Automated Classification of 4 Diseases

Author:
Pranav Prakash Chandra

April 10, 2024

Goal: The main goal of this method is to perform the classification of four different diseases (Control Patient(1); Parkinson's Disease(2); Demetia with Lewy bodies(3) and Progressive Supranuclear Palsy(4)) from the provided 40 relative cerebral blood flow (rCBF) images and dopamine availability (DAT) image dataset.

Predicted Output: [3, 2, 3, 3, 1, 1, 1, 2, 3, 1, 1, 1, 2, 1, 1, 4, 1, 1, 1, 3, 1, 3, 3, 2, 3, 4, 2, 1, 2, 1, 1, 2, 2, 1, 1, 3, 3, 1, 1, 2, 4]

Approach

The first step towards the classification problem was to extract relevant features from the dataset. The first step towards extracting the features was to use the given VOItemplate data and the specified important labels ([10,11,16,17,18,19,28, 29,34,35,39,40,80,90,105,120]) to extract a reduced VOI from the input images in both the rCBF and SBR datasets. From the extracted VOI of the datasets, the following statistical features were extracted.

- The Mean of the intensity of volumes
- The Standard Deviation of the intensity of volumes
- The Max intensity of the volumes

Each of the statistical features were extracted from each separate volume in the reduced VOI. To increase the feature space and explore the frequency of intensities in the data, the extracted reduced VOI dataset was converted in to the frequency domain with the help of fft and was zero shifted and the magnitude of the frequencies were extracted. From these magnitudes, the following statistical spectral features were derived:

- The Mean of the magnitudes
- The Standard Deviation of the magnitudes
- The Max intensity of the magnitudes
- The Min intensity of the magnitudes

To further enhance the feature space, histograms of the original images were taken and a region of the histogram between [3,100) was trimmed. The mean and the standard deviation were calculated from the clipped histogram bins to encapsulate the distribution of the histogram values.

The spatial, spectral and histogram statistical features from different regions were selected as they consisted of distinguishing features for the different diseases.

These feature space was Trained with an Ensemble of Neural Networks with different combination of train validation splits and the output was aggregated using voting method of the most popular vote. The Neural Network was chosen for its ability to fit the complex non-linear features. In this case, the neural network adapted to the minute frequency and intensity details of the data and performed well in the training and validation data. An ensemble of 5 models were trained and the majority vote was considered for the final prediction.

The models all had training accuracy of 1.0 and average validation accuracy of 0.8