

The TMUCRD-mapped OMOP-OHDSI is used to build a CRLM prediction model with an AUC of 0.868.

Title: Predicting Colorectal Cancer Liver Metastases using TMUCRD-Mapped OMOP CDM

Background: Liver Metastases is one of the most prevalent malignancies that cause mortality among colorectal cancer patients where 25–30% affected by the malignancy worldwide. Predicting the occurrence of colorectal cancer liver metastases (CRLM) could benefiting physician on deciding the diagnoses. We create a prediction model by using machine learning on the electronic health record to discover susceptible characteristics for secondary malignant neoplasm of liver cancer in colorectal cancer patients.

Figure 1: Cohort selection process

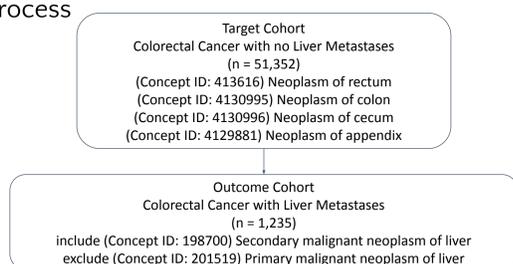


Table 2: Baseline demographic age range, gender, and ethnicity

Covariate	N = 1,235 (100%)
Gender	
Male	674 (54.57%)
Female	561 (45.43%)
Age Range	
25-29	5 (0.40%)
30-34	3 (0.34%)
35-39	19 (1.54%)
40-44	56 (4.53%)
45-49	84 (6.80%)
50-54	118 (9.55%)
55-59	149 (12.06%)
60-64	167 (13.52%)
65-69	171 (13.85%)
70-74	131 (10.61%)
75-79	103 (8.34%)
80-84	106 (8.58%)
85-89	74 (5.99%)
90-94	34 (2.75%)
Ethnicity	
Not Hispanic or Latino	1,235 (100%)

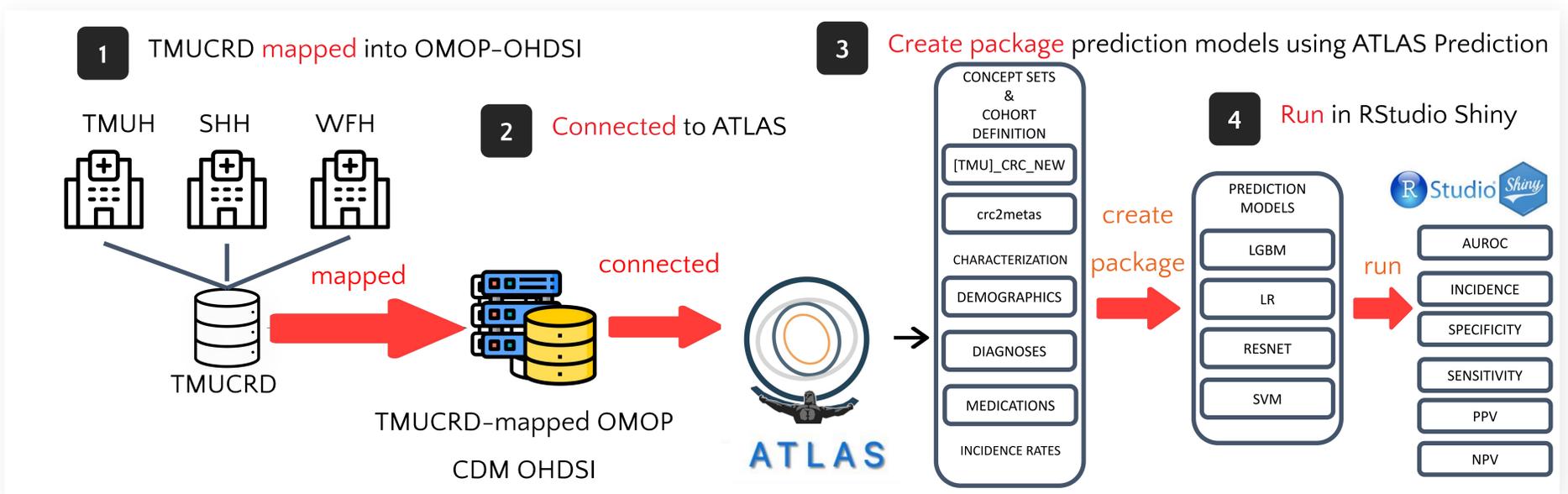
Table 1: Performance of machine learning algorithms

Algorithm	AUROC	Incidence	Specificity	Sensitivity	PPV	NPV
LightGBM	0.868	0.518%	97.8%	5.6%	1.3%	99.5%
Logistic Regression	0.850	0.518%	97.3%	33.3%	6%	99.6%
Residual Network	0.830	-	-	-	-	-
SVM	0.212	0.518%	96%	0%	0%	99.5%

Figure 2: ROC Plot, Precision Recall Plot, F1 Score Plot, Box Plot, Calibration Plot, and Demographic Plot of LGBM Train



Methods



Limitation: Lack of specific cancer parameters (e.g: cancer stages, cancer site, and cancer size) on Cohort Definition on ATLAS, might hindrance our model performances to achieve higher performance.



Yudha E. Saputra, Daniel C.A. Nugroho, Muhammad Solihudin Muchtar, Jason C. Hsu

