Liver is a common organ target of metastasis for colorectal cancer, which indicates the possibility of nonmetastasis cancer. We create a prediction model by using machine learning on the electronic health record to discover susceptible characteristics for liver cancer in colorectal cancer patients. We found from 152 Included Concepts generated for our [TMU]_CRC concept sets, while in [TMU]_Metastases_secondary_cancer generated 28 Included Concepts. We found 12132 records in outcomes. Ada Boost algorithm showed the highest AUROC of 0.74, followed by MLP (0.718), Decision Tree (0.676), KNN (0.664), Random Forest (0.650), LR (0.5), and NB (0.494).

Our model showed that predicting hepatic malignancy in colorectal cancer patients using TMUCRD-mapped OMOP-OHDSI data is possible and achieved a good prediction result with Ada Boost algorithm. We intend to discover other malignancies in the colorectal cancer patients cohort, along with other data partners.

This study used TMUCRD-mapped OMOP-OHDSI data to create and assess machine-learning models for predicting hepatic malignancy in colorectal cancer patients. TMUCRD Mapped OMOP CDM consists of 3,649,627 patients from 1 January 2004 - 30 December 2020, which was derived from original TMUCRD data where the data pooled from Taipei Medical University affiliated hospitals (Shuang Ho Hospital, Wan Fang Hospital, and Taipei Medical University Hospital) that already mapped using OMOP CDM. Generated using the Incidence Rates feature using ATLAS with target cohort [TMU]_CRC_New and outcome HCC, from 39,316 patients, the proportion of [+/-] per 1,000 persons is 41.64. We use ATLAS to create the cohort and R Studio to create the prediction model. We use Concept Sets, Cohort Definition, and Prediction features in ATLAS. Our concept name consists of Primary malignant neoplasms of rectum, neoplasms of rectosigmoid junction, neoplasms of colon, neoplasms of anus, neoplasms of anal canal, and malignant tumor of rectosigmoid junction. Our covariates demographic, diagnoses, and medication. We implement algorithms consist of Logistic Regression (LR), Ada Boost, Multi Layer Perception (MLP), Random Forest, Naive Bayes (NB), K-Near Neighboor (KNN) and Decision Tree.

**Figure 1.** HCC in CRC Prediction Model Concept

This model showed that predicting hepatic malignancy in colorectal cancer patients using TMUCRD-mapped OMOP-OHDSI data is possible and achieved a good prediction result with Ada Boost algorithm. We intend to discover other malignancies in the colorectal cancer patients cohort, along with other data partners.

Contact: d931109002@tmu.edu.tw

**Figure 2.** HCC in CRC Prediction model Threshold, Incidence, sensitivity, specificity, PPV, and NPV

**Figure 3.** HCC in CRC Prediction model ROC and Calibration Plot