

This project involves a text prediction task on medical text using Medical Transcriptions [1]. Apply clinical text classification adapting ClinicalBERT [2] and Bio_ClinicalBERT [3] to the implementation in [4]. Show the confusion matrix, accuracy, precision, recall and F1-score for each category class. Draw also AUC (Area Under the Curve) and PR (Precision-Recall) curves [5]. Additionally choose one of the three customized ClinicaBERT versions in [6] as given in [7], [8], [9]. Compare the results of three models and explain the details in a report.

[1] <https://www.kaggle.com/datasets/tboyle10/medicaltranscriptions/data>

[2] <https://huggingface.co/medicalai/ClinicalBERT>

[3] https://huggingface.co/emilyalsentzer/Bio_ClinicalBERT

[4] <https://www.kaggle.com/code/ritheshsreenivasan/clinical-text-classification>

[5] <https://medium.com/nwamaka-imasogie/clinicalbert-using-deep-learning-transformer-model-to-predict-hospital-readmission-c82ff0e4bb03>

[6] <https://www.cambridge.org/core/journals/natural-language-engineering/article/lightweight-transformers-for-clinical-natural-language-processing/BEF81FDE6E12B9DC5AD4906AE67CDDEB>

[7] <https://huggingface.co/nlpie/distil-clinicalbert>

[8] <https://huggingface.co/nlpie/tiny-clinicalbert>

[9] <https://huggingface.co/nlpie/clinical-miniALBERT-312>