DRP analysis model evaluation

Inhoud

[Model results 3](#_Toc103861328)

[Statistics explanation 3](#_Toc103861329)

[SVC Model: 3](#_Toc103861330)

[SVC Model with same sized data samples: 3](#_Toc103861331)

[SVC Model with same sized data samples and simplified categories: 3](#_Toc103861332)

[What is the performance? 3](#_Toc103861333)

# Model results

When evaluation the performance of a model we usually look at the evaluation metrics. The ones I use are: accuracy, precision, recall and the f1 score.

### Statistics explanation

**Accuracy** or Classification Accuracy as it is also known. Is the ratio of number of correct predictions to the total number of input samples.

**Precision** is the number of correct positive results divided by the number of positive results predicted by the classifier. *This metric shows the percentage of instances the Model predicted that there was DRP present compared to how many times it was correct. This is used to check for false positives, which could lead to patients that do not require aid getting redirected to a professional.*

**Recall** is the number of correct positive results divided by the number of all relevant samples (all samples that should have been identified as positive).

**F1 score**

In most cases scoring a higher percentile means more correctly guessed images, and a lower score means less correctly guessed. This is not always necessarily a good thing, for example a high accuracy could mean that your data set is biased and requires cross validation.

However generally scoring high in all statistics is a positive. Here were the evaluation scores for the recent models.

## SVC Model:

img

## SVC Model with same sized data samples:

img

## SVC Model with same sized data samples and simplified categories:

Afbeelding met tekst

Automatisch gegenereerde beschrijving

# What is the performance?

As you can see in the images above, the models are basically guessing. An accuracy of 53% just means the model is guessing and gets lucky a few times. This performance is somewhat disappointing as I was hoping for a score closer to 70% for the initial results.

# Why is the performance so low?

There could be many different explanations as to why the score is this low. Low amount of data, low quality of data and the data is too similar to each other are a few that come to mind. I base this analysis on that the scores are close to