Facial Expression Recognition: Predicting Early Signs of Psychopathological Disease Among Children

Anthony Dang andang@ucsd.edu
Varun Dinesh vdinesh@ucsd.edu
Chaninder Rishi crishi@ucsd.edu
Tauhidur Rahman trahman@ucsd.edu

Abstract

Information from emotions can greatly benefit researchers as it can be used as an informative predictor for mental health symptoms. Many researchers may find themselves needing to collect emotional data from facial expressions, however, there is no universal standard. While there are big companies with apps in this field already, they do not offer a solution that emphasizes privacy and convenience. We developed a web app that researchers can use to easily gather data about emotions for research purposes. The app works as a skeleton that can plug in machine learning models that predict emotions from facial expressions.

1 Introduction

Understanding mental health is a large challenge in the modern day. Mental health disorders are on the rise and so research on the topic is growing as well. Studying children's mental health shows promising results in the field so, we aim to develop machine learning models for adolescent brain and cognitive development with facial expression data. We want to be able to model the mental health of children with facial expression data, in the hope of being able to predict possible symptoms the child may have. We aim to develop models that are able to predict a person's mental health as they grow up based on the data we have on their childhood. We built a facial expression emotion detector to help aid us in our goals.

Our project roadmap has evolved since we last wrote our proposal. Our plan now is to build off an existing research paper from UMass.

We will implement this entire activity into an web app so that the original authors can streamline their research process and collect better, high-quality data. In the web app, the user will be presented with a series of options from which they will have to make a selection. The game is explained more in-depth below.

Here is a general step-by-step plan of how the game will work:

1. Show user 3 different cakes and wait for their selection
2. If they make a selection:
    - 2 second delay
    - Show a response (positive or negative)
    - Use our model to classify their reaction to the response
    - Have a 2 second delay (trial rest)
    - Self-rating selection (emojis)
3. If they make a selection:
    - 2 second delay
    - Show a response (positive or negative)
    - Use our model to classify their reaction to the response
    - Have a 2 second delay (trial rest)
    - Self-rating selection (emojis)
3. If they don't make a selection within 7 seconds, re-prompt the user.
4. Incorporate a 20 second delay between each task.
5. Repeat the above 8-10 times.

Here is a reference link that explains other details regarding the game

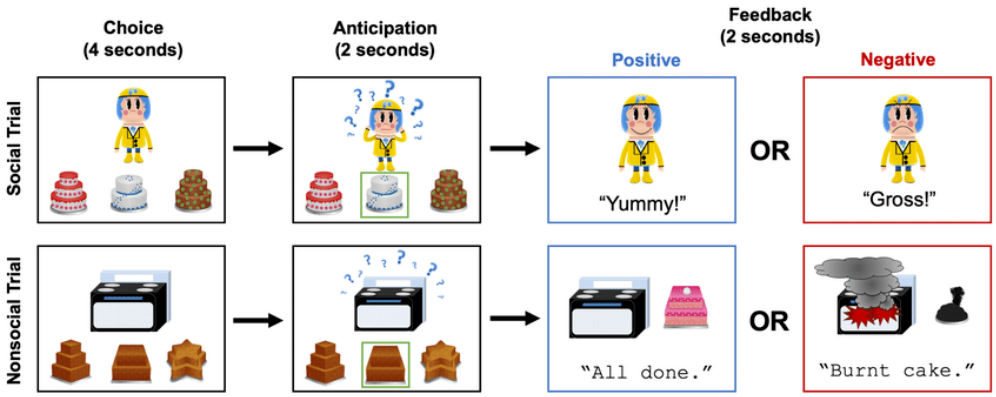

Figure 1: Incredible Cake Kids

2 Methods

We used the FER-2013 dataset, which contains facial expression images with the emotions as labels. We created multiple baseline models that could accurately predict emotions from facial expressions of a general population. As we are not pioneers in this field, we looked to other researchers and existing models already. The reason we started with facial expressions of a general population is because we wanted a very general model before we fine-tune to other datasets. Then, we used models which were lite-weight to plug into our web app we developed. The app was developed in JavaScript.

Figure 2: FER-2013 Dataset

Then, we fine-tuned our models to The Dartmouth Database of Children's Faces. The dataset of Children's Faces comprises images featuring 40 male and 40 female models aged between 6 and 16 years. The models are photographed against a black backdrop, wearing black

bibs and hats to conceal their hair and ears. The photographs are taken from five distinct camera angles and capture eight various facial expressions. Independent raters evaluated the models, ranking them based on the overall credibility of their poses.

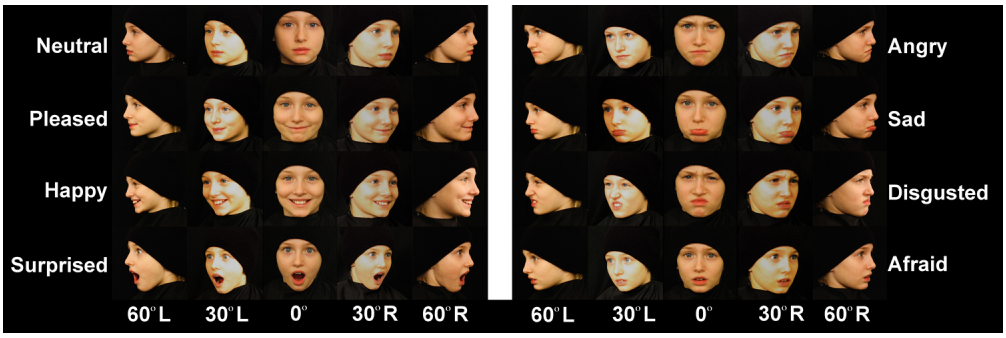

Figure 3: Dartmouth Dataset

3 Results

Our first baseline was heavily adapted from a user on Kaggle, named *Drcapa*. This model was able to attend a 54% accuracy on the test set. The model uses a simple convolutional neural network with four ReLu activation layers, MaxPool layers, and a SoftMax activation layer.

Our second baseline model was using an op-for-op Pytorch re implementation of the Vision Transformer architecture from Google and loading a pre-trained model in order to develop a training pipeline for our data. The Vision Transformer involved resizing the training data images, converting to Tensor, and normalizing the image. A Pytorch data loader was used to pre process the training data and was trained on a vision transformer model with pre-trained weights. The last layer was removed in order to preserve the embeddings which were then converted to Tensor in order to train a logistic regression model on the embeddings in order to classify the seven different emotion classes.

Our third baseline model was created using the MobileNetV2 in TensorFlow. We used transfer learning to adapt the image recognition model to this new task of detecting facial expressions. We created some custom layers (including convolution pooling, MaxPooling, and dropout layers) in hopes to improve the performance of the model and make it more generalizabile on unseen data. This model did not perform as well as our first baseline model, attaining a testing accuracy of 22%. As part of the process of developing this model, we did try another model which was made up of the following layers: Conv2D, MaxPooling2D, Flatten, Dropout, and Dense. This model attained 95% accuracy but only 30% validation accuracy.

Deepface is a Python framework for analyzing human faces. It's functionality includes facial recognition, expression analysis, and more. We are building our models using this pre-existing framework. The labels for the facial expressions vary slightly across our 2 datasets (FER and Dartmouth), so we first adapted for this. Then, we used Deepface's emotion recognition functionality on the Dartmouth dataset.

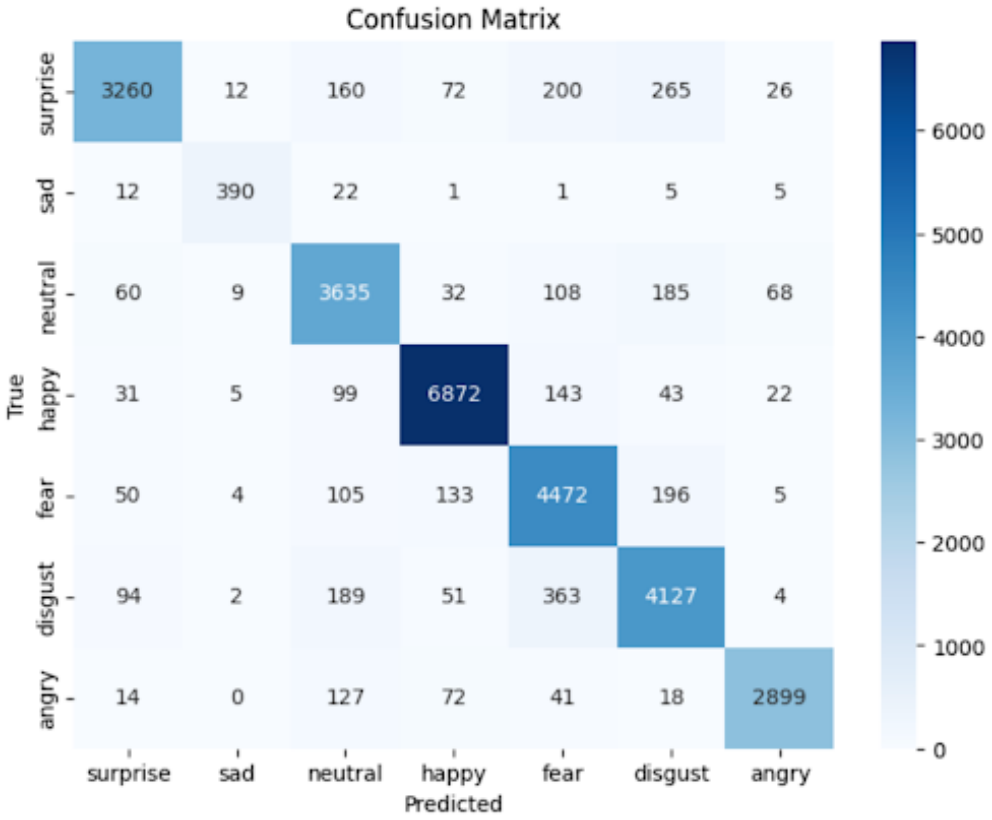

Figure 4: Confusion Matrix

4 Discussion

The baselines we created are important because it'll aid us in our Quarter 2 Project. We plan to use these models to fine-tune it to the facial expressions of children. Unfortunately, the second baseline model took too long to run the pre-processing on the entirety of the training data using the Vision Transformer due to memory constraints. The third model also didn't seem to deliver too high of an accuracy. These results indicate that we should look into other models for this tasks and possibly research models that have already been used to perform this tasks. We can then fine-tune and improve those existing models.

The ethical considerations revolving this project which detects children's emotions using facial expression neural network models includes being completely transparent and accountable for the potential limitations and risks for the models we created. There are potential

biases and unfairness in the models we used, as the datasets used to train children facial expression images only included certain races, and left out others. There are an over representation of Caucasian faces and an under representation of faces from minority groups. This creates an inherent bias in the model where children who are, for example, black or Asian may not have their facial expression be accurately detected due to not enough images of Asian and black children's faces being present in the two datasets.

5 Conclusion

Children facial expression data is crucial to predicting the mental health of their future. The models we have right now to detect these facial expressions serve as a very good starting point, but they need some more improvement. We may have to use more data for each type of expression that we are trying to classify and we may need to use models that are simpler (less parameters, less layers). One problem we are noticing is that our models tend to over-fit to training data even after just a few epochs. This may be an indication that our model is too complex to begin with. After improving these models and getting a bit more decent accuracy, we will work on implementing these models into an app so that they can detect these emotions in real-time.

5.1 Links

- Website
- GitHub