

# Young Adults' Instagram Posts and Depressive Moods: A Study in Mexico in the Wild

Iván A. Encinas-Monroy<sup>1</sup>, Jessica Beltrán<sup>2</sup>, Luis H. Sánchez<sup>3</sup>, Luis Felipe-Rodríguez<sup>1</sup>,  
Adrián Macías<sup>1</sup>, Cynthia B. Pérez<sup>4</sup>, Manuel Domitsu<sup>1</sup>, Luis A. Castro<sup>1,\*</sup>

<sup>1</sup> Instituto Tecnológico de Sonora, Departamento de Computación y Diseño,  
Mexico

<sup>2</sup> Universidad Autónoma de Coahuila, Centro de Investigación en Matemáticas Aplicada,  
Mexico

<sup>3</sup> Instituto Politécnico Nacional,  
Centro de Investigación y Desarrollo de Tecnología Digital,  
Mexico

<sup>4</sup> Instituto Tecnológico de Sonora, Unidad Guaymas,  
Mexico

ivanalex@msn.com, jessicabeltran@uadec.edu.mx,  
lh.sg26@gmail.com, {luis.rodriguez, adrian.macias, cynthia.perez,  
manuel.domitsu}@itson.edu.mx, luis.castro@acm.org

**Abstract.** Patterns of use of social networking sites like Instagram can be indicators of the mental state of users. Of particular interest to the HCI community are those markers and patterns useful for inferring the mental health of users experiencing depressive episodes or moods. Detecting individuals' depressive moods through their typical Instagram activity remains a challenge due to the diversity of the content posted. Previous research often focuses on retrieving content of hashtags related directly to depression for analysis. Thus, although based on real posts, results can be highly biased. Analyzing all user posts in individuals' day-to-day lives can yield ecologically valid findings, but it is challenging. We conducted an observational study aimed at detecting the depressive moods of users from their Instagram posts. We analyzed text, images, and posting behavior using two approaches: inferential statistics and machine learning. Our results indicate that the time of day and the hue levels of a posted image could lead to the detection of depressive moods. Furthermore, our machine-learning approach yielded up to 65% of

accuracy. Although our study yields ecologically valid findings, several challenges remain to be addressed due to the heterogeneity of the dataset, as it typically happens in real-world studies.

**Keywords.** Social networking sites, depressive mood detection, Instagram, machine learning, behavior analysis, image analysis, text analysis, transfer learning.

## 1 Introduction

Depression is one of the leading causes of disability around the world, affecting about 300 million people [42]. Detecting and treating depression in young people is therefore paramount. Young people tend to seek help and support on social networking sites (SNS) through externalization [16, 2], where they share how they feel about certain situations or topics.

In these situations, their friends and relatives can help them by showing support or care [3].

Instagram posts describing antidepressant use have increased exponentially from 2010 to 2018 [14]. Instagram acknowledges the importance of this topic by showing a message when a user searches for images related to certain keywords such as depressed (see Figure 1).

Users have reported five primary social and psychological motives for using Instagram: social interaction, archiving, self-expression, escapism, and peeking [22]. Images and text shared in posts can be analyzed to unravel patterns that can signal the presence of depression, such as the preferences of colors [6], certain topics through images or captions [2, 3], the use of certain words [38], image filters [31], or explicitly expressing depressive symptoms [20].

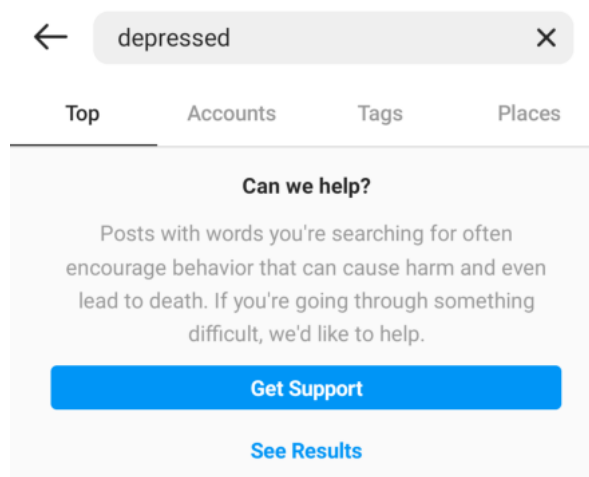
Social media markers have been reported as a valid way to detect depression [20, 31]. On Instagram, these markers significantly differ between depressed people and nondepressed people, which include the number of Instagram followers, frequency of Instagram use or content of messages, and filters [20, 31].

A qualitative analysis of depression-related posts on Instagram revealed different types of disclosures individuals make in the context of depression-tagged posts [3]. In [17], authors found that people with depressive symptoms are more likely to prefer the use of Twitter over Instagram and Facebook.

Also, in [24] was reported that more frequent Instagram use had associations with greater depressive symptoms when users reported a high proportion of strangers followed. Finally, negative social comparison has been one of the reasons for depressive symptoms [24, 26].

With regards to Facebook, having few Facebook friends and mutual friends, posting frequently, and using few location tags are positively correlated with depressive symptoms [20]. Nonetheless, other works report that Facebook can work as a protective factor against depressive symptoms [17].

In this work, we focus on the in-the-wild usage of Instagram to study depressive moods since it



**Fig. 1.** Screenshot of the Instagram mobile app when looking for images related with keyword depressed

is an SNS that has been popular with younger generations.

In addition, Instagram's nature of promoting oneself and telling others about what is happening during your day contrasts with other discussion-oriented SNSs (e.g., Twitter) [28], which makes Instagram a suitable platform to analyze users' content and investigate whether it can reveal their moods.

As opposed to previous works that use surveys [24, 23, 22, 17, 26, 30, 33] or analyze posts filtered by depression-related hashtags such as #depression or #depressed [2, 3, 14, 29, 1], in this work we developed a tool to collect Instagram posts of young adults (i.e., our participants) and ask them to answer the PANAS-X inventory each time they posted.

Next, we analyzed those images, text, and posting behavior to associate them with depressive moods as users go by in their daily lives. We used two approaches: inferential statistics and machine learning.

Analyzing in-the-wild posts can be challenging since there may be posts that are completely unrelated to the posters' moods. In this regard, our study yields ecologically valid results and understandings of how SNS are used in real life [4, 7].



Fig. 2. Snapshots of the web-based app (In spanish)

## 2 Related Work

The interest in identifying users' internal states through Instagram posts has increased in the last decade [24, 23, 22, 17, 26, 30, 33, 2, 3, 14, 29]. Observational studies typically use self-report data to find associations between psychological inventories and behavior [24, 23, 22, 17, 26, 30, 33]. Other works, however, have focused on analyzing Instagram posts seeking behavior patterns from images, text, and emoticons. In this section, we describe works that focus on the analysis of text, images, and posting behavior in SNS.

### 2.1 Text Analysis in SNS

Text analysis tools and techniques have been increasingly used to get insights into the users' internal states (e.g., mood and emotions), or other psychological traits [38, 18]. Text

analysis has been particularly used to explore and predict different mental disorders through posts on SNSs [9, 11].

Previous works have shown that certain keywords in SNS posts can be used to identify individuals with mental disorders. For instance, these types of posts include hashtags such as #depression, #anxiety or #suicide, among other words that may be related to mental disorders such as the names of antidepressants [14] or the name of the disorder itself [9, 10, 27].

For instance, in [10] it was proposed a lexicon of depressed users on Twitter and found that some recurrent themes were related to symptoms (e.g., anxiety, withdrawal, severe, delusions), disclosure (e.g., fun, play, helped, god), treatment (e.g., medication, side-effects, doctor, doses), and relationships (e.g., home, woman, she, him). Authors from [34] showed that college students with depression often use more personal singular pronouns at the moment of writing. Although



**Fig. 3.** Random sample of 50% of the images posted by individuals who were categorized in the minimal, mild, moderate, and severe classes. Images have been intentionally blurred for privacy

this study was carried out with written essays, these findings could potentially be extrapolated to SNS posts.

Another approach often used is sentiment analysis since keywords and words used by users with mental disorders are often charged with negative emotions [11, 43]. For example, it has been reported that some words such as issues, bad, or anxiety could be used to predict the jump from depression to suicidal thoughts [11].

Some of these words had high frequency like other less negative ones such as make, around, time, when, where, and others. However, this could raise some concerns as people might not be completely honest on SNS by purposely undermining their own negative feelings so others

do not feel bad for them [12] or by expressing more positive emotions than they are actually experiencing since it might attract more attention to the post [37].

Text analysis per se can be challenging, but it can be more difficult if analysis is carried out without further context. For a more precise interpretation of the users' moods, more data associated with the users' moods at the time of posting is required.

## 2.2 Image Analysis on SNS

The content of shared images can include data about the user's interests and potentially about their mental state or mood. Examples of these

**Table 1.** Face data extracted by Google cloud vision API grouped by Beck's class

Beck's Class	Users (N)	Images (N)	Faces (N)	Faces (Max N)	Faces (Mean)	Positive Emotion (Mean)	Negative Emotion (Mean)
Minimal	17	151	138	10	0.9139	25.0066	14.7615
Mild	5	65	33	10	0.5076	18.0461	12.1538
Moderate	7	53	67	8	1.2641	27.0377	20.4905
Severe	6	56	55	10	0.9821	18.6428	21.6250

**Table 2.** Face data extracted by Google cloud vision API grouped by the PANAS-X main categories

PANAS-X Category	Images (N)	Faces (N)	Faces (Max N)	Faces (Mean)	Positive Emotion (Mean)	Negative Emotion (Mean)
Positive	180	184	10	1.0222	28.2722	13.5888
Negative	45	35	8	0.7777	15.3777	29.0222
Neutral Positive	44	32	6	0.7272	20.4772	20.6818
Neutral Negative	56	42	6	0.7500	13.2857	11.6785

types of content include the number of faces (i.e., individuals) in the image, the predominant color, and the types of objects, among others.

The content of Instagram images has been analyzed and linked to different aspects of users. For example, when studying the relationship between personality traits and gender, and the images posted, researchers found a link between extraversion and gender of users [21].

Likewise, [36] extracted objects using Microsoft Azure Cognitive Services with the aim of classifying images into thousands of categories, such as car, city, interior, and others, and used them to determine the age and gender of users. The relationship between mental health and shared photos has also been studied.

For example, in [27] studied the relationship between several visual attributes of the images such as color, themes or emotions, and self-disclosures of Instagram users related to their mental health. Images shared on SNSs have also been used for detecting depression.

In [31] analyzed the content of images and data such as the number of posts per day or the number of likes to compare nondepressed and depressed individuals. Features such as the number of faces

in the photos and color properties were related to depression.

However, the dataset entries were observed and labeled through crowdsourcing with Amazon's Mechanical Turk, which could have introduced a bias since emotions can be interpreted differently by third parties.

Moreover, the user who posted the image might have a salient emotion at the moment of posting, which could have been missed by annotators since they are only looking at the image.

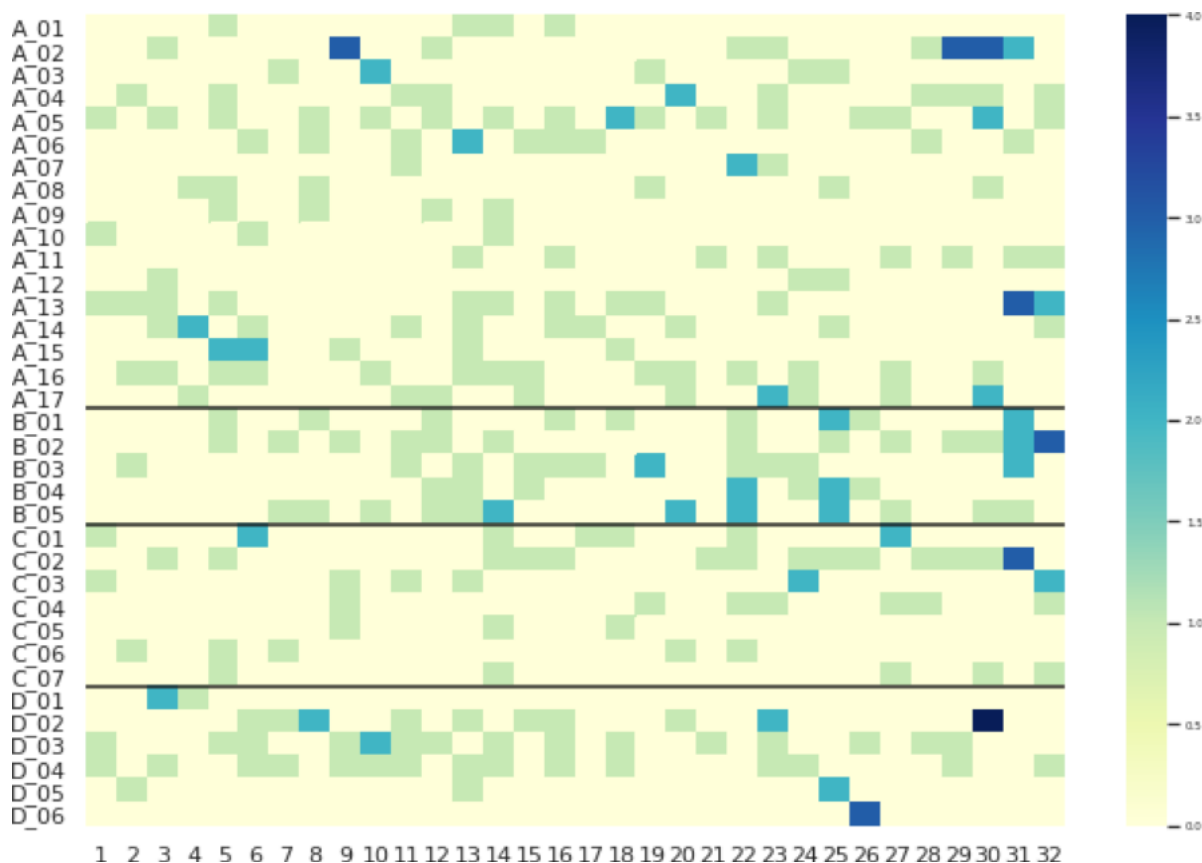
Then, having third parties annotate the images of others can be a challenge to derive adequate findings, since different cultures or experiences can shape or bias the annotations.

In another study, [46] used a deep regression network (deemed *DepressNet*) to analyze faces since they might indicate a depressive disorder. However, apart from selfies, users on Instagram typically post different types of images, such as landscapes, artwork, or pets, which makes it difficult to deploy in the wild.

Finally, in [8] used multimodal data from Instagram posts, including the content of images, text, and user's behavior, to detect users with depressive moods. For the image analysis,

**Table 3.** Number of images with animals and sketches grouped by the Beck's class

Beck's Category	Animals	Sketches	% of Total Images by Beck's Category
Minimal	14	16	19.87
Mild	10	12	33.85
Moderate	4	2	11.32
Severe	9	2	19.64
Total	37	32	84.68



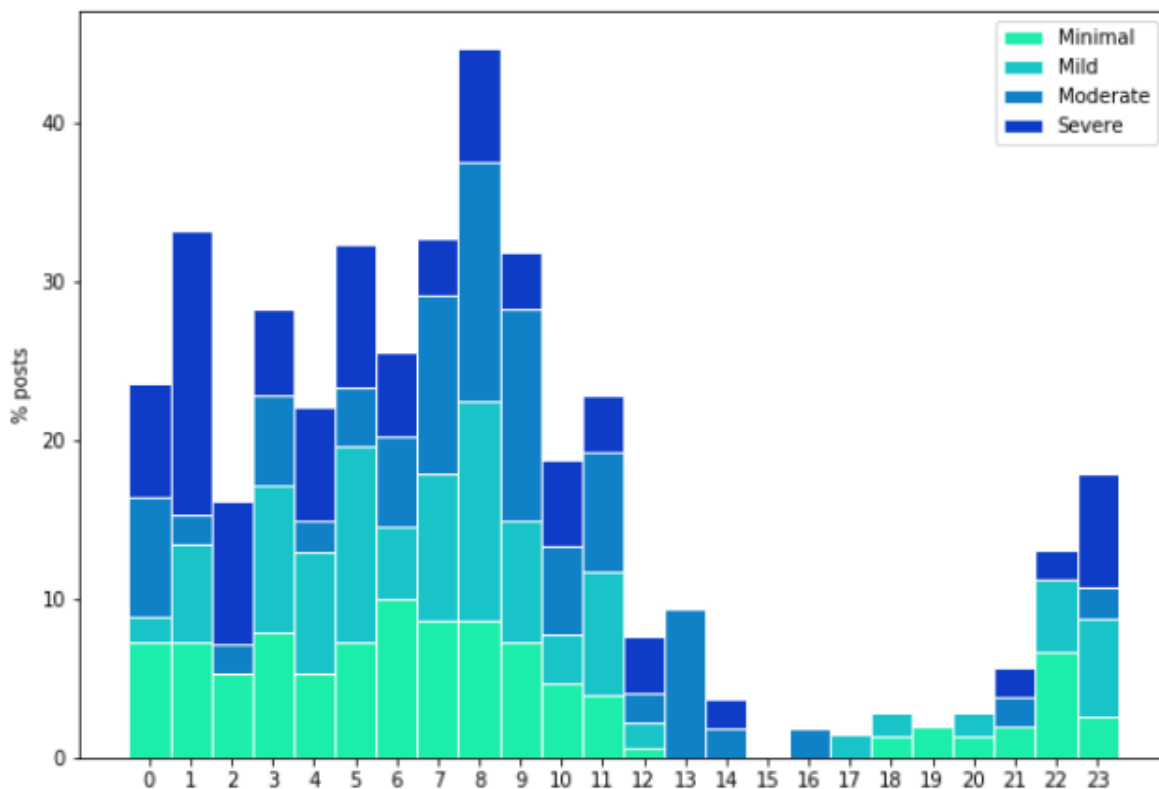
**Fig. 4.** Posting activity from all users throughout the duration of the study (x-axis = days; y-axis = participants)

used the AlexNet Convolutional Neural Network for transfer learning to get a prediction score of depressive images.

Afterward, they merged individual predictions for image, text, and behavior for an overall prediction. Although the dataset was obtained from real users, they retrieved the dataset by searching

specific depression keywords on Instagram users' profiles, which could have biased the results toward individuals who self-describe as depressed.

In general, there is a need for conducting studies that analyze data coming from individuals with their typical behavior on SNS, i.e., ecologically valid findings. Moreover, research must consider a



**Fig. 5.** Posts binned per the hour of day (x axis = time of day; y axis = number of posts)

wider range of posts from users [29] as opposed to selecting posts with specific hashtags (e.g., #depression), which can bias the results and our understanding of the manner in which these aspects take place in the real world.

### 3 Methods

We carried out an observational study to collect users' Instagram posts from which we analyzed text, images, and posting behavior (i.e., time of post).

In this section, we describe the participants, research procedure, dataset, and data preprocessing.

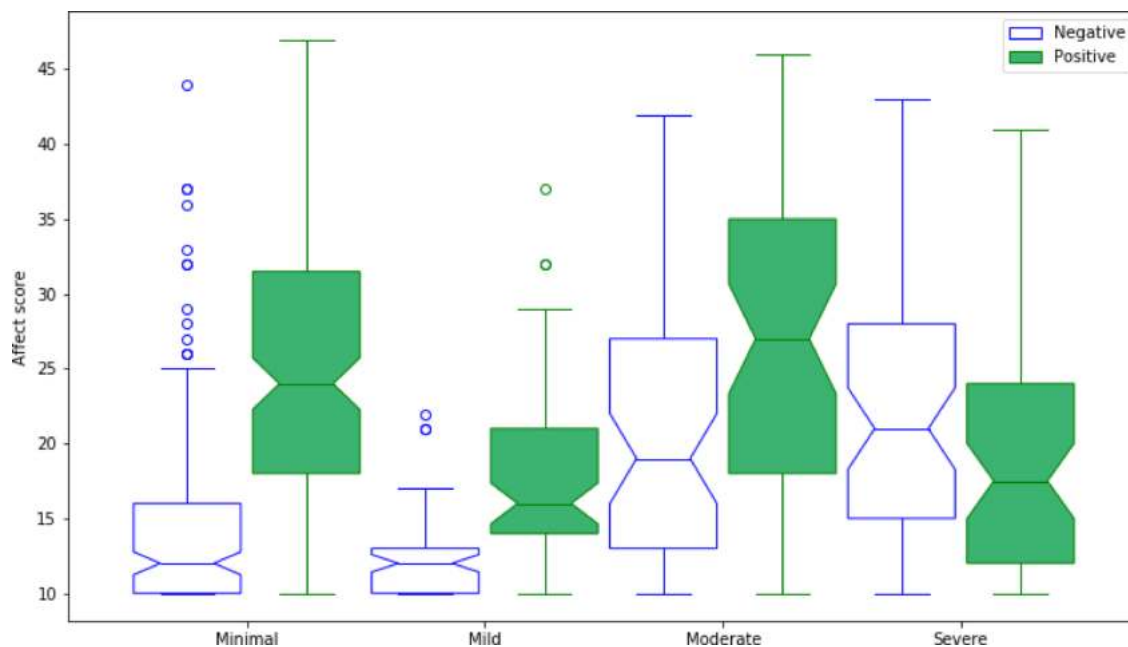
#### 3.1 Participants

We used a convenience sampling method to recruit participants. The invitations to participate in the study were sent through electronic media such as WhatsApp or Facebook Messenger.

We recruited 50 individuals from Northwest Mexico, from which 35 participants (13 male) remained until the end of the study.

Our participants were, on average, 23.51 years old ( $SD = 3.36$ ), ranging 19-40 years old. Sixteen (16) of them (45.71%) were university students.

34% of the participants said they were regular users of Instagram. All our participants were native speakers of Spanish. All participants signed an informed consent. No monetary incentive was given.



**Fig. 6.** General negative affect (GNA) and general positive affect (GPA) from PANAS-X across the Beck's classes

### 3.2 Instruments

We used the following instruments to obtain data in the wild.

#### 3.2.1 PANAS-X

We used the validated Spanish version of the PANAS-X [32], which has 46 items to measure the positive affect and negative affect using a 5-point Likert-like scale (1 = Lightly or nothing; 5 = Always).

The original English version consists of 60 items [41]. The PANAS-X shows two different kinds of categories for both ends of the valence spectrum: General Positive Affect (GPA) and General Negative Affect (GNA), as well as the Basic Positive Affect (BPA) and Basic Negative Affect (BNA). GPA and GNA are directly related to the results of the more commonly used PANAS [41], as they are composed of the same items.

Their basic counterparts are composed of different kinds of items that are only present in PANAS-X and reach other kinds of emotions like fear, sadness, guilt, hostility, joviality, self-assurance, and attentiveness.

The rationale behind PANAS-X's positive and negative affect is that people are able to feel both kinds of emotions at the same time in the same high or low intensity levels.

It is possible to detect intense joy or happiness while also detecting strong feelings of sadness or anger through these questionnaires. Such cases can be related to being confused about what one can feel about certain situations.

#### 3.2.2 Beck's Depression Inventory

Beck's depression inventory (BDI) [35] is a Spanish version of a psychological test used to evaluate the depressive symptomatology of people.

The BDI has 21 items designed to assess the severity of symptoms of depression in adults and adolescents. The BDI has a score range of 0 to 63, depending on the option selected by the person.

This score helps researchers and health professionals categorize the level of depression according to people's symptoms in order to identify its intensity or evaluate its therapeutic progress.



**Table 4.** Pearson's R correlation between PANAS-X affect scores, and SentiStrength and Google AutoML

		<b>SentiStrength Positive Affect</b>	<b>SentiStrength Negative Affect</b>	<b>Google AutoML Natural Language Value</b>
GPA	r	0.037	-0.004	0.101
	N	325.0	325.0	325.0
	p	0.256	0.474	<b>0.035</b>
GNA	r	-0.069	-0.086	-0.108
	N	325.0	325.0	325.0
	p	0.106	0.061	<b>0.025</b>

**Table 5.** Contingency table showing the number of posts per week

<b>Beck's Class</b>	<b>Week 1</b>	<b>Week 2</b>	<b>Week 3</b>	<b>Week 4</b>	<b>Week 5</b>	<b>Total</b>
Minimal	32	36	27	28	28	<b>151</b>
Mild	5	16	10	21	13	<b>65</b>
Moderate	10	9	8	16	10	<b>53</b>
Severe	13	16	8	12	7	<b>56</b>
<b>Total</b>	<b>60</b>	<b>77</b>	<b>53</b>	<b>77</b>	<b>58</b>	<b>325</b>

### 3.2.3 Web-Based App for Data Collection

We developed a web-based app (Figure 2) to retrieve the users' latest posts using their Instagram (IG) handle. Every time the users posted on IG, our web-based app retrieved the posted image, text, and date.

Below the retrieved image from the IG post, the app displayed the PANAS-X questionnaire so that our participants could rate the types of emotions felt at the moment of posting that particular image.

This strategy differs from previous approaches since it provides not only the data linked to the post (e.g., image, text, timestamp) but also about their emotions at the time of posting.

### 3.3 Data Collection Protocol

Thirty-five individuals participated in this study who received a demographic questionnaire and Beck's depression inventory (BDI) in Spanish [35], which they responded to online without supervision.

First, we explained to the participants the general purpose of the study. Also, we asked them about their frequency of use of IG and IG handle. Of the 35 participants, 17 obtained a Beck score that falls in the category of minimal depression

(score 0-13), 5 in mild depression (score 14-19), 7 as moderate depression (score 20-28), and 6 with severe depression (score 29 or more). The procedure was as follows:

- For 32 days, participants had to use Instagram as they would typically use it. We suggested our participants post 4 times per week, although this was not compulsory.
- After each post, they were asked to answer the PANAS-X that corresponded to that publication using our Web-based app.
- In the event that the participant posted and did not answer the corresponding PANAS-X after a few hours, one of the authors sent a reminder via the WhatsApp or Instagram messaging service.

## 4 Data Preprocessing and Feature Extraction

A total of 325 entries were posted by the 35 participants. On average, each user posted 9.28 times (SD = 4.94) throughout the duration of the study. From the 325 images, 151 were posted by participants in the category of minimal depression,

**Table 6.** Contingency table showing the number of posts per time of day

Beck's class	Morning	Afternoon	Night	Total
Minimal	85	5	61	<b>151</b>
Mild	44	2	19	<b>65</b>
Moderate	35	7	11	<b>53</b>
Severe	27	1	28	<b>56</b>
<b>Total</b>	<b>191</b>	<b>15</b>	<b>119</b>	<b>325</b>

**Table 7.** Results with the 16-day time window

Classifier	Accuracy	Precision	Recall	F1-Score
SVM	0.65	0.65	0.55	0.55
RF	0.62	0.49	0.66	0.46
LR	0.65	0.65	0.55	0.55

according to Beck's depression inventory, 65 by participants in the mild class, 53 by those in the moderate class, and 56 by participants who were categorized as severely depressed. Figure 3 shows a random sample of the images posted by users from all depression categories.

From the 325 entries, 46 entries consisted of image-only publications. From these, 18 entries came from participants categorized in the minimal class, 19 from participants in the mild class, 2 from participants in the moderate class, and 7 from those classified as severely depressed, according to Beck's depression inventory.

From the 279 posts that included text, they included a mean text description of 11.59 words (SD = 19.91). Each text record consisted of the following: post ID (Integer), IG handle (string), timestamp (Integer), text description (string), image URL (string), and type of post (string: carousel, video, or image).

In summary, the dataset consisted of 325 image files, 279 text records, and 325 46-tuple vectors, i.e., one PANAS-X answer per post. To analyze the content of posts, we extracted features by preprocessing the data using state-of-the-art tools.

#### 4.1 Text Processing

For studying the link between the text of an IG post and the emotions reported by the participants, we used two different tools that identify the

general emotion from the given text and also provide additional information that can be related to depression, such as the amount of singular or plural pronouns, which has been reported to be relevant in depressed students [34].

The first tool was the Google AutoML Natural Language [13], which delivers the magnitude and value score of the identified emotion in the text, pronoun count, first-person pronouns, first-person singular pronouns, plural pronouns, and first-person plural pronouns.

The second tool used was the Spanish version of SentiStrength [40], original version by [39], which delivers the negative and positive scores from the identified emotion in the text. In the case of SentiStrength, we removed emoticons since the tool is unable to detect them and can only interfere the analysis.

Finally, we also computed the number of characters and words, the ratio of the number of pronouns over the number of words, the ratio of first-person pronouns over total pronouns, and the ratio of plural pronouns over total pronouns. In total, we obtained 14 features from the text.

#### 4.2 Image Processing

To extract features from the image dataset, we used the state-of-the-art Automated Machine Learning (AutoML) Vision by Google Cloud Platform<sup>1</sup>, which is an implementation of AutoML for image classification and object detection.

It consists of an Application Programming Interface (API) that offers machine learning models that assign labels and detects objects in images, and it can also be used to train personalized models of machine learning. In total, we obtained 9 features from images. Color perception has been suggested as a marker of mood [5], where grayer and darker colors are related to depressive moods.

We included the dominant Red (R), Green (G), and Blue (B) colors of each photo provided by the Google Cloud service. Also, the levels of hue and saturation have been of interest to researchers as they could possibly indicate levels of sadness or depression [31].

<sup>1</sup>cloud.google.com/

**Table 8.** Results with the 7-day (week) time window

Classifier	Accuracy	Precision	Recall	F1-score
SVM	0.62	0.61	0.53	0.55
RF	0.51	0.47	0.43	0.43
LR	0.59	0.58	0.59	0.58

**Table 9.** IG dataset content

Class	Training	Validation
Predominantly depressive	138 images	35 images
Barely depressive	121 images	30 images

Drawing on this, the mean values of Hue, Saturation, and Value (HSV) were retrieved from the images through the scikit-image library<sup>2</sup>.

We obtained relevant object labels from images that can be used to relate to a depressive mood. For example, the label face, since according to [reece2017instagram, the number of faces reflects a greater social interaction, which can be related to less tendency to depression. Table 1 and Table 2 show data related to faces and their emotions as detected by Google Cloud Vision API.

Since the dataset features images with humans, pets, landscapes, drawings or anything the user felt like posting, we also obtained labels animal, if there were animals in the photo, and sketch, since we identified that several images were drawings, cartoons or sketches. Table 3 shows the data with this type of content.

### 4.3 Behavior Data Processing

Mental status can be related to posting behavior, such as the time or frequency of posting activity. In Figure 4, we show a heatmap with the posting activity from all our participants.

A darker color indicates more publications during each day. For the purposes of this figure, we grouped the users by Beck's class in the Y axis (A<sub>xx</sub> = minimal depression, B<sub>xx</sub> = mild, C<sub>xx</sub> = moderate, D<sub>xx</sub> = severe).

Figure 5 shows the time of day in which participants posted the most, grouped by the participants' severity of depression.

<sup>2</sup>scikit-image.org/

It can be seen that there is a sharp increase of activity from 10 PM to 5 AM in our participants, particularly those categorized as severely depressed. Most posts occurred from midnight to noon across all groups. From behavior processing, we obtained 3 features: date, time, and time between consecutive posts in minutes.

## 5 Results

In this section, we present the results of analyzing data collected from the observational study to detect depressive moods of users from their Instagram posts using two approaches:

1. Inferential statistics.
2. Machine learning techniques.

### 5.1 Inferential Statistics Approach

#### 5.1.1 Relating the Severity of Depression and the Emotions Linked to a Post

Detecting depressive moods in IG posts where there are no clear signals of sadness, such as gloomy pictures or certain keywords, can be challenging. For this, we first need to identify the types of emotions related to a particular post, i.e., the emotions experienced by the participants at the moment of posting.

For this, we used t-student tests to explore the relationship between the level of depression as approximated by the Becks' inventory (administered at the beginning of the study), and the general positive/negative emotions through the PANAS-X (administered each time the user posted).

As shown in Figure 6, for the General Positive Affect (GPA), participants with depressive moods (i.e., mild, moderate, and severe) had lower scores (GPA = 20.98, SD = 8.84) when compared with users with minimal depressive moods (GPA = 25.00, SD = 9.22), which was statistically significant ( $t = 4.01$ ,  $df = 312$ ,  $P < 0.0001$ ).

As for the General Negative Affect (GNA), participants with depressive moods had a significantly higher score (GNA = 17.74, SD = 6.51) when compared with users with

**Table 10.** MART dataset content across the two classes

Class	Paintings
Negative	183 images
Positive	317 images

minimal depressive moods (GNA = 14.76, SD = 8.09), which was statistically significant ( $t = -3.67$ ,  $df = 321$ ,  $P < 0.0001$ ). Also, participants with severe depression had a mean GPA of 18.64 (SD = 7.30), while the rest of the classes had an average GPA score of 23.72 (SD = 9.35) ( $t = -4.49$ ,  $df = 97$ ,  $P < 0.0001$ ).

On the other hand, participants with severe depression also reported higher scores of the GNA (mean = 21.62; SD = 8.41) when compared with the average GNA (mean = 15.26; SD = 6.87) of the rest of the participants ( $t = 5.3$ ,  $df = 71$ ,  $P < 0.0001$ ). This suggests that the PANAS-X helps discriminate negative and positive emotions associated with a particular post. Even more, these feelings seem to concur with Beck's classes.

### 5.1.2 Text Analysis

We obtained the Pearson's R correlation between the reported PANAS-X's GPAs and GNAs and the scores from the SentiStrength and Google AutoML sentiment analysis tools (Table 4). As shown in Table 4, there is a positive correlation between the GPA and the Google AutoML Natural Language Value ( $r = 0.101$ ,  $N = 325$ ,  $p = 0.035$ ).

Also, there is a negative correlation between Google AutoML Natural Language Value and the GNA ( $r = -0.108$ ,  $N = 325$ ,  $p = 0.025$ ), contrary to what was expected. Although weak, these correlations show that there is a relationship between the GPA and the GNA and the valence detected by Google AutoML and that there is an association between the affect related to a particular post and the affect of the text.

Also, SentiStrength's results were not statistically significant, but it is interesting that with both tools, the GNA was slightly more correlated with the text than the GPA, which could potentially suggest that people are more expressive about their negative feelings rather than their positive ones. Still, this is inconclusive.

### 5.1.3 Image Analysis

One of our interests is to better understand certain markers that describe depressive moods in Instagram posts. [31] reported that people with depression are more interested in "bluer" or "darker blue" images in comparison with nondepressed. Since hue levels are directly related to the blue color of an image.

We conducted an analysis of images posted by our participants. The hue levels from the images posted by participants with depressive moods (mean = 0.40, SD = 0.09) were higher than those with minimal depressive moods (mean = 0.37, SD = 0.09).

The mean difference was 0.03, which was statistically significant ( $t = -2.13$ ,  $df = 316$ ,  $p = 0.017$ ). This means that participants with depression generally posted images that were slightly "bluer". Also, there were fewer faces in the photos posted by participants with depressive moods (mean = 0.89, SD = 1.71) than in those by nondepressed (mean = 0.91, SD = 1.44). The mean difference was 0.02, which was not significant ( $t = -0.13$ ,  $df = 323$ ,  $p = 0.45$ ).

### 5.1.4 Behavior Analysis

As shown in Figure 4, there are slight differences between the groups. The group in the mild class seems to have more activity than the rest of the groups. A chi-square test was used to explore the association between the severity of depression and the number of posts per week. Table 5 shows the contingency table. There was no significant association between the severity of depression of the participant and the number of posts per week ( $X^2 = 13.35$ ,  $df = 12$ ,  $p = 0.34$ ).

Previous works have reported signals of possible posting patterns in depressed individuals [10]. That is, depressed individuals tend to be more active at night than non-depressed ones. For this analysis, we binned posts into three times of day (morning: 4:00 AM - 12:59 PM, afternoon: 1:00 PM - 7:59 PM, night: 8:00 PM - 3:59 AM). The chi-square test was used to explore the association between the severity of depression (i.e., Beck's class) and the number of posts per time of the day.

**Table 11.** IG and MART: Merged dataset

Class	Training	Validation
Predominantly Depressive	285 images	71 images
Barely Depressive	373 images	94 images

Following Table 6, there was a significant difference between Beck's class and the time of day at the moment of posting ( $X^2 = 20.85$ ,  $df = 6$ ,  $p < 0.01$ ). Therefore, the evidence collected suggests that posting time is associated with Beck's class in which our participants were categorized.

## 5.2 A Classic Machine Learning Approach

We used a classic machine learning approach to classify the participants' Beck's class based solely on the data shared in their posts. For this, we represented each participant with a vector that we used for training models. Afterward, we evaluated the models on unseen vectors. Since we have a small dataset, we used N-fold cross-validation for the evaluation.

### 5.2.1 Target Classes

We used binary target classes. For this, we combined mild, moderate, and severe Beck's classes into a single class deemed as Predominantly depressive and the other remaining class minimal deemed as Barely depressive.

This arrangement categorized 17 subjects within the class Barely depressive and 18 within the class Predominantly depressive. However, we removed the data from participants A\_01, A\_07, A\_09, A\_10, A\_12, C\_05, D\_01, D\_05, and D\_06 since they posted infrequently and spent about two weeks without posting, making feature vectors invalid in all these cases.

This exclusion leads to 12 subjects in Barely depressive class and 14 in Predominantly depressive class.

### 5.2.2 Feature Vectors

We preprocessed each post and obtained the 26 features described in Section 4 such as magnitude and value of text emotion, count of pronouns, number of faces in the image, time of posting, and others. Since a single post is not enough to characterize a person with depression, we aggregate data from several posts to obtain a 52-tuple vector composed of the Mean and Standard Deviation (SD) per feature.

We run experiments with two different time windows to aggregate features and construct the vectors. The first time window comprised 16 days, which produced 50 vectors. Lastly, a 7-day time window resulted in 116 vectors. We performed a correlation analysis across the vector elements, and dropped those highly correlated.

For instance, the SD of the number of characters had a 99% correlation with the SD of the number of words, so we only kept one of those. The correlation threshold to drop features was set to 90%, thus the feature vector size was reduced to 37.

### 5.2.3 Binary Classification

We used the normalized 37-feature vectors to train machine learning models using the Python Scikit-learn library<sup>3</sup>. We run experiments with the following algorithms: Support Vector Machines (SVM), Random Forest (RF), and Logistic Regression (LR) using 10-fold cross-validation. We run a hyperparameter grid search in all algorithms, and we report the best results.

The hyperparameter search space was {C: 0.1, 1, 10, 30, 40, 50; Gamma: 0.5, 1, 5, 10} for SVM, {C: 0.001, 0.1, 1, 10, 30, 40, 100; Penalty: "l1", "l2"; Solver: "liblinear", "saga"} for LR, and {emax depth: 30, 50, 60, 80, 100; minimum samples per leaf: 2, 3, 5, 8; minimum samples per split: 4, 7, 8, 10; number of estimators: 60, 80, 90, 100, 150} for RF.

Additionally, we evaluated using different top k features that resulted from the feature selection process during cross-validation. We next compared the results using Accuracy, Precision,

<sup>3</sup>scikit-learn.org/

**Table 12.** IG and Art photo: Merged dataset

Class	Training	Validation
Predominantly Depressive	459 images	142 images
Barely Depressive	404 images	125 images

Recall, and F1-score evaluation metrics per 16 days (Table 7) and 7-day (Table 8). It can be seen that SVM and LR outperform RF. In the case of a 7-day time frame, SVM had higher Accuracy and Precision but lower Recall than LR.

The best features for the 16-day window (Table 7) were text Google value score average, word count in first person voice average, animal flags average, value of the image, red intensity average, green intensity, and blue intensity average.

The best hyperparameters for the 16-day window were  $C=30$  and  $\gamma = 0.001$  for SVM; max depth=40, max features=1, and n estimators=80 for RF; and  $C=0.01$ , penalty=12, and solver=liblinear for LR. In addition, the best features for the 7-day window (Table 8) were the text Google value score average, word count in first person voice average, pronouns ratio average, animal flags average, and value of the image.

The best hyperparameters for the 7-day window were:  $C=10$  and  $\gamma = 5$  for SVM; max depth=20, max features=4 and n estimators=100 for RF; and  $C=1$ , penalty=11 and solver= liblinear for LR. We can see that features that produce better results are similar across all classifiers. In fact, the features “animal flags average” and “value” remain relevant across all classifiers.

### 5.3 A Deep Learning Approach

We also utilized deep learning (DL) algorithms to learn about the potential to discriminate between depressed and nondepressed individuals based on images. We used the same aforementioned classes: Predominantly depressive and Barely depressive. Due to a small dataset, we used transfer learning, which is a DL approach where learned features in a pre-trained model with larger datasets are transferred to a second network with other target tasks and data.

The transfer learning approach has proven to be powerful when there are data in a target network [45]. The base model used for the transfer learning was pre-trained with the ImageNet dataset<sup>4</sup>, which contains more than 14 million images within 1, categories.

The architecture used is ResNet50 [15], from which the weights of the pre-trained model were transferred to our target classification task, removing the output layer of ResNet50 and adding the connected layers of classification by image. We used the library Tensorflow<sup>5</sup> for the experiments. We performed four different experiments in this DL approach.

For the 4 experiments, we used transfer learning with the model with ResNet50 architecture. The key difference among the 4 experiments was the datasets used. Since we have a small dataset, we sought to augment it with similar datasets labeled from people with depression or labeled as positive or negative images.

As mentioned, our participants shared images with artistic content and drawings besides photos of people and landmarks. We found datasets with these types of content that are labeled as positive and negative, which we next describe.

In total, we used 3 datasets for these four experiments (E1, E2, E3, E4), merging some of them for the experiments: a) the dataset we collected (i.e., IG dataset); b) a dataset from abstract paintings from the MART museum from Italy (i.e., MART dataset); c) and a dataset from artistic images showing emotions (i.e., Art photo dataset). We next list the experiments:

- E1: ResNet50 with IG
- E2: ResNet50 with IG + MART
- E3: ResNet50 with IG + Art photo
- E4: ResNet50 with IG + MART + Art photo

During these experiments, the training across the 4 experiments was made through 45 epochs. We tested with dropout regularization to avoid overfitting.

<sup>4</sup>[www.image-net.org/](http://www.image-net.org/)

<sup>5</sup>[www.tensorflow.org/](http://www.tensorflow.org/)

**Table 13.** IG, MART, and Art Photo: Merged dataset

Class	Training	Validation
Predominantly Depressive	597 images	188 images
Barely Depressive	641 images	203 images

In the following sections, we describe each of the following experiments with more details. The results of those will be shown in the following pages.

### 5.3.1 E1: ResNet50 with IG

As mentioned, we used 2 classes: Predominantly depressive and Barely depressive. For E1, the IG dataset was split into training data (80%) and validation data (20%). The data distribution was done as shown in Table 9.

### 5.3.2 E2: ResNet50 with IG + MART

In [44] applied a methodology to classify 500 abstract paintings from the MART museum from Italy. They used 2 classes: Positive perception and Negative perception.

They performed a statistical classification in which 100 participants reported their first impression of the abstract painting on a 1 – 7 scale (1 = Negative perception; 7 = Positive perception).

Authors defined those with average scores lower or equal to 4 as the negative class and those with average scores above 4 as the positive. The distribution of classes is shown in Table 10.

In E2, we merged the IG dataset with the dataset created by [44], where the samples from the Negative class were merged with the samples from the Predominantly depressive class.

Similarly, the samples in the Positive class were merged with the samples of the Barely depressive class. Table 11 shows the resulting dataset split into two: training (80%) and validation (20%).

### 5.3.3 E3: ResNet50 with IG + Art Photo

Authors from [25] proposed a methodology to extract emotional features of images, like color, composition, hue, and others, with the aim of classifying images using the presented emotion. They used a dataset with 806 artistic images.

These images were obtained through an image hosting website using the search terms: Amusement, Awe, Contentment, Excitement as positive emotions, and Anger, Disgust, Fear, Sad to represent negative emotions.

The images were created by artists who sought to awaken a specific emotion to the viewer through the color manipulation, illumination, composition, and so forth.

Like the previous example, we merged these images with the IG dataset, taking the images with negative class as Predominantly depressive and the images with positive class as Barely depressive. The total of images with configuration of 80% for training and 20% for validation is shown in Table 12.

### 5.3.4 E4: ResNet50 with IG + MART + Art Photo

The last dataset used is a combination of the three datasets previously described. This resulted in a new dataset with images close to those an average user can post on Instagram (i.e., real, artistic, or abstract images). The total of images with a configuration of 80% training and 20% validation is shown in Table 13.

### 5.3.5 Classification Results with DL

Figure 7 and Figure 8 show the training and validation accuracy from the four experiments. As we can see, the results are lower than 0.60 accuracy across all the experiments in both the training and validation sets. This experiment provides an understanding of the potential of using the algorithms in this problem using all the data for training and validation.

For further exploration, we trained and validated using only the MART dataset, not including the Instagram data. The accuracy results are 1 in training and 0.75 in validation, showing overfitting but, most importantly, showing that adding the

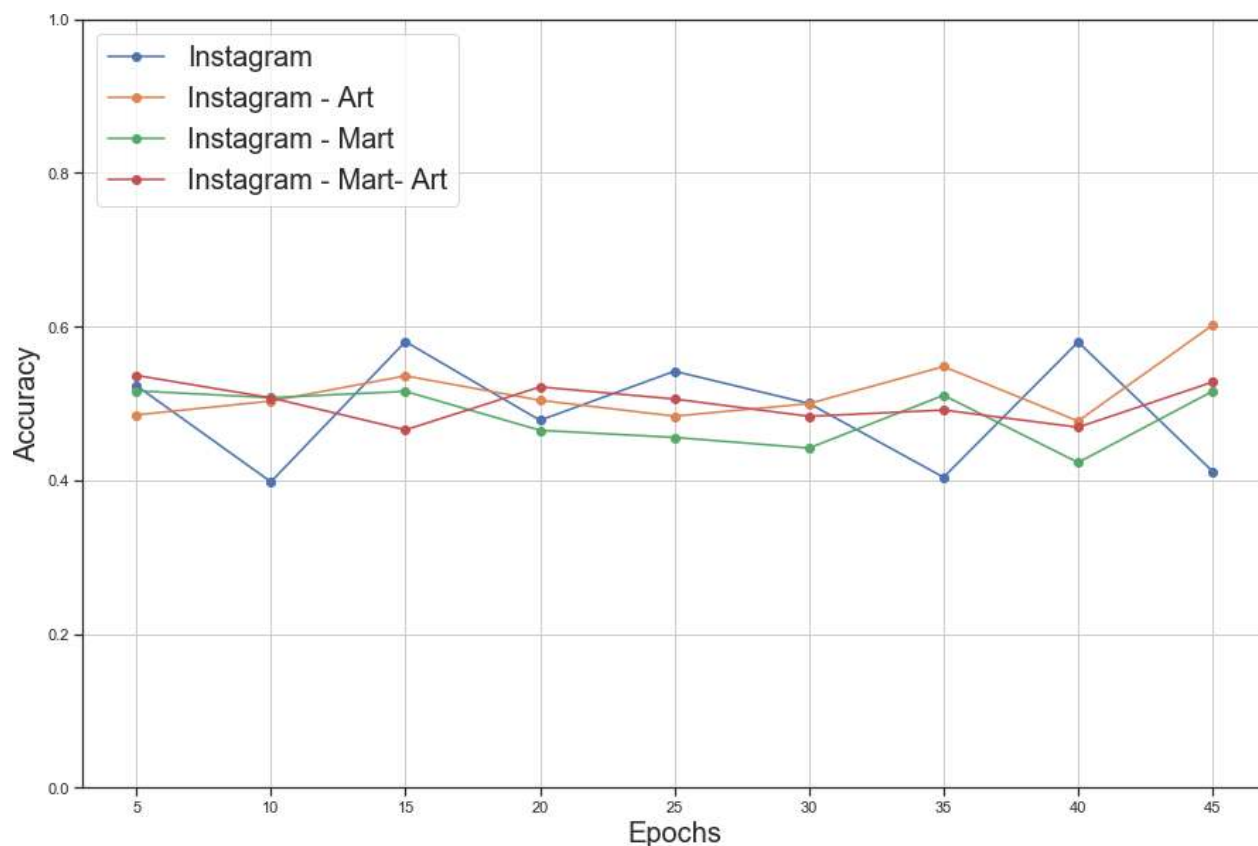


Fig. 7. Training accuracy through 45 epochs

Instagram datasets significantly decreased results. This can be caused due to different reasons, such as the image distributions from both datasets being different or the labels positive or negative not aligning well with the classes from the Beck's depression inventory.

In general, just like most available datasets, both the MART dataset and the Art photo dataset were labeled a posteriori by third parties and do not properly reflect the emotions when a photo is taken or posted by an individual.

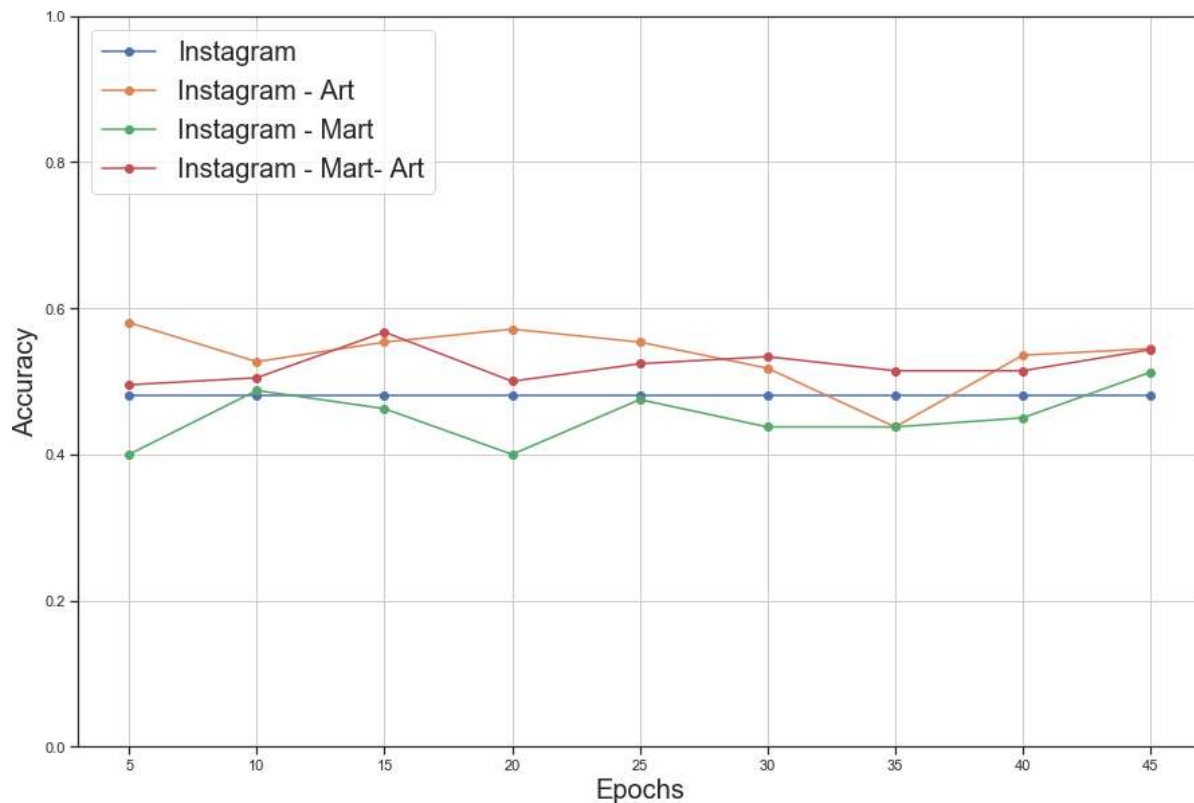
On the contrary, our Instagram dataset was collected in the wild, where our participants were categorized according to the Beck's depression inventory, who might be privately dealing with their condition, showing subtle signs through their posts while not being truly explicit about it. The results obtained through deep learning are not conclusive.

We believe that these results can be improved by combining more information from the posts, such as the image captions. However, we wanted to show these results, which can illustrate the difficulty of this problem. More studies of this kind are needed to deepen the understanding of this problem and the way the DL can help tackle it.

## 6 Discussion

There are several aspects that can be of interest to the HCI community, and that can be used in real-world applications for monitoring patients who have been diagnosed with certain disorders, such as depressive disorders. First, from our results, we can see that the PANAS-X can help identify emotions related to a particular post, having found that those with depressive moods have a higher





**Fig. 8.** Validation accuracy through 45 epochs

GNA and a lower GPA than those who barely have depressive moods. Also, as expected, those with prominent depressive moods (i.e., severe depression) have a higher GNA and lower GPA when compared with all other Beck's classes.

Even when this aspect was asked every time a participant posted an image on Instagram, these emotions seemed to be generalized as they spanned across the entire set of participants, i.e., there was an actual difference between the groups.

As for the images published, we found that participants who have prominent depressive moods generally posted images that are bluer, that is, the prominent underlying color is blue.

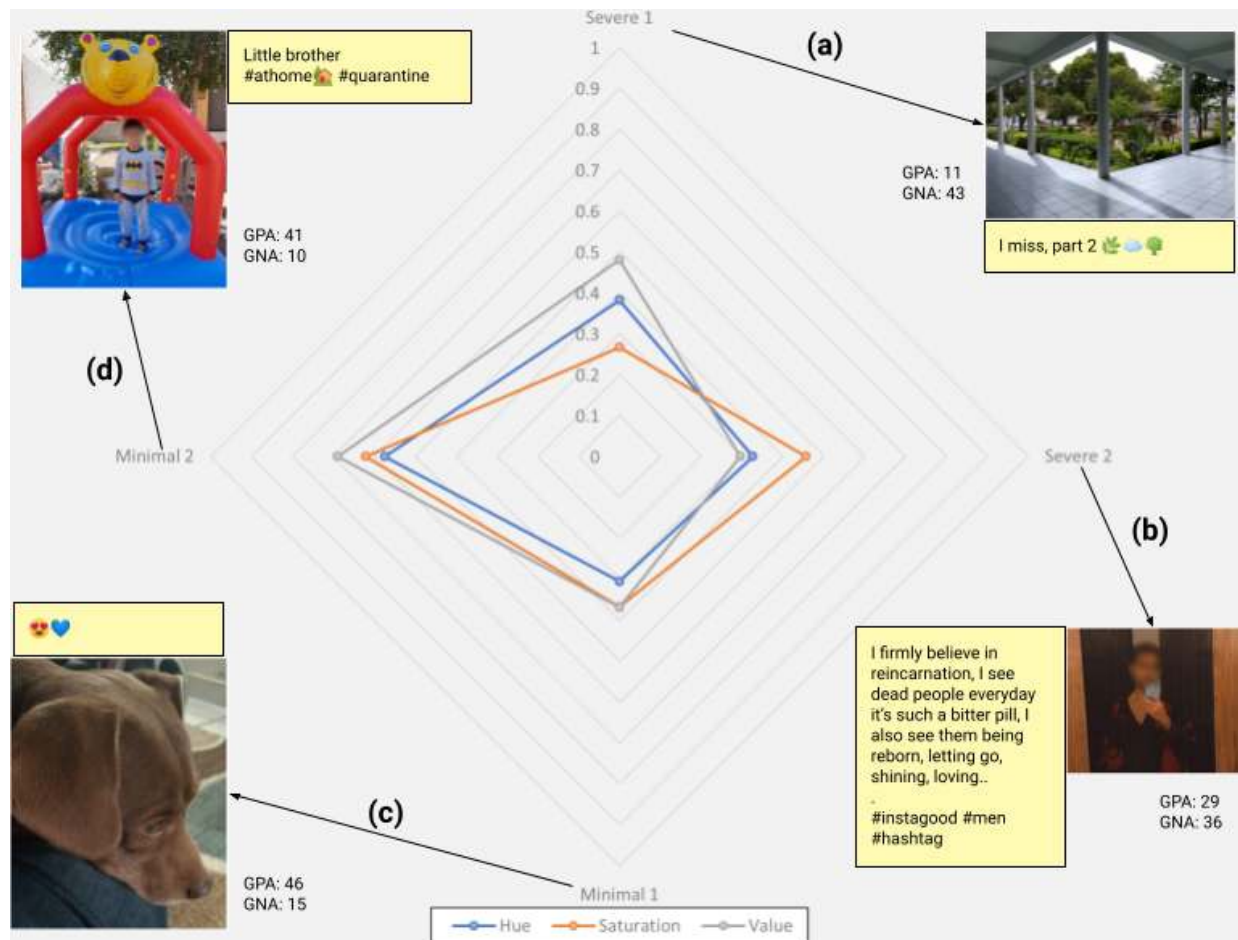
This concurs with other studies that report that blue or grayscale colors are preferred by depressed individuals [6]. The results from our study help us understand that these findings can also be replicated in Mexican young adults (i.e., our participants).

Although this is not conclusive due to statistical power, our results suggest that this is an interesting avenue to explore. Also, from the posting behavior, the number of posts by our participants (i.e., posting behavior) did not vary across all the study between groups.

The time of day, however, seems to be a good discriminant between those with prominent depressive behaviors and those who were classified as minimally depressed.

The obtained results with machine learning are higher than chance, which proves the feasibility of using these types of algorithms to be potentially used to detect depression with reservations.

However, better ways to discriminate the types of posts are paramount for designing and training appropriate machine learning models. Possibly, for this particular problem, we may need multi-modal classification from data coming from various sources, rather than focusing only on IG data.



**Fig. 9.** Posts from four participants: (a) Severe 1: GPA=11, GNA=43; (b) Severe 2: GPA=29, GNA=36; (c) Minimal 1: GPA=46, GNA=15; (d) Minimal 2: GPA=41, GNA=10. The maximum value for GPA and GNA is 50. Faces have been blurred for privacy

Still, our results are valuable in that is one of the first studies trying to understand whether in-the-wild behaviors such as the use of IG can be used to infer certain behaviors of interest.

### 6.1 Ecological Validity

The in-the-wild nature of this study can be double-edged. On the one side, the findings derived from the data can be ecologically valid [4, 7], since the dataset we collected is similar to what young adults are posting on Instagram in their day-to-day lives.

Using hashtagged posts can cause filter bias during analysis [29]. To avoid this, in this work, we considered all types of posts by users during a four-week window, which can be difficult for algorithms to discriminate due to the noise derived from heterogeneous content published by users.

One of the challenges in this work is the diversity in the data across various levels of depression, as categorized by Beck's depression inventory. For instance, an image shared by a person with severe depression might be similar in terms of colors or lighting to an image shared by a person without any particular sign of depression.

At the same time, people with the same severity of depression might share unrelated or opposite images. To illustrate the heterogeneity of the dataset, Figure 9 shows the HSV levels and GPA/GNA data related to four posts from different users who were categorized either as minimal or severe, according to Beck's depression inventory.

We also show the actual images, and the caption associated, which have been translated to English by the authors. Following, Figure 9a and Figure 9b, we can see that Hue and Value (ie., brightness) are similar, but not in terms of Saturation. Also, both GPA are relatively low (GPAa=11, GPAb=29) and both GNA are high (GNAa=43, GNAb=36).

We want to highlight that both images were posted by participants categorized as severely depressed. Interestingly, these images yielded similar values to Figure 9c in terms of hue, saturation, and value (HSV). In other words, images Figure 9a and Figure 9b look alike to their polar opposite, Figure 9c.

Presumably, one expected type of image with high GNA and low GPA from our severely depressed participants is Figure 9c, which is dark and essentially colorless. This image strongly differs from Figure 9d, even when both have similar values of GPA (GPAc=46; GPAd=41) and GNA (GNAc=15, GNAd=10). Taking all of these aspects into account, using machine learning approaches for classifying these sorts of images is not trivial.

In most related work, the challenge associated with using complete datasets from the day-to-day lives of users has not been accounted for since datasets generally originate from homogeneous groups (e.g., online communities for depression support) or commonly used hashtags on SNS (e.g., #depression).

Since posts can have similar characteristics stated from the moment that they are posted, much more homogeneous data within target classes can be obtained, and higher variations between classes can have a significant positive impact on the performance of machine learning approaches. As we have seen in this work, having a more diverse dataset has a negative impact on the performance, especially due to the size of the dataset.

## 6.2 Limitations and Scope

One limitation of the present study is the quantity of data collected. For the statistical and machine learning approaches, it is highly beneficial to have large datasets (in this case, posts and users) to be able to yield better results. Therefore, this study could have benefited from a larger sample of users over a longer period.

## 7 Conclusion

We presented an analysis of posts collected in the wild from users of Instagram. From our results, we can conclude that identifying depressive moods from Instagram posts can be challenging since participants typically post about their inner states as much as about their interests, preferences, hobbies, or even memes.

Images do not necessarily relate to feelings or emotions, but may also be associated with situations and interpersonal strategies (e.g., social status). The results of this work can be summarized as follows: some behaviors can be potentially used to discriminate depressed from nondepressed users, such as the time of posting and the hue color of the images.

The in-the-wild nature of this study yielded ecologically valid results, but further user context could be useful for adequate results in classification. As seen when merged with other datasets, our dataset noisy.

Future work includes collecting more data from more participants and over a longer period, performing experiments with different combinations of information gathered from posts such as text, images, and perhaps additional context such as filters or location.

Also, we suggest creating subcategories of images such as people, locations, pets, or other predefined categories, which could help increase classification performance, but with additional overhead.

## Acknowledgments

We thank our participants for their valuable time and effort. This work was partially funded with a scholarship provided to the first and third authors, and the Instituto Tecnológico de Sonora through the PROFAPI program.

## References

1. **Ahmed, A., Aziz, S., Toro, C. T., Alzubaidi, M., Irshaidat, S., Serhan, H. A., Abd-Alrazaq, A. A., Househ, M. (2022).** Machine learning models to detect anxiety and depression through social media: A scoping review. *Computer Methods and Programs in Biomedicine Update*, Vol. 2, pp. 100066. DOI: 10.1016/j.cmpbup.2022.100066.
2. **Andalibi, N., Ozturk, P., Forte, A. (2015).** Depression-related imagery on instagram. *Proceedings of the 18th ACM Conference Companion on Computer Supported Cooperative Work and Social Computing*, pp. 231–234. DOI: 10.1145/2685553.2699014.
3. **Andalibi, N., Ozturk, P., Forte, A. (2017).** Sensitive self-disclosures, responses, and social support on instagram: The case of #depression. *Proceedings of the ACM Conference on Computer Supported Cooperative Work and Social Computing*, pp. 1485–1500. DOI: 10.1145/2998181.2998243.
4. **Andrade, C. (2018).** Internal, external, and ecological validity in research design, conduct, and evaluation. *Indian Journal of Psychological Medicine*, Vol. 40, No. 5, pp. 498–499. DOI: 10.4103/ijpsym.ijpsym\_334\_18.
5. **Barrick, C. B., Taylor, D., Correa, E. I. (2002).** Color sensitivity and mood disorders: biology or metaphor?. *Journal of affective disorders*, Vol. 68, No. 1, pp. 67–71. DOI: 10.1016/S0165-0327(00)00358-X.
6. **Carruthers, H. R., Morris, J., Tarrier, N., Whorwell, P. J. (2010).** The Manchester color wheel: Development of a novel way of identifying color choice and its validation in healthy, anxious and depressed individuals. *BMC Medical Research Methodology*, Vol. 10, No. 1. DOI: 10.1186/1471-2288-10-12.
7. **Carter, S., Mankoff, J., Klemmer, S. R., Matthews, T. (2008).** Exiting the cleanroom: On ecological validity and ubiquitous computing. *Human-Computer Interaction*, Vol. 23, No. 1, pp. 47–99. DOI: 10.1080/07370020701851086.
8. **Chiu, C. Y., Lane, H. Y., Koh, J. L., Chen, A. L. P. (2020).** Multimodal depression detection on instagram considering time interval of posts. *Journal of Intelligent Information Systems*, Vol. 56, No. 1, pp. 25–47. DOI: 10.1007/s10844-020-00599-5.
9. **Coppersmith, G., Harman, C., Dredze, M. (2014).** Measuring post traumatic stress disorder in twitter. *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 8, No. 1, pp. 579–582. DOI: 10.1609/icwsm.v8i1.14574.
10. **De-Choudhury, M., Gamon, M., Counts, S., Horvitz, E. (2013).** Predicting depression via social media. *Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media*, Vol. 36, No. 1–2, pp. 168–169. DOI: 10.3109/01460862.2013.798190.
11. **De-Choudhury, M., Kiciman, E., Dredze, M., Coppersmith, G., Kumar, M. (2016).** Discovering shifts to suicidal ideation from mental health content in social media. *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pp. 2098–2110. DOI: 10.1145/2858036.2858207.
12. **Dibble, J. L., Levine, T. R. (2013).** Sharing good and bad news with friends and strangers: Reasons for and communication behaviors associated with the mum effect. *Communication*

- Studies, Vol. 64, No. 4, pp. 431–452. DOI: 10.1080/10510974.2013.770407.
13. **Google (2017)**. AutoML for large scale image classification and object detection. [ai.googleblog.com/2017/11/automl-for-large-scale-image.html](https://ai.googleblog.com/2017/11/automl-for-large-scale-image.html).
  14. **Gupta, R., Ariefdjohan, M. (2020)**. Mental illness on instagram: a mixed method study to characterize public content, sentiments, and trends of antidepressant use. *Journal of Mental Health*, Vol. 30, No. 4, pp. 518–525. DOI: 10.1080/09638237.2020.1755021.
  15. **He, K., Zhang, X., Ren, S., Sun, J. (2016)**. Deep residual learning for image recognition. *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778. DOI: 10.1109/cvpr.2016.90.
  16. **Hunt, J., Eisenberg, D. (2010)**. Mental health problems and help-seeking behavior among college students. *Journal of Adolescent Health*, Vol. 46, No. 1, pp. 3–10. DOI: 10.1016/j.jadohealth.2009.08.008.
  17. **Jeri-Yabar, A., Sanchez-Carbonel, A., Tito, K., Ramirez-del-Castillo, J., Torres-Alcantara, A., Denegri, D., Carreazo, Y. (2018)**. Association between social media use (twitter, instagram, facebook) and depressive symptoms: Are twitter users at higher risk?. *International Journal of Social Psychiatry*, Vol. 65, No. 1, pp. 14–19. DOI: 10.1177/0020764018814270.
  18. **Jianqiang, Z., Xiaolin, G. (2017)**. Comparison research on text pre-processing methods on twitter sentiment analysis. *IEEE Access*, Vol. 5, pp. 2870–2879. DOI: 10.1109/access.2017.2672677.
  19. **Kiersz, A., Allan, A. (2019)**. Suicide is gen Z's second-leading cause of death, and it's a worse epidemic than anything millennials faced at that age. [www.businessinsider.com/cd-c-teenage-gen-z-american-suicide-epidemic](http://www.businessinsider.com/cd-c-teenage-gen-z-american-suicide-epidemic).
  20. **Kim, J., Uddin, Z. A., Lee, Y., Nasri, F., Gill, H., Subramanieapillai, M., Lee, R., Udovica, A., Phan, L., Lui, L., Iacobucci, M., Mansur, R. B., Rosenblat, J. D., McIntyre, R. S. (2021)**. A systematic review of the validity of screening depression through facebook, twitter, instagram, and snapchat. *Journal of Affective Disorders*, Vol. 286, pp. 360–369. DOI: 10.1016/j.jad.2020.08.091.
  21. **Kim, Y., Kim, J. H. (2018)**. Using computer vision techniques on instagram to link users' personalities and genders to the features of their photos: An exploratory study. *Information Processing and Management*, Vol. 54, No. 6, pp. 1101–1114. DOI: 10.1016/j.ipm.2018.07.005.
  22. **Lee, E., Lee, J. A., Moon, J. H., Sung, Y. (2015)**. Pictures speak louder than words: Motivations for using Instagram. *Cyberpsychology, Behavior, and Social Networking*, Vol. 18, No. 9, pp. 552–556. DOI: 10.1089/cyber.2015.0157.
  23. **Lin, L. Y., Sidani, J. E., Shensa, A., Radovic, A., Miller, E., Colditz, J. B., Hoffman, B. L., Giles, L. M., Primack, B. A. (2016)**. Association between social media use and depression among U.S. young adults. *Depression and Anxiety*, Vol. 33, No. 4, pp. 323–331. DOI: 10.1002/da.22466.
  24. **Lup, K., Trub, L., Rosenthal, L. (2015)**. Instagram #Instasad?: Exploring associations among instagram use, depressive symptoms, negative social comparison, and strangers followed. *Cyberpsychology, Behavior, and Social Networking*, Vol. 18, No. 5, pp. 247–252. DOI: 10.1089/cyber.2014.0560.
  25. **Machajdik, J., Hanbury, A. (2010)**. Affective image classification using features inspired by psychology and art theory. *Proceedings of the 18th ACM International Conference on Multimedia*, pp. 83–92. DOI: 10.1145/1873951.1873965.
  26. **Mackson, S. B., Brochu, P. M., Schneider, B. A. (2019)**. Instagram: Friend or foe? The application's association with psychological well-being. *New Media and Society*, Vol. 21, No. 10, pp. 2160–2182. DOI: 10.1177/1461444819840021.

27. **Manikonda, L., De-Choudhury, M. (2017).** Modeling and understanding visual attributes of mental health disclosures in social media. *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pp. 170–181. DOI: 10.1145/3025453.3025932.
28. **Manikonda, L., Meduri, V. V., Kambhampati, S. (2021).** Tweeting the mind and instagramming the heart: Exploring differentiated content sharing on social media. *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 10, No. 1, pp. 639–642. DOI: 10.1609/icwsm.v10i1.14819.
29. **McCosker, A., Gerrard, Y. (2020).** Hashtagging depression on instagram: Towards a more inclusive mental health research methodology. *New Media and Society*, Vol. 23, No. 7, pp. 1899–1919. DOI: 10.1177/1461444820921349.
30. **Pittman, M., Reich, B. (2016).** Social media and loneliness: Why an instagram picture may be worth more than a thousand twitter words. *Computers in Human Behavior*, Vol. 62, pp. 155–167. DOI: 10.1016/j.chb.2016.03.084.
31. **Reece, A. G., Danforth, C. M. (2017).** Instagram photos reveal predictive markers of depression. *EPJ Data Science*, Vol. 6, No. 1. DOI: 10.1140/epjds/s13688-017-0110-z.
32. **Robles, R., Páez, F. (2003).** Estudio sobre la traducción al español y las propiedades psicométricas de las escalas de afecto positivo y negativo (panas). *Salud Mental*, Vol. 26, No. 1, pp. 69–75.
33. **Rozgonjuk, D., Pruunsild, P., Jürimäe, K., Schwarz, R. J., Aru, J. (2020).** Instagram use frequency is associated with problematic smartphone use, but not with depression and anxiety symptom severity. *Mobile Media and Communication*, Vol. 8, No. 3, pp. 400–418. DOI: 10.1177/2050157920910190.
34. **Rude, S., Gortner, E. M., Pennebaker, J. (2004).** Language use of depressed and depression-vulnerable college students. *Cognition and Emotion*, Vol. 18, No. 8, pp. 1121–1133. DOI: 10.1080/02699930441000030.
35. **Sanz, J., Perdigón, A. L., Vázquez, C. (2003).** Adaptación española del inventario para la depresión de Beck-II (BDI-II): 2. Propiedades psicométricas en población general. *Clínica y Salud*, Vol. 14, No. 3, pp. 249–280.
36. **Song, J., Han, K., Lee, D., Kim, S. W. (2018).** “Is a picture really worth a thousand words?”: A case study on classifying user attributes on Instagram. *PLoS One*, Vol. 13, No. 10, pp. e0204938. DOI: 10.1371/journal.pone.0204938.
37. **Stieglitz, S., Dang-Xuan, L. (2013).** Emotions and information diffusion in social media—sentiment of microblogs and sharing behavior. *Journal of Management Information Systems*, Vol. 29, No. 4, pp. 217–248. DOI: 10.2753/mis0742-1222290408.
38. **Tausczik, Y. R., Pennebaker, J. W. (2009).** The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, Vol. 29, No. 1, pp. 24–54. DOI: 10.1177/0261927x09351676.
39. **Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., Kappas, A. (2010).** Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology*, Vol. 61, No. 12, pp. 2544–2558. DOI: 10.1002/asi.21416.
40. **Vilares, D., Thelwall, M., Alonso, M. A. (2015).** The megaphone of the people? Spanish SentiStrength for real-time analysis of political tweets. *Journal of Information Science*, Vol. 41, No. 6, pp. 799–813. DOI: 10.1177/0165551515598926.
41. **Watson, D., Clark, L. (1999).** The PANAS-X: Manual for the positive and negative affect schedule-expanded form. Psychology Publications. DOI: 10.17077/48vt-m4t2.
42. **World Health Organization (2021).** World mental health day. [www.who.int/news-room/fact-sheets/detail/depression](http://www.who.int/news-room/fact-sheets/detail/depression).

43. **Yang, Y., Jia, J., Zhang, S., Wu, B., Chen, Q., Li, J., Xing, C., Tang, J. (2014).** How do your friends on social media disclose your emotions?. Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 28, No. 1. DOI: 10.1609/aaai.v28i1.8740.
44. **Yanulevskaya, V., Uijlings, J., Bruni, E., Sartori, A., Zamboni, E., Bacci, F., Melcher, D., Sebe, N. (2012).** In the eye of the beholder: Employing statistical analysis and eye tracking for analyzing abstract paintings. Proceedings of the 20th ACM International Conference on Multimedia, Association for Computing Machinery, pp. 349–358. DOI: 10.1145/2393347.2393399.
45. **Yosinski, J., Clune, J., Bengio, Y., Lipson, H. (2014).** How transferable are features in deep neural networks? Vol. 2, pp. 3320–3328.
46. **Zhou, X., Jin, K., Shang, Y., Guo, G. (2020).** Visually interpretable representation learning for depression recognition from facial images. IEEE Transactions on Affective Computing, Vol. 11, No. 3, pp. 542–552. DOI: 10.1109/taffc.2018.2828819.

*Article received on 11/01/2023; accepted on 15/03/2024.*

*\* Corresponding author is Luis A. Castro.*