1202P

A Deep Learning Approach Using Routine Pathology Images to Guide Precision Medicine in Metastatic CRC

Chaitanya Parmar[1], Albert Juan Ramon[1], Patricia Raciti[1], Kristopher Standish[1], Sanjib Chowdhury[1], Supriya Shah[1], Fernando Cruz-Guilloty[1], Joshua C Curtin[1], Mahadi Baig[1], Robert Schnepp[1], Xuesong Lyu[1], Zayed Albertyn[1], Christopher Moy[1], Joel Greshock[1], Lev Demirdjian[1]

Johnson & Johnson Innovative Medicine[1]

Key Takeaway

Computer Vision based AI algorithms trained from CRC Whole Slide Images can not only enhance, but also potentially provide robust alternatives to conventional testing methods, thereby reducing the burden on testing and accelerating the delivery of therapies to patients

Conclusions

- Computer Vision algorithms trained on CRC Whole Slide Images can robustly predict the presence of RAS/RAF, MSI and CMS-1/2/3/4 mutations
- At current levels of performance, these algorithms, when tuned for high sensitivity, can reduce burden on testing by **~58-73%** across biomarkers as pre-screening tools
- Models can highlight specific regions in the Whole Slide Images which provide deeper insights/interpretability into morphologies associated with these biomarkers

Please scan QR code

Poster

Narrated poster video

Supplementary material

https://www.congresshub.com/Oncology/ESMO2024/Colorectal/Parmar

Introduction

Survival outcomes of patients (pts) with metastatic colorectal cancer (CRC) have improved in recent decades, reflecting advances in targeted treatments, diagnostics, and biomarker (BM)-driven pt selection. Notably, microsatellite instability (MSI), RAS, and RAF have shown to be predictive of outcomes to specific targeted therapies, and the gene-expression based Consensus Molecular Subtypes (CMS) have demonstrated predictive value across stages of CRC. Nevertheless, factors such as socioeconomic disparities may preclude pt segments from access to molecular testing, while cost and technical challenges make RNA-based signatures difficult to implement in clinical practice. Alternative methods for BM detection, such as analysis of pathology images obtained via routine clinical care, can support subsequent clinical decision making. We developed a deep-learning (DL) algorithm to infer RAS/RAF/MSI and CMS biomarkers using only H&E digital pathology images.

Methods

Current methods of genetic testing are time consuming and expensive. Here, we propose a workflow where we can leverage Computer Vision AI algorithms that can analyze H&E images and detect the presence on certain biomarkers. Such algorithms can be set up to work with existing testing methods as pre-screening tools to reduce the burden on testing or can be potentially set up as alternatives.

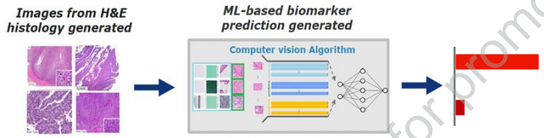

Figure 1: Using AI models to augment testing

Training a Foundation Model

AI models when trained directly on histopathology images tend to learn biases that stem from different scanners/staining intensities, resulting in a drop in performance on unseen test data. A foundation model is a large model trained on vast amounts of raw unlabeled data to generate rich generalizable representations of images that can be used to train different types of models in the downstream. Our Foundation Model is trained on ~55k WSIs from 32 tumor types from various scanners/sites.

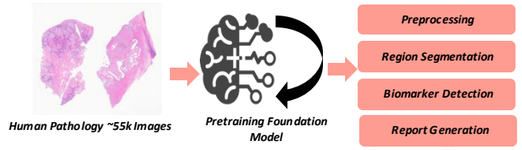

Figure 2: A Foundation model can be used as a starting point to train different types of models in the downstream

Datasets & Models

Training Data

We leveraged the publicly available TCGA dataset; specifically, the TCGA-CRC and TCGA-STAD datasets. There were a total of 597 samples from which the prevalences for BRAF, KRAS, NRAS and MSI were ~3.8%, ~14.6%, ~3.8% and ~4.5% respectively. For CMS, the dataset was set up for for multiclass classification. The prevalences for CMS-1,2,3,4 were ~14.8%, 42.4%, 13.9% and 28.8%, respectively. Any samples that were deemed ambiguous as per the CMScaller[1] algorithm were removed. For each model, the dataset was split into 80% training and 20% testing while keeping the biomarker prevalences maintained. We use cross-validation for the 80% training split and use the model with the highest AUC on the test set to report performance.

Binary Classification Models	
Biomarker	*Prevalence*
BRAF	3.8%
KRAS	14.6%
NRAS	3.8%
MSI	4.5%

Multiclass Model	
Biomarker	*Prevalence*
CMS-1	14.8%
CMS-2	42.4%
CMS-3	13.9%
CMS-4	28.8%

Table 1: Datasets

Network Architecture

Given every whole slide image has many artifacts such as background region, pen marks, blurs, folds etc., it is important to remove such regions before training the algorithms so that the models do not capture any unknown biases in the system. Hence, a preprocessing step is performed to identify all such regions and remove them before training the models.

Post preprocessing, the usable tissue region is divided into non-overlapping 224x224 patches. Each WSI has ~4000 such patches. The Foundation Model is then used to generate representations from these patches. These representations are then used to train an attention-based aggregation model(one for each biomarker separately) for making the final prediction.

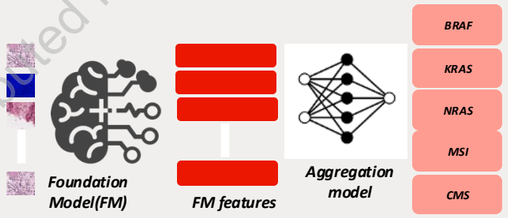

Figure 3: Network Architecture

Results

Downstream Models

For all the binary classification models, an operating threshold for achieving 90% sensitivity is selected to obtain the potential reduction rate in testing for each of these models.

BRAF - Our BRAF binary classification model achieved a cross-validation AUC of ***0.82 +/- 0.01***. On the test set, an AUC of ***0.85*** was achieved, translating to a potential ***reduction in testing by 58% while maintaining an 100% sensitivity on the test set***

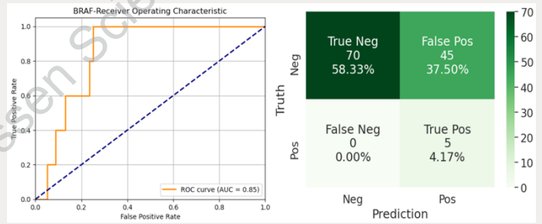

Figure 4 – BRAF Performance

KRAS - Our KRAS binary classification model achieved a cross-validation AUC of ***0.91 +/- 0.03***. On the test set, an AUC of ***0.86*** was achieved, translating to a potential ***reduction in testing by 73% while maintaining an 84% sensitivity on the test set***

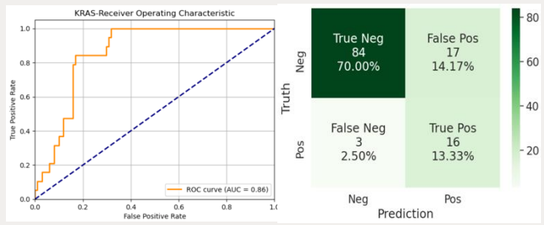

Figure 5 – KRAS Performance

NRAS- Our NRAS binary classification model achieved a cross-validation AUC of ***0.83 +/- 0.07.*** On the test set, an AUC of ***0.87*** was achieved, translating to a potential ***reduction in testing by 68% while maintaining an 100% sensitivity on the test set***

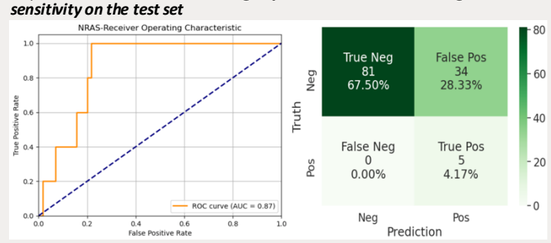

Figure 6 – NRAS Performance

MSI- Our MSI binary classification model achieved a cross-validation AUC of ***0.90 +/- 0.01***. On the test set, an AUC of ***0.87*** was achieved, translating to a potential ***reduction in testing by 73% while maintaining an 80% sensitivity on the test set***

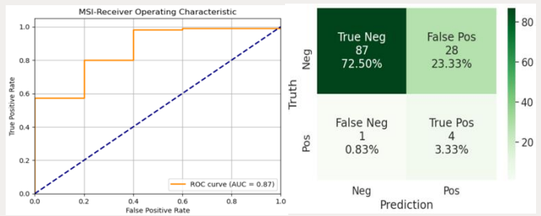

Figure 7 – MSI Performance

CMS - Our CMS multiclass classification model achieved a cross-validation AUC of ***0.79 +/- 0.03***. On the test set, an AUC of ***0.82*** was achieved.

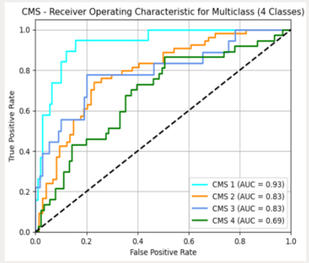

Figure 8 – CMS Performance

Interpretability through Segmentation – All models can highlight regions in the Whole Slide Image which are directly related to the areas of interest that the model focuses on to make a prediction. Segmentation model cab ne used to quantify different types of morphologies in these regions such as Tumor, Stroma, Lymphocytes, Necrosis. to further gain insights into which histological patterns drive specific biomarkers.

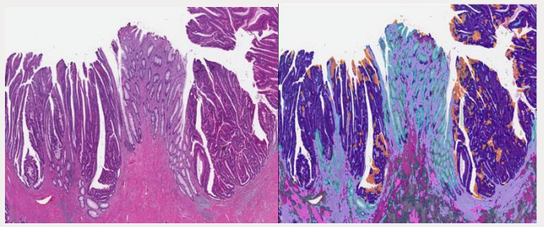

Figure 9 – Morphological Interpretability; Left – CRC region, Right – Segmentation overlay to highlight morphologies

References:
1 Eide, P.W., Bruun, J., Lothe, R.A. *et al.* CMScaller: an R package for consensus molecular subtyping of colorectal cancer pre-clinical models. *Sci Rep* **7**, 16618 (2017). https://doi.org/10.1038/s41598-017-16747-x

Colorectal Cancer