Identifying Emotional Trends with Image Processing via Machine Learning:

The Su-Backstrom-William-Palecek (SuBaWiPa) Distress Detector

TEAM MEMBERS

Matthew Su

Grace Backstrom

Mary William

Alexander Palecek

Abstract

The COVID-19 pandemic has touched everyone, everywhere. Not only has this virus physically injured millions, it has emotionally scarred as many (for which there is no vaccine). It is now that mental health and wellbeing be forefront as isolation and distancing continue to harm those it protects from infection. Given the rise of screen time and web usage, a computer-based background process that can positively impact users is perfectly suited to address these issues. MATLAB is a powerful software with many tools that can be implemented to monitor the emotional state of users. Utilizing four add-ons, MATLAB can train a neural network, access a webcam, take photos, classify the emotion in the photo, and search for trends that could indicate when someone could benefit from a phone call or a visit with a counselor. The Su-Backstrom-William-Palecek Distress Detector is a single MATLAB script with an additional tandem file that can perform the necessary functions to ensure users that need help will get it. The SuBaWiPa classifier is capable of correctly identifying 38.1% of emotions and turning that data into actionable prompts and reports useful for the end user as well as professionals. Improvements in accuracy, scope, and specificity are attainable in the near future, and the possible applications are limitless.

Introduction

During the COVID-19 pandemic, humans face a unique challenge of remaining physically distant from social contacts, yet still being expected to maintain the same level of productivity as before the pandemic limited in-person interactions. The lack of contact with friends and the pressure to maintain productivity are a perfect recipe for emotional burnout. Both those with and without previously diagnosed mental illnesses may struggle to manage their mental health.

To this end, the purpose of this project was to create a system that monitors a person's emotional wellbeing. This program takes a picture of the user's face, classifies their mood and identifies trends in the user's mood over a period of time. The analysis of the user's mood may be used as a personal metric, or data from multiple users may be anonymously compiled to detect trends across broad demographics. Additionally, if a user appears to be consistently distressed, then the user will be prompted to connect with a professional or a predetermined contact. The user will then have an opportunity to discuss their emotional wellbeing and receive strategies to improve their mental health.

This approach is uniquely suited to the situation at hand. The pandemic is directly related to a significant rise in computer usage, even as app usage remains relatively steady.[6] Users are logging on to work, learn, network, shop, chat with others, and play at record numbers. Thus, tapping into that screen time to monitor the mental wellbeing of users is well founded.

It was hypothesized that this implementation would be able to effectively classify a user's emotions and detect a trend over a period of time. However, because the resources of this project are limited to whatever an undergraduate student may access, the system was not predicted to be entirely accurate in classifying emotions. Computational resources, availability of usable datasets, and time were limiting factors in this project.

Methods

Overall Process

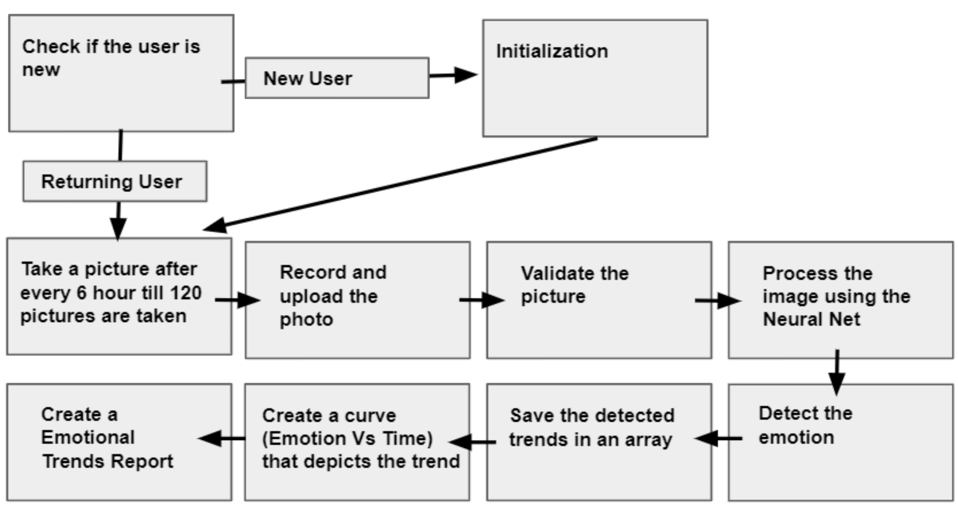

Figure 1. The block diagram of the system is shown, indicating distinct paths taken for new users and returning users. Each path culminates in the formation of a report listing the user's predicted emotions.

Facial Identification and Processing

In order to identify the user's face, a combination of the MATLAB Support Package for USB Webcams and the Computer Vision Toolbox was used. The Webcam support camera accessed the user's webcam, took a snapshot of the area visible to the webcam, and saved the image. The webcam was then turned off to reduce the obtrusiveness of the program. (See the takePicture function.) From there, the image was processed to determine whether or not the image was valid for evaluating the user's mood. To detect a face within the image, a CascadeObjectDetector object was created with the Computer Vision Toolbox. The CascadeObjectDetector object used the Viola-Jones algorithm to detect a bounding box of a face. The Viola-Jones algorithm utilized machine learning and integral images in order to detect simple features[3]. If the bounding box of a face was not detectable, or if more than one face was detected, the image was discarded. Otherwise, the program proceeded to crop the image based on the bounding box, then resized the image to 48x48 pixels and grayscaled the image, such that the new image matched the style of the images used to train the neural network. (See the processPicture function.)

Emotional Classification

Emotions can be expressed in many different ways. Your posture, facial expressions, and speech all contribute to the feelings you're expressing. Paul Ekman's universal emotions studied explains that there are 7 distinct emotions: anger, contempt, disgust, fear, sadness, and surprise[5]. Each facial expression of the emotions have distinct features that express the emotions. Anger can be recognized by eyebrows pulled together, eyes open wide, and lips tightly pressed together. Happiness can be observed by raised cheeks, pulled back lips, and narrowed eyes. These are key features used to classify facial expressions with emotions.

Data

To reduce the amount of time that would otherwise be spent manually selecting and classifying images, a pre-existing set of images was procured. The FER-2013 dataset consists of several thousand (28709 training, 3589 testing) images of human facial expressions[1]. The images are sorted into folders based on the perceived emotion that the subject expresses – anger, disgust, fear, happiness, neutrality, sadness, or surprise. Each image is 48 by 48 pixels large and is in grayscale. Each image is focused on the subject's face, such that the subject's background and other extraneous objects (hair, hats, etc.) are not a consideration.

Neural Network

The neural network was created and trained using MATLAB's Deep Learning Toolbox. The training images were sorted into folders based on their perceived emotion. Then, the images were loaded into an image datastore. 80% of the images were used to train the neural network. The network itself was comprised of 23 layers[2].

The first layer was an image input layer set to input 48 by 48 pixel grayscale images. The image input layer normalized the data by subtracting the mean of all the training data from each training image. The second layer was a 2-D convolutional layer with 16 filters each sized 5x5 and an amount of padding that ensured the input and the output were of the same size. By maintaining the size of the input throughout the network, this method of padding prevented the data from losing information too quickly. By using 16 5x5 filters, the neural network recognizes 16 feature maps containing broad features. The third and fourth layers were, respectively, a batch normalization layer and a rectified linear unit (ReLU) layer. The batch normalization layer normalized the data, while the ReLU layer set any negative values to 0. Using these two layers in conjunction reduced the time taken to train the network. The fifth layer was a 2-D max pooling layer which downsampled the data.

The second through fifth layers were repeated four more times, with the exception of the max pooling layer, which was excluded from the final iteration. However, the 2-D convolution layer was changed to a 3x3 filter, and the amount of filters was doubled in each repetition (32 filters, 64 filters, etc.). As the size of the filters decreased and the number of filters increased, the neural network was able to construct more feature maps with more specific features. A fully connected layer with an output size of 7 was constructed for classifying 7 main emotions. A softmax layer and a classification layer completed the neural network.

A neural network was trained using the specified layers, 15 epochs, and the Adam optimizer. 15 epochs ensured that the training set was iterated over a sufficient number of times to assess the data, without taking an excessive amount of time to train the network. (See neuralNet.m.)

Statistical Analysis

The accuracy of the program can be assessed with the following approach: via a modified program, three unique users can collect and save 29 images. Each user replicates each of the seven Ekman emotions that were trained into the classifier in four separate instances. These images are then labeled with their predicted emotion and the accuracy is determined.

Report Generation

For generating a report at the end of seven days, we implemented the MATLAB Report Generator. The Report Generator APIs (DOM and Report) were used to specify and format report content. Firstly, the Report and DOM APIs were imported and the document type was set to Pdf in the report container. Then we add components like Title, chapters and tie them back into the report container to generate a detailed report. To enhance UX, we added the name and phone number of the primary support, provided by the user, to the report. This was made possible by collecting this information during the initialization step of the code.

Results

The pictures taken are processed into small, 48x48 pixel, grayscale images. They are difficult to resolve without zoom modifications or other visual assistants. To better understand the data, larger, representative versions of the files are collected in Figure 2. These photos are an assortment of correctly and incorrectly identified emotions.

Figure 2. Enlarged photos collected during classifier analysis. Pictured (respectively) are a correctly identified happy photo from User 2, an incorrectly identified happy picture from User 1, and an incorrectly identified sad photo from User 3.

Using the trained neural network classifier, various photos of users were examined and the emotional state of the individual was predicted. Three unique users participated. Subjects partially emulated each of the seven Ekman core emotions. The results are collected and presented in Tables 1 - 3.

Table 1. Twenty-eight photos - four of each of the seven Ekman emotions - are collected from User 1 and the classifier predicts the emotion of the user. Correct predictions are colored green and incorrect predictions are red.

Actual Emotion	Neutral	Disgust	Fear	Surprise	Happy	Anger	Sad
Image 							
Prediction	neutral	disgust	neutral	neutral	happy	anger	anger
Image 							
Prediction	neutral	neutral	neutral	neutral	neutral	anger	anger
Image 							
Prediction	neutral	neutral	anger	anger	neutral	anger	anger
Image 							
Prediction	neutral	fear	fear	neutral	happy	anger	neutral

The accuracy of the classifier is assessed in Table 4. Individual analysis is given to each emotion, producing an interesting pattern.

Table 2. Twenty-eight photos - four of each of the seven Ekman emotions - are collected from User 2 and the classifier predicts the emotion of the user. Correct predictions are colored green and incorrect predictions are red.

Actual Emotion	Sad	Neutral	Fear	Anger	Disgust	Surprise	Happy
Images 							
Prediction	Sad	Neutral	Neutral	Neutral	Surprise	Neutral	Happy
Images 							
Prediction	Neutral	Neutral	Fear	Anger	Neutral	Neutral	Happy
Images 							
Prediction	Sad	Neutral	Fear	Neutral	Neutral	Sad	Happy
Images 							
Prediction	Neutral	Neutral	Sad	Neutral	Neutral	Neutral	Happy

Table 3. Twenty-eight photos - four of each of the seven Ekman emotions - are collected from User 3 and the classifier predicts the emotion of the user. Correct predictions are colored green and incorrect predictions are red.

Actual	Sad	Neutral	Fear	Anger	Disgust	Happy	Surprise
images							
prediction	Sad	Fear	Fear	Happy	Fear	Happy	Happy
images							
Prediction	Fear	Neutral	Angry	Sad	Fear	Happy	Sad
images							
Prediction	Angry	Fear	Angry	Sad	Happy	Happy	Fear
images							
Prediction	Angry	Fear	Sad	Fear	Sad	Happy	Sad

As shown by the heat map in Table 4, the average happy prediction is 83.33 +/- 16.67% accurate. This is the highest such average, closely followed by the average correct neutral predictions at 75.00 +/- 25.00% accuracy.

Table 4. Assessing the accuracy of the neural network classifier. Prediction accuracies by emotion are listed and summarized. A heat map is applied to the average correct score.

	Sad Accuracy	Neutral Accuracy	Fear Accuracy	Anger Accuracy	Disgust Accuracy	Surprise Accuracy	Happy Accuracy	Total Accuracy
User 1	0.0%	100.0%	25.0%	100.0%	25.0%	0.0%	50.0%	42.9%
User 2	50.0%	100.0%	50.0%	25.0%	0.0%	0.0%	100.0%	46.4%
User 3	25.0%	25.0%	25.0%	0.0%	0.0%	0.0%	100.0%	25.0%
Average	25.00%	75.00%	33.33%	41.67%	8.33%	0.00%	83.33%	38.10%
Standard Error	14.43%	25.00%	8.33%	30.05%	8.33%	0.00%	16.67%	6.63%

Conversely, the lowest prediction accuracies are found from surprise and disgust, averaging 0.00 and 8.33% accurate respectively. Out of the 84 total photos collected, the Su-Backstrom-William-Palecek Distress Detector correctly identified thirty-two emotions for an average accuracy of 38.1%.

The predicted emotions are used to plot a distress curve. A threshold is set at anger. Any predicted emotions on or below the distress threshold are counted toward the distress score, as determined by Equation 1.

$$\frac{[data\ points\ at\ or\ below\ distress\ threshold]}{[total\ data\ points]} = distress\ score \tag{1}$$

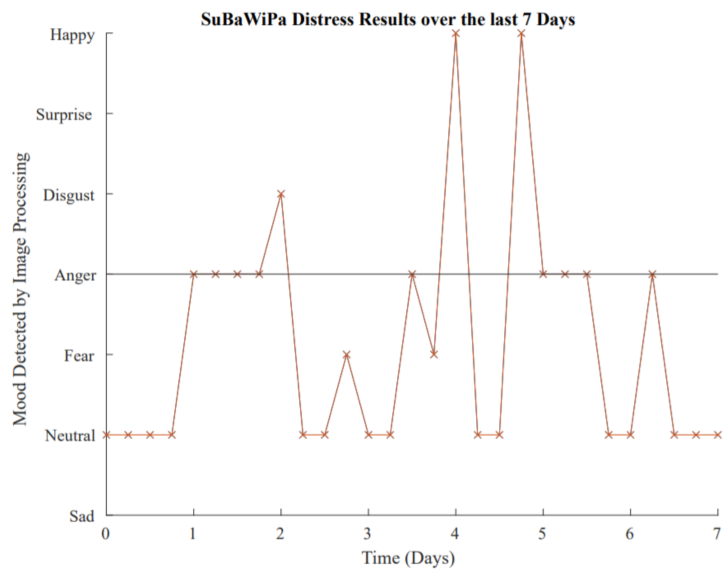

Figure 3. The distress curve generated from User 1. Twenty-six data points lie on or below the distress threshold of the twenty-nine total values for a distress score of 90.0%.

The SuBaWiPa Distress Detector flags any score greater than or equal to 50.0% and prompts the user to seek assistance to boost their happy outcomes and drive down their distress score.

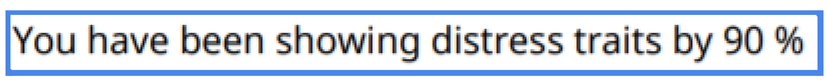

Figure 4. The report provided to User 1 encouraged them to seek assistance based on their distress score above.

Discussion

Emotions occur automatically and are not chosen. Happiness had the highest accuracy of all emotions classified in the code. When testing for the accuracy of the code, the user would attempt to replicate the facial expressions for the emotion and compare the facial expression produced with the facial expression recognized by the code. Compared to emotions such as surprise or disgust, happiness was a much easier facial expression to produce as disgust and surprise are emotions that occur less frequently.

Accessibility to programs similar to SuBaWiPa are difficult to access through MATLAB due to the amount of add ons required to run the programs. Many of the programs online did not provide a list of add-ons required to run certain functions making it difficult to compare the accuracy of the results.

Currently the program detects depressive traits by analyzing the amount of times the user has shown anger, fear, sadness and neutral every 6 hours over the course of a week. Although this method is a good start in detecting these depressive traits, it is also important to analyze the frequency of mood swings. Mood swings are very common in people experiencing depressive symptoms. In order to accurately detect these mood swings, the program would need to capture and analyze images more frequently and record the amount of times the user changes from a positive mood to a negative mood within the course of the week.

It is important to inform the user what information the program will be using. Images of the user, the user's surroundings, and contact information are all private information that can be leaked if handled incorrectly. The program first provides a term of service to inform the user what types of information will be recorded in order to perform its function. All the information can be stored without the use of the internet to ensure none of the private information can be leaked online. The program also stores images and contact information in a location that the user chooses to ensure privacy on their own device. With data beaches being very common, it is important to make sure any information stored has a strong security system or make sure that the information stored is offline.

It is important to note that the dataset used for training the network included thousands of images. These thousands of images included different subjects, angles of faces, and ranges of expression. While emotions can be broadly categorized, there is a level of subjectivity within these categories. One person's smile may be another person's neutral expression. Additionally, some images within the dataset were drawn from stock photo image sites and other image databases. Images from these databases may err on the "dramatic" side. However, because the neural net was found to be lacking in accuracy, it can be concluded that the dataset used to train the network was not suitable for the goals of this project. Because the user is to be observed when they are working on their computer, it is likely that the user will be focused on their screen. When concentrating on a task, their expression is not likely to be dynamic. In contrast, the dataset utilizes images of humans with highly expressive faces. In the future, the neural network could be strengthened by training it with a different dataset. This dataset would have to take into account that most users' expressions would be neutral by default, with microexpressions indicating a change in emotional state. By tailoring the neural net to classify less expressive individuals, the program may be more effective at tracking the emotional state of a user who is focused on their device.

Recording a user's demographic with their emotional trends can be useful in finding what demographics need mental health resources. This information can be sent anonymously to trustworthy organizations such as the World Health Organization in order to improve the wellbeing of those in that demographic. Areas with high amounts of covid related deaths or horrific crimes can take a toll on the mental health of those around that community. It is important to detect these trends and help provide resources to those who need the space to process their emotions.

The program's function would be useful in a clinical setting. Oftentimes nurses can not always check on their patients. This program would help nurses and doctors monitor the trends in the emotional state of their patients. With more information being provided, patients can receive the proper care and support needed to improve their wellbeing. However, the data collected by the program must be approved by the patient before use.

Because of limitations on time and resources, there were several features that could not be implemented. The most important feature of note is the user's ability to directly contact their preferred method of support. Currently, when the program detects a depressive trend in the user, the program simply adds a message on the report of the user's emotional trend that suggests the user should contact their preferred method of support. While helpful, the program should endeavor to be as proactive as possible to ensure the safety and wellbeing of the user. Having a button that directly connects the user to a therapist, a friend, or a family member may be more effective at encouraging the user to seek help, as opposed to only a written suggestion. In a similar vein, if the user displays a consistent pattern of concerning behavior, and if the user has consented to having their emotional state being observed, then it may be helpful to have a means of updating healthcare personnel in real time. Then, if the user experiences a crisis, medical professionals will be prepared to assist them.

There are many features that are needed in order to improve the accuracy of the program. Depressive traits are very difficult to detect since there are many different symptoms that are present. Symptoms such as lack of motivation, changes in sleep or appetite are not able to be detected using the program. A feature that asks the user about these traits may prove beneficial in improving the accuracy of the interpretation of the trends found by the program. With an accuracy of 38.1% there are many improvements needed regarding the neural network used as well as the quality of the images captured. These future changes will improve the accuracy and the validity of the results.

References

1. Sambare, M. (2019, July). FER-2013, Version 1. Retrieved from https://www.kaggle.com/msambare/fer2013.
2. Specify Layers of Convolutional Neural Network. (n.d.). Retrieved from https://www.mathworks.com/help/deeplearning/ug/layers-of-a-convolutional-neural-network.html
3. Viola, P., & Jones, M. (n.d.). Rapid object detection using a boosted cascade of simple features. *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001.* doi:10.1109/cvpr.2001.990517
4. MATLAB Report Generator. (n.d.). Retrieved from https://www.mathworks.com/help/rptgen/
5. Universal Emotions. (n.d.). Retrieved from https://www.paulekman.com/universal-emotions/
6. Koeze, E. and Popper, N. (2020, April 7). The Virus Changed the Way We Internet. The New York Times. Retrieved from http://www.nytimes.com