Computerized Art Therapy Assessment for Depression

Efrat Gilad
Antioch University
Seattle, WA

Ran Gilad-Bachrach, Mary Czerwinski
Microsoft Research
Redmond, WA

ABSTRACT

In this paper we discuss the use of computerized art therapy assessments to screen for mental disorders. These types of assessments have several advantages: They are cross cultural, they can collect both psychological and physiological aspects that manual assessments cannot measure, and they can act as interventions.

We have developed DFEATS, a digital version of the Formal Elements Art Therapy Scales (FEATS), a depression assessment tool. A large study using a crowd sourcing platform revealed that subjects express strong and significant preference for the use of the art therapy assessment tool over the use of traditional questionnaires. Moreover, users who suffered from higher levels of depression expressed even stronger preference for the art therapy assessment over the use of questionnaires. However, DFEATS did not provide sensitive enough detection of depression. Therefore, we conclude that there are great benefits from using art therapy as a source for interventions in the digital world, while more studies are needed to find ways to best develop digital art therapy assessments.

Author Keywords

Art Therapy; Depression; Behavioral Changes

ACM Classification Keywords

J.4 Computer Applications → Social and Behavioral Sciences → Psychology

INTRODUCTION

Art therapy has many advantages as a therapeutic and diagnosis tool for mental disorders. For example, it relies less on language as a mean of communication and therefore can cross cultural and developmental barriers that are harder to cross with other techniques. Over the years, several art therapy assessment tools have been developed, for example the Person Picking an Apple from a Tree (PPAT) [1] together with the Formal Elements Art Therapy Scales (FEATS) [2]. The PPAT is a task in which a subject is asked to draw on paper a person picking an apple from a tree. The subject is given 12 colored pens to complete this task. After that, the drawing is assessed using FEATS and analyzed to come up with the diagnosis of different mental disorders, most notably depression.

This assessment method does not scale well to large populations due to the manual labor associated with analyzing the drawings and the expertise needed to complete this task. Therefore, we have developed a digital version of this assessment. In the digital version, the drawing task is performed on a touch screen computer and the drawing is analyzed using machine learning tools. This method has many advantages. Since the analysis is done by a software algorithm, the inter-rater differences no longer play any role. Not only does this method avoid the subjectivity of the manual assessment, it also allows us to collect information that is not available when the assessment is administered manually. For example, we can measure the speed in which lines are drawn, the number of pen strokes, and the variability in stroke length.

This study has multiple objectives: We were interested in verifying the technical feasibility of providing such art therapy assessments at a large scale. At the same time, we were also interested in subjects' reactions to such tools and how they would evaluate them when compared to traditional, questionnaire-based assessment. Finally, we were interested at the accuracy by which depression can be recognized from these automated assessment tools.

We have developed a web based tool to administer the digital version of the PPAT and we tested it using Amazon's Mechanical Turk (mTurk). In the study, subjects were asked to complete both a depression questionnaire and the drawing assessment. We found an overwhelming subjective preference for the drawing assessment over the questionnaire. Even more remarkable was the fact that subjects suffering from depression expressed even greater preference for the drawing assessment tool, and therefore, we suggest that it could have a positive effect if used as an intervention.

We have extracted 130 different features from the drawings we collected in mTurk and used machine learning techniques to analyze them. We have found that there is a correlation between these features and symptoms of depression. However, the correlation is insufficiently strong to detect depression with high accuracy.

Figure 1: A screen shot of the drawing tool.

We conclude that there is a great interest in art-therapy assessments and interventions, and in their digital manifestations. Current technology allows for using art therapy at scales never before attempted in order to help a large population in need.

In the rest of this document we provide a short description of the study and the results. We refer the interested reader to Gilad, 2015 [3] for more details about this work.

THE DIGITAL ASSESMENT TOOL

In this section we discuss how the manual assessment was turned into a digital one. The manual assessment contains two components. In the first step, subjects are provided with a piece of paper and 12 pens and are asked to draw a person picking an apple from a tree (PPAT). In the second step, the drawing is analyzed using the Formal Elements Art Therapy Scales (FEATS).

We have created a web-based tool that allows drawing using a touch screen. Our goal was to make the tool simple to use, as close as possible to paper and 12 pens. Therefore, we kept this structure: the drawing tool allowed subject to pick between 12 colors, the same colors used in the PPAT task. We did not provide options, such as mixing colors, or controlling the width of the lines. Therefore, the interface was very simple to use as is evident from the responses of the subjects to the usability study on which we elaborate later on. **Error! Reference source not found.** shows a screen shot of the drawing tool.

While drawing, the tool collected all the strokes users made in order to allow feature extraction used to analyze the drawing. Upon completion, all of this information was collected securely for analysis on an Azure blob storage. Figure 2 shows two examples of drawings collected by our system.

In order to analyze the data and attempt depression detection, we extracted 130 different features describing the drawing and the process of generating it. These features include, for example, the percentage of the canvas that is covered, and the number of lines from each color. These features were inspired by the different scales in the manual assessments, the FEATS. However, there are some notable differences: Some features are easy to extract in the digital form but hard from the manual version of the assessment, such as the number of lines used, and the speed in which they were drawn. On the other hand, a few characteristics of the drawing are hard to compute without a human, for example, the level of realism in the drawing.

Figure 2: Two examples for drawings collected by our system. The subjects were asked to draw a person picking an apple from a tree.

In order to detect depression, we used trained machine learning algorithms. We have tested different learning algorithms: Logistic Regression [4], Random Forests [5], Boosted Trees [6], and DART [7]. We used Principal Component Analysis (PCA) [8] to reduce the dimensionality of the data, given the small sample available for training (we have collected 147 drawings.

EXPERIMENT DESIGN

We have recruited subjects participants using Amazon Mechanical Turk (mTurk) for this study. Subjects were required to be at least 18 years old and resident of the US. We also asked participants to use a device with a large touch screen such as laptop or tablet. Participants were asked to complete the following steps:

1. sign an informed consent form
2. complete a demographics questionnaire
3. complete a screening questionnaire [9]
4. complete the Zung Self-Rating Depression Scale (ZSDS) [10]
5. use the drawing tool to draw a person picking an apple from a tree
6. complete a usability questionnaire

Since most people do not suffer from depression, we have used an importance sampling method. Hence, not all participants were asked to perform the drawing task. Instead, all participants with a score of 50 or above in the ZSDS were considered to be suffering from a mild or severe level of depression and therefore they were asked to use the drawing tool. Each such participant was time matched with another that scored <50 in the ZSDS (not depressed). We paid $0.50 to any participant that completed all the steps in the study and added a bonus of $2.10 for those invited to perform the drawing task.

1220 workers participated in the study. From this pool, we have collected 147 drawings, 62 of them from people suffering from mild or sever levels of depression, and 85 drawings from people who did not suffer from depression.

MAIN RESULTS

Under the assumption that the drawings can tell about mental disorders such as depression, we have trained classifiers to distinguish between the drawings of people with no signs of depression and people with signs of depression. The most accurate classifier used both the features from the drawing as well as the demographic information and the answers to the usability questionnaire. This classifier has an Area Under the Curve (AUC) of 0.79. Most of the contribution comes from the demographics information (gender, marital status, etc.). If we use only the demographic information, a classifier can achieve an AUC of 0.78. However, if we restrict the classifier to use only the features extracted from the drawings, the highest AUC achieved is 0.68. This shows that the drawing does reveal information about the mental state; however, this information is insufficient to be used as a screening mechanism (an AUC closer to 1 would be needed for that).

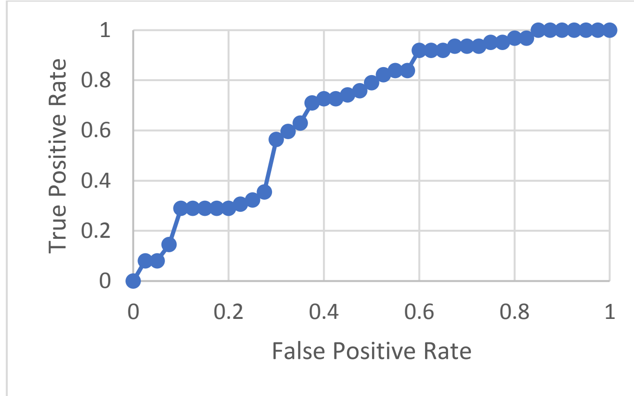

Figure 3: Receiver Operator Curve (ROC) for the predictor of depression symptoms.

A closer examination of the contribution of individual features reveals that the most predictive features measure the use of color in the picture. Examples of such features include the amount of greenish color in the picture, and the percentage of lines using the turquoise color. While the difference in these features was not statistically significant, the fact that out of the 11 most contributing features, 10 were related to color, is very unlikely under the null hypothesis that the use of these features is uncorrelated with the state of depression (p-value=0.00012 using exact Fischer test).

After performing the drawing task, we asked the users to answer a 13 question usability questionnaire. Some questions focused on the simplicity/complexity of use of the drawing tool while other questions asked how relaxing the subjects found the different activities. Another set of questions focused on their preference between the drawing assessment and the questionnaire. We used a 5 point Likert scale (2=strongly-agree, 0=neutral,-2=strongly-disagree)

Users found the drawing tool easy to use (score=1.01, p-value<0.0000001, exact binomial test). When asked if they will need technical support to use this tool, they expressed strong disagreement (score=-1.48, p-value<0.001, exact binomial test). Similar answers were obtained when asked about other usability issues. Moreover, on all these questions, participants that were suffering from signs of depression consistently found the tool easier to use. While the differences are not statistically reliable, on some questions they are trending toward significance.

We found similar patterns when we asked people about their preference between the drawing assessment and the questionnaire. For example, when we asked "If I had to choose between filling out a questionnaire or a drawing assessment I would prefer the drawing assessment" the average score of the result was 0.95 (agree, p-value<0.001, exact binomial test). Again, subjects with depression gave higher scores than those with no signs of depression. Similar results were found when we asked if the drawing task was relaxing.

Out of the 13 statements in our usability questionnaire, users did not express a significant preference for only one statement: “"I found filling out the questionnaire assessment relaxing”.

Out of the 13 statements, 4 statements probed on the preference between the two assessment methods. We combined them together to form a scale measuring preference for the drawing assessment. Using this scale we conclude that users prefer the art therapy assessment over the questionnaire (score=0.59). Participants with signs of depression expressed even stronger preference for the art-therapy assessment (p-value<0.03 using chi-square test).

DISCUSSION

Our study shows that there is great interest in a digital form of art-therapy. While the digital art-therapy assessment was not sufficient to screen for depression, participants found it to be a useful affective intervention. While the assessment worked as an intervention in general, it was significantly more effective for participants with signs of depression.

We also saw that the wide availability of touch screen computers allows delivering art-therapy interventions and assessment using the internet, which allows scaling up to the large population of people facing mental disorders, such as depression. Therefore, designers of interventions for mental support applications should consider art-therapy as a source of effective interventions.

REFERENCES

[1] V. Lowenfeld, The Nature of Creative Activity, New York, NY: Harcourt, Brace, 1939.

[2] L. Gantt and C. Tabone, Formal elements art therapy scale: The rating manual, Morgntown, WV: Gargoyle Press, 2012.

[3] E. Gilad, "Computerized FEATS: An empirical study," Antioch University, Seattle, 2015.

[4] James, G., Witten, D., Hastie, T., & Tibshirani, R, An introduction to statistical learning, NY, New York: Springer, 2013.

[5] L. Breiman, "Random forests," *Machine Learning,* vol. 45, no. 1, pp. 5-32, 2001.

[6] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Annals of Statistics,* vol. 29, no. 5, pp. 1189-1232, 2001.

[7] Korlakai Vinayak, R., & Gilad-Bachrach, R., "DART: Dropouts meet Multiple Additive Regression Trees," in *The Eighteenth International Conference on Artificial Intelligence and Statistics*, 2015.

[8] I. Jolliffe, Principal component analysis, New York, NY: Springer Velag, 2002.

[9] Arbisi, Paul A.; Ben-Porath, Yossef S., "An MMPI-2 infrequent response scale for use with psychopathological populations: The Infrequency-Psychopathology Scale," *Psychological Assessment,* vol. 7, no. 4, pp. 424-431, 1995.

[10] . J. Biggs, . L. Wylie and . V. Ziegler, "Validity of the Zung Self-rating Depression Scale," *The British Journal of Psychiatry,* pp. 381-385, 1978.