In this project, I approached the problem of CT scan image classification to identify if a person is infected by SARS-CoV-2 through the analysis of his/her CT scans. The dataset consists of 1252 CT scans that are positive for COVID-19 and 1230 CT scans for patients non-infected by SARS-CoV-2. I downloaded this data from the provided Google Drive link and proceeded with the project.

Firstly, I had to resize the images to a fixed size. As the images were in different sizes, I used the Pillow library to resize them to 224x224 pixels.

Then, I performed data augmentation on the images using the ImageDataGenerator class from the Keras library. I used five arguments for data augmentation, namely, rotation range, width shift range, height shift range, shear range, and zoom range. Data augmentation is a technique used to create new training examples by augmenting the existing data with modified versions of images. This helps to prevent overfitting of the model by providing a more diverse training set.

I used the ResNet-50 model for training the CT scan image classification model. ResNet-50 is a deep neural network architecture that is widely used in computer vision tasks due to its high performance. I trained the model using binary cross-entropy loss and Adam optimizer. I also included early stopping and model checkpoint to ensure that the model does not overfit and to save the best model during training.

After training the model, I performed predictions on the test set and evaluated the performance of the model using multiple performance metrics such as precision, recall, and f1-score. The precision metric measures the proportion of true positive predictions among all positive predictions, while recall measures the proportion of true positive predictions among all actual positive instances. The f1-score is the harmonic mean of precision and recall, and provides a balanced measure of model performance.

In conclusion, I approached the problem statement of CT scan image classification for detecting COVID-19 by resizing the images to a fixed size, performing data augmentation, training the model using ResNet-50 architecture, and evaluating the model's performance using multiple metrics. This project highlights the potential of artificial intelligence methods for identifying COVID-19 patients through the analysis of their CT scans, which could help in the early detection and treatment of the disease.