Question 4

The final validation accuracy is: 0.66286
The final test accuracy is: 0.66949

Ensemble process:
we use three neural network models to implemented bagging ensemble. We first randomly sample three sample with replacement from out training data set. Then we train three different neural network independently for each training sample. These three neural network are complete independent and can run individually. After all models are trained, we use them to make prediction separately, finally we take the average of each of their predictions as our final prediction.

Better or Not:
No, the bagging model is nearly the same performance as the single neural network model, so it doesn't improve the performance.

Reason:
Ensembling the same model which train on different data subset has lack model diversity, thus it does not always improve the model performance. Also small training subset could be another problem, when the training set is small, there could be a issue that training subset are even smaller so that each model are not well trained, which result in poor performance.