Predicting Hepatic Carcinoma from Colorectal Cancer Patient using Machine Learning with TMUCRD Mapped-OMOP-OHDSI

Yudha E. Saputra[1]; Daniel C.A. Nugroho[2]; Muhammad Solihuddin Muhtar[3]; Jason C. Hsu[4]

[1] International Ph.D. Program in Biotech and Healthcare Management, College of Management, Taipei Medical University
[2] Graduate Institute of Biomedical Informatics, College of Medical Science and Technology, Taipei Medical University
[3] Graduate Institute of Data Science, College of Management, Taipei Medical University
[4] College of Management, Taipei Medical University


Yudha E. Saputra


Daniel C.A. Nugroho


Muhammad Solihuddin Muhtar


Jason C. Hsu

Background

Liver is a common organ target of metastasis for colorectal cancer, which indicates the possibility of nonmetastasis cancer[1,2]. 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.

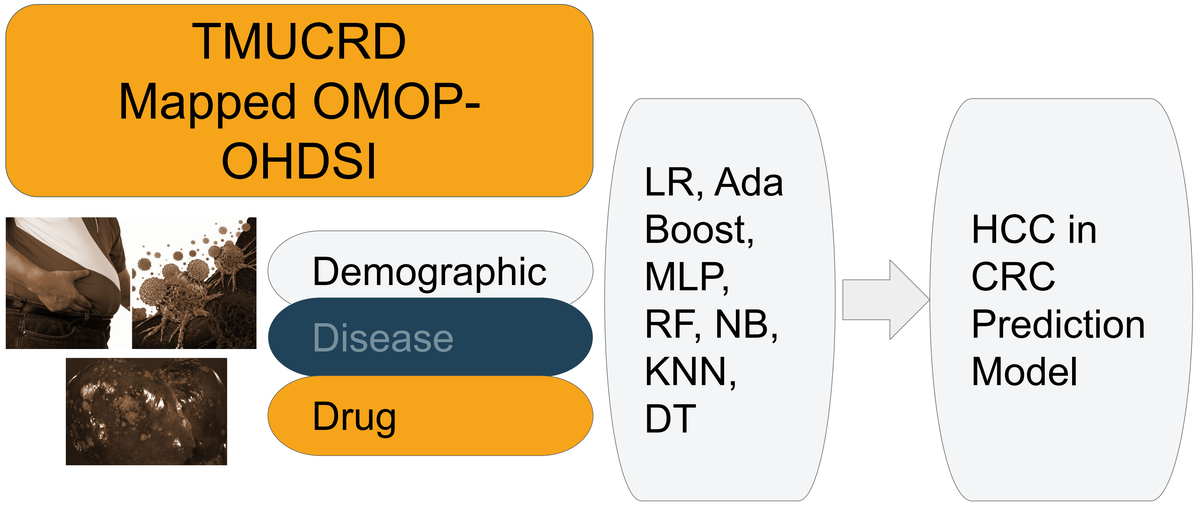

Figure 1. HCC in CRC Prediction Model Concept

Methods

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.

Results

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).

Algorithms	AUROC
LR	0.500
Ada Boost	0.740
MLP	0.718
RF	0.650
NB	0.494
KNN	0.664
DT	0.676

Table 2. HCC in CRC Prediction model algorithm comparisons

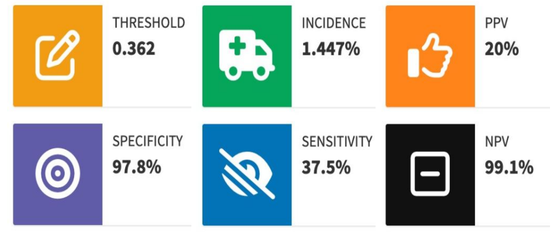

Figure 2. HCC in CRC Prediction model Threshold, Incidence, sensitivity, specificity, PPV, and NPV

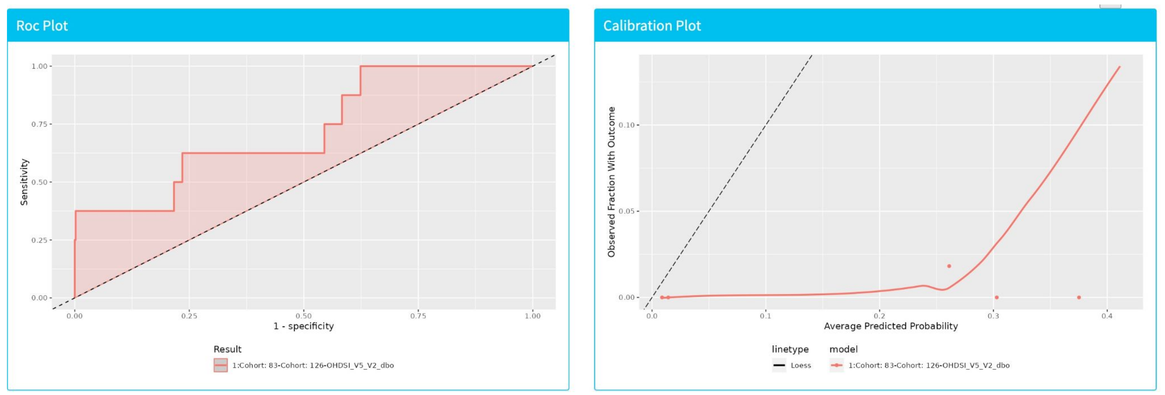

Figure 3. HCC in CRC Prediction model ROC and Calibration Plot

Conclusions

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.

References

1. Rompianesi G, Pegoraro F, Ceresa CD, Montalti R, Troisi RI. Artificial intelligence in the diagnosis and management of colorectal cancer liver metastases. World journal of gastroenterology. 2022;28(1):108-122. doi:10.3748/wjg.v28.i1.108
2. Shuwen H, Xi Y, Qing Z, Jing Z, Wei W. Predicting biomarkers from classifier for liver metastasis of colorectal adenocarcinomas using machine learning models. Cancer Medicine. 2020;9(18). doi:10.1002/cam4.3289

Contact: d931109002@tmu.edu.tw