

Vision Language Models for Novel Art Therapy Evaluation in Schizophrenia

Ivan Nenchev^{1,2} , Karin Dannecker³ , Maren Rabe¹, Marie Jeschke⁴, and Christiane Montag¹ 

¹ Department of Psychiatry and Psychotherapy, Charité at St. Hedwig Hospital, Charité – Universitätsmedizin Berlin, Berlin, Germany

² Berlin Institute of Health at Charité – Universitätsmedizin Berlin, Berlin, Germany

³ Weißensee Kunsthochschule Berlin MA Art Therapy Programme, Berlin, Germany

⁴ Galerie ART CRU Berlin, Berlin, Germany

Abstract

Traditional methodologies for evaluating visual artistic output in art therapy remain rare and time-intensive, creating barriers to systematic assessment of therapeutic progress. This study presents the first application of multimodal dense embeddings for longitudinal evaluation of art therapy outcomes in individuals with schizophrenia. We analyzed 168 art therapy images produced by 14 participants with schizophrenia using CLIP (Contrastive Language-Image Pretraining) embeddings. CLIP embeddings successfully captured meaningful semantic patterns, with real images showing significantly greater semantic dispersion than spatially randomized controls. Longitudinal analysis revealed progressive semantic diversification over time, with significant increases in semantic distance between consecutive images ($\beta = 0.284$, $p = 0.001$) and cumulative semantic drift from first images ($\beta = 0.336$, $p < 0.001$). Individual differences analysis showed high variability in volume metrics spanning several orders of magnitude ($M = 1.13 \times 10^{11}$, $SD = 2.05 \times 10^{11}$), indicating highly individual semantic exploration patterns. Vision language models provide a novel and objective methodology for evaluating the progression of art therapy that reveals systematic patterns of semantic evolution during treatment. The progressive semantic diversification observed suggests that art therapy facilitates expanding creative expression and psychological exploration over time. The substantial individual differences in semantic exploration patterns indicate potential for personalized treatment approaches based on creative trajectory analysis. This methodology offers promising applications for systematic art therapy assessment, treatment monitoring, and personalized intervention strategies in clinical practice.

Keywords: art therapy, schizophrenia, CLIP embeddings, vision language models, semantic analysis, longitudinal assessment

1 Introduction

Art therapy (AT) is the therapeutic use of visual art. Through drawing, painting, sculpting, and other media, visual works are created and processes are initiated that serve as symbolic equivalents of experiences, emotions, thoughts, and fantasies [3]. These artworks help patients to perceive, understand, and communicate their conscious and unconscious conflicts and difficulties [4]. The aesthetic-therapeutic process activates a person's imaginative, cognitive, integrative, and motor capacities. AT supports emotional expression, stimulates thought processes and the ability to

Ivan Nenchev, Karin Dannecker, Maren Rabe, Marie Jeschke, and Christiane Montag. "Vision Language Models for Novel Art Therapy Evaluation in Schizophrenia." In: *Computational Humanities Research 2025*, ed. by Taylor Arnold, Margherita Fantoli, and Ruben Ros. Vol. 3. Anthology of Computers and the Humanities. 2025, 294–305. <https://doi.org/10.63744/sttzqxdNWsq1>.

© 2025 by the authors. Licensed under Creative Commons Attribution 4.0 International (CC BY 4.0).

organize experience, and deepens empathy, introspection, relational capacities, and communication skills. The tangible and enduring nature of artistic works are particularly effective therapeutic factors, as they help patients develop a stronger sense of identity, autonomy, self-efficacy, and self-worth. As a therapeutic modality, AT serves as a valuable complement to traditional biological and psychological treatments and is implemented across a broad range of patients with psychiatric [10] and somatic conditions [25].

Growing empirical evidence supports AT's clinical efficacy across diverse populations and conditions. A systematic review and meta-analysis of randomized clinical trials found that visual AT was associated with therapeutic benefits across several outcomes, though methodological quality varied among studies [11]. Specific populations have demonstrated particularly promising results. Meta-analyses have shown beneficial effects for individuals with post-traumatic stress disorder [23], large effect sizes for group art therapy interventions among children and adolescents [14], and significant reductions in anxiety symptoms in pediatric populations [25]. For individuals with schizophrenia, the evidence presents a more nuanced picture. While the NICE Guidelines indicate that arts therapies are effective in reducing negative symptoms compared to control conditions [17], and several studies suggest positive effects [15; 16], the overall evidence base remains inconclusive [1].

AT has been increasingly influenced by growing digitization [26], presenting both opportunities and challenges for practitioners and clients. While digital technologies can enhance client involvement and broaden creative boundaries, they also introduce risks of exclusion due to reduced access and varying levels of digital competence among participants. The potential roles of artificial intelligence (AI) in AT are rapidly evolving and multifaceted. A recent survey by Jütte et al. [12] of 56 art therapists found that 30% frequently or occasionally use AI in their practice with patients. AI's therapeutic role exists on a spectrum, ranging from partner in co-creative processes [5; 20] to curator of personalized visuals with therapeutic intent [24]. This spectrum also encompasses varying levels of autonomy, from supportive tools that augment human creativity to autonomous agents capable of independent therapeutic decision-making [27]. Interestingly, AI approaches for the evaluation of AT imagery have not been explored yet. While most AI research in AT has concentrated on enhancing the therapeutic process itself, AI approaches for evaluating artwork have remained largely unexplored. We believe that AI methods hold significant potential for advancing imagery assessment.

1.1 Evaluation of art therapy process

The exploration of the artworks and creative processes provides the art therapist with valuable diagnostic insights and guides the formulation of therapeutic goals and the ongoing conduct of therapy. Nevertheless, traditional methodologies for the standardized evaluation of the visual artistic output in AT [2; 6; 8; 21] focus on rating the artistic output on several predefined scales. These approaches remain rare and are often time-intensive, creating barriers to systematic assessment of therapeutic progress. In addition to this, several attempts have been made to develop approaches for automatic assessment. For example, Gengenbach et al. [7] demonstrated that a convolutional neural network could match human ratings of artworks. More recently, Kim et al. [13] employed a ResNet model to successfully classify between healthy controls and individuals with anxiety, depression, and schizophrenia based on their mandala drawings.

Contemporary computer vision and multimodal vision-language models present novel methodologies for image evaluation that could revolutionize art therapy assessment. CLIP (Contrastive Language–Image Pretraining), trained on 400 million image-text pairs from publicly available internet sources, employs contrastive learning to associate images with corresponding text [19]. This approach makes CLIP general-purpose and zero-shot capable across diverse downstream tasks, including aesthetic evaluation [9]. In this paper, we propose a novel method to evaluate the images

produced by patients during AT using a longitudinal analysis based on multimodal embeddings. To our knowledge, this represents the first application of multimodal embeddings for evaluating images produced during art therapy sessions. This study addresses the following research questions:

1. To what extent do CLIP embeddings capture meaningful semantic content in AT imagery?
2. How does semantic similarity between consecutive artworks change over the course of AT sessions?
3. How do individual participants differ in their exploration and occupation of the visual semantic space?

2 Materials and methods

2.1 Longitudinal corpus of images produced during AT

The study is based on a secondary analysis of visual material produced during a randomized controlled trial (RCT) evaluating the efficacy of psychodynamic group art therapy for individuals with schizophrenia or related psychotic disorders [16]. The study was ethically approved and conducted in accordance with the Declaration of Helsinki. The intervention consisted of 12 group sessions delivered over six weeks, with each session lasting 90 minutes. Group size varied between 3 and 6 participants. Sessions took place in a purpose-designed studio space located within the psychiatric clinic, intended to provide a creative environment. The therapeutic model followed a non-directive framework. Participants were free to choose their materials, techniques, and subjects without guidance or thematic prescription. They were encouraged to find their own visual language and proceed at their own pace. The art therapist’s interventions were aimed to supporting the artistic process and facilitating participants’ engagement with their imagery as well as the verbal reflection about the art work and its potential meaning, both individually and within the group context. At the end of each session, all artworks were systematically photographed and cataloged. This process yielded a longitudinal image corpus documenting the unfolding of individual and group expression across the treatment period. This dataset serves as the basis for the current AI-based reanalysis.

2.2 Participants

Of the 16 participants who completed the intervention, 2 were excluded from the reanalysis because they primarily used clay, producing three-dimensional works that could not be reliably captured through two-dimensional photographs. The remaining 14 participants (female, $n = 8$; male, $n = 6$, 12 diagnosed with paranoid schizophrenia (ICD-10 F20.0), one diagnosed with schizoaffective disorder (F25.2) and one with acute polymorphic psychotic disorder (F23.0)) produced a total of 164 artworks during the intervention (mean = 11.7, SD = 7.5). Due to the non-directive approach and the allowance for participants to work at their own pace, there was considerable variation in productivity, ranging from 3 to 30 pieces per individual. The basic sociodemographic and clinical characteristics of the sample are summarized in Table-1.

2.3 Feature Extraction and Formal Analysis

For the formal analysis of the artwork images, feature extraction was performed using the OpenAI CLIP model, specifically the clip-vit-base-patch32 architecture, accessed via the Hugging Face transformers library [19]. This model encodes visual content into 512-dimensional embeddings that capture semantic and stylistic features of the images. All computations were carried out on an NVIDIA A100 GPU within the high-performance computing (HPC) cluster environment of our

| | Mean | Std | Min | 25% | 50% | 75% | Max |
|-----------------------------|--------|--------|-------|--------|--------|--------|--------|
| Age (years) | 36.69 | 10.83 | 24.96 | 28.78 | 33.44 | 42.16 | 58.29 |
| Duration of illness (years) | 11.19 | 10.59 | 0.13 | 5.32 | 8.28 | 14.47 | 40.29 |
| Clorpromazin Equivalents | 377.82 | 237.40 | 80.00 | 200.00 | 300.00 | 587.50 | 800.00 |
| GAF | 38.07 | 13.83 | 10.00 | 30.00 | 37.50 | 47.25 | 60.00 |
| Symptom severity (BPRS) | 34.21 | 10.58 | 19.00 | 26.25 | 33.50 | 41.75 | 53.00 |
| Number of images | 11.71 | 7.50 | 3.00 | 6.25 | 11.00 | 13.00 | 30.00 |

Table 1: Descriptive statistics of demographic, clinical, and artistic variables in the patient sample. Values represent mean, standard deviation, and distributional parameters (minimum, quartiles, maximum) for age, duration of illness, chlorpromazine equivalents, Global Assessment of Functioning (GAF), symptom severity (BPRS), and number of images created.

hosting institution. These embeddings served as the basis for subsequent quantitative analyses of the visual material.

Since artworks from people with schizophrenia have not yet been evaluated with CLIP embeddings, we employed a permutation-based validation framework to assess whether these embeddings capture meaningful semantic patterns beyond low-level visual features such as color and texture. We calculated the empirical coherence of the embedding space as the mean cosine distance from individual image embeddings to the group centroid. To test whether observed clustering patterns reflected genuine semantic content rather than superficial visual similarities, we generated a null distribution through 1,000 permutation iterations. In each iteration, we spatially altered the original images by randomly shuffling pixel values while preserving the overall color distribution, then re-encoded these spatially randomized images through the CLIP model. If CLIP embeddings primarily captured color or texture information, "scrambled" images would yield similar coherence values to the originals. Conversely, if embeddings captured higher-order semantic features, pixel-permuted images would show significantly different (typically higher) coherence values, indicating loss of meaningful structure. Statistical significance was assessed by comparing observed coherence against the null distribution. To visualize the embedding space structure, we performed t-distributed Stochastic Neighbor Embedding (t-SNE) dimensionality reduction on the 512-dimensional CLIP embeddings and plotted the resulting two-dimensional representation with actual images positioned at their corresponding coordinates. This visualization approach allowed for qualitative assessment of whether semantically similar images clustered together in the reduced space, providing complementary evidence to the quantitative permutation test that subsequent analyses were based on semantic content rather than confounding low-level visual features.

To quantify intra- and interindividual variation in image embeddings, we computed pairwise cosine distances between all embeddings using `sklearn.metrics.pairwise.cosine_distances` [18]. Each embedding was labeled by participant ID. We then categorized distances as either within-subject or between-subject and compared these groups using one-way ANOVA and variance ratio analysis.

To examine longitudinal trends in visual semantic change, we computed two sets of cosine distance scores based on the CLIP embeddings for each participant. The first set captured the similarity between the participant’s first image and each of their subsequent images, providing a measure of cumulative semantic drift over time. Each d_j captures the semantic distance between image j and image $j + 1$, enabling longitudinal modeling or step-wise analysis rather than aggregating over the entire sequence.

$$\text{Distance to next image} = 1 - \frac{x_j \cdot x_{j+1}}{\|x_j\| \|x_{j+1}\|}, \quad \text{for } j = 1, \dots, n - 1$$

The second set measured the similarity between each image and the one immediately following it, reflecting local semantic continuity across sessions.

$$\text{Distance from first image} = 1 - \frac{x_1 \cdot x_{j+1}}{\|x_1\| \|x_{j+1}\|}, \quad \text{for } j = 1, \dots, n - 1$$

These similarity scores $\{d_1, d_2, \dots, d_{n-1}\}$ served as outcome variables in two separate linear mixed-effects models. In both models, image position in the sequence was included as a fixed effect to estimate change over time, while participant identity was modeled as a random intercept to account for inter-individual variability.

In a final exploratory analysis, we aimed to estimate how individual participants behave within the abstract visual semantic space defined by their embeddings. To capture individual-level differences in how semantically broad or narrow these representational spaces are, we characterized the spread and volume of each participant’s embeddings. Unlike measures such as mean cosine similarity, which assess semantic coherence by quantifying directional alignment between vectors, this approach focused on distributional dispersion—i.e., how widely embeddings are spatially distributed—using metrics that are sensitive to both direction and magnitude. To facilitate interpretability and reduce computational complexity, high-dimensional embeddings were first projected into a lower-dimensional space using Principal Component Analysis (PCA). For each participant, we then extracted their PCA-reduced vectors and computed two complementary diversity metrics: spread, defined as the sum of standard deviations across all PCA dimensions, and volume, calculated as the product of the peak-to-peak (max–min) ranges along each axis, estimating the size of the hyperrectangle enclosing the embedding distribution. Together, these metrics provide a scale-sensitive characterization of individual variability in semantic space exploration. All computations were performed using the Python libraries scikit-learn and statsmodels [18; 22].

3 Results

3.1 CLIP Embedding Validation

To validate that CLIP embeddings capture meaningful semantic patterns rather than superficial visual features in the dataset of AT images produced by people with schizophrenia, we conducted a permutation-based statistical analysis. The observed coherence of real images in CLIP embedding space was 0.2290 (mean cosine distance to centroid), significantly higher than the null distribution generated from spatially scrambled images (mean = 0.0896, SD = 0.0010; $p = 0.0099$). This significant difference demonstrates that randomly altered images clustered much more tightly in embedding space than semantically intact images. In addition to this, the t-SNE dimensionality reduction visualization of the 512-dimensional CLIP embeddings corroborated the quantitative findings. Real images were distributed across distinct regions of the two-dimensional space, with semantically similar images (e.g., landscapes, portraits, abstract expressions, flowers and plants) forming loose clusters. This spatial organization in the reduced dimensional space provided qualitative confirmation that CLIP embeddings successfully captured meaningful semantic relationships within the AT image collection.

We observed a highly significant effect of subject identity on embedding similarity: within-subject distances were significantly lower than between-subject distances (ANOVA: $F(1, 9739) = 271.24$, $p < 0.00001$). However, the variance of within-subject distances was similar to that of between-subject distances (variance ratio = 0.96), suggesting that while embeddings are reliably more similar within individuals, the dispersion of distances is comparable across both groups.



Figure 1: t-SNE visualization of the dataset.

3.2 Longitudinal Analysis of Visual Semantic Evolution

We computed the semantic distance between consecutive images for each participant by calculating the cosine distance between pairs of embeddings, which were sorted by participant ID and image sequence number, and missing values for the last image in each sequence were handled appropriately. Both the predictor (image position) and outcome (cosine distance) variables were standardized using z-score normalization prior to modeling.

A linear mixed-effects model was then fitted, predicting standardized semantic distance between two consecutive embeddings from standardized image position, with participant included as a random intercept. The model showed a significant positive association between image position and semantic distance ($\beta = 0.284$, $SE = 0.085$, $z = 3.33$, $p = 0.001$), indicating that the semantic difference between consecutive images increased as the intervention progressed. Random intercept variance ($\sigma^2 = 0.310$) reflected moderate variability between participants.

To examine broader semantic drift over time, we calculated the cosine distance between each image and the participant’s first image. Embeddings were standardized, and a linear mixed-effects model was fitted with image position as a fixed effect and participant as a random intercept. The model revealed a significant positive effect of image position on semantic distance from the first image ($\beta = 0.336$, $SE = 0.080$, $z = 4.19$, $p < 0.001$), suggesting a cumulative increase in visual-semantic deviation as the intervention progressed. The intercept was not significant ($\beta = 0.115$, $p = 0.499$). Between-participant variance was estimated at 0.315, indicating moderate inter-individual variability in the degree of semantic change.

3.3 Semantic Space Exploration

Analysis of participants’ exploration within the CLIP embedding space revealed substantial individual differences in both the breadth and scope of semantic territories covered. The spread metric, which quantifies the distributional dispersion of embeddings across PCA dimensions, showed relatively consistent values across participants ($M = 34.7$, $SD = 4.1$, $range = 28.1 - 41.4$). This

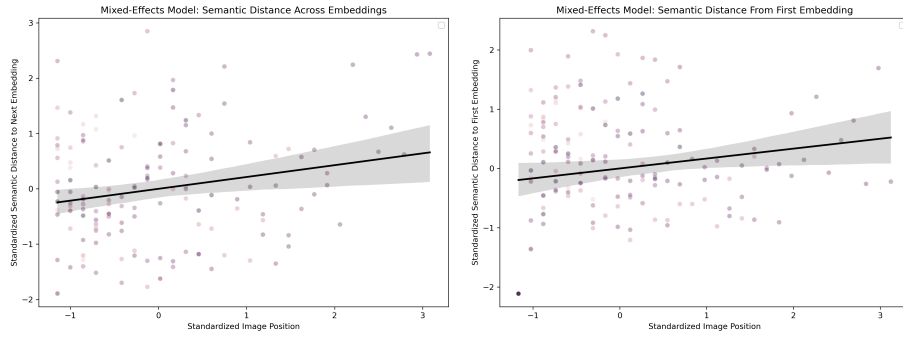


Figure 2: Semantic distance patterns in artwork over the course of art therapy intervention.

suggests that while participants varied in the specific semantic content they explored, the overall extent of their exploration within the reduced dimensional space was fairly uniform.

In contrast, the volume metric—representing the size of the hyperrectangle enclosing each participant’s embedding distribution—demonstrated much greater variability ($M = 1.13 \times 10^{11}$, $SD = 2.05 \times 10^{11}$). The volume scores spanned several orders of magnitude, from 9.41×10^7 to 7.36×10^{11} , indicating dramatic differences in the semantic space occupied by different participants. Figure 3 shows barplots for spread and volume of semantic space exploration across all participants (panels A-B), alongside t-SNE visualizations of image sequences for four exemplary participants (panels C-F). These examples illustrate contrasting patterns of semantic exploration: participants 42 and 33 demonstrate extensive semantic territory exploration with images distributed across large volumes of semantic space, while participants 21 and 35 show more constrained exploration with images remaining relatively tightly clustered within smaller semantic territories. Each point represents an individual artwork, with connecting lines indicating temporal sequence progression through the therapeutic intervention.

4 Discussion

This study represents the first application of multimodal embeddings for evaluating artistic output in AT, addressing a critical methodological gap. Traditional assessment methods for visual artistic expression remain rare and time-intensive [2; 6; 8; 21], creating barriers to systematic evaluation of therapeutic progress.

Our validation approach demonstrates that CLIP embeddings capture high-level semantic content rather than superficial visual features. When semantic structure was destroyed through pixel-level permutations while preserving color information, the resulting images produced highly similar embeddings that clustered together. In contrast, real art therapy images with diverse semantic content exhibited much greater dispersion across the embedding space. The embeddings also captured unique artistic styles of individual participants.

These validation results establish that our analyses are based on genuine semantic content rather than visual artifacts, supporting CLIP embeddings as a robust foundation for understanding thematic patterns in therapeutic artistic expression.

The longitudinal analysis reveals compelling patterns in the therapeutic process. The progressive increase in semantic distance between consecutive images suggests that art therapy facilitates expanding creative expression over time, with participants exploring increasingly diverse semantic territories rather than repeating similar visual themes.

This expanding creativity may reflect the therapeutic process as patients become more comfortable expressing varied emotions, enhanced cognitive flexibility enabling exploration of diverse visual concepts, and deepening self-exploration as individuals access broader psychological terri-

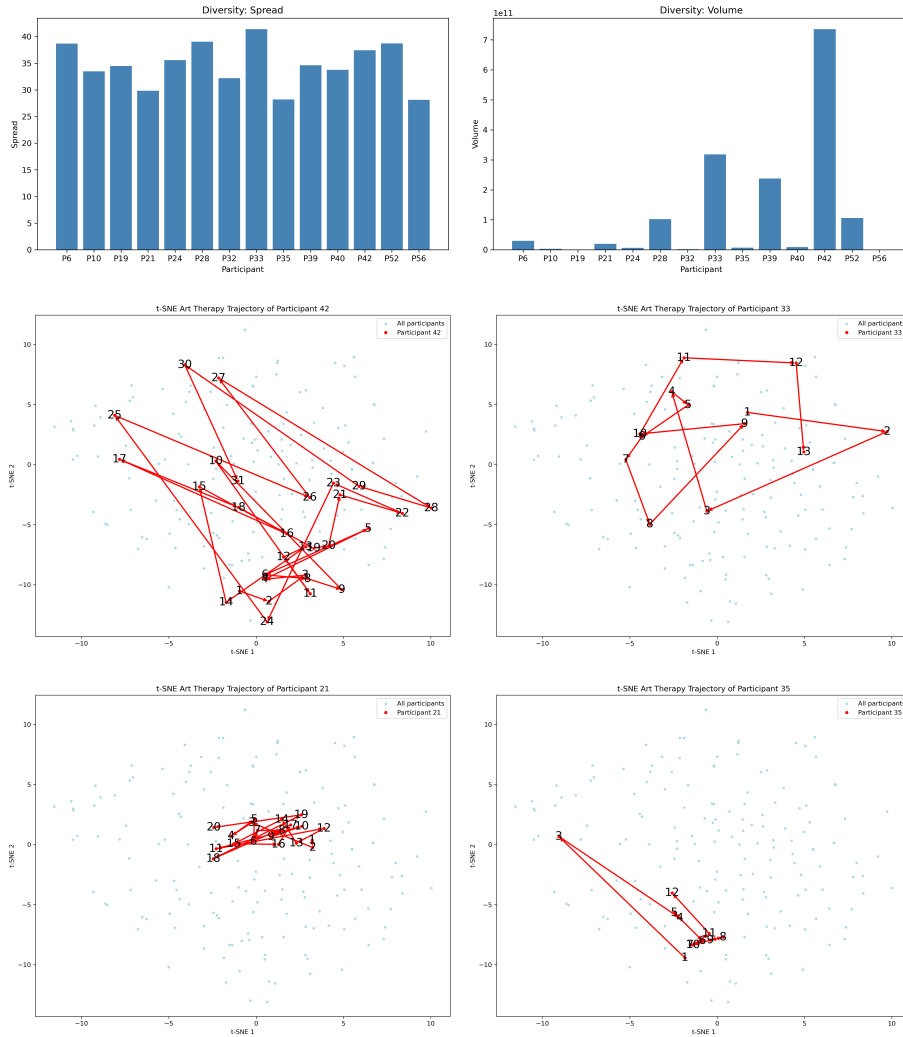


Figure 3: Individual differences in semantic space exploration during art therapy.

tories through artistic expression.

We also examined how participants occupy abstract visual semantic space throughout their therapeutic journey. The contrasting patterns between spread and volume metrics suggest that participants differed primarily in the range extremes of their semantic exploration rather than in overall distributional characteristics. While most participants showed similar patterns of dispersion around their central semantic tendencies (as indicated by consistent spread values), a subset demonstrated markedly expanded exploration boundaries (as captured by the highly variable volume metric). This finding indicates that individual differences in art therapy may be characterized less by different styles of exploration and more by differences in the boundaries of semantic territories accessed during the therapeutic process.

CLIP embeddings offer significant advantages over traditional evaluation methods by providing objective, scalable, and reproducible measures of artistic expression without requiring extensive training or subjective interpretation. This could facilitate systematic documentation of therapeutic progress, enable larger-scale research studies, and support real-time treatment monitoring.

Our findings may hold particular relevance for schizophrenia treatment, a domain in which cognitive flexibility and self-expression are often impaired. The progressive semantic diversification observed in participants' artwork may indicate therapeutic gains, such as enhanced cognitive

flexibility, greater expressive range, and improved adaptive functioning. However, it remains to be established how changes in CLIP ratings relate to variations in psychopathological symptoms or cognitive performance. At this stage of research, it cannot yet be determined whether these changes correspond to clinical improvement or deterioration. Longitudinal studies integrating standardized clinical and cognitive assessments are needed to elucidate the relationship between semantic variability in artwork and therapeutic outcomes.

This work opens several promising directions for future research and clinical application. The findings suggest the potential for developing automated assessment tools that could make art therapy evaluation more objective, accessible, and scalable, thereby broadening the reach and impact of creative therapies in clinical settings. Real-time monitoring of semantic exploration could offer clinicians objective feedback on therapeutic progress. The methodological framework introduced here could also be extended to other forms of creative expression, such as music therapy or creative writing, offering a broader perspective on the role of artistic processes in mental health treatment.

5 Conclusions

This study demonstrates that multimodal embeddings provide a powerful new methodology for evaluating artistic expression in AT. The progressive semantic diversification observed in participants' artwork provides objective evidence for therapeutic change and creative growth over time. The substantial individual differences in semantic exploration patterns suggest opportunities for personalized treatment approaches based on creative trajectory analysis. These findings establish a foundation for more systematic, objective, and scalable evaluation of art therapy interventions, with potential applications across diverse clinical populations and therapeutic settings.

6 Limitations

This study has several limitations. First, the dataset is relatively small, which may constrain the generalizability and robustness of the findings. Second, the sample consists exclusively of participants diagnosed with psychotic disorders, limiting the applicability of results to other clinical or non-clinical populations. Third, the study lacked a control group, making it difficult to disentangle the effects of AT from other factors such as spontaneous symptom fluctuation or nonspecific therapeutic influences. Additionally, while we demonstrated significant semantic changes over time, we did not directly correlate these changes with clinical outcomes or symptom measures. Future research should examine whether semantic trajectory patterns predict treatment response or clinical improvement. Finally, the analysis was based on a single model (CLIP-ViT-B/32); future research should compare the performance of alternative vision–language models to assess consistency and potential model-specific biases.

7 Acknowledgments

I.N. is a participant in the BIH Charité Digital Clinician Scientist Program funded by the Charité – Universitätsmedizin Berlin and the Berlin Institute of Health at Charité (BIH). The authors acknowledge the Scientific Computing of the IT Division at the Charité - Universitätsmedizin Berlin for providing computational resources that have contributed to the research results reported in this paper.

References

- [1] Attard, Angelica and Larkin, Michael. “Art therapy for people with psychosis: a narrative review of the literature”. English. In: *The Lancet Psychiatry* 3, no. 11 (Nov. 2016). Publisher: Elsevier, pp. 1067–1078. DOI: 10.1016/S2215-0366(16)30146-8.

- [2] Cohen, Barry M., Hammer, Jeffrey S., and Singer, Shira. "The diagnostic drawing series: A systematic approach to art therapy evaluation and research". In: *The Arts in Psychotherapy*. Special Issue Assessment in the Creative Arts Therapies 15, no. 1 (Mar. 1988), pp. 11–21. ISSN: 0197-4556. DOI: 10.1016/0197-4556(88)90048-2. (Visited on 06/29/2025).
- [3] Dannecker, Karin. *Psyche und Ästhetik: die Transformationen der Kunsttherapie*. ger. 4., durchgesehene Auflage. Berlin: Medizinisch Wissenschaftliche Verlagsgesellschaft, 2021. ISBN: 978-3-95466-579-2.
- [4] Dannecker, Karin. "Why do some people see the unseen?" en. In: *GMS Journal of Arts Therapies – Journal of Art-, Music-, Dance-, Drama- and Poetry-Therapy* 6 (Feb. 2024). DOI: 10.3205/jat000034.
- [5] Du, Xuejun, An, Pengcheng, Leung, Justin, Li, April, Chapman, Linda E., and Zhao, Jian. "DeepThInk: Designing and probing human-AI co-creation in digital art therapy". In: *International Journal of Human-Computer Studies* 181 (Jan. 2024), p. 103139. DOI: 10.1016/j.ijhcs.2023.103139.
- [6] Elbing, Ulrich, Hölzer, Michael, Danner-Weinberger, Alexandra, and Wietersheim, Jörn von. "Reliabilität und Validität des Instruments „DoKuPro – Dokumentation Kunsttherapeutischer Prozesse“". In: *Musik-, Tanz- und Kunsttherapie* 20, no. 1 (Jan. 2009), pp. 1–7. DOI: 10.1026/0933-6885.20.1.1.
- [7] Gengenbach, Thomas and Schoch, Kerstin. "ARTificial intelligence raters. Neural networks for rating pictorial expression". en. In: *Journal of Science and Technology of the Arts* 14, no. 1 (Apr. 2022). Number: 1, pp. 49–71. DOI: 10.34632/jsta.2022.10196.
- [8] Hacking, S., Foreman, D., and Belcher, J. "The descriptive assessment for psychiatric art. A new way of quantifying paintings by psychiatric patients". eng. In: *The Journal of Nervous and Mental Disease* 184, no. 7 (July 1996), pp. 425–430. DOI: 10.1097/00005053-199607000-00005.
- [9] Hentschel, Simon, Kobs, Konstantin, and Hotho, Andreas. "CLIP knows image aesthetics". English. In: *Frontiers in Artificial Intelligence* 5 (Nov. 2022). Publisher: Frontiers. DOI: 10.3389/frai.2022.976235.
- [10] Hu, Jingxuan, Zhang, Jinhuan, Hu, Liyu, Yu, Haibo, and Xu, Jinping. "Art Therapy: A Complementary Treatment for Mental Disorders". English. In: *Frontiers in Psychology* 12 (Aug. 2021). Publisher: Frontiers. DOI: 10.3389/fpsyg.2021.686005.
- [11] Joschko, Ronja, Klatte, Caroline, Grabowska, Weronika A., Roll, Stephanie, Berghöfer, Anne, and Willich, Stefan N. "Active Visual Art Therapy and Health Outcomes: A Systematic Review and Meta-Analysis". In: *JAMA Network Open* 7, no. 9 (Sept. 2024), e2428709. ISSN: 2574-3805. DOI: 10.1001/jamanetworkopen.2024.28709.
- [12] Jütte, Lennart, Wang, Ning, Steven, Martin, and Roth, Bernhard. "Perspectives for Generative AI-Assisted Art Therapy for Melanoma Patients". en. In: *AI* 5, no. 3 (Sept. 2024). Number: 3 Publisher: Multidisciplinary Digital Publishing Institute, pp. 1648–1669. DOI: 10.3390/ai5030080.
- [13] Kim, Seong-in, Kim, Kee-Eung, and Song, Seunghwan. "Exploring artificial intelligence approach to art therapy assessment: A case study on the classification and the estimation of psychological state based on a drawing". In: *New Ideas in Psychology* 73 (Apr. 2024), p. 101074. DOI: 10.1016/j.newideapsych.2024.101074.

- [14] Kong, Soyeon and Han, KyeongA. “A meta-analysis of the effectiveness of group art therapy for children and adolescents of multicultural families: A focus on Korea”. In: *Children and Youth Services Review* 161 (2024). Place: Netherlands Publisher: Elsevier Science, pp. 1–11. ISSN: 1873-7765. DOI: 10.1016/j.chidyouth.2024.107646.
- [15] Montag, Christiane and Dannecker, Karin. “A neglected area in schizophrenia treatment and research: the efficacy of art therapy: Results of a pilot randomized controlled trial and qualitative study”. In: *Arts Therapies and New Challenges in Psychiatry*. Num Pages: 26. Routledge, 2017.
- [16] Montag, Christiane, Haase, Laura, Seidel, Dorothea, Bayerl, Martin, Gallinat, Jürgen, Herrmann, Uwe, and Dannecker, Karin. “A pilot RCT of psychodynamic group art therapy for patients in acute psychotic episodes: feasibility, impact on symptoms and mentalising capacity”. eng. In: *PloS One* 9, no. 11 (2014), e112348. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0112348.
- [17] “Overview | Psychosis and schizophrenia in adults: prevention and management | Guidance | NICE”. eng. Publisher: NICE. Feb. 2014. URL: <https://www.nice.org.uk/guidance/cg178> (visited on 07/12/2025).
- [18] Pedregosa, F. et al. “Scikit-learn: Machine Learning in Python”. In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.
- [19] Radford, Alec et al. “Learning Transferable Visual Models From Natural Language Supervision”. arXiv:2103.00020 [cs]. Feb. 2021. DOI: 10.48550/arXiv.2103.00020.
- [20] Schmutz, Yannis Valentin, Kravchenko, Tetiana, Souissi, Souhir Ben, and Kurpicz-Briki, Mascha. “Integrating Generative AI into Art Therapy: A Technical Showcase”. arXiv:2412.03287 [cs]. Dec. 2024. DOI: 10.48550/arXiv.2412.03287.
- [21] Schoch, Kerstin, Gruber, Harald, and Ostermann, Thomas. “Measuring art: Methodical development of a quantitative rating instrument measuring pictorial expression (RizbA)”. In: *The Arts in Psychotherapy* 55 (Sept. 2017), pp. 73–79. DOI: 10.1016/j.aip.2017.04.014.
- [22] Seabold, Skipper and Perktold, Josef. “statsmodels: Econometric and statistical modeling with python”. In: *9th Python in Science Conference*. 2010.
- [23] Wang, Jiahua, Zhang, Bo, Yahaya, Rosliza, and Abdullah, Azizah binti. “Colors of the mind: a meta-analysis of creative arts therapy as an approach for post-traumatic stress disorder intervention”. In: *BMC Psychology* 13, no. 1 (Jan. 2025), p. 32. ISSN: 2050-7283. DOI: 10.1186/s40359-025-02361-4.
- [24] Yilma, Bereket A., Kim, Chan Mi, Ludden, Geke, Rompay, Thomas van, and Leiva, Luis A. “The AI-Therapist Duo: Exploring the Potential of Human-AI Collaboration in Personalized Art Therapy for PICS Intervention”. In: *International Journal of Human-Computer Interaction* 0, no. 0 (), pp. 1–14. DOI: 10.1080/10447318.2025.2487859.
- [25] Zhou, ShiShuang, Yu, MeiHong, Zhou, Zhan, Wang, LiWen, Liu, WeiWei, and Dai, Qin. “The effects of art therapy on quality of life and psychosomatic symptoms in adults with cancer: a systematic review and meta-analysis”. eng. In: *BMC complementary medicine and therapies* 23, no. 1 (Dec. 2023), p. 434. DOI: 10.1186/s12906-023-04258-4.
- [26] Zubala, Ania, Kennell, Nicola, and Hackett, Simon. “Art Therapy in the Digital World: An Integrative Review of Current Practice and Future Directions”. English. In: *Frontiers in Psychology* 12 (Apr. 2021). Publisher: Frontiers. DOI: 10.3389/fpsyg.2021.600070.

- [27] Zubala, Ania, Pease, Alison, Lyszkiewicz, Kacper, and Hackett, Simon. “Art psychotherapy meets creative AI: an integrative review positioning the role of creative AI in art therapy process”. English. In: *Frontiers in Psychology* 16 (Mar. 2025). Publisher: Frontiers. DOI: 10.3389/fpsyg.2025.1548396.