VISUAL SENTIMENT MINING IN CONTEMPORARY ART REVIEWS

R. Shobana [1] ✉ iD, Dr. Poonam Shripad Vatharkar [2] ✉ iD, Dr. Sonia Riyat [3] ✉ iD, Dukhbhanjan Singh [4] ✉ iD, Meeta Kharadi [5] ✉ iD, Ila Shridhar Savant [6] ✉

[1] Associate Professor, Department of Computer Science and Engineering, Aarupadai Veedu Institute of Technology, Vinayaka Mission's Research Foundation (DU), Tamil Nadu, India
[2] Assistant Professor, MES Institute of Management and Career Courses (IMCC), SPPU, Pune, Maharashtra, India
[3] Professor, Department of Management, Arka Jain University, Jamshedpur, Jharkhand, India
[4] Centre of Research Impact and Outcome, Chitkara University, Rajpura- 140417, Punjab, India
[5] Assistant Professor, Department of Fashion Design, Parul Institute of Design, Parul University, Vadodara, Gujarat, India
[6] Department of Artificial intelligence and Data Science, Vishwakarma Institute of Technology, Pune, Maharashtra, 411037, India
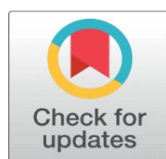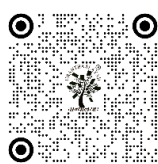
ABSTRACT

Visual sentiment mining is a recent field of interdisciplinary study with potential to use computational methods to analyze the emotional reactions to visual art. Written by critics, curators and viewers, contemporary art reviews are full of rich affective meanings, which are not only visual stimuli but also cultural context. But the current sentiment analysis methods are mainly text based or geared towards general images and thus cannot capture the subtle emotional connotations inherent in works of art. The paper presents a single visual sentiment mining model that is specially developed towards modern art reviews, through the concurrent analysis of the images of works of art and the textual content of the reviews. Simultaneously, language models based on transformers are used to derive fine-grained sentiment and emotion signals in art reviews and detect the tone of the subjectivity and metaphorical language and the intent of a critique. A professional dataset is multimodal and is made as a combination of high-resolution images of artwork and professionally written reviews and annotations of emotion. These are rigorous preprocessing, cross-modal alignment, and optimized hyperparameter supervised learning as part of the experimental approach. The analysis is performed based on conventional measurements (accuracy and F1-score) as well as a suggested index of emotion coherence to understand the correspondence between the visual and textual moods.

Keywords: Visual Sentiment Mining, Contemporary Art Reviews, Multimodal Analysis, Computer Vision, Natural Language Processing, Affective Computing

1. INTRODUCTION

Modern art exists within a complicated cultural area where meaning is not predetermined but is constantly negotiated by visual representation, references to context and critical discussion. Critic, curator, scholar and audience art reviews are essential in creating the perception, meaning and value of artworks. Such reviews are impressively subjective, which is a combination of aesthetic commentary, cultural commentary, and subjective emotional evaluation.

The interpretation of the feeling underlying such reviews is thus the key of examining reception, influence, and impact in the present-day art ecologies. Due to the booming development of digital platforms, online exhibitions, and art discourse mediated by social media, the amount of visual art objects and the subsequent textual commentary on them has grown exponentially, offering new possibilities and challenges to computational analysis. Visual sentiment mining is an attempt to computeually recognize and decipher visual emotion as expressed in visual representations and the textual accounts of these representations. Although sentiment analysis has been rather successful in such fields as product reviews, social media analytics, and news opinion mining, the direct research of art has been little examined. Artistic sentiment can be elusive, unclear, metaphorical, and culturally specific and hard to measure by traditional polarity-based or otherwise generic emotion classifiers Guo et al. (2020). In contrast to conventional photographs or even adverts, works of art purposefully use abstraction, symbolism, colour theories, and compositional tension to create multiple emotional reactions. Likewise, the language of art reviews often contains subtle language, irony, allusions, and emotional metaphors, which are not accessible to conventional methods of processing natural language. Current computer vision methods of sentiment analysis have been mostly concerned with the detection of fundamental emotions in facial expressions, scenes, color distributions, commonly trained on data that do not relate to artistic practice Hung et al. (2024).

Similarly, sentiment models typically work best with explicit opinionated language as opposed to interpretive criticism when the language is text-based. Therefore, such models find it difficult to reconcile visual affect and critical sentiment in terms of applying them to art reviews. The aspect of this gap that exists is that domain-specific visual sentiment mining frameworks are necessary, which combine art theory, affective computing, and multimodal machine learning. The appearance of the deep-learning paradigm, especially the convolutional neural networks (CNNs), vision transformers (ViTs), and the language models based on transformers, offers effective means of overcoming this problem Dhar and Bose (2022). These models can learn high level semantic, stylistic and contextual representations both of images and text. Used together in a multimodal environment, they allow the simultaneous examination of both visual components (e.g., harmony of colors and texture, space arrangement and composition, and recurring motifs), and textual ones (e.g., tone and emotional description and evaluative language). This integration is particularly applicable to the modern art, in which an emotional meaning can be formed often through the interactions of what is perceived and its description. There are practical and theoretical consequences of visual sentiment mining in the current art reviewing Purohit (2021). In the case of curators and galleries, sentiment-aware analytics may be used to assist in exhibition design, audience engagement analysis, and decision-making by the curator. To critics and researchers, computational tools may help in determining the prevailing emotional patterns, shifts in style, and patterns of reception over time, space or artistic movements.

2. BACKGROUND AND LITERATURE REVIEW
2.1. SENTIMENT ANALYSIS IN COMPUTER VISION: FOUNDATIONAL CONCEPTS

Computer vision Sentiment analysis is a branch of computer vision that is concerned with detecting and decoding affective states or emotional expressions directly based on visual information. The initial methods were based on low-level handcrafted features like color histograms, brightness, saturation, texture, and edge density and were motivated by psychological theories that warm colors are related to positive emotions and darker colors are related to negative affect. These techniques offered the foundation of affective image analysis and could not provide the ability to detect high-level semantics and contextual meaning Rodríguez-Ibánez et al. (2023). The introduction of deep learning brought about the development of the convolutional neural networks (CNNs) that highly developed the visual sentiment analysis using hierarchical representations that capture local patterns as well as global structures. CNN model showed better results in identifying emotion labeled images and scene level sentiment. Most recently, sentiment modeling has also been advanced using vision transformers (ViTs) that better represent the long-range dependencies and global image representations in complex visual scenes Chen et al. (2024). Besides the generic feature learning, there have been affective descriptors like color harmony, visual balance, contrast and compositional symmetry to more closely match the computational features to human emotional perception.

2.2. ART THEORY PERSPECTIVES ON EMOTIONAL INTERPRETATION

Emotion has been a key aspect of aesthetic experience in art theory, where the work of art is no longer seen as an object in visual terms but as a system of expression that can produce an emotional effect. The study of visual aspects of color, line, texture, rhythm, and composition that create emotional effect under the formalist approaches are independent of narrative content. Conversely, expressionist and phenomenological theories emphasize on the emotional intent of the artist and embodied response of the viewer and conclude that meaning is created by way of perceptual and emotional experience Park et al. (2024). The contribution of symbols and cultural references and historical background to the creation of emotional interpretation is further emphasized by semiotic and iconographic methods. In this perspective, emotions are not universal cues but constructed in culture that relies upon common visual languages and interpretive schemes Filieri et al. (2022). The modern theory of art builds on these concepts but recognizes ambiguity, many meanings, and involvement of the audience. The audience is urged to form emotional interpretations, grounded in subjective memory, social identity, and circumstantial context that can contribute to the occurrence of different affective interpretations of the same piece of art. This compatibility manifests in art criticism in the use of metaphorical language, subtle tone, and judgmental speech as opposed to the use of direct emotion terms. These theoretical implications are major problems to computational sentiment analysis which in most cases assumes discrete and universally known emotions Sykora et al. (2022).

2.3. EXISTING DATASETS, METHODS, AND LIMITATIONS IN ART-FOCUSED SENTIMENT MINING

The number of studies in art-oriented sentiment mining is relatively small because of the lack of domain-specific data and the handling of emotional annotation in art. The available datasets tend to upscale generic affective image datasets, like emotion-tagged images or social-networking images, which do not have the stylistic richness and conceptual richness of works of art. Other works have been creating small art collections through matching museum images to crowdsourced labels or reduced emotion representations, however such annotations are often subjective and not consistent and are often culturally biased Chatterjee et al. (2021). Art reviews and critiques are simply present on the textual level less frequently than standard sentiment corpora, not to mention that their language is more complicated than traditional NLP models can handle. Although these methods prove the sentiment prediction to be a viable task, they tend to use rough categories of emotions and fail to explicitly match visual and textual affect Jeon et al. (2023). Table 1 provides a review of the studies on multimodal sentiment analysis and visual sentiment analysis in art. Among the essential drawbacks, one can distinguish the lack of the dataset size, poor cross-modal grounding, low interpretability, and poor art-discussion-specific evaluation metrics.

Table 1

Table 1 Related Work on Visual and Multimodal Sentiment Analysis in Art and Visual Media				
Domain	**Dataset Type**	**Core Methodology**	**Sentiment**	**Limitations**
Art Images	Abstract paintings	Handcrafted visual features	Valence-based sentiment	Limited scalability, no text
Social Media Images Nguyen and Nguyen (2023)	Flickr images	CNN + sentiment tags	Polarity sentiment	Not art-specific
Affective Images	Emotion-labeled photos	CNN-based emotion classifier	Discrete emotions	Lacks artistic abstraction
Visual Style Xu et al. (2019)	Flickr style dataset	CNN style features	Implicit affect cues	No explicit sentiment labels
Artwork Analysis	WikiArt	CNN + aesthetic features	Emotion categories	Ignores textual reviews
Text Sentiment	Review corpora	Transformer-based NLP	Valence–arousal	No visual grounding
Multimodal Emotion Shaik Vadla et al. (2024)	Social media posts	Multimodal deep fusion	Multi-label emotions	Non-curated datasets
AI & Art	Generative art datasets	CNN + interpretability	Affective attributes	Not review-based

Image–Text Emotion	Image caption datasets	Attention-based fusion	Emotion consistency	Limited art semantics
Museum Analytics Wang et al. (2021)	Museum collections	Deep visual embeddings	Visitor emotion inference	No critic sentiment
Art Recommendation	Online galleries	Hybrid ML models	Mood-based recommendation	Shallow sentiment labels
Visual Affect	Affective image sets	ViT-based modeling	Valence–arousal	Generic imagery
Multimodal Reviews Nguyen et al. (2024)	Product reviews	Transformer fusion	Sentiment polarity	Not artistic discourse
Contemporary Art Reviews	Curated art & critic reviews	CNN + ViT + NLP Transformers	Multidimensional emotion + coherence	

3. PROPOSED FRAMEWORK FOR VISUAL SENTIMENT MINING
3.1. DATASET CONSTRUCTION FROM CURATED ART REVIEWS AND ARTWORK IMAGES

The suggested framework starts with a high-quality multimodal dataset that will be constructed, which will consist of images of contemporary artwork and a critical review. Artworks are obtained through an edited collection of digital archives, galleries, museum collections, and peer-reviewed exhibition catalogs to have stylistic variety in media, genre and artistic trend. The image of the artwork is taken in high resolution to capture the finer-grained information of the visual feature like texture, brushwork, materiality, and spatial arrangement. Visual sentiment mining is multimodal as illustrated in Fig. 1. Textual data is composed of written art reviews written by professionals and curators and chosen to draw out subtle emotional interpretation as opposed to casual commentary.

Figure 1
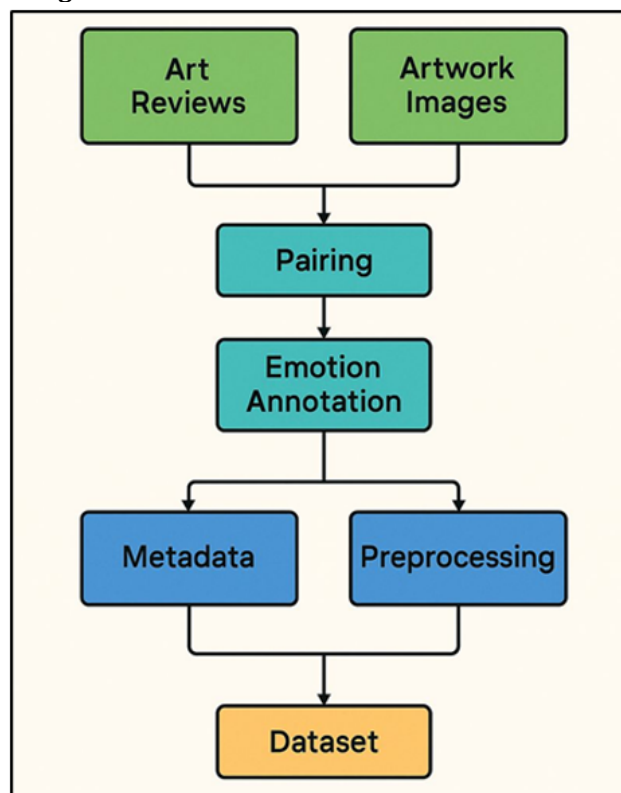
Figure 1 Multimodal Dataset Construction Pipeline for Visual Sentiment Mining in Contemporary Art Reviews

In a bid to guarantee semantic alignment, only reviews which contained personal mentions towards the chosen artworks are kept. Metadata such as background about artists, date of composition, medium, context of exhibition and cultural background are included to aid in contextual analysis. One of the data preprocessing options is the image normalization, resolution standardization, text cleaning and sentence level segmentation.

3.2. VISUAL FEATURE EXTRACTION USING CNNS, VITS, AND AFFECTIVE DESCRIPTORS

Visual feature extraction is developed in order to grasp perceptive and affective attributes of art pieces. Convolutional neural networks are used to unlearn global features and learn local features like edges, the texture, brushstroke, and material features that assist in emotional perception. Ready-made CNN structures are fine-tuned on the curated art metric to change representations in natural images to artistic ones. Simultaneously, vision transformers are applied so as to simulate the world relationships throughout the artwork so that the compositional balance, space hierarchy and the symbolic arrangement are captured that affect affective response. ViTs are especially successful at the self-attention mechanism when it comes to abstract and non-photorealistic art styles. Deep visual representations are augmented with hand-designed affective descriptors, which are based on art theory and psychology by computing color harmony indices, saturation and contrast, symmetry scores, texture complexity and visual entropy. These characterizations give direct correlations between computer characteristics and the emotional feelings of people. The fusion of features of CNN, ViT, and affective features into a unified image representation is known as the feature fusion strategies whereby attention-based weighting is adopted. This is a hybrid solution between the expressive power and interpretability, which makes the model obtain the low-level sensory cues as well as the high-level aesthetic structures applicable to sentiment analysis in modern art.

3.3. TEXTUAL SENTIMENT EXTRACTION FROM CRITIC REVIEWS USING NLP TRANSFORMERS

The sentiment extraction of text is concerned with modeling the interpretive and nuanced language of art criticism. NLP models that utilize transformers are adopted because they are able to get long-range dependencies, context meaning, and metaphorical meanings. The word processors are trained on a set of filtered art reviews to be customized to the specifics of critical writing knowledge. The sentence-level and paragraph-level embeddings are created to capture the evaluative statements made locally and the tone in general. The special emphasis is placed on affective phrases, descriptive adjectives, and comparative constructs which express emotional judgment in an indirect manner. Art-relevant emotion lexicons are added to domain-specific emotion lexicons to provide model attention to emotion salient tokens. Besides the normal sentiment polarity, the framework makes predictions of multi-label emotion distributions that are consistent with valence-arousal or art-adapted emotion categories. In order to achieve a better correspondence between visual sentiment and the visual feature, cross-attention mechanisms are subsequently introduced to pair the textual emotion-cue with that of a visual feature.

4. EXPERIMENTAL SETUP AND METHODOLOGY
4.1. DATASET PREPROCESSING, ANNOTATION PROTOCOLS, AND TRAINING PIPELINE

The experimental design starts with a methodical pre-treatment of both visual and text data to be able to be consistent and reliable. The images of the artwork are reduced to a standardized resolution, color-balanced, and enhanced with the use of specific transformations like rotation, scaling, and mild cropping to enhance the model generalization without the distortion of an artistic meaning. Textual reviews are subjected to cleaning procedures such as token normalization, elimination of non-informative symbols, segmentation of sentences and lemmatization without any deletion of stylistic marks that can be used to express sentiments. The annotation protocols will create the necessary degree of subjectivity and consistency through the incorporation of expert annotations with schematic rules. Scholars and art trained annotators categorize emotional dimensions according to a valence arousal system with added discrete emotion categories adapted to art situations. These models are then combined into a multimodal architecture having synchronous batching and intermodal matching.

4.2. MODEL ARCHITECTURES, HYPERPARAMETERS, AND OPTIMIZATION

The proposed framework uses a modular framework that integrates visual and sentiment models based on text, and a multimodal fusion layer. The visual branch combines a small-scale convolutional neural network with the localized feature extraction task and vision transformer with the global compositional understanding task. The textual branch employs a language model in the form of a transformer model changed to art-critical speech. Each of the two branches outputs the results that are then embedded in a common embedding space via fully connected layers. Figure 2 represents

optimized architecture pipeline which allows visual sentiment mining. The model uses cross-attention mechanisms to weight visual and textual cues dynamically depending on the emotional relevance of such cues.
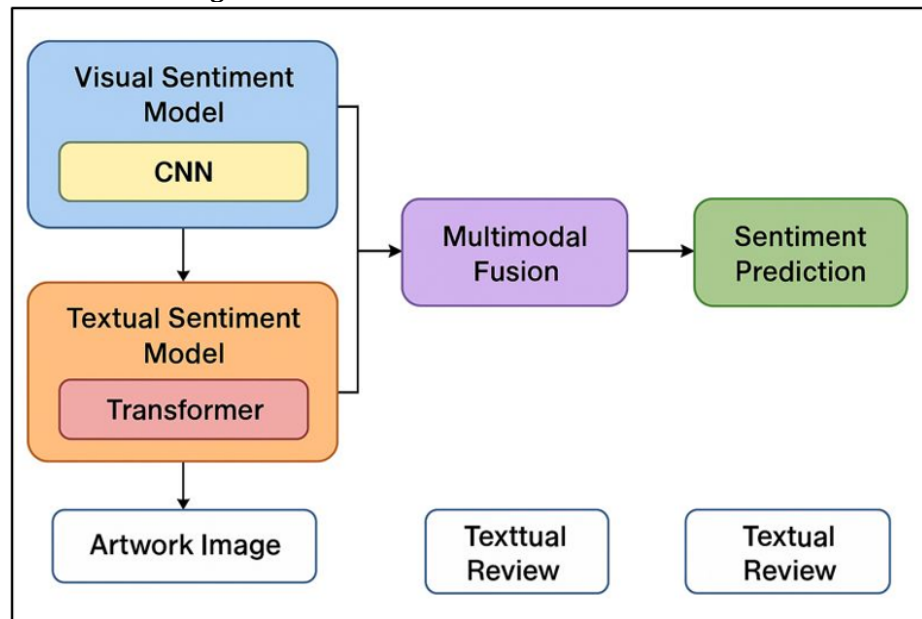
Figure 2

Figure 2 Architecture and Optimization Pipeline for Visual Sentiment Mining

The major hyperparameters are the learning rate, batch size, the dimensionality of the embedding, attention heads, and dropout rates that are optimized using both grid and Bayesian search approaches. Pretrained and newly-initialized layers are given separately imposed learning rates in order to stabilize the fine-tuning. Adaptive gradient-based methods are used to optimize it, and gradient clipping is applied to prevent instability of multimodal training. To enhance generalization, they use the regularization methods, dropout, weight decay, and data augmentation.

4.3. EVALUATION METRICS: ACCURACY, F1-SCORE, EMOTION COHERENCE INDEX

The measurements of visual sentiment mining models should be measured in both predictive and cross-modal emotional consistency. General measures of classification like accuracy are applied to estimate general correctness of sentiment or emotion category prediction. Since the information in class imbalance and subjectivity (in art sentiment data) are prevalent, the F1-score is chosen to bring out precision and recall in the emotion classes. To provide equal evaluation of less common categories of emotions, the macro-averaged F1-scores are provided. In addition to traditional measures, a sentiment representation index of emotion consistency between visual and textual representations is proposed as well. The index is used to quantify the difference between the expected distributions of emotion in the image and the corresponding review with the help of distance or correlation-based functions in the shared embedding space. When there are greater values of coherence it means that there is a greater correspondence between visual affect and critical interpretation. Also, the qualitative error analysis is performed to explore the instances of poor coherence, showing interpretive difference or ambiguity of annotation. Multimodal and unimodal baselines are compared using statistical significance test. A combination of these evaluation measures gives a broad understanding of the model effectiveness and interpretability as well as compatibility with human emotional comprehension.

5. RESULTS AND ANALYSIS

The experimental findings indicate that the suggested multimodal visual sentiment mining system is effective in comparison to unimodal baselines in all measurement parameters. Models that used CNN and ViT visual representations and transformer-based textual embeddings had a greater accuracy and macro F1-score, which means that they are more robust in dealing with a variety of emotional categories. The emotion coherence index also affirmed a high degree of

compatibility of visual affect and critical sentiment, which proved the success of cross-modal fusion. The qualitative analysis indicated that the framework has effectively helped to capture the subtle emotions that are represented in color harmony, compositional tension, and metaphorical language. The instances of low performance were mainly linked to the most abstract works of art and intentionally ambiguous reviews, accentuating inherent variability of interpretation, instead of the lack of models.

Table 2

Table 2 Comparative Performance of Unimodal and Multimodal Models				
Model Configuration	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-Score (%)**
Visual Only (CNN)	68.7	66.9	65.4	66.1
Visual Only (ViT)	71.3	69.8	68.2	69
Text Only (Transformer)	74.6	75.2	73.1	74.1
CNN + Text (Late Fusion)	78.9	79.4	77.8	78.6

Table 2 findings show that there are distinct performance disparities between unimodal and multimodal strategies of visual sentiment mining in modern art reviews. Of unimodal models, the text-only transformer performs best (74.6% and 74.1% accuracy and F1-score, respectively), which implies that critic reviews have more and explicit emotional clues, which can be more readily modeled than only visual ones.

Figure 3
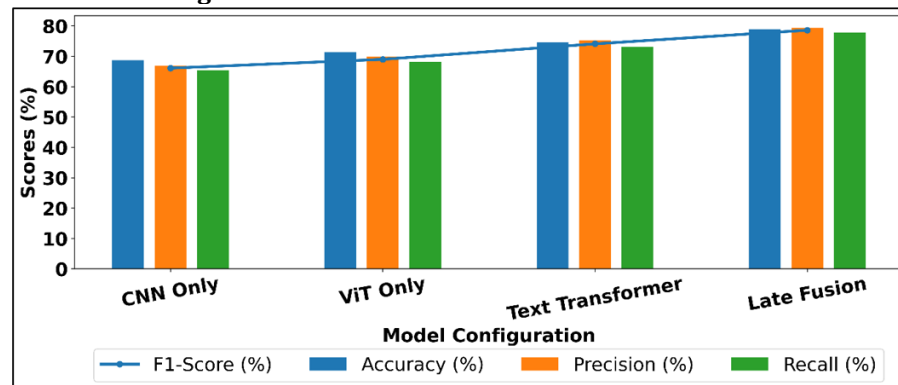
Figure 3 Comparison of Accuracy, Precision, Recall, and F1 Across Model Configurations

Accuracy, precision, recall, F1 are compared across models as shown in Figure 3. Comparatively lower scores are achieved by the visual-only models, with CNN-based model showing the F1-score of 66.1% and ViT-based model advancing it to 69.0% which is indicative of the benefit of a more global contextual modeling over the artistic composition and affect. Figure 4 presents the comparison of evaluation metrics of the visual, text, multimodal models.

Figure 4

Figure 4 Evaluation Metrics Across Visual, Text, and Multimodal Models

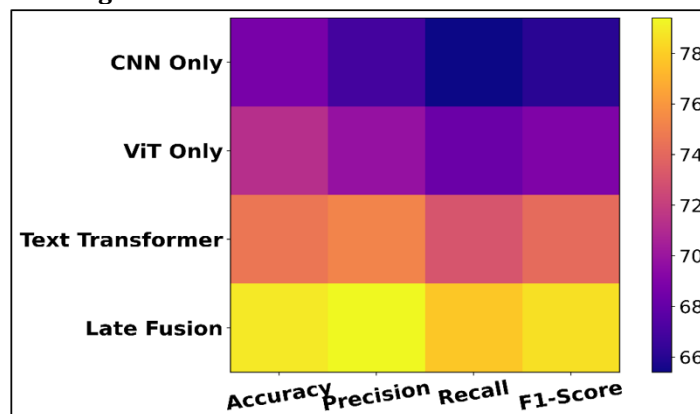But both of the visual models are confined by the abstract and symbolic characteristics of visual arts, in which emotions are unspoken. The CNN + Text late fusion model score a significant improvement of 78.9% accuracy and 78.6% F1-score proves the complementary characteristics of visual and textual modalities. The fact that the performance increased by more than 4.5 percent in F1-score, when compared to the optimal unimodal model, speaks volumes of the significance of cross-modal integration.

Table 3

Table 3 Emotion-Level Classification Performance of the Proposed Model			
Emotion Category	**Precision (%)**	**Recall (%)**	**F1-Score (%)**
Joy / Delight	88.4	86.9	87.6
Melancholy	83.7	81.5	82.6
Tension / Anxiety	86.1	84.3	85.2
Serenity / Calm	89.2	87.6	88.4

As shown in Table 3, the proposed multimodal model is performing remarkably well with the overall high score in various categories of emotions, which provides evidence of its strength in attaining subtle affective reactions in modern art reviews. Serenity/Calm has the best precision (89.2%), and F1-score (88.4), implying that more reliable detection is made of visually and linguistically consistent harmony, balance, and soft tonal language emotions. Joy/Delight scores are also high since the F1-score is 87.6 indicating that both in terms of the color usage and the positive critical descriptors, there are clear affective cues. Figure 5 presents the comparison of precision, recall, and F1-score in comparison with emotion categories.

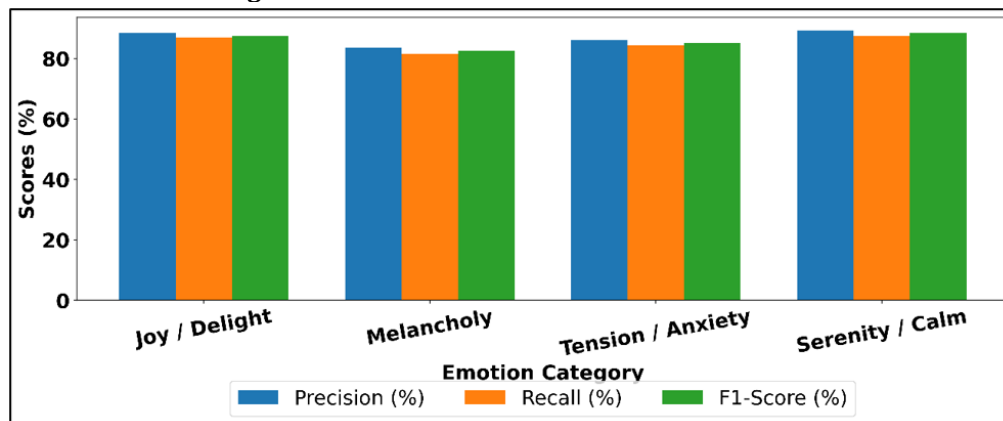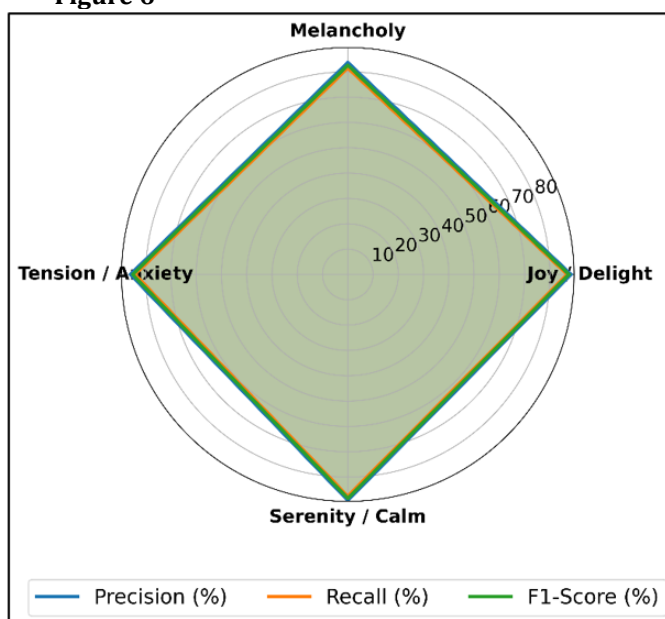**Figure 5**

Figure 5 Comparison of Precision, Recall, and F1-Score Across Emotion Categories

By contrast, in Melancholy, precision and recall are relatively lower, as sadness or introspection is expressed in art and in criticism in a very subtle and often metaphorical manner. The radar visualization of emotion classification performance metrics indicates results as demonstrated in Figure 6.

Figure 6

Figure 6 Radar Visualization of Emotion Classification Metrics

Tension / Anxiety balance between the accuracy (86.1) and the recall (84.3) is quite high, which means that the visual dissonance and language of conflict or distress are well recognized. On the whole, the range of differences across categories supports the role of emotional clarity and interpretive ambiguity in classification performance and proves the fact that the offered framework can be used to model explicit and complex expressions of emotions with strong reliability.

6. APPLICATIONS AND IMPLICATIONS
6.1. AUTOMATED ASSISTANCE FOR ART CRITICS AND CURATORS

Visual sentiment mining is able to be an intelligent support system to art critics and curators, enhancing, as opposed to removing, human interpretation. Technologies can be used to process large sets of art and reviews to determine dominating emotional patterns, changes in the tone of critique, and affective tendencies across exhibitions or eras. To the critics, emotional mind-reading devices can be helpful in comparative writing by pointing out emotional differences between artists, movements or story-lines of individual curators, allowing writing of better quality and, crucially, reflection. Dashboards of emotional distributions in suggested exhibitions can be useful to the curators, enabling them to create emotionally consistent or intentionally contrasting display sequences. This is also the case with such systems, where discrepancies between visual sentiment and critical reception can be indicated, and further interpretative analysis can be performed. Notably, these tools are not finalizing, they merely offer systematized information that helps the curator to make decisions. The framework will adhere to the practice of transparency and contextual reasoning through providing explainable connections between visual characteristics, verbal indicators, and expected feelings, which is significant to scholarly practices.

6.2. SENTIMENT-AWARE ART RECOMMENDATION AND DIGITAL EXHIBITIONS

Emotionally aware image recognition would allow the creation of more emotionally intelligent art recommendation systems and digitized exhibition experiences. The conventional recommendation engines commonly use stylistic similarity, popularity of artists, or history of user interactions but ignore the emotional aspects such that they are very important to the audience engagement. Recommendation systems can also be used to propose an artwork that matches the emotional inclination, moods, or reflective objectives of a viewer by using visual and textual sentiment mining. Sentiment-based curation in digital exhibitions and virtual museums can dynamically group works of art in emotionally engaging stories, which is more immersive and translative. As an example, the exhibitions may be organized according

to the affective themes like tension, serenity, or melancholy instead of the stylistic categories. The interactive interfaces can enable the visitors to browse the collections using an emotional journey that creates a personalized and empathetic experience. On the institutional level, the sentiment analytics can be used to inform the exhibition design, strategy of online engagement, and outreach to the audience by identifying the effects that various emotional tones have on the viewing behavior.

6.3. TOOLS FOR ARTISTS TO ASSESS EMOTIONAL RECEPTION OF THEIR WORKS

To artists, it can be difficult to comprehend the emotional interpretation of their work by the critics and the audience, especially when it is on a large scale or digital format. Visual sentiment mining offers reflective tools which can give us an aggregated interpretable understanding of emotional reception without degrading artistic meaning into simplistic scores. Reviewing, examining the exhibition texts, and other related visual elements will allow artists to see the trends of how feelings like tension, warmth, or ambiguities were interpreted on a regular basis. This feedback may assist in self-reflection, experimentation and communication strategies in art and still maintain autonomy of the creative process. Emotion-sensitive technologies can also assist artists in determining the disconnect between desired expression of emotion and the perceived response, inspiring a conversation and not dictatorship. These tools can be useful in learning and professional development in contexts such as critique sessions, portfolio assessment, and classical longitudinal analysis of artistic development. Notably, it is significant that ethical deployment focuses on the use of consent, transparency, and a contextual explanation so that it cannot be misunderstood or over-rule on automated decisions.

7. CONCLUSION

The paper has proposed a theoretical framework of visual sentiment mining in the context of modern day art review, identifying the drawbacks of the generic sentiment analysis tools in the context of artistic discourse. The proposed system by incorporating rich visual feature detection, transformer-based textual sentiment classification, and multimodal fusion systems involves the intricate interaction between visual expression and critical interpretation. The refined data and art-sensitive annotation procedures were used to make sure that emotional labels were interpreted by their fine-tuning instead of mere simplistic polarity, consistent with the approach of computational analysis and art-theoretical views. The importance of multimodal models was also proved by experimental evaluation that establishes that multimodal models out-perform image-only and text-only approaches by a considerable degree of margin. The introduction of emotion coherence index did give a significant measure of cross-modal alignment that may add more insights to the old accuracy-based measure. Such findings indicate that the development of emotional sense in the modern art is neither generated by the independent modes but by the dynamics of their interaction both in the cultural and contextual sense. In addition to the technical contributions, the study emphasizes the practical implications on the critics, curators, institutions as well as artists. Emotionally knowledgeable tools have the potential to aid in curatorial planning, improve online exhibition experiences and give reflective feedback regarding how the emotional reception, without closing off the interpretive openness. Notably, the framework is created to support the human expertise, but not to substitute it, which supports the transparency, explainability and ethical use.

CONFLICT OF INTERESTS

None.

REFERENCES

Chatterjee, S., Goyal, D., Prakash, A., and Sharma, J. (2021). Exploring Healthcare/Health-Product E-Commerce Satisfaction: A Text Mining and Machine Learning Application. Journal of Business Research, 131, 815–825. https://doi.org/10.1016/j.jbusres.2020.10.043

Chen, D., Hu, Z., Tang, Y., Ma, J., and Khanal, R. (2024). Emotion and Sentiment Analysis for Intelligent Customer Service Conversation using a Multi-Task Ensemble Framework. Cluster Computing, 27(3), 2099–2115. https://doi.org/10.1007/s10586-023-04073-z

Dhar, S., and Bose, I. (2022). Walking on Air or Hopping Mad? Understanding the Impact of Emotions, Sentiments and Reactions on Ratings in Online Customer Reviews of Mobile Apps. Decision Support Systems, 162, Article 113769. https://doi.org/10.1016/j.dss.2022.113769

Filieri, R., Lin, Z., Li, Y., Lu, X., and Yang, X. (2022). Customer Emotions in Service Robot Encounters: A Hybrid Machine–Human Intelligence Approach. Journal of Service Research, 25(4), 614–629. https://doi.org/10.1177/10946705221103937

Guo, J., Wang, X., and Wu, Y. (2020). Positive Emotion Bias: Role of Emotional Content from Online Customer Reviews in Purchase Decisions. Journal of Retailing and Consumer Services, 52, Article 101891. https://doi.org/10.1016/j.jretconser.2019.101891

Hung, H. Y., Hu, Y., Lee, N., and Tsai, H. T. (2024). Exploring Online Consumer Review–Management Response Dynamics: A Heuristic–Systematic Perspective. Decision Support Systems, 177, Article 114087. https://doi.org/10.1016/j.dss.2023.114087

Jeon, J., Kim, E., Wang, X., and Tang, L. R. (2023). Predicting Restaurant Hygiene Ratings: Does Customer Review Emotion and Content Matter? British Food Journal, 125(10), 3871–3887. https://doi.org/10.1108/BFJ-01-2023-0011

Nguyen, H. T. T., and Nguyen, T. X. (2023). Understanding Customer Experience with Vietnamese Hotels by Analyzing Online Reviews. Humanities and Social Sciences Communications, 10, Article 1–13. https://doi.org/10.1057/s41599-023-02098-8

Nguyen, N., Nguyen, T. H., Nguyen, Y. N., Doan, D., Nguyen, M., and Nguyen, V. H. (2024). Machine Learning-Based Model for Customer Emotion Detection in Hotel Booking Services. Journal of Hospitality and Tourism Insights, 7(4), 1294–1312. https://doi.org/10.1108/JHTI-03-2023-0166

Park, H., Jiang, S., Lee, O. K. D., and Chang, Y. (2024). Exploring the Attractiveness of Service Robots in the Hospitality Industry: Analysis of Online Reviews. Information Systems Frontiers, 26(1), 41–61. https://doi.org/10.1007/s10796-021-10207-8

Purohit, A. (2021). Sentiment Analysis of Customer Product Reviews Using Deep Learning and Comparison with Other Machine Learning Techniques. International Journal for Research in Applied Science and Engineering Technology, 9(1), 233–239. https://doi.org/10.22214/ijraset.2021.36202

Rodríguez-Ibánez, M., Casánez-Ventura, A., Castejón-Mateos, F., and Cuenca-Jiménez, P. M. (2023). A Review on Sentiment Analysis from Social Media Platforms. Expert Systems with Applications, 223, Article 119862. https://doi.org/10.1016/j.eswa.2023.119862

Shaik Vadla, M. K., Suresh, M. A., and Viswanathan, V. K. (2024). Enhancing Product Design through Ai-Driven Sentiment Analysis of Amazon Reviews Using BERT. Algorithms, 17(2), Article 59. https://doi.org/10.3390/a17020059

Sykora, M., Elayan, S., Hodgkinson, I. R., Jackson, T. W., and West, A. (2022). The Power of Emotions: Leveraging User-Generated Content for Customer Experience Management. Journal of Business Research, 139, 614–629. https://doi.org/10.1016/j.jbusres.2022.02.048

Wang, J., Zhao, Z., Liu, Y., and Guo, Y. (2021). Research on the Role of Influencing Factors on Hotel Customer Satisfaction Based on BP Neural Network and Text Mining. Information, 12(3), Article 99. https://doi.org/10.3390/info12030099

Xu, X., Liu, W., and Gursoy, D. (2019). The Impacts of Service Failure and Recovery Efforts on Airline Customers' Emotions and Satisfaction. Journal of Travel Research, 58(6), 1034–1051. https://doi.org/10.1177/0047287518789285