Name: Nguyen Quang Khoi

**Part 1: Classification of Facial Expressions** 1.2.1. The three facial expressions to be classified



Figure 1: Three images from the dataset

1.2.2 Target expressions distribution

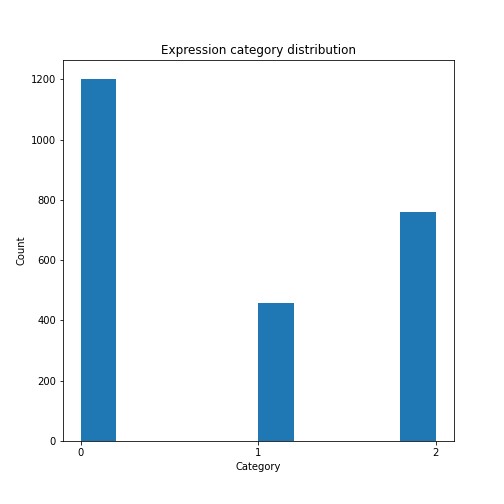


Figure 2: Expression category distribution

1.3.1 Implementing a base KNN (k = 1) as a base model

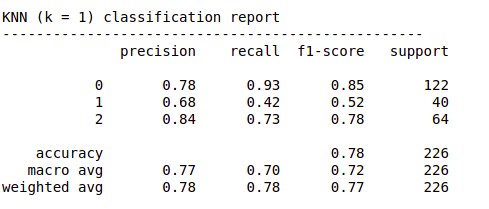


Figure 3: Accuracy and other metrics of the baseline KNN (k = 1) model

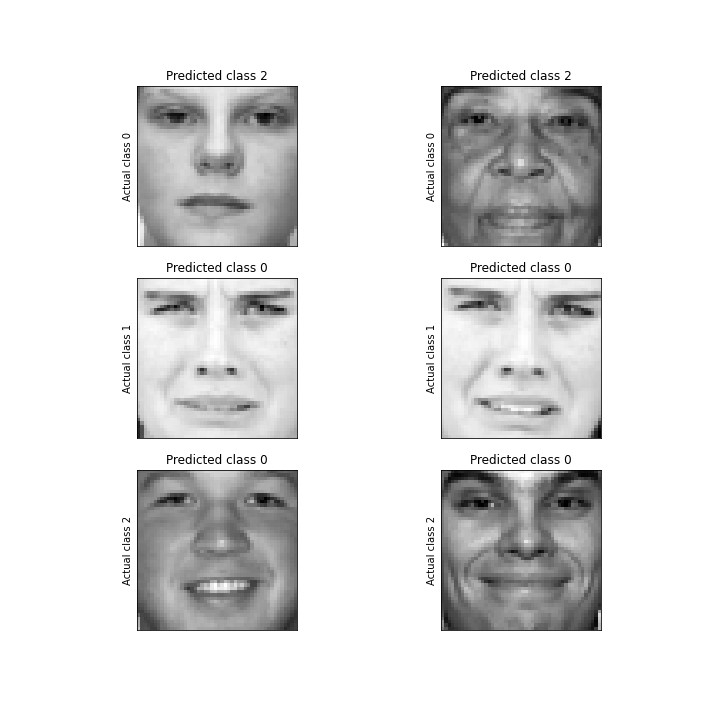


Figure 4: Two mis-classified images for each class

1.3.2 **Accuracy of KNN model with different k values**

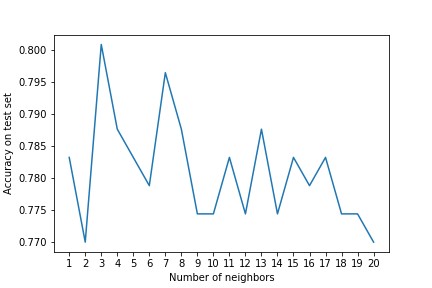


Figure 5: Best hyperparameters for the baseline model

1.3.3 **Implementing different models and various feature selection techniques**

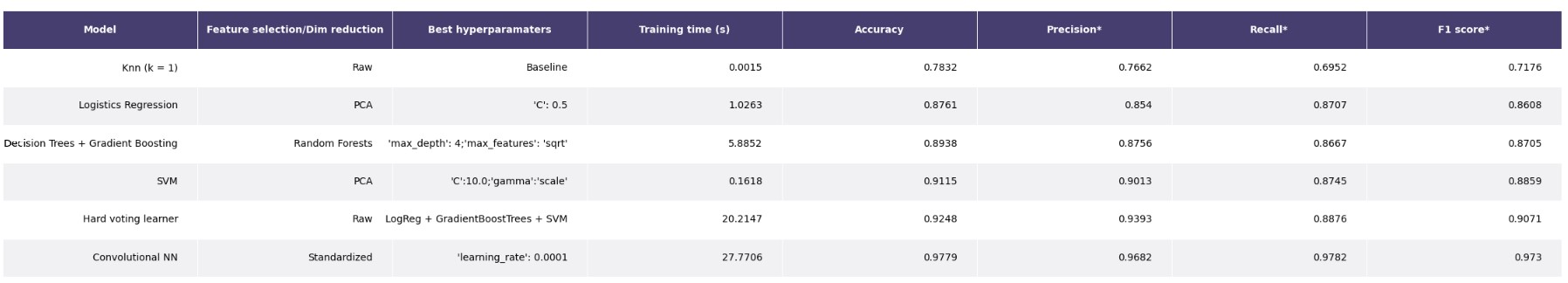


Figure 6: Results of different model

\* Macro average

Note: The table is quite big, so you need to zoom in to check it. I also include the table as a png file in the submission folder. I used TensorFlow to build my neural network. I also used google colaboratory to train some of my models. Therefore, there is some google colaboratory setup code in the notebook.

Models used: KNN, Logistic regression, Decision Trees, SVM, Hard voting, and CNN.

The best classifier is the Convolutional Neural Network with learning rate of 1e-3. The picture below is the architecture.

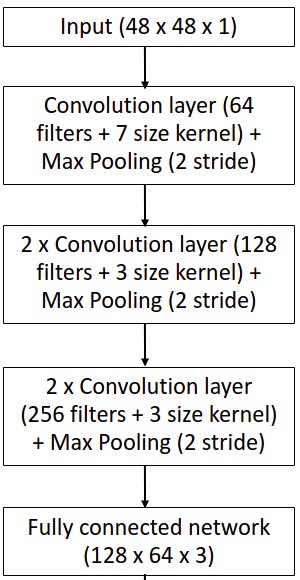
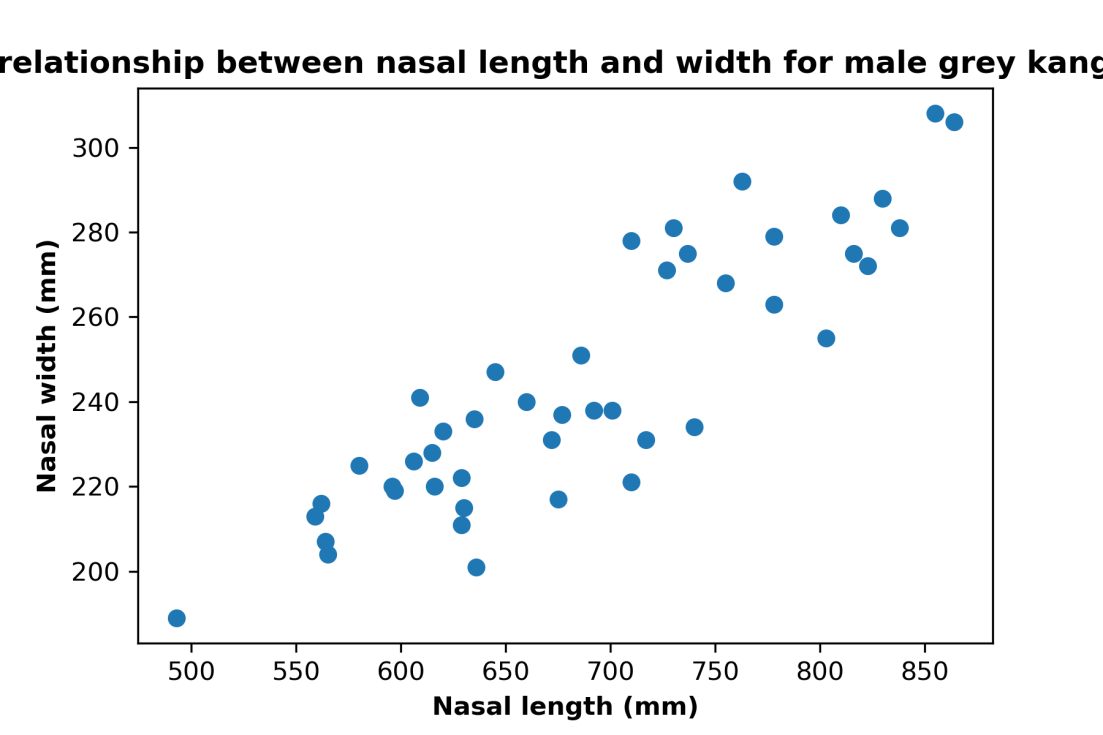


Figure 7: CNN architecture

**Part 2: Regression to estimate the width of a grey kangaroo’s nose**



A screenshot of a computer

Description automatically generated with low confidence

SVR is the best regressor with the highest R^2 score (0.7209)

2.4. There are missing data in the dataset, testing with two imputation methods

|  |  |
| --- | --- |
| Imputation method | R^2 score on a Linear Regression model |
| Mean imputation | 0.5375 |
| KNN (k = 3) imputation | 0.6289 |

The best data imputation method is KNN (k = 3) imputation.