Dataset link - https://github.com/microsoft/MS-SNSD/tree/master/clean\_test