ekman, who proposed that these emotions are biologically innate and universally expressed across cultures.

Despite the foundational understanding of basic emotions, real-world emotional expressions are more complex. Complex emotions, such as optimism, contempt, or awe, combine elements of basic emotions and are heavily influenced by context and personal experiences. For example, optimism might encompass aspects of joy and interest, making it challenging to categorize using simple labels. This complexity presents significant challenges for both human annotators and automated systems.

Researchers have recently developed various models and tools to improve emotion recognition. One such model is the wheel of emotions, proposed by Robert Plutchik, which categorizes emotions into primary, secondary, and tertiary levels, illustrating the complexity and interrelationships between different emotions. This model helps to visualize how basic emotions can blend to form complex emotions, providing a more nuanced understanding of emotional expressions.

Several studies have highlighted the difficulties associated with emotion recognition. For instance, Barrett et al. [1] emphasized that emotions are not universally expressed in the same way across different contexts and cultures, which adds another layer of complexity to emotion recognition, requiring models that can adapt to diverse expressions and interpretations. Additionally, advancements in technology have led to the development of automated emotion recognition systems, which use machine learning algorithms to analyze facial expressions, vocal tones, and physiological signals.

However, these systems often struggle with the same complexities that challenge human annotators. For example, a study by Calvo and D'Mello] found that automated systems are less accurate in recognizing complex emotions than basic emotions, highlighting the limitations of current technologies. In our study, we focus on the challenges that human annotators face in recognizing emotions from the CMU dataset, which contains a diverse array of images designed to evoke specific emotional responses. Despite being provided with a predefined set of basic emotions, many annotators struggled with images depicting complex or compound emotions. This struggle underscores the need for more sophisticated models and tools to capture the nuances of human emotional expression.

Additionally, this dataset is unbalanced. As a result of the conducted surveys determining emotions, the basic ones prevailed - i.e. joy, sadness, and expectation. Having such a dataset can be problematic, but various resampling methods [10] can be used to improve it. In the analyzed case, some of the classes have a very low level of support, but as shown in [11], current machine learning algorithms obtain satisfactory results also in the case of this type of problem.

The images provided offer a glimpse into the intricate nature of emotion recognition. In Figure 1a, a facial expression is displayed that could be characterized as a blend of joy and anticipation, while Figure 1b portrays an expression that may convey anger with an underlying hint of disgust. These examples emphasize the difficulty of categorizing emotions that do not fit comfortably into basic categories.

a) b)

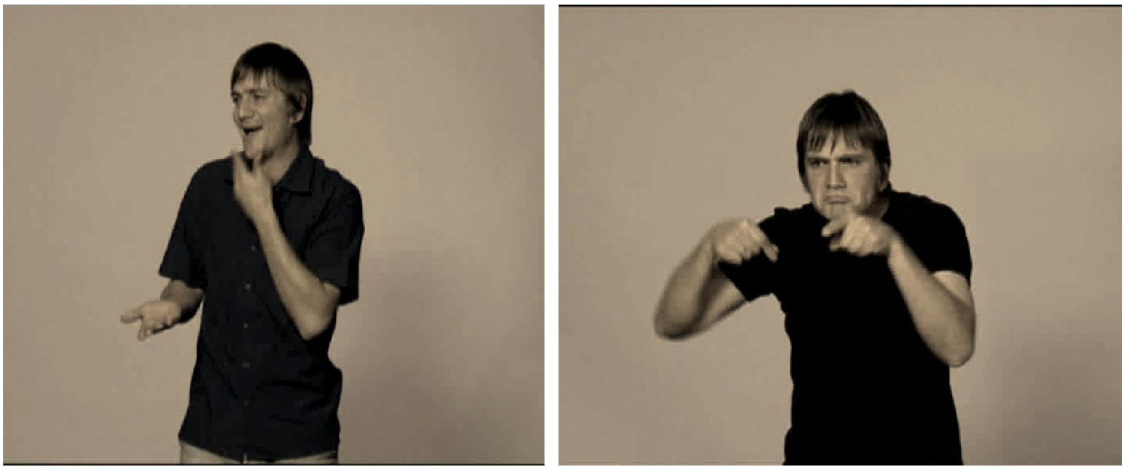

Figure 1: a) Joy and Anticipation, b) Anger and Disgust [9]

It is crucial to acknowledge the challenges and limitations of current emotion recognition methods to develop more precise and dependable systems that capture the full range of human emotions. This understanding not only propels research in psychology but also bolsters applications in artificial intelligence and human-computer interaction. In our study, we conducted a comprehensive survey involving 220 students, who annotated images from the CMU dataset to delve deeper into these challenges.

3.Existed Models for Emotion Categorization

Understanding the intricacies of human emotions has long been a topic of interest for academics and researchers. Over the centuries, various techniques and models have been developed to classify and comprehend the wide range of emotional experiences. Among these models are the Wheel of Emotions, and the Circumplex Model, each providing distinct insights into the structure and dynamics of human emotions.

The Wheel of Emotions published in [6] is one notable method for categorizing emotions. Created by psychologist Robert Plutchik, the model portrays emotions as interconnected entities that are organized into primary, secondary, and tertiary categories. Placed on a circular diagram, emotions are positioned relative to one another based on their similarity and intensity.

Figure 2: Robert Plutchik's Wheel of Emotions [6]

For instance, primary emotions like joy and trust are located opposite secondary emotions such as disgust and sadness. Secondary emotions emerge from combinations of primary ones, while tertiary emotions further refine these blends. The Wheel of Emotions offers researchers a comprehensive framework for examining the intricate web of human emotional experiences.

In contrast to discrete categorizations, the Circumplex Model [4] of Emotion, developed by James A. Russell, views emotions as continuous and dynamic processes within a circular space defined by two orthogonal axes: valence and arousal. Valence represents the positivity or negativity of an emotion, while arousal indicates its level of physiological activation. Emotions closer to the center of the circle are less intense and more neutral, whereas those toward the perimeter are more intense and distinctive. This model allows researchers to explore the subtle nuances and intricate interactions between affective states. It has applications in various domains, including emotion regulation, interpersonal communication, and clinical psychology, offering valuable insights into how individuals navigate their emotional landscapes.

By visualizing emotions along these dimensions, the Circumplex Model provides a detailed and nuanced perspective, aiding in emotion categorization, predicting behavioral responses, and designing effective computing systems. For example, emotions like excitement and happiness are high in both valence and arousal, while sadness is low in arousal but negative in valence. This framework helps to understand and predict how emotions influence behavior and interaction.

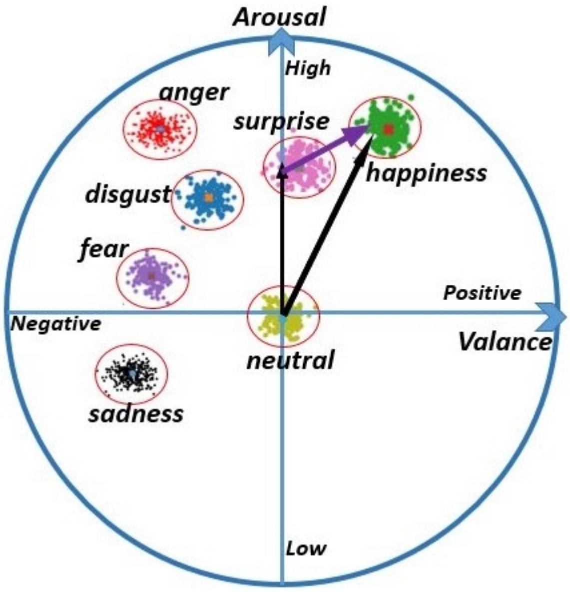

Figure 3: The Valence-Arousal Model [2]

Klaus Scherer developed a cognitive appraisal model called the Component Process Model [8]. The CPM is founded on evolutionary theory and thus views each appraisal as having evolutionary significance (e.g., preventing death, advancing reproductive goals). The CPM states that cognitive appraisal is a process in which we continuously appraise and reappraise our environment. In Fig. 4, the effect of the model - impact on effects, impact on effect classifications.

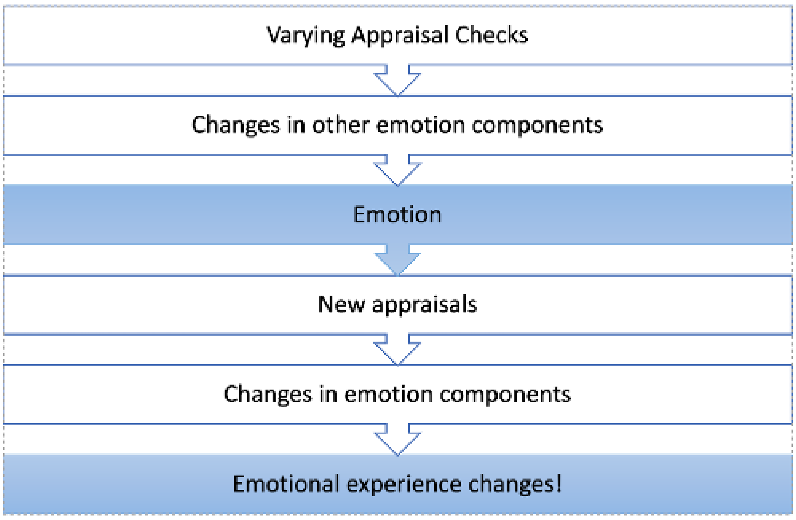

Figure 4: Influence of Appraisals on Emotion Experience [8]

The CPM divides appraisal into four different stages: 1) relevance check, 2) implications check 3) coping potential check, and 4) normative significance evaluation. Stage 1 occurs earliest in the emotional experience, whereas check 4 occurs last. At each step, several cognitive appraisal dimensions occur, with step 1 including the more primitive, universal appraisals and step 4 including the later, more cultural appraisals [7].

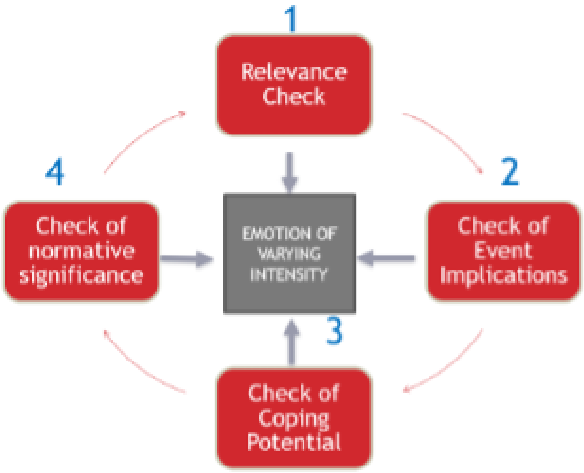

Figure 5: Component Process Model (CPM) Appraisal Checks [7]

Figure 5 shows the four appraisal checks. Each appraisal check contains specific cognitive appraisal dimensions, which Scherer wrote in [8].

4.Exploring Challenges in Emotion Categorization

The challenges faced in categorizing human emotions are numerous, primarily due to the intricate and multifaceted nature of emotional experiences. One of the main obstacles is the ambiguity and overlap of emotional states, as emotions often manifest as complex constructs with varying intensities and blending across different categories. This makes it difficult to accurately identify and categorize emotions, especially in situations where individuals exhibit mixed or conflicting emotional expressions.

Another significant challenge confronting efforts to categorize emotions is the quality of the data used. Noise from various sources, including environmental factors, individual variability in expression, and measurement inaccuracies, acts as a significant barrier. This interference obscures the underlying emotional signals, injecting uncertainty and inaccuracies into the categorization process, which compromises the reliability and validity of emotion recognition systems and hinders their ability to accurately interpret and classify emotional states.

The granularity of emotion categories represents another critical issue confronting researchers and practitioners in the field. Categories that are either too broad or too narrow in scope can impede the efficacy of classification algorithms and diminish the utility of emotion recognition applications. While overly broad categories may fail to capture the subtleties and nuances of specific emotional states, excessively narrow categories risk oversimplifying the complexity of human emotions, limiting their discriminative power and practical applicability.

The process of categorizing emotions is further complicated by cultural variability, as emotions are not only shaped by individual differences but also by sociocultural contexts. Cultural norms, values, and socialization practices influence the expression, interpretation, and evaluation of emotions, giving rise to cultural-specific patterns and nuances in emotional experiences. As a result, emotion recognition systems must contend with the challenge of accounting for cultural diversity and adapting to cross-cultural differences in emotional expression and perception.

To address these challenges, researchers and practitioners in the field of emotion categorization must develop robust methodologies, leverage advanced computational techniques, and integrate interdisciplinary insights from psychology, neuroscience, anthropology, and computer science. By doing so, advancements in emotion categorization have the potential to enhance our understanding of human emotions, enrich the capabilities of emotion recognition systems, and foster more nuanced and culturally sensitive approaches to studying and interpreting emotional experiences.

5.Methodology

The methodology section describes the steps taken to gather and analyze the data.

5.1. Dataset Description

The CMU Panoptic Dataset [3], developed by Carnegie Mellon University, is an extensive resource for research in computer vision, human-computer interaction, and robotics. Captured in a specialized studio equipped with over 500 cameras arranged in a geodesic dome and a Vicon system with 60+ infrared cameras, this dataset provides high-resolution video and precise 3D skeletal data of human movements. It encompasses a wide array of activities, including individual actions, social interactions, and object interactions, featuring diverse participants to ensure demographic diversity. With detailed annotations for tasks such as 2D and 3D pose estimation and action recognition, the dataset's time-synchronized data streams allow for multimodal analysis.

This dataset was chosen due to its richness and variety, containing 593 images that capture a broad spectrum of human expressions and interactions. The CMU Panoptic Dataset supports applications in human pose estimation, action recognition, social interaction analysis, behavioral studies, and robotics, making it an invaluable tool for advancing research in human behavior analysis. Access to the dataset is generally granted for academic and research purposes without requiring agreement to terms of use.

5.2. Survey Design and Implementation

A systematic approach was followed to design and conduct the survey, ensuring the quality and reliability of the data collected. The CMU Panoptic Dataset was cleaned by removing duplicated images to reduce the dataset to 440 unique images. This step was crucial to ensure the integrity and accuracy of the survey results. Subsequently, a suitable environment was prepared for conducting the survey, ensuring that participants could easily interact with the survey platform. This included setting up user-friendly interfaces and providing clear instructions.

To efficiently manage the survey, the 440 images were divided into ten separate surveys, making the process more manageable for participants and reducing the likelihood of fatigue, thereby ensuring more accurate responses. Following Robert Plutchik's Wheel of Emotions, a set of basic emotions was selected for participants to choose from. Although many images depicted compound emotions, the aim was to determine if participants could recognize and categorize the basic emotions present in the images. For each image, participants were also asked if they were confident in their responses, providing additional data to assess their confidence in emotion recognition.

220 participants were recruited from our university, and the purpose and importance of their participation were explained to ensure a diverse and representative sample. During the survey, assistance and guidance were provided, instructing participants on how to use the platform and encouraging confident responses. This support was crucial to ensure that participants felt comfortable and understood the survey process.

After the survey was completed, the data was cleaned by removing incomplete responses and retaining only the complete ones for analysis. This ensured the dataset was comprehensive and reliable. By following these steps, a thorough and systematic approach was ensured in conducting the survey, enabling the gathering of high-quality data for research on emotion recognition and categorization.

5.3. Analysis Techniques

In the research analysis, a confusion matrix was employed to explore participants' perception and categorization of emotions from images. The confusion matrix served as the primary tool, illustrating how well participants identified emotions compared to the actual emotions depicted in the images. This matrix effectively highlighted patterns of misclassification between similar or related emotions, such as joy and anticipation, or disgust and anger.

By focusing on the confusion matrix a comprehensive examination of participants' abilities to recognize and categorize emotions was enabled, revealing valuable insights into the nuances and challenges of emotional perception from visual stimuli.

6.Results and discussion

The creation of a confusion matrix served to visually illustrate the challenges in emotion recognition. This matrix displays the frequency of misclassifications among various emotional categories, highlighting the overlap and ambiguity between them. The confusion matrix (Figure 6) specifically emphasizes the common misclassifications of Joy and Anticipation, Disgust and Anger, Disgust and Sadness, and Sadness and Surprise. To comprehend these misclassifications, the framework of Robert Plutchik's Wheel of Emotions (Figure 2) was consulted.

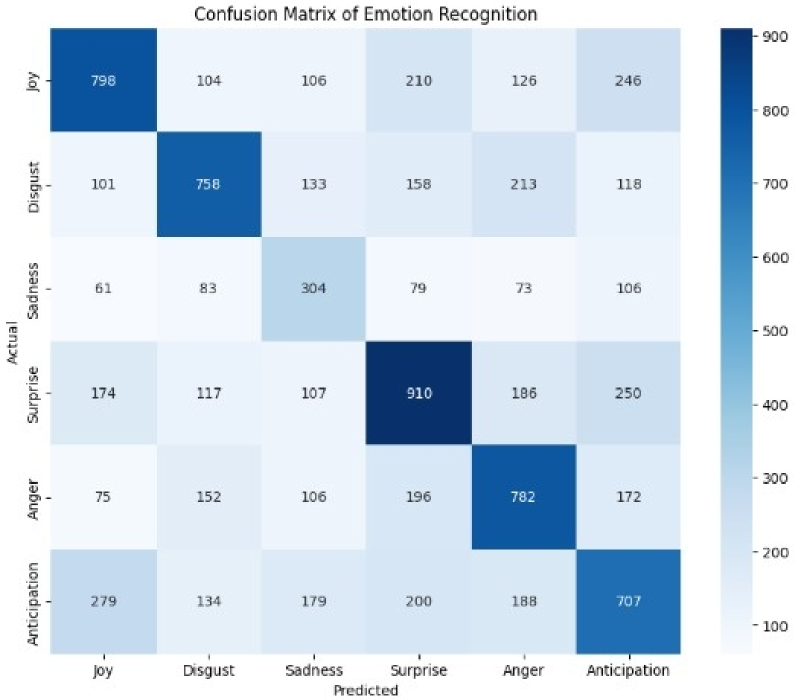

Figure 6: Confusion Matrix Figure of Emotion Misclassifications

The confusion matrix uncovers several significant patterns in the realm of emotion recognition:

1. Joy and Anticipation
   - Matrix Insights: The matrix demonstrates that Joy is frequently misclassified as Anticipation, and vice versa, with notable counts in the off-diagonal positions.
   - Emotional Relationships: According to Robert Plutchik's Wheel of Emotions, Joy and Anticipation are closely connected and frequently combine to form the compound emotion of Optimism. The close relationship between these emotions implies that individuals might exhibit facial expressions or vocal tones that blend these emotions, making it difficult for emotion recognition models to differentiate between them.
   - New Insights: This suggests that emotion recognition systems may need to consider Joy and Anticipation as part of a broader category or utilize additional context to disambiguate these emotions in practical applications.
2. Disgust and Anger

- Matrix Insights: According to the matrix, Disgust and Anger have a high misclassification rate. This may be due to the psychological and expressive similarities between these emotions.
- Emotional Relationships: The emotion wheel shows that Disgust and Anger are adjacent and can combine to form Contempt. This proximity in the emotional spectrum suggests that the physical and vocal cues for these emotions are often similar, leading to higher misclassification rates.
- New Insights: Improving training data and feature selection for emotion recognition models could lead to better accuracy in distinguishing between Disgust and Anger. Emphasizing the distinguishing features between these emotions in training could be beneficial.

3. Disgust and Sadness
    - Matrix Insights: The confusion matrix also reveals a frequent misclassification between Disgust and Sadness. Although these emotions are distinct, they share some commonalities in expression.
    - Emotional Relationships: On the emotion wheel, Disgust and Sadness can combine to form emotions such as Remorse. This relationship indicates that the boundary between these emotions can be fluid, leading to confusion.
    - New Insights: Refining emotion recognition algorithms by incorporating more nuanced features that differentiate between sadness and disgust, possibly considering context or secondary emotional cues, could lead to better accuracy.
4. Sadness and Surprise
    - Matrix Insights: Significant misclassifications are occurring between Sadness and Surprise, suggesting shared features that may confound recognition systems.
    - Emotional Relationships: The emotion wheel illustrates that these emotions can lead to Disappointment, which might add to the confusion.
    - New Insights: For emotion recognition systems, it could be advantageous to incorporate situational context or temporal patterns to better differentiate between emotions that are often confused, such as Sadness and Surprise.

The confusion matrix effectively showcases the difficulties and subtleties in recognizing and categorizing emotions from images. Frequent misclassifications between closely related emotions, such as Joy and Anticipation or Disgust and Anger, emphasize the need for more advanced emotion recognition models. Integrating contextual information and refining training data to accentuate distinctive traits between similar emotions can enhance model precision. Moreover, comprehending the connections between emotions as depicted in the emotion wheel can guide the development of more efficient recognition systems.

7.Conclusion

In this research, the intricacies of human annotators' emotion classification were thoroughly investigated, ultimately revealing considerable obstacles. The confusion matrix analysis underscored frequent errors in distinguishing between closely related emotions, thereby emphasizing the inherent complexities involved in accurate classification.

These complexities stem from factors such as cultural discrepancies and individual experiences. Nonetheless, additional research in this domain is essential. Future studies will concentrate on the development of real-time emotion classifiers using video sequences, the evaluation of various classification techniques, and their application to diverse datasets.

In summary, while noteworthy advancements have been achieved, the recognition of emotions from images remains a challenging yet critical area of research that demands ongoing attention and refinement.

References

[1] Barrett, L. Feldman, et al: The experience of emotion, Annu. Rev. Psychol. 58.1, 2007, pp. 373-403.

[2] A. Bhattacharjee et al: On the performance analysis of APIs recognizing emotions from video images of facial expressions, 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), IEEE, 2018.

[3] R. Calvo, S. D'Mello: Affect detection: An interdisciplinary review of models, methods, and their applications, IEEE Transactions on affective computing 1.1, 2010, pp. 18-37.

[4] C. Coulacoglou, D. H. Saklofske: The assessment of family, parenting, and child outcomes, Psychometrics and psychological assessment, 2017, pp.187-222.

[5] H. Karimova: The Emotion Wheel: What It Is and How to Use It, Emotional Intelligence, https://positivepsychology.com/emotion-wheel/ 2017,

[6] R. Plutchik: A general psychoevolutionary theory of emotion, Theories of emotion, Academic press, 1980, pp. 3-33.

[7] K.R. Scherer: Appraisal processes in emotion: Theory, methods, research, In K.R. Scherer, A. Schorr, & T. Johnstone (Eds.), Oxford University Press, 2001, p. 94.

[8] K. R. Scherer et al: Emotion inferences from vocal expression correlate across languages and cultures, Journal of Cross-Cultural Psychology, 32(1), 2001, pp. 76–92.

[9] T. Simon et al: Hand keypoint detection in single images using multiview bootstrapping, Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, 2017.

[10] M. Tomaszewski, J. Osuchowski: Effectiveness of Data Resampling in Mitigating Class Imbalance for Object Detection, CEUR Workshop Proceedings, Volume 3628, pp. 79 - 92, 2023 3rd International Workshop on Information Technologies: Theoretical and Applied Problems, ITTAP 2023, https://www.scopus.com/inward/record.uri?eid=2s2.0-85184360977&partnerID=40&md5=b18119edd8a969177a80c11496844e22

[11] M, Tomaszewski, P. Michalski, J. Osuchowski: Evaluation of Power Insulator Detection Efficiency with the Use of Limited Training Dataset. Appl. Sci. 2020, 10, 2104. https://doi.org/10.3390/app10062104