Bipolar Mania Eye Image Classification

Jessica Wheeler  
*Southern Methodist University, jswheeler@smu.edu*

Jean Jecha  
*Southern Methodist University, jjecha@smu.edu*

Manjula Kottegoda  
*Southern Methodist University, mkttegoda@smu.edu*

Sharon Teo  
*Southern Methodist University, steo@smu.edu*

Eric C. Larson  
*Southern Methodist University, eclarson@lyle.smu.edu*

Follow this and additional works at: [https://scholar.smu.edu/datasciencereview](https://scholar.smu.edu/datasciencereview)

Part of the [Mental and Social Health Commons](https://scholar.smu.edu/datasciencereview), and the [Social and Behavioral Sciences Commons](https://scholar.smu.edu/datasciencereview)

Recommended Citation

Wheeler, Jessica; Jecha, Jean; Kottegoda, Manjula; Teo, Sharon; and Larson, Eric C. (2018) "Bipolar Mania Eye Image Classification," *SMU Data Science Review*, Vol. 1 : No. 1 , Article 1. Available at: [https://scholar.smu.edu/datasciencereview/vol1/iss1/1](https://scholar.smu.edu/datasciencereview/vol1/iss1/1)

This Article is brought to you for free and open access by SMU Scholar. It has been accepted for inclusion in SMU Data Science Review by an authorized administrator of SMU Scholar. For more information, please visit [http://digitalrepository.smu.edu](http://digitalrepository.smu.edu).
Mania in the Eyes

Jean Jecha¹, Manjula Kottegoda¹, Sharon Teo¹, Jessica S. Wheeler¹, Eric C. Larson¹, and Julie A. Fast²

¹Southern Methodist University, Dallas, Texas
²Bipolar Consultant, BP Magazine author

{jjecha@smu.edu, mkottegoda@smu.edu, steo@smu.edu, jswheeler@smu.edu, eclarson@lyle.smu.edu, juliefastcoaching@gmail.com}

Abstract. In this paper, we present a novel method to detect mania in bipolar individuals that utilizes pictures of the individual’s eyes. Both mania and depression can be detrimental to the individual and in some cases, can be life threatening. It is often difficult for an individual to identify that they are in a state of hypomania or mania. To understand if automated methods for detecting mania are possible, we created a study website and curated a dataset of eye images from individuals with bipolar disorder. These images were labeled as either manic, depressed or stable, where manic images were further broken down into euphoric or dysphoric. Using convolutional neural networks, random forest, and logistic regression, we were able to produce models detecting euphoric mania with accuracy rates averaging between 23% and 73%. Although our results may have promising accuracies, further investigation of the precision and recall values indicated overfitting in the models.

1 Introduction

Bipolar disorder is believed to be a genetic brain disorder that affects a person’s ability to self-regulate their mood. Individuals diagnosed with bipolar have mood swings involving both lows (bipolar depression) and highs (hypomania or mania) [1]. Extreme highs are referred to as full-blown mania, where a person feels very “up” or “high” and may engage in high-risk behaviors [12]. Hypomania, on the other hand, is less intense and involves having more energy, but is not as severe of a mood swing as full-blown mania. For the rest of this paper we will use the word mania to reference both hypomania and mania. Along with depression and mania, symptoms of anxiety, restlessness, irritation, and focus problems are common in individuals with bipolar disorder [12]. Mania can consist of two types: euphoric and dysphoric. Euphoric mania is a feeling of intense excitement and happiness where everything feels wonderful, beautiful, fantastic, and expansive. Dysphoric mania is commonly associated with extreme feelings of agitation, restlessness, or “wired” activities [14]. Figure 1 summarizes some of the characteristics of mania in general and those of euphoric versus dysphoric mania.
Fig. 1. Venn Diagram showing the shared characteristics of both types of mania that define a manic state as well as the characteristics of euphoric and dysphoric mania.

Bipolar depression may follow a manic episode for many individuals [14]. Thus, early detection of a manic episode is useful for forecasting a possible depressed state. Moreover, mania can involve dangerous, impulsive, pleasure-seeking, and high-risk behaviors that disrupt daily functioning [12]. Common manic behaviors include poor financial investments, sexual indiscretions, shopping sprees, and drug/alcohol abuse—all of which can destroy an individual’s work, school, or social relationships [12]. Early detection of mania can therefore alert an individual to follow a specific management plan. The knowledge that an individual is experiencing a manic state would also be useful for the caregivers, healthcare providers, and even family members to be aware of. However, many bipolar individuals, especially those who experience hypomania, find it difficult to recognize that they are having a manic episode [14]. Therefore, methods other than self-identification are needed to detect mania in individuals. One method of detecting manic episodes comes from anecdotal evidence that changes in the eyes reveal when a bipolar individual is experiencing mania [5]. It is unclear if this anecdotal evidence of changes in the eyes is physically measurable or if individuals are projecting their own conclusions about detecting mania from an individual’s eyes. In this paper, we present a novel methodology for both the collection of eye images of bipolar individuals and the analysis of those images to determine if the individual is experiencing mania.
Following a basic primer on bipolar disorder in Section 2, we discuss related work in Section 3, examining the effectiveness of monitoring mental health using mobile automatic systems as well as studies that analyze eye image data. In Section 4, we introduce our first contribution which is a data collection methodology wherein eye images and survey data are collected while keeping the participants’ identity anonymous. Participants could go to the study’s custom website, hosted by Amazon Web Services (AWS) to crop and upload their photos and indicate their mental state at the time the photo was taken. Manic, stable, and depressed photos were collected, where stable and depressed states were used solely as tools for more effective classification of mania. In a period of under two months, 147 photos were collected, 54% manic, 25% stable, and 21% depressed. Machine learning methods of Keras and Artificial Neural Networks (ANN) were applied as well as more traditional methods such as logistic regression and random forests. We performed four convolutional neural network (CNN) analyses, breaking up the data in different ways and average accuracies ranging from 23% to 73% resulted. When the data was sub-grouped as euphoric mania versus all “other,” the highest accuracies resulted in the fourth CNN model. We used that grouping to create a logistic regression and a random forest model, which resulted in an average accuracy of 70% and 65%, respectively. However, since the dataset was so small and 31% of the images came from one participant, overfitting was very likely. Thus, we constructed confusion matrices that show the numbers of True Positives, False Positives, False Negatives, and True Negatives. Using this information, precision and recall were calculated and revealed that our relatively high accuracies were misleading. Because machine learning algorithms require thousands of observations to work effectively, our results remain inconclusive due to the small volume of our dataset.

2 A Primer on Bipolar Disorder

Bipolar disorder exists on a spectrum with the most common types being bipolar I and bipolar II. The types are defined by the length, frequency, and pattern of episodes of mania and depression [11]. Individuals with bipolar I experience extreme manic episodes as well as periods of depression [11]. Individuals with bipolar II, which is the most common type of bipolar disorder, experience a high frequency of major depression and hypomania [11]. Since bipolar II highs are not as extreme, it can often be misdiagnosed as major depression. Major depression is what mood disorders have in common and is experienced more often than mania in bipolar individuals [13]. An individual with bipolar depression is more likely to have suicidal thinking and behavior [13]. Between 25% and 50% of patients with bipolar disorder attempt suicide at least once [2].

According to the National Institute of Mental Health, bipolar disorder is estimated to affect 5.7 million adult Americans, or about 2.6% of adults age 18 and older [10]. The illness can start at any age from early childhood to as late as 40’s and 50’s with a median age of 25 [10]. Bipolar can affect both men and women equally and is seen in all races, ethnic groups, and social classes [10]. Although bipolar affects men and women equally, research shows that three times as many women than men experience
rapid cycling, wherein an individual goes through four or more manic, hypomanic, or depressed episodes in a year [10].

2.1 Diagnosis and Treatment

Bipolar disorder can often be misdiagnosed because it has a wide range of symptoms and behaviors [14]. Most often the patient or provider doesn’t recognize the highs of mania and patients will usually seek treatment when experiencing major depression [14]. As a result, the patient may be diagnosed with major depression since the provider is not aware of the individual’s mania. Studies have shown that 10% to 25% of individuals experiencing bipolar depression will be misdiagnosed as having major depression [13]. Individuals who are incorrectly diagnosed can receive incorrect treatment for bipolar disorder that can lead to episodes of mania or other problems [13]. Also, individuals may not realize that their highs are not the norm and view the highs as being more productive [14]. Hypomania most often goes unrecognized or unreported since the highs are not as extreme as full-blown mania, thus generating a misdiagnosis for individuals that have bipolar II disorder [11]. Having a method of detecting mania from an individual’s eyes can potentially be another tool aiding in the treatment of bipolar disorder.

The DSM-5 (Diagnostic and Statistical Manual of Mental Disorders) characterizes a manic episode by a distinct and abnormal state of elevated, expansive, or irritable mood occurring for at least one week [12]. Hypomania is characterized by the same abnormal state of elevated, expansive, or irritable mood that lasts at least four consecutive days [12]. As mentioned previously, mania can be detrimental to an individual’s work, school or personal relationships. A person having a manic episode is likely to engage in impulsive, pleasurable, or high-risk behaviors (e.g., over spending, sexual indiscretions, poor financial investments, drug and alcohol abuse) [12]. Individuals having manic episodes do not always recognize their mania so having a means or method of detecting these manic episodes can help the individual significantly in treating their bipolar disorder. Successful treatment of bipolar disorder entails treating both mania as well as depression. The most common treatment for bipolar disorder is a blend of medications and psychotherapy. [12]. Common medications used for the treatment of bipolar disorder would include mood stabilizers, antidepressants, and atypical antipsychotics [12]. In some individuals, antidepressants alone can trigger a manic state and a mood stabilizer will most likely be needed to stabilize their emotions [12]. Along with medications, psychotherapy is another effective tool used in the treatment of bipolar disorder. One of the more common types of psychotherapy is Cognitive Behavioral Therapy that helps the individual to change negative or harmful thoughts. Some therapy treatments are family focused teaching coping strategies, communication and problem-solving methods. Psychoeducation is another tool for teaching the individual with bipolar disorder about their conditions to help identify their trigger points and impending mood swings so that they can seek treatment before reaching a full-blown episode [12]. Having an application that can detects mania would be a great compliment to the treatment plan for bipolar disorder.
3 Related Work

Early detection of changes in mental health for those who are diagnosed with mental illnesses is crucial for effective intervention, treatment, or self-management [15]. With the ubiquitous presence of smartphones, monitoring mental health and well-being using various applications has gained in popularity [16]. One self-management app called FOCUS allows participants to rate their clinical status and functioning and provides illness self-management and interventions accordingly [17]. FOCUS has received positive feedback where participants found the app useful [18], demonstrating the effectiveness of managing mental health with the use of mobile applications. However, there are currently no applications specific to managing mania for bipolar disorder and this paper takes a step in bridging that gap by analyzing the effectiveness of using machine learning to classify eye images according to the type of mania experienced.

Being able to automatically detect behavioral changes that are strong indicators of mental health changes is extremely important especially since those experiencing those changes are usually not aware of their behaviors [20]. Cohn et al. [19] were able to detect depression with 80% accuracy in those undergoing treatment for major depression using automatically measured facial actions and vocal prosody. To capture facial and vocal expression measurements, they used manual FACS, active appearance modeling, and pitch extraction [19]. Their findings show the feasibility of automatic detection for depression as well as the possibilities around automated facial image analysis. Maxuni et al. [20] successfully classified stratified levels of bipolar disorder using speech monitoring. Our goal is to use similar technology to measure mania in the eyes.

An application being developed by Wang et al., called CrossCheck, seeks to provide a way to passively monitor behaviors of those with schizophrenia [15]. Using various features of a smartphone, such as the phone’s accelerometer, GPS/location, microphone, and usage, they were able to track different activities related to indicators of mental health with statistical significance (such as depression, hearing voices, worrying about being harmed, sleep patterns) [15, 3]. Traditionally, surveys are used to assess mental well-being and doctor visits after-the-fact and hospital visits, resulting in late detection [15, 3]. Therefore, the work of Wang et al. [15] has shown that the use of passive tracking in smartphone technology can lead to better healthcare. Perhaps the algorithm presented in this paper can be implemented in a passive monitoring system or application to detect the presence of mania using physical changes in the eyes to pinpoint mania. This type of detection system may be successful in preventing the full-blown effects of mania and may help relieve the depression that may occur after a manic episode, ultimately leading to better healthcare by using passive monitoring.

At the core of our study is the algorithmic methodology that utilizes photos taken on a smartphone to analyze eye image data. The feasibility of using smartphones to analyze changes in the eyes is demonstrated by Larson et al. [7, 8] in the creation of
Traditionally, the use of expensive lab devices such as pupillometers and infrared eye-trackers were required to detect sub-millimeter fluctuations in eye pupil size [7]. These fluctuations in pupil size are connected to the cognitive load one experiences, or the amount of effort required to process information [7]. The PupilWare and PupilWare-M systems sought to measure pupillary response using smartphone cameras as an indicator in assessing an individual’s ability to process information [7, 8]. For PupilWare-M, the algorithm created from PupilWare was updated to work in real time on a smartphone [8]. The work of Larson et al. [7, 8] provides a framework for our study in terms of the technology used to analyze eye image data gathered from a smartphone.

In recent years, the field of deep learning has demonstrated state-of-the-art performance in many domains such as image classification and recognition, natural language processing, and medical bioinformatics. Convolutional neural networks (CNN) are essential tools for deep learning in the image processing domain. For example, results of the ImageNet object identification competition have improved from 28% error rate to 2% error rate since the introduction of deep convolutional network architectures. [24]. In our research, we use a simple convolutional neural network architecture to automatically identify mania associated with bipolar disorder through cropped images of individual’s eyes.

4 Data Collection

A custom secure website was created for this study in order to collect the required eye image data as well as related metadata necessary to categorize images from individuals with bipolar disorder. Potential subjects were gathered using several methods: leveraging renowned author Julie A. Fast's bipolar community following, social media such as Facebook and Instagram posts, creation of a Facebook page dedicated to the study, flyers distributed throughout SMU campus and other locations, and articles in a Psychology Today blog and Bipolar Magazine. Amazon web services (AWS) were used to host the website and securely store all the data collected. Several factors influenced the decision to build upon the AWS offerings. These factors included the security standards of AWS, the ability to set up an infrastructure within a matter of hours, the low-cost compute capability, the flexibility of storage sizing options, the ease of collaboration, and the auto scalability features. AWS Beanstalk was used for the web component of the deployment and included capacity provisioning aspects as well as load balancing and application health monitoring. EC2 instances were provisioned as part of the Elastic Beanstalk setup to host the web applications. The application needed a backend to hold the images that were uploaded and secure S3 buckets were used as a solution to perform this task. Lastly, metadata of answers to the questionnaire presented on the website were stored in an RDS MySQL database. The architecture used for this study’s data collection website is shown in Figure 2.
Fig. 2. The structure and organization of the study’s website. The web application was hosted on Elastic Beanstalk and connects to an RDS database and application storage S3 buckets, all on the AWS cloud.

Participation in this study was completely voluntary and only those 18 and older were invited to participate. No information regarding the identity of a participant was collected, only demographics such as diagnosis, age group, and sex were collected. Figure 3 shows a flow diagram of the website’s architecture. Included in the homepage was information about the study and instructions on how to submit photos. The creation of a username was necessary to submit photos although no password was required in order to protect an individual's anonymity. Participants were asked to create an “anonymous” username that did not include any identifying information but that could be something they would remember if they chose to submit multiple photos. If an individual chose to participate in the study, a consent form agreement was required before they could create a username and upload their photos. Once an anonymous username was created, participants were asked to upload an image where they could then crop the eyes only. Returning participants could log in directly on the homepage to get to the Photo Upload page (see Figure 3).

After uploading an image, information was gathered as to what mental state the participant was in at the time the photo was taken in the Uploaded Image Survey. The options were manic, stable, or depressed. With mania as our main goal for classification, we utilized stable and depressed photos solely as tools for more effective classification of mania. If the mania option was chosen, a mental state questionnaire appeared (see Appendix). This questionnaire provided 14 statements presented as checkboxes: 6 statements describing euphoric mania, 6 statements describing dysphoric mania, and 2 statements describing psychosis. Statements indicating euphoric mania described the feeling of grandiosity, that anything in life was possible, that a person felt more talkative and social than usual. Statements
describing dysphoric mania included the feeling of an aggressive energy, risky decision making, and a lack of empathy. Symptoms of psychosis include hallucinations or delusions, unwarranted suspicion of others, less emotional expression [39]. Information regarding any medication use, illicit drug use, and the use of a flash were also collected as they could affect the appearance of the eyes (redness, pupillary response, etc.).

![Flow diagram of website design. From the Homepage, a participant can create a username where they enter information about their diagnosis, age group, and sex and give informed consent. After creating a username, they are directed to the Image Upload Page. A participant may also sign in directly from the Homepage and is directed to the Image Upload Page. Once a user crops and uploads their photo, they are brought to the Uploaded Image Survey page where they can describe their mental state at the time the photo was taken and indicate any medications or drugs used.](image)

**Fig. 3.** Flow diagram of website design. From the Homepage, a participant can create a username where they enter information about their diagnosis, age group, and sex and give informed consent. After creating a username, they are directed to the Image Upload Page. A participant may also sign in directly from the Homepage and is directed to the Image Upload Page. Once a user crops and uploads their photo, they are brought to the Uploaded Image Survey page where they can describe their mental state at the time the photo was taken and indicate any medications or drugs used.

## 5 Data Preprocessing and Exploratory Data Analysis

In a period of just under two months, 147 images total were collected through the study’s website. The images and metadata (image details plus answers to the questionnaire) were imported into a Jupyter notebook using boto3. Images were cropped manually, as needed, to capture the eyes only and then all images were resized to 100 pixels X 300 pixels. For some images this introduced warping, as the
cropping of the eyes for each image is not consistent (it is dependent on the resolution of image that a participant used and also how the eyes were oriented in the image). Even so, it is common to use this resizing method in the convolutional network community [25, 26, 27, 28, 29].

There were 3 options available to users in the Uploaded Image Survey: manic (including hypomania), depressed, or stable. A state label was created in order to categorize each. Figure 4 shows the percentages of images collected for each category: 54% of the images collected were categorized as manic, 25% were stable, and 21% were depressed.

![Image](image.jpg)

**Fig. 4.** Image states of data collected in just under two months from the study’s website. The Uploaded Image Survey has three options for a participant to specify their mental state during the time the photo was taken: manic, depressed, or stable. Note that the manic option includes hypomania.

As mentioned previously, there were a total of 14 statements presented as checkboxes for participants to choose from when they described the image uploaded as manic. These statements were used to categorize a participant’s mania as euphoric (6 statements) or dysphoric (6 statements) and there were 2 statements that indicated psychosis. The presence of psychosis would be useful to gather further information as to how often psychosis occurs and how it may affect mania in the eyes. If any of the 6 statements indicating euphoric mania were chosen, these were placed in a variable called `IsEuphoric`. The same was done with the variables `IsDysphoric` and `IsPsychosis`.

We wanted to explore the distribution of statement answers for euphoric and dysphoric mania statements to see if any occurred more than others. Figure 5 shows the frequencies of each. For dysphoric mania statements, the highest percentage of
participants answers involved being uncomfortable in one’s body. This was followed by making “dangerous or unusually aggressive and risky decisions such as intense road rage or picking fights.” For euphoric mania answers, there is a relatively even mixture of all of the statements, with an individual “feeling upbeat and had a lot of fun” as the highest occurring. Close behind that statement was the statement of a person being “happy, talkative, more social, and the life of the party.” It is clear that euphoric and dysphoric mania manifest themselves in different ways so we wanted to see if this possibly affects the appearance of a bipolar individual’s eyes or if we could possibly group both types of mania into one mania category and see if the results could be generalized.

![Breakdown of Questionnaire Answers](image)

**Fig. 5.** The breakdown of answers categorized as euphoric mania versus dysphoric mania. The highest percentage of euphoric mania answers involved the individual feeling upbeat and having a lot of fun. The highest percentage of dysphoric mania participant answers involved discomfort being in their own body.

Since statements were presented as checkboxes, we expected there may be a mix of euphoric and/or dysphoric and/or psychosis answers. Although our main goal was to categorize images as either euphoric or dysphoric mania, it is interesting to see what
different combinations were chosen. For this part of the study, we focused on euphoric or dysphoric and left the effect of psychosis for future research.

As predicted, there were participants who chose a mixture of euphoric mania, dysphoric mania, and psychosis statements. The distribution of categories and image labels are displayed in Figure 6. There were 42 euphoric mania, 12 euphoric/dysphoric, 12 dysphoric/psychosis, 9 euphoric/dysphoric/psychosis, and only 5 dysphoric.

![Fig. 6. The breakdown of mania subcategories: these labels were determined based on a participant’s answers to the mania questionnaire. There were 6 euphoric mania statements, 6 dysphoric mania statements, and 2 statements indicating psychosis. Frequencies of the different combinations are shown in this figure.](image)

### 6 Machine Learning Models and Training Methods

The resized images were normalized by dividing each value by the maximum pixel intensity (on our images, 255). We also employed image expansion to our training sets. Expansion helps to guard against overfitting to the training data by randomly changing the properties of the input image. We achieved this using the ImageDataGenerator class within Keras. Using this package, we manipulated the training images to randomly rotate 5 degrees, translate heights and width by up to 10%, and make horizontal flips of the images. These random perturbations to the images are applied at every iteration of training the Keras model. Therefore, it is much less likely that the model learns to simply “memorize” the pixels of input images because they are constantly and randomly adjusted.

We employed the package Keras [30, 31, 32] with a TensorFlow backend [33] in order to create and train our CNN models. Since the input data on Keras models are trained on numpy arrays only, we converted the resized images into a large numpy array. We also investigated algorithms using more traditional machine learning models such as random forests and logistic regression. These models were
implemented using the scikit-learn API [34]. The dataset was split into training and testing sets using 10-fold cross validation. K-fold cross validation works by splitting the training dataset into k subsets and takes turns training models on all subsets except one which is held out and used for testing [40]. The model performance is evaluated on this held out validation dataset [40]. This process is repeated until all subsets are given an opportunity to be in the test set [40]. The performance measure is then averaged across all the models that were created [40].

For our initial analysis, we used a straightforward and simple convolutional network architecture. This general architecture uses convolutional layers followed by subsampling layers (pooling). The convolutional layers are “trained filters” that are learned during the training of the architecture. They include hyper parameters such as the filter dimensions and the number of filters to learn. We used 3x3 filters throughout the architecture since more complicated filters (i.e., 5x5) can be comprised of smaller filters in multiple layers. Moreover, recent work [38] has shown that small filter sizes in multiple layers tend to be better descriptors in image classification applications. We utilized a total of four filters. The number of filters was determined through a grid search for each layer. Max pooling was our method of subsampling because it has been shown to be superior to other pooling methods in a variety of applications [35]. The number of layers used as well as hyper parameters associated with each layer are shown in Figure 7.

We chose the final layer to be softmax nonlinearity and the loss function to be categorical cross entropy. The loss layer specified the penalization of the deviation between predicted and true labels [37]. These parameters were more favorable since they resulted in faster convergence as compared to mean squared error [37]. For optimization, we chose the ADAM method, which tracks the direction of movement of previous gradient updates and applies normalization to future gradient updates over a sliding window. This method has been shown empirically to be more efficient than most other gradient descent methods for neural network models [36].
7 Results and Analysis

The different types of mania may manifest in differing ways, and as such we performed a total of 4 CNN analyses, breaking up the data in distinctive ways (see Figure 8). The first analysis created a CNN model wherein the dataset was divided into euphoric mania, dysphoric mania, all other combinations of mania including psychosis into “manic-other”, depressed, and stable states. The second analysis created another CNN model where all subsets of mania were combined into one “mania” category and compared to stable and depressed states. The third analysis also combined all subsets of mania into one “mania” category and also grouped both stable and depressed under “other” to create CNN model 3. The fourth analysis, or fourth CNN model, grouped only euphoric mania in one category and all of the rest of the other types of data into “other.” A summary of the frequencies of each category can be seen in Figure 8.

Fig. 7. CNN Architecture. A 300 X 100 image is input into a convolutional neural network for feature extraction. First subsampling occurs, then convolution, followed by another round of subsampling and convolution to create an output.

Fig. 8. State summaries for each of the four analyses performed. This figure shows how the data was divided for each analysis and placed into each resulting model.
As mentioned previously, 10-fold cross validation was used in generating all of our models. We also created a logistic regression model as well as a random forest model. The resulting accuracies for all of the models generated can be seen in Figure 9. Our first CNN model, or analysis 1, was the lowest performing, whereas CNN models 2 and 3 resulted in about the same average accuracy of 50%. CNN model 4 was the highest performing with an average accuracy of around 73%. We decided to further explore the grouping of CNN model 4, wherein euphoric mania was observed against all other data in “other” and built logistic regression and random forest models. The logistic regression model had an accuracy range that reached 86% but averaged around 70%. Lastly, the random forest model resulted in an average accuracy of around 65%.

![Fig. 9. Accuracy Rates of the models created. CNN are the different convolutional neural network models for the four analyses performed. LR = logistic regression model. RF = random forest model. The different CNN models resulted in a wide range of accuracies. CNN 4, LR, and RF resulted in the highest accuracies.](image)

To explore the accuracies further, confusion matrices were created using the logistic regression and random forest models. Because this is a small dataset, an accuracy paradox may be present wherein accuracy alone cannot be fully trusted. One significant problem that occurs with smaller datasets involves high variance wherein overfitting is much harder to avoid compared to larger datasets. Overfitting is “the production of an analysis which corresponds too closely or exactly to a particular set of data, and may therefore fail to fit additional data or predict future observations.
reliably” [41]. It is notable to mention that 31% of the collected images were from one participant. Thus, observing precision and recall values would provide more insight than accuracies alone.

A confusion matrix displays the number of True Positives on the top left corner, False Negatives on the top right, False Positives on the bottom left, and True Negatives on the bottom right. The true label is shown on the Y axis and the predicted label is shown along the X axis. True Positives are the cases where we predicted euphoric manic and they were euphoric manic. True Negatives were where we predicted an image was “other” and they truly were “other.” False Positives are where we predicted an image was euphoric manic but it was truly “other.” Lastly, False Negatives were when an image was predicted as “other” but was truly euphoric manic.

As can be seen in our logistic regression confusion matrix in Figure 10, the number of these True Positives in our training set was 4. The True Negative rate was the highest at 23. Although our algorithm correctly classified a higher number of “other” images, this does not give us very useful information. There were 8 False Positives in the logistic regression model, indicating that euphoric manic was present when it really was not. Lastly, there were 10 False Negatives where “other” was predicted but the image was truly euphoric manic. For the random forest model, there were also a high number of False Negatives and True Negatives. There were only 2 True Positives out of a total of 14 Positive classes.

![Confusion Matrix for Logistic Regression and Random Forest](image)

**Fig. 10.** Confusion Matrix for Logistic Regression and Random Forest

To get a more detailed look at the quality of our algorithm, precision and recall were calculated. Precision is the number of True Positives over the total number of positive class values (ie. True Positives + False Positives) [42]. Precision can be seen as the measure of a classifier’s exactness [42]. The precision of our logistic regression model is 33%, which is extremely low. This tells us that our results are not that useful, despite the 70% average accuracy of the model. This makes sense since there was a high number of True Negatives, where images labeled as “other” were predicted and
correctly classified. The relatively high average accuracy but low precision is indicative of overfitting. The precision of the random forest model, (confusion matrix in Figure 10) was 25%. This means that only 25% of the time euphoric manic was actually classified correctly. (See Table 1)

Recall is a measure of the completeness of a classifier [42]. It is the number of True Positives divided by the number of True Positives plus the number of False Negatives. Recall tells us that out of the 33% in logistic regression model or 25% in random forest model where euphoric manic was correctly classified, only 28.6% or 14.6% (respectively) of the euphoric manic images were in the testing dataset.

Table. 1. Precision and Recall percentages for Logistic Regression and Random Forest

<table>
<thead>
<tr>
<th>Model</th>
<th>Precision</th>
<th>Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logistic Regression</td>
<td>33.3%</td>
<td>28.6%</td>
</tr>
<tr>
<td>Random Forest</td>
<td>25.0%</td>
<td>14.3%</td>
</tr>
</tbody>
</table>

Both logistic regression and random forest models resulted in high False Positive and True Negative rates. In our case, False Positive rates indicated that an image was predicted to be “other” but was truly euphoric manic, which is not good. Having a high True Negative rate in our case also did not provide a lot of information as it means that “other” images were correctly classified as so. Lastly, the True Positive rate, where euphoric manic images were correctly classified was so low that it was difficult to determine whether these same results could be scaled. Overall, taking a closer look at precision and recall gave us better insight as to how the models really performed than merely looking at the resulting accuracies as accuracies are often misleading, especially with small datasets.

8 Conclusion

In less than eight months, we were able to construct a data collection website for this study, collect 147 images from the bipolar community, and analyze our dataset using various machine learning techniques. We performed 4 analyses using a CNN model with 10-fold cross validation using the Keras package. The four CNN models grouped the data in different ways and resulted in 23% to 73% average accuracies. The fourth CNN model, wherein euphoric mania was observed against all “other” images performed the best. As such, we used that grouping to create a logistic regression and a random forest model. The logistic regression model resulted in an average accuracy of 70% and the random forest model resulted in an average accuracy of 65%. However, since the dataset was so small, with only 147 total images, and
with 31% of those images coming from one person, overfitting was likely. Observing precision and recall values gave us more insight as to how the models really performed despite the accuracies that resulted. Precision in the logistic regression model was 33.3% and recall was 28.6%, which tells us that only 33.3% of the time euphoric mania was classified correctly. Because the counts were so low, these results remain inconclusive. Random forest had precision of 25% but its recall was low at 14.3%, indicating a high number of False Negatives wherein euphoric manic was classified as such when it truly was not, which was ultimately evidence of overfitting.

Because machine learning algorithms need thousands of data points to work effectively, our results remain inconclusive due to the low volume of data in our dataset. However, this is just the beginning of the study and the website is continually collecting data. Once more data is collected, our algorithm can be reexamined to see whether or not mania could be successfully classified using machine learning. Also, psychosis may be revisited to learn how often this occurs with mania and how it may affect mania in the eyes. If enough data is gathered from individuals, algorithms can be examined to perform different subsets of the data for a personalized calibration model. Further research and development of our algorithms would be needed to show whether mania can be successfully detected and potentially implemented in a smartphone application to aid in the treatment for individuals with bipolar disorder.
References


Appendix

Mental State Questionnaire
Please check the boxes that represent how you felt during the time of the picture.
| ☒ I felt upbeat and had a lot of fun. |
| ☒ People told me I was cruel or that I was not being empathetic. |
| ☒ Being in my body felt uncomfortable and restless. |
| ☒ I felt that anything in life was possible. |
| ☒ I felt compelled to discuss my amazing ideas with everyone and get them involved in my beautiful project. |
| ☒ I felt like the world around me was not real. |
| ☒ I felt an aggressive energy, as if I wanted to drive fast, punch walls, or give people a piece of my mind. |
| ☒ I felt that normally difficult tasks were easy and enjoyable. |
| ☒ I felt that everything in my life was wrong and I had to leave immediately no matter who got hurt. |
| ☒ I was happy, talkative, more social, and the life of the party. |
| ☒ I made dangerous or unusually aggressive and risky decisions such as intense road rage or picking fights. |
| ☒ I acted suspicious, irritated and mean. |
| ☒ I felt wonderful in my body. |
| ☒ I was worried that people were talking about me. |

Please check the box that represents your diagnosis:
- Bipolar one
- Bipolar two
- Bipolar non specified
- Schizoaffective
- I don’t know

Check the medications you were taking at the time of the picture:
- SSRI and SNRI Antidepressants
- Anti-psychotics
- Anti-Anxiety medication
- Mood Stabilizer
- ADD/ADHD medications
- Other: ____________________

Optional but helpful to the study: Check the substances you were using at the time of the picture. Please know this information will be kept confidential.
- Marijuana
- Hallucinogenics (such as LSD or magic mushrooms)
- Meth
- Cocaine
- Opioids (such as Vicodin, Oxycontin, heroin)
- Alcohol
- Steroids

Other Information: Was a flash used in this photo?
- Yes
- No
- I don’t know