

# Weekly report

Fan Zhang  
a1738078

## •What you have done for the project during the week individually

1. Refactor the code of ProcessCFilesWithCodeSensor.py, ProcessRawASTs\_DFT.py, BlurProjectSpecific.py and make the input and output matches the description of the paper.
2. Still working on the LSTM.py, for now I have figured how does the code split out the validation data and test data from the source data.  
They used the train\_test\_split() from sklearn.model\_selection.\_split. As an example, the FFmpeg data set, which contains 191 vulnerable functions samples and the code splits them into 3 different sections with 124, 38 and 29 sample sizes. And counts their nonzero values. I have totally understand the process of the above steps and generated the right output of these steps.

## •What you plan to do for the project next week individually

1. Since I have found that the code stops right before the 6<sup>th</sup> step of the code which is the activations step, I will keep looking the code here and try to figure out how to generate the right final output of the trained models.
2. Need to find a new related paper
3. Modify the current GitHub page
4. Transfer our current Gannt chart to Jira

## •List the issue you encounter and give reasons if applicable

1. Not familiar with the tensorflow tool and lack of knowledge of the deep learning methods mentioned in the paper.
2. The paper doesn't describe the code process very clearly.
3. There are many commented code within the file, I am not sure if they are useful or not.