# CNN-Based Emotion Classification in Visual Art for Therapeutic and Creative Applications

Zofia Ruta
*Faculty of Mathematics and Computer Science*
*Adam Mickiewicz University*
Poznam, Poland
email: zofrut1@st.amu.edu.pl

Valentina Franzoni*
*Department of Mathematics and Computer Science*
*University of Perugia*
Perugia, Italy
email: valentina.franzoni@unipg.it

***Abstract*—Emotion recognition from artworks has the potential to enhance the experience of art exhibitions, where emotions conveyed by artworks can enhance the viewer's experience with synchronised lighting, music, and multimedia elements. Integrating emotion detection technology and applications to the art experience enlarges the way of perceiving and embracing art, leading to personalized therapy applications (e.g., art therapy). We used Convolutional Neural Networks and Transfer Learning to detect emotions in paintings, comparing three state-of-the-art models with different characteristics. A prototype application has been developed to show the classification capability of the best-performing model. The results highlight the effectiveness of our approach, particularly for binary classification, in real-world applications, such as adaptive art exhibitions and real-time art therapy tools. Challenges, such as dataset limitations and the subjective nature of emotions in art, were addressed through careful dataset integration and preprocessing, as well as the use of transfer learning to optimize performance. This work introduces applications of CNN in art therapy, immersive art experiences, and beyond, by demonstrating the potential of combining datasets and applying advanced deep learning techniques to emotion recognition in art, from enhancing art experiences to supporting emotional analysis in other creative industries.**

***Keywords-Emotion detection; CNN; Transfer learning; Art emotion recognition; Multimodal art augmentation; WikiArt; ArtEmis; art emotion dataset; supervised classification; cognitive behavioral analysis.***

## I. INTRODUCTION

Integrating emotion detection technology into the art experience offers an innovative way to transform how viewers engage with art. By detecting the emotions evoked by paintings, art exhibitions can be enhanced with synchronized multimedia elements, such as lighting, music, and digital media [1][2], creating immersive and dynamic environments that respond to the emotional content of the artwork. This approach goes beyond traditional static displays to offer viewers an emotionally tailored experience that has the potential to redefine the relationship between art and technology.

Emotion recognition in visual art could also offer significant value in art therapy [3]. In therapeutic settings, art is often used as a medium for individuals to express emotions that may be difficult to verbalize. Detecting and analyzing emotions in artwork can provide therapists with deeper insights into their clients' emotional states, allowing for more personalized therapeutic interventions. On the other hand, artworks conveying specific emotions can be used by the therapist to elicit an emotion in the client, as a starting point for narrative medicine, which, for the time needed to identify proper art pieces or produce them, usually can be applied only after a session, or using expensive tools, such in the case of Virtual Reality [4]. The integration of automated emotion recognition with art therapy could enhance the therapeutic process, supporting both therapists and clients in exploring complex emotional landscapes through visual art, with real-time support, offering valuable tools providing objective insights to complement subjective interpretations [5].

Emotion detection technology applied to artworks has broad societal impacts, such as helping stakeholders (e.g., in marketing and politics) to understand how visual stimuli elicit emotional responses from viewers. As cognitive-behavioral theory suggests, emotions and cognitive processes are closely linked [6][7], and analyzing how visual art impacts emotions can provide important insights into human behavior and decision-making. To avoid misuse that could lead to manipulation or bias [8][9], this application should be strictly linked with an ethical evaluation.

Despite such promising applications, the field of emotion detection in art is underdeveloped. Most previous works on image emotion analysis mainly used landmark-based element recognition; however, these features are vulnerable and not invariant to the different arrangements of elements [10]. This issue can be solved using techniques based on principle-of-art features including balance, emphasis, harmony, variety, gradation, and movement, which experiments are based on peer-rated abstract paintings. While Convolutional Neural Networks (CNNs) and transfer learning have been successfully applied to emotion classification in other domains, their application to artistic works remained underexplored, and challenged by the lack of sufficiently large labeled datasets [11]. In this work, by merging the WikiArt Emotion [12] and ArtEmis [13] datasets, we have addressed this limitation and created a more comprehensive and balanced dataset, improving the data quality, and allowing for more effective fine-tuning of deep learning models. Recent research is exploring the use of Large Language Models (LLMs) and Generative Neural Networks to enhance this process, requiring large computational capabilities or costly schedules for training a new neural model [14].

Most of the papers in the state of the art, which are based on artistic datasets, present social and artistic photographs, instead of paintings. There are some papers where WikiArt Emotion dataset is used, but only to evaluate models trained on datasets that present realistic images of faces (e.g., the FER-2013 dataset) [15] or realistic, non-artistic images [16]. In such works, researchers often rely on two prominent emotion models, the Ekman model [17] and the Mikels model [18] or a simplified binary classification into positive and negative emotions (i.e., sentiment analysis) [19].

Among previous works, we can highlight some technical reports applying CNNs to artworks for sentiment analysis. In [20], where researchers used also only artworks the best accuracy they achieved with CNNs pre-trained on ImageNet was 56%, but they didn't publish the dataset they used (thus, avoiding applicability and direct comparison), and considered only three sentiment classes (i.e., positive, negative, neutral). The main limitations highlighted by the authors relate to labels' noise which highly depends on the labeler, and interpretation difficulties by humans since people can respond differently to stroke edges, color tones, and objects of paintings. In [21], the best-achieved accuracy was 73%, and emotions have been again limited to a binary process of positive-negative sentiment analysis. Researchers used there the QuickShift algorithm in data preparation to simplify the image dataset, improving accuracy only for some art styles. When handling highly schematic work, such as minimalist paintings, the proposed methodology was highly unsatisfactory given the over-simplification of the images after processing with the QuickShift algorithm. As emphasized by the authors, another limitation is the number of existing datasets that contain a significant number of images for automation processes with emotions associated with humans.

In addition, there is an ongoing debate about the ability of machine learning models to accurately classify emotions in art, given its subjective nature [22][23]. Some researchers argue that deep learning models, which are often trained on structured data, such as photographs, may struggle to interpret the abstract and interpretive qualities of art [24][25]. Others, however, suggest that with the right data and methods, including Convolutional Neural Networks (CNNs), emotion detection in the art scenario can be meaningful and effective [26][27]. This study contributes to this debate by exploring both binary (pleasant/unpleasant) and multi-class classification and assessing their feasibility and limitations in the context of visual art. Our results demonstrate the effectiveness of binary classification in detecting emotional content, with multiclass classification offering additional insights despite being more challenging.

The limitations of our current approach mainly involve the quality of available datasets.

The rest of the paper is structured as follows. In section II, materials and methods are detailed, in particular the dataset collection and preprocessing, the architecture of the model, the operations implemented for training and optimization, and the metrics used for evaluation. In section III, results are shown and discussed, comparing the performance of the neural networks in the study. Finally, in section IV, conclusions are drawn, and future work is proposed to enhance the application and overcome current limitations.

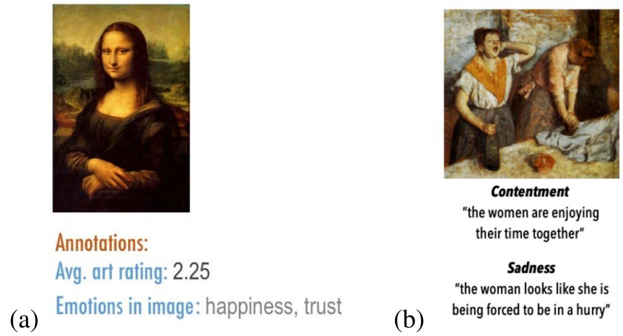


FIGURE 1. EXAMPLE OF EMOTIONAL ANNOTATIONS OF PAINTINGS IN THE ORIGINAL DATASETS: (A) WIKIART EMOTION; (B) ARTEMIS.

## II. MATERIALS AND METHODS

All experiments were performed on a workstation equipped with an NVIDIA Tesla V100 GPU with 32 GB of VRAM. The models were implemented in PyTorch, and additional libraries, such as Scikit-learn, were used for performance evaluation. Code and scripts to replicate the experiments will be made available upon reasonable request: all our scripts are fully documented to facilitate replication of the experiments. In the following paragraphs, we are going to detail the dataset merging and preprocessing, the architecture of the model, the training and optimization phases, and the evaluation metrics used for the two aims of binary and multiclass classification.

### A. Dataset Collection and Preprocessing

For this study, two publicly available datasets were used: the WikiArt Emotion Dataset [12] and the ArtEmis Dataset [13]. The WikiArt Emotion Dataset incudes 2,129 annotated paintings selected from the WikiArt collection, with emotions labeled using Paul Ekman's six basic emotions: anger, disgust, fear, happiness, sadness, and surprise. The ArtEmis dataset was introduced as a large-scale dataset of emotional reactions to images along with language explanations of these chosen emotions. It contains emotional annotations of 80,000 artworks from the WikiArt platform, automatically categorized by Ekman's six basic emotions, together with an explanatory phrase. Figure 1 shows an example of emotional annotation for each original dataset.

The datasets were merged to create a more comprehensive and balanced set of images, normalizing labels to the six emotional states from the Ekman model. Our merged labeled dataset includes 4,120 images for emotion classification. The final distribution of the dataset across the six emotion classes is shown in Table I. The merging of these datasets resulted in an improved balance across all six classes, with no significant overrepresentation of any single emotion. This balanced distribution ensures that the model receives sufficient training data for each emotion, improving the model's ability to classify emotions more accurately.

All images were preprocessed by resizing them to a uniform size of 224x224 pixels to meet the input requirements for the CNN models. Additionally, standard normalization techniques were applied to ensure that the pixel value distributions were consistent with the expectations of deep learning models.

TABLE I. DISTRIBUTION OF EMOTION CLASSES AFTER MERGING WIKIART EMOTION AND ARTEMIS DATASETS

| Emotion | Number of Samples | Percentage (%) |
|---|---|---|
| Anger | 438 | 10.63% |
| Disgust | 700 | 16.99% |
| Fear | 567 | 13.76% |
| Happiness | 1044 | 25.34% |
| Sadness | 637 | 15.46% |
| Surprise | 734 | 17.82% |
| **Total** | **4120** | **100%** |

### B. Model Architecture

We applied convolutional neural networks to the task of emotion recognition in visual art. Three pre-trained models were used: Visual Geometry Group (VGG16), MobileNet V2, and Inception V3. These models were fine-tuned using transfer learning, where the final fully connected layers were retrained on the merged dataset to classify images into pleasant/unpleasant emotions (binary classification) and six basic emotions (multi-class classification). The choice of models is based on their proven effectiveness in image classification tasks, especially in domains with limited data [26], [28].

### C. Training and Optimization

The learning rate and the optimizer play critical roles in the training and convergence of deep learning models. For this study, the training was performed using the Adam optimizer [21] with a learning rate of 0.0001 and a batch size of 32.

The Adam optimizer was selected as the primary optimization algorithm due to its proven effectiveness in handling sparse gradients and dynamically adapting learning rates during training. This adaptability is particularly useful for complex tasks, such as emotion recognition in visual art, where the gradient landscape can be highly non-linear and difficult to navigate. Adam was complemented by Stocastic Gradient Descent (SGD) – particularly effective in cases where the model is simple and the dataset is large – a robust choice for problems where generalization is important, and by RMSprop to address the issue of SGD's sensitivity to the choice of learning rate by introducing a moving average of the squared gradients, which allows the learning rate to remain effective throughout training.

The learning rate was set at 0.0001 for most experiments, based on empirical testing and its suitability for fine-tuning pre-trained CNN models. A smaller learning rate ensures that the fine-tuning process does not disrupt the pre-trained weights excessively while allowing gradual adjustment to the new dataset. This choice is critical for transfer learning tasks where the models are already trained on large-scale datasets and only require refinement for domain-specific tasks.

The choice of a lower learning rate combined with the Adam optimizer thus reflects careful experimental design, balancing the need for precise model adjustments with the computational efficiency required for training deep networks on moderately sized datasets.

The models were trained for 50 epochs, and early stopping was implemented to avoid overfitting. Cross-entropy loss was used as the loss function for both binary and multiclass classifications. An 80/20 train-test split was applied to the dataset. Function parameters in Python has been adapted to classify on unbalanced classes.

Performance metrics, such as accuracy, precision, recall, and F1 score were tracked during training. Such settings have been tested and chosen experimentally.

### D. Evaluation Metrics

For the binary classification task (pleasant/unpleasant emotion), Accuracy, Precision, Recall, and F1-score are used to evaluate model performance. For multiclass classification, Accuracy is used as a performance metric, and a confusion matrix is generated to analyze the model's ability to discriminate between Ekman's six basic emotions (i.e., anger, disgust, fear, happiness, sadness, surprise).

## III. RESULTS AND DISCUSSION

The binary classification task focused on predicting whether a painting evokes a pleasant (i.e., happiness, surprise) or unpleasant (i.e., anger, disgust, fear, sadness) emotion. The model, fine-tuned on the merged WikiArt Emotion and ArtEmis datasets, showed promising results, especially with the InceptionV3 model, which outperformed the other classifiers.

### A. Comparison of Neural Network Performance

To evaluate the performance of the three deep learning models (VGG16, MobileNetV2, and InceptionV3) on the task of emotion detection in paintings, we evaluated their accuracy, precision, recall, and F1-score using different optimizers and learning rates, as visible in Table II, where results show InceptionV3 achieving the highest accuracy (in bold, the best result for each Classifier).

TABLE II. COMPARISON OF DEEP LEARNING MODEL PERFORMANCE (VGG16, MOBILENETV2, AND INCEPTIONV3) ON EMOTION CLASSIFICATION TASKS USING TRANSFER LEARNING.

| Classifier | Optimizer | 0.001 | 0.01 |
|---|---|---|---|
| 3*InceptionV3 | adam | 41.26% | 41.38% |
| | rmsprop | 39.56% | 36.29% |
| | sgd | 40.05% | **44.54%** |
| 3*MobileNetV2 | adam | 21.60% | 16.75% |
| | rmsprop | 28.76% | 13.96% |
| | sgd | 37.99% | **40.78%** |
| 3*VGG16 | adam | 41.88% | 41.38% |
| | rmsprop | **42.11%** | 40.53% |
| | sgd | 32.77% | 33.86% |

- **VGG16**: The best accuracy achieved by VGG16 was **42.11%** when trained with the **RMSprop** optimizer at a learning rate of **0.001**. Although it performed well compared to MobileNetV2, its precision, recall, and F1 score were lower than those of InceptionV3, especially in distinguishing emotions, such as anger and sadness.

- **MobileNetV2**: MobileNetV2 showed considerable variability in performance. The highest accuracy recorded for MobileNetV2 was **43.2%** when using the **Adam** optimizer with a learning rate of **0.0001**. However, its precision and recall were not as consistent, and it generally underperformed compared to InceptionV3 in classifying emotions across the dataset, thus it has not been included in Table II.
- **InceptionV3**: Of the three models, **InceptionV3** showed superior performance, with the highest accuracy of $\sim 45\%$ achieved with the **SGD** optimizer and a learning rate of **0.01**. InceptionV3 also showed the best balance of precision, recall, and F1 score, especially for emotions, such as happiness, surprise, and fear. While it struggled slightly with anger and sadness, it still outperformed the other models in these categories.

InceptionV3 with Adam optimizer and learning rate $0.01$ achieved the best performance. The better performance of **InceptionV3** is evident not only in its overall accuracy but also in its ability to generalize better across different emotions, making it the most reliable model for emotion detection in paintings, in our context.

### *B. Results Discussion for InceptionV3*

In the following paragraphs, we will discuss the results for the InceptionV3 model, which performed best among the tested models (VGG16, MobileNetV2, and InceptionV3).

*1) Binary classification Results:* For the binary task of classifying emotions as pleasant (e.g., happiness, surprise) or unpleasant (e.g., anger, disgust, fear, sadness), the InceptionV3 model achieved an accuracy of 71%. Overall, the model correctly distinguished between pleasant (happiness, surprise) and unpleasant (anger, disgust, fear, sadness) emotions. Misclassifications primarily occurred in borderline cases where emotions, such as surprise and fear, shared overlapping visual cues. E.g., artworks depicting surprise often share intensity and ambiguity, which the model occasionally interprets as fear, which is acceptable, being surprise a critical emotion in its compatibility with both pleasant and unpleasant classes. Subtle emotional cues in serene or reflective artwork may have led the model to associate sadness with positive emotions, especially if the color palette or composition evoked calmness. Misclassifications visible in Table III suggest that the binary classification task, while relatively straightforward, can be influenced by subjective and ambiguous cues within the artwork.

The following points summarize the most relevant results based on each evaluation metric:

- **Accuracy:** The highest accuracy for binary classification was achieved using the **InceptionV3** model with an accuracy of 71%. This result was measured consistently across the test set, demonstrating reliable classification of positive and negative emotions.
- **Precision, Recall, F1-Score:** All three metrics (**Precision**, **Recall**, and **F1-score**) were recorded at 71%, indicating balanced performance across positive and negative classes.
- **Confusion Matrix:** The confusion matrix (see Table III) showed that most misclassifications occurred between emotions that were borderline or ambiguous.

TABLE III. CONFUSION MATRIX FOR BINARY CLASSIFICATION USING THE INCEPTIONV3 MODEL.

| | **Predicted Pleasant** | **Predicted Unpleasant** |
|---|---|---|
| **Actual Pleasant** | 78.98% | 21.02% |
| **Actual Unpleasant** | 27.11% | 72.89% |

The binary classification task highlights the feasibility of emotion detection in visual art when the emotional states are grouped into categories for pleasant and unpleasant emotions.

*2) Multiclass Classification Results:* The multiclass classification task was designed to predict one of Ekman's six basic emotions (anger, disgust, fear, happiness, sadness, surprise). Results for this task were more variable due to the increased complexity of the emotional categories. Table IV shows the model's accuracy varies across classes (correct classifications on the diagonal are highlighted in italics), with challenges noted for anger and sadness.

- **Accuracy:** The highest multiclass accuracy achieved was $\sim 45\%$, with the **InceptionV3** model outperforming both VGG16 and MobileNetV2. The relatively lower accuracy compared to the binary task reflects the challenge of emotion detection in visual art, where emotions are often subjective and nuanced.
- **Confusion Matrix:** The confusion matrix for the multiclass classification showed that the model was more accurate at recognizing some emotions, such as *surprise, happiness, and fear*, but struggled with others, such as *anger* and *sadness*. The overlap between these emotions suggests that they share similar visual cues, making them harder to distinguish. Although the dataset was more balanced after merging the WikiArt Emotion and ArtEmis datasets, there was still a slight skew, with emotions like *happiness* and *surprise* slightly more represented than others like *anger* and *fear* (see Table I). This distribution allowed for more consistent performance across emotion classes, but some of the variance in performance can be attributed to these minor imbalances. In particular, the confusion matrix from the results (see Table IV) shows that particular emotions, such as *anger* and *sadness*, were harder for the model to discriminate. While this could be partly due to similar visual cues, the lower representation of *anger* in the dataset may have contributed to this challenge. *Happiness* was often correctly classified due to its distinct bright and vivid visual cues, such as warm colors and joyful scenes. However, it was occasionally over-represented, potentially due to its relatively higher frequency in the dataset. *Surprise*, while distinguishable in some cases, was misclassified as fear or happiness depending on the accompanying visual elements. This



Copyright (c) IARIA, 2025. ISBN: 978-1-68558-239-5 

TABLE IV. Confusion matrix for multiclass classification of Ekman's six basic emotions.

| *Actual/ Predicted* | **Anger** | **Disgust** | **Fear** | **Happiness** | **Sadness** | **Surprise** |
|---|---|---|---|---|---|---|
| **Anger** | *50.23%* | 4.57% | 6.85% | 2.28% | 22.83% | 13.24% |
| **Disgust** | 2.57% | *81.43%* | 3.57% | 3.57% | 5.71% | 3.14% |
| **Fear** | 10.58% | 7.94% | *67.02%* | 3.00% | 7.05% | 4.41% |
| **Happiness** | 1.44% | 0.96% | 1.92% | *84.29%* | 5.75% | 5.65% |
| **Sadness** | 7.06% | 5.49% | 3.92% | 9.11% | *64.36%* | 10.05% |
| **Surprise** | 3.41% | 2.04% | 2.04% | 5.72% | 3.81% | *82.97%* |

reflects the inherent ambiguity of surprise as an emotion, which can lean toward positive or negative interpretations.

## IV. Conclusions

This study demonstrates the effectiveness of Convolutional Neural Networks (CNNs) for emotion recognition in visual art, specifically applying VGG16, MobileNetV2, and InceptionV3 models fine-tuned using a combination of the WikiArt Emotion and ArtEmis datasets. Among the models tested, InceptionV3 proved to be the most reliable, particularly for binary classification (pleasant/unpleasant), with an accuracy of 71% and balanced performance across metrics. Although multiclass classification yielded lower accuracy due to the nuanced and subjective nature of emotions in art, InceptionV3 still performed reasonably well, especially in recognizing happiness, surprise, and fear.

The approach presented here highlights the potential of using deep learning models for applications in art therapy and immersive art experiences. By integrating these models with transfer learning, we addressed the challenge of limited labeled data and improved the system's ability to effectively classify emotions. Our results highlight the benefits of combining multiple datasets to improve emotion detection in art and promote a more interactive and emotionally engaging experience in artistic environments.

Future work could explore the inclusion of larger, more diverse datasets and further refine the classification capabilities, especially for complex emotions. For example, targeted data augmentation strategies (e.g., brightness adjustments, hue shifts) could help simulate the variability in emotion expression and improve model generalization. Also incorporating additional datasets or generating synthetic data [23] using generative models could help to balance classes, enhancing the representation of underrepresented emotions like anger and fear. Regarding the classification model, combining visual features with textual descriptions (e.g., artist statements or viewer annotations) could provide complementary information to improve emotion classification.

## References


[1] K. Mondloch, "The influencers: Van gogh immersive experiences and the attention-experience economy," *Arts*, vol. 11, no. 5, 2022, ISSN: 2076-0752. DOI: 10.3390/arts11050090.

[2] A. Sobitan and P. Vlachos, "Immersive event experience and attendee motivation: A quantitative analysis using sensory, localisation, and participatory factors," *Journal of Policy Research in Tourism, Leisure and Events*, vol. 12, no. 3, pp. 437–456, 2020. DOI: 10.1080/19407963.2020.1721638.

[3] J. Czamanski-Cohen and K. L. Weihs, "The role of emotion processing in art therapy (repat) intervention protocol," *Frontiers in Psychology*, vol. 14, p. 1 208 901, 2023.

[4] I. Hacmun, D. Regev, and R. Salomon, "The principles of art therapy in virtual reality," *Frontiers in Psychology*, vol. 9, p. 2082, 2018.

[5] Z. Liu, C. Yao, Q. Wang, and F. Ying, "A study on the emotional responses to visual art," in *International Conference on Entertainment Computing*, Springer, 2023, pp. 450–460.

[6] R. Reisenzein, "Cognitive theory of emotion," in *Encyclopedia of Personality and Individual Differences*, V. Zeigler-Hill and T. K. Shackelford, Eds. Cham: Springer International Publishing, 2020, pp. 723–733.

[7] G. Hauke, C. Lohr-Berger, and T. Shafir, "Emotional activation in a cognitive behavioral setting: Extending the tradition with embodiment," *Frontiers in Psychology*, vol. 15, p. 1 409 373, 2024.

[8] G. Biondi *et al.*, "Ethical design of artificial intelligence-based systems for decision making," *Frontiers in Artificial Intelligence*, vol. 6, p. 1 250 209, 2023.

[9] V. Franzoni, J. Vallverdù, and A. Milani, "Errors, biases and overconfidence in artificial emotional modeling," in *IEEE/WIC/ACM International Conference on Web Intelligence-Companion Volume*, 2019, pp. 86–90.

[10] S. Zhao *et al.*, "Exploring principles-of-art features for image emotion recognition," in *Proceedings of the 22nd ACM International Conference on Multimedia*, ser. MM '14, Orlando, Florida, USA: Association for Computing Machinery, 2014, pp. 47–56.

[11] E. Cetinic, T. Lipic, and S. Grgic, "Learning the principles of art history with convolutional neural networks," *Pattern Recognition Letters*, vol. 129, pp. 56–62, 2020.

[12] S. Mohammad and S. Kiritchenko, "Wikiart emotions: An annotated dataset of emotions evoked by art," in *Proceedings of the eleventh international conference on language resources and evaluation (LREC 2018)*, 2018.

[13] P. Achlioptas, M. Ovsjanikov, K. Haydarov, M. Elhoseiny, and L. J. Guibas, "Artemis: Affective language for visual art," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 11 569–11 579.

[14] P. Kumar, "Large language models (llms): Survey, technical frameworks, and future challenges," *Artificial Intelligence Review*, vol. 57, no. 9, p. 260, 2024.

[15] 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)*, IEEE, 2020, pp. 1–9.

[16] C. Redies, M. Grebenkina, M. Mohseni, A. Kaduhm, and C. Dobel, "Global image properties predict ratings of affective pictures," *Frontiers in psychology*, vol. 11, p. 953, 2020.

[17] S. Ceccacci, A. Generosi, L. Giraldi, and M. Mengoni, "Emotional valence from facial expression as an experience audit tool: An empirical study in the context of opera performance," *Sensors*, vol. 23, no. 5, p. 2688, 2023.

[18] C. Montenegro *et al.*, “A dialogue-act taxonomy for a virtual coach designed to improve the life of elderly,” *Multimodal Technologies and Interaction*, vol. 3, no. 3, p. 52, 2019.

[19] M. Baldoni, C. Baroglio, V. Patti, and C. Schifanella, “Sentiment analysis in the planet art: A case study in the social semantic web,” *New Challenges in Distributed Information Filtering and Retrieval: DART 2011: Revised and Invited Papers*, pp. 131–149, 2013.

[20] W. Tan *et al.*, “Cnn models for classifying emotions evoked by paintings,” Technical Report, SVL Lab, Stanford University, USA Tech. rep., Tech. Rep., 2018.

[21] C. González-Martín, M. Carrasco, and T. G. Wachter Wielandt, “Detection of emotions in artworks using a convolutional neural network trained on non-artistic images: A methodology to reduce the cross-depiction problem,” *Empirical Studies of the Arts*, vol. 42, no. 1, pp. 38–64, 2024.

[22] A. Elgammal, B. Liu, D. Kim, M. Elhoseiny, and M. Mazzone, “The shape of art history in the eyes of the machine,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, 2018.

[23] S. Zhao *et al.*, “Affective image content analysis: Two decades review and new perspectives,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 10, pp. 6729–6751, 2021.

[24] P. Lelièvre and P. Neri, “A deep-learning framework for human perception of abstract art composition,” *Journal of vision*, vol. 21, no. 5, pp. 9–9, 2021.

[25] E. J. Crowley and A. Zisserman, “In search of art,” in *Computer Vision-ECCV 2014 Workshops: Zurich, Switzerland, September 6-7 and 12, 2014, Proceedings, Part I 13*, Springer, 2015, pp. 54–70.

[26] T. M. Tashu, S. Hajiyeva, and T. Horvath, “Multimodal emotion recognition from art using sequential co-attention,” *Journal of Imaging*, vol. 7, no. 8, 2021.

[27] R. Pereira *et al.*, “Systematic review of emotion detection with computer vision and deep learning,” *Sensors*, vol. 24, no. 11, 2024.

[28] T. Chutia and N. Baruah, “A review on emotion detection by using deep learning techniques,” *Artificial Intelligence Review*, vol. 57, no. 8, p. 203, 2024.


 