MoM (Review Meeting 3)

Date and Time: **13th March 2024, 8 pm to 9.30PM IST.**

**Meeting Attendees**:

* David de Hilster, Hugo Watanuki, Dr Shobha G, 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 :** Present the work related to “Building NLP pipeline for Electronic Health Records” and understand the future steps.

*Discussion/Demonstration Done:*

1. Valid dataset for the project was collected from hospital during Feb. To begin with, we had 200 patients’ brain EHR. To carry out Named Entity Recognition (NER), Comprehend Model has been built. Also, one more model namely **biomedical\_ner\_all** which uses SoftMax activation at end of LSTM and negative log loss function is tried. Trained up to 10 epochs.
2. The team will still explore NLP+ML model, validate and then finalize.
3. Though compression of brain MRI from DICOM to JPG is very challenging, students extracted 4 patients MRI scan from hospital. Each patient had approximately 2000 MRI images. The remaining patients’ MRI scans will be collected by the end of March 2024.
4. During March beginning, team found another Tumor Dataset which has 4 classes of brain diseases (Meningitis, Glioma, pituitary, Parkinson’s). During Feb, team got Alzheimer 80,000 image in JPG format which falls into 4 categories.
5. Nikhil used data augmentation to change position, shift, rotation, rescale, shear, Nikhil used Google inception model to reduce computational cost. He also came across VGGNet transfer learning model.

*Recommendations:*

1. Hugo said to list each model and note down precision. Also asked about NLTK scope as it was mentioned in proposal.
2. Also minimize varying all parameters to save time. Understand the shape/distribution of the data. Check whether the trained data shape is similar to hospital data.
3. He also suggested training on our own data collected from the hospital. Use GNN bundle version 2 on HPCC Systems.
4. David said it’s going fine.

*Actions to be Taken:*

1. Can use either Microsoft Project or Excel spreadsheet tool to track project status.
2. Note down each activity and files w.r.t to person and timeline.
3. Update timeline, justification and challenges on GitHub repository.