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

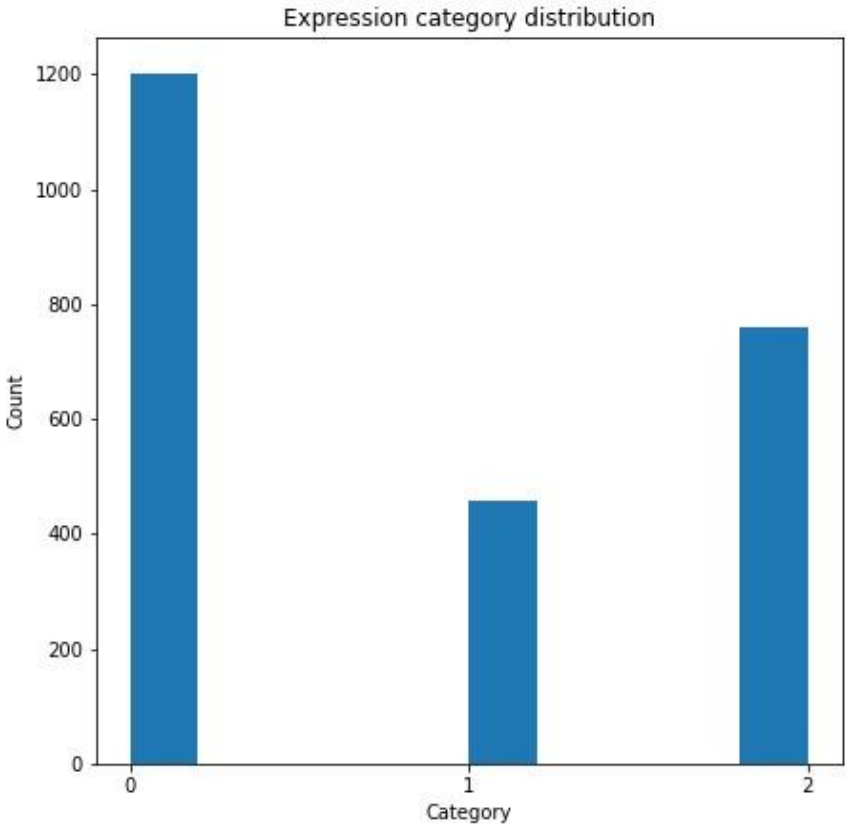


Figure 2: Expression category distribution

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

KNN (k = 1) classification report

	precision	recall	f1-score	support
0	0.78	0.93	0.85	122
1	0.68	0.42	0.52	40
2	0.84	0.73	0.78	64
accuracy			0.78	226
macro avg	0.77	0.70	0.72	226
weighted avg	0.78	0.78	0.77	226

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

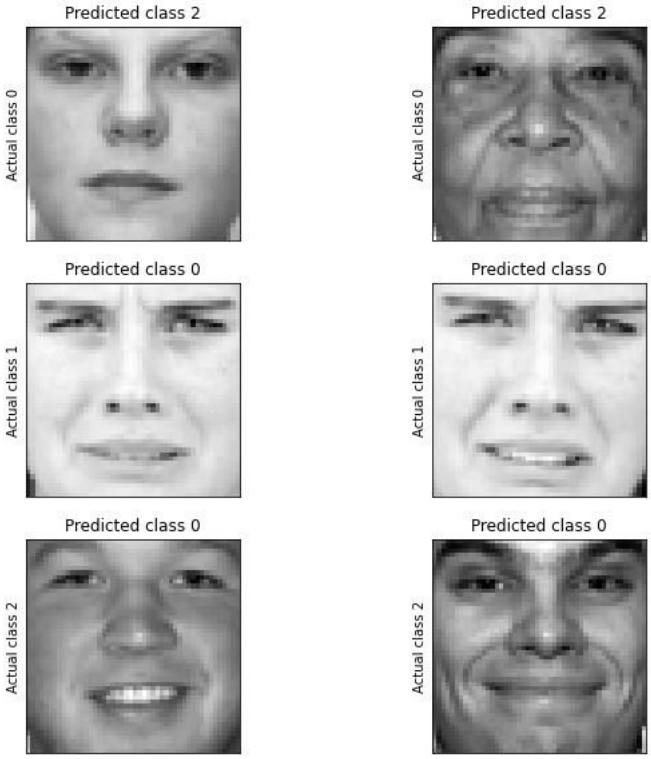


Figure 4: Two mis-classified images for each class

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

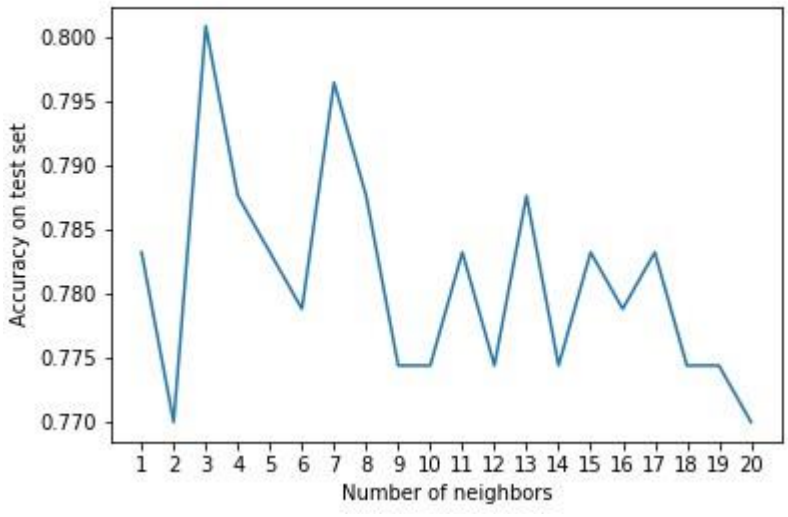


Figure 5: Best hyperparameters for the baseline model

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

Model	Feature selection/Dim reduction	Best hyperparamaters	Training time (s)	Accuracy	Precision*	Recall*	F1 score*
Knn (k = 1)	Raw	Baseline	0.0015	0.7832	0.7662	0.6952	0.7176
Logistics Regression	PCA	'C': 0.5	1.0263	0.8761	0.854	0.8707	0.8608
Decision Trees + Gradient Boosting	Random Forests	'max_depth': 4;'max_features': 'sqrt'	5.8852	0.8938	0.8756	0.8667	0.8705
SVM	PCA	'C':10.0;'gamma':'scale'	0.1618	0.9115	0.9013	0.8745	0.8859
Hard voting learner	Raw	LogReg + GradientBoostTrees + SVM	20.2147	0.9248	0.9393	0.8876	0.9071
Convolutional NN	Standardized	'learning_rate': 0.0001	27.7706	0.9779	0.9682	0.9782	0.973

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.

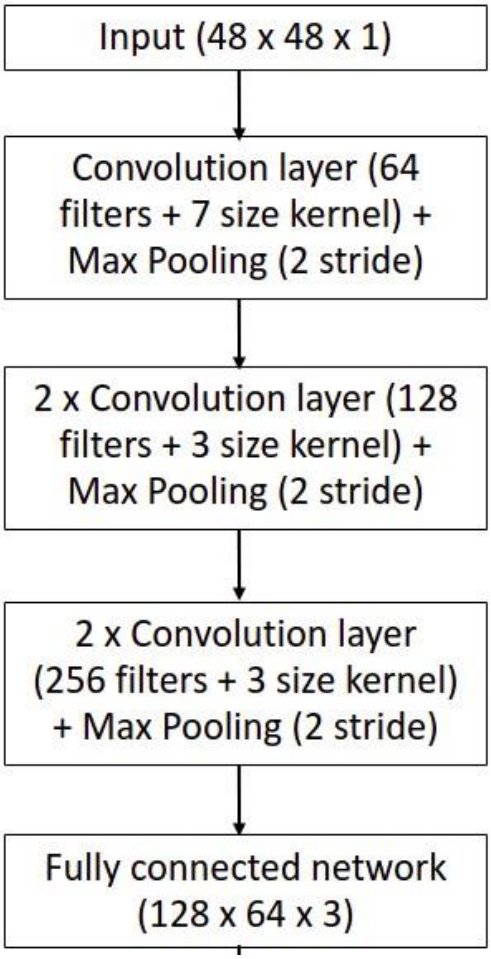
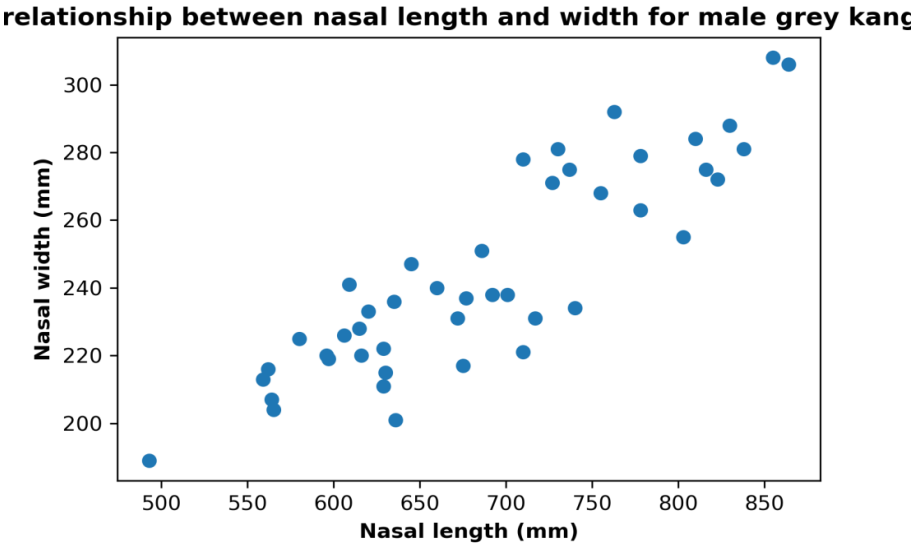
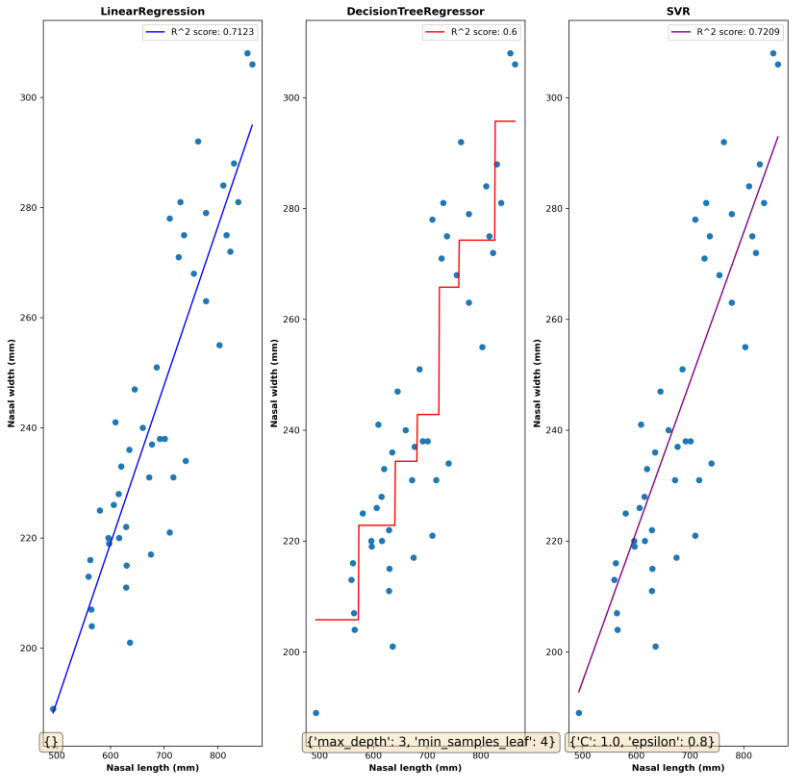


Figure 7: CNN architecture

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





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.