Q4)

After Finetuning our model, we can see a slight decrease in WER on cv-valid-dev to 0.1580 from 0.2913. This tells us that the model has successfully learned new features related to the common-voice dataset. To further improve the model accuracy, we can train on larger samples of the dataset using Cloud services like Google collab or Databricks, which I have not in this example. To improve model scalability and generablilty, we could import other datasets and choose an appriorate sampling method to expose the model to a larger variance of data. Additionally, if data gathering is an issue, we can use data augmentation methods to try to improve the WER like time stretching, pitch shifting, noise injection etc. Furthermore, if our datasource is a live-stream of audio data, we can set periodic finetuning on the new audio data to compare the performance with the baseline model so that the model will always be up to date.

With these methods, we will be able to scale the model and deploy it.