Inference time for the different models at different test runs