Facial Gesture Recognition
Faculty Supervisor: Dr. T. Aaron Gulliver
SENG 499 | Summer 2020

**Design**
The system follows a pipeline which breaks the process into 4 steps:
input -> face detection & feature extraction -> classification -> result
The face detection & feature extraction is done using a python library face-recognition that is powered by dlib [1]. It would detect faces and return xy coordinates for facial features.

Early research indicated the mouth was a feature to focus on. The main features that were focused on were the upper and lower lip data. The chosen data points returned from the library were normalized to a box around the mouth. The resultant data points were classified by a SVM [3] which was trained on a dataset of ~1500 images. Finally, the system was deployed in a docker environment.

![Facial detection example](image)

**Results**
The project yielded successful predictions for smiling and a neutral expression while predictions for frowning were less than satisfactory.

![Normalized Confusion Matrix](image)

Our first attempt attempt failed completely in identifying frowning as shown in figure 5. Researching further led us to a scientific article for reliable, low resolution images, [2] which led to a second attempt with a larger dataset was able to identify some frowns but still did not meet our standard as shown in figure 6. This is mainly due to the data returned by the face-recognition library, we could partially address this issue by training our classification models on datasets of a specific user.

**Topic**
Facial expressions are an important aspect of human communication that can express emotion or attitude. While humans are capable of identifying numerous expressions, interpreting these expressions digitally is challenging but could facilitate more accessible digital communication or provide aid to those with communication challenges.

**Goals**
The main goal of the project was to be able to accept input as still images and identify basic facial gestures with an accuracy of 70%. The main gestures to identify were smiling, frowning, and a neutral expression. There was a stretch goal that was to be able to accept a live webcam feed.

**Discussion**
The chosen method was successful in reliably identifying gestures for smiling and neutral. However, it produced poor performance with frowning due to design assumptions. This is due to the model the library uses which returns features for the entire face and only has 24 coordinates for the mouth. Due to this, the coordinates returned for frowning and neutral were almost identical. There are many avenues for improvement which include more facial features, an evolving classification model, and a more specific feature model.

![Facial detection example](image)

**Conclusion**
The project aimed at designing a system capable of processing input images and classifying the facial gesture with an accuracy of >70%. The project relied on 2 major components: Manual training & validation, and the implemented service. The manual training & validation included a manual classification of images, processing those images into gesture datasets, training the classification model and validating the accuracy of the model. The implemented service would accept an image, extract facial features, and use the classification model to predict the person's gesture.

**References**