The problem statement for this project is CT Scan Image Classification, where the aim is to develop an artificial intelligence method that can identify if a person is infected by SARS-CoV-2 through the analysis of his/her CT scans. The dataset contains 1252 CT scans that are positive for COVID-19 and 1230 CT scans for patients non-infected by SARS-CoV-2, totaling 2482 CT scans. The data has been collected from real patients in hospitals from Sao Paulo, Brazil.

To start, the first step was to download the data from the provided link. The images in the dataset were in different sizes, so a fixed size was selected on which to work. This would ensure that the model would not have any difficulty in processing the images of different sizes. The next step was data augmentation, which is the process of creating new training samples from the existing data by performing random transformations. At least 5 arguments were used in the data augmentation process, such as width_shift_range and height_shift_range.

The model selected for this project was ResNet, and from this, any layer model could be picked such as ResNet-18, ResNet-34, ResNet-50, ResNet-101, ResNet-110, ResNet-152, ResNet-164, ResNet-1202. The next step was to train the model, and early stopping and model checkpoint were both used. Early stopping was used to stop the training process if the model was not improving, while model checkpoint was used to save the best model after each epoch.

Finally, prediction was made, and multiple performance metrics were mentioned. These performance metrics included accuracy, precision, recall, F1 score, and confusion matrix. The model's accuracy was measured, and it was found to be good. It was also noted that precision was high, meaning that the model made fewer false-positive predictions. Recall was also high, indicating that the model made fewer false-negative predictions.

In summary, this project aimed to develop an artificial intelligence method to identify if a person is infected by SARS-CoV-2 through the analysis of his/her CT scans. The dataset contained 1252 CT scans that were positive for COVID-19 and 1230 CT scans for patients non-infected by SARS-CoV-2, totaling 2482 CT scans. The project used data augmentation and ResNet for training the model. Multiple performance metrics were used to evaluate the model, and the accuracy, precision, recall, F1 score, and confusion matrix were measured.