MoM (Review Meeting 5)

Date and Time: **16th May 2024, 7.30PM to 9PM IST.**

**Meeting Attendees**:

* David de Hilster, Hugo Watanuki, Dr Jyothi Shetty, Dr. Vishalakshi Prabhu, Nihar Mandahas, Skanda P R, Pratheek Rao, Manvith L B, Arya Hariharan, Eshan Mathur, Prashant Ronad, Nikhil Vasu.

**Agenda 1**: Present the findings and work done in the past few weeks related to the Legal Assistance Tool and understand future steps.

*Discussion/Demonstration Done:*

*Recommendations:*

*Actions to be Taken:*

**Agenda 2**: Present the work related to “Building NLP pipeline for Electronic Health Records” and understand the future steps.

*Discussion/Demonstration Done:*

1. Demographic of patients listed. Preprocessed in such a way that Symptom, Duration, Gender and Affected Organ are labelled and displayed when the whole text is given as input.
2. Passing medical data through ESP and getting the response. However, cannot copy and paste Python code into ECL.
3. Getting NULL response from Roxie server. If Roxie does not work on PC, check the configuration (oracle box or windows)
4. Why NULL response? Is it from Java, ESP, or Roxie? Hence suggested to look at Log files from Roxie server. {#var/log/HPCCSystems/myroxie}. Check memory log.
5. Alternatively, if ESP is not talking to Roxie, check ESP logs.
6. Nikhil said it takes 7 to 8 hours to train the MRI images when the dataset size is around 7000. Hugo suggested to try 100/200 records at a time. However, Nikhil says the problem in maintaining the ratio when downscaling.
7. For MRI, Nikhil is using VGG16 model, its imported, not from scratch. Using TensorFlow library.
8. Nikhil asked if model is in Table format, can it be passed through the Roxie? However, Table is broken to Records by Roxie. To raise this question in Developers Forum (JIRA/Stack Overflow).
9. Prashant asked if the speed is dependent on nodes in Roxie? It does but also on other factors.

*Recommendations:*

1. What are the upcoming plans? GNN is used for text or image?
2. When are we using Embedded Pytorch for NLP? Its not there in university cluster.
3. Hugo asked to use TensorFlow library on university cluster but make sure code is functional.
4. What is the model representation way for MRI training? Is it table/collection of vector/math formula)? Hugo suggested to store model as binary file.
5. David had to drop out of meeting to attend other tasks. But he said “good work so far”.

*Actions to be Taken:*

1. Try the models on cloud environment (reproduce in more systems)
2. Finetune both NLP models for EHR.
3. Minimize overfitting, train on other MRI’s.