Poster ID: 131P
Category: Miscellaneous
Colon Cancer

Disclosure of Potential Conflicts of Interest:
The Presenter of this study states that there is nothing to declare.

Machine Learning-Aided Colorectal Cancer Detection Based on Circulating Tumor Cells

Paz A.[1], Iliopoulos A.[2], Apostolou P.[2], Papasotiriou I.[3]

[1] Immunology and Regenerative medicine, White Clinic private practice, Algés, 1495-131, Portugal
[2] Research Genetic Cancer Centre S.A., Florina, 53100, Greece
[3] Research Genetic Cancer Centre International GmbH, Zug, 6300, Switzerland

Background: Colorectal cancer (CC) constitutes one of the most prevalent types of cancer, with extremely high mortality rate. Diagnosis of CC is characterized by low accuracy and high invasiveness. Thus, the development of accurate, noninvasive CC detection methods is a necessity, but also very challenging. Machine learning (ML) models can improve pathologists' diagnostic accuracy, specificity, and sensitivity. This study presents preliminary results concerning ML classifiers discriminating between healthy and CC samples, using data of circulating tumor cells (CTCs) and their expression profile.

Methods: A dataset was generated based on 20 biomarkers including CTCs enumeration and protein expression (e.g. CD44, CD133, SOX2, OKT4, Nanog, MET, CD34, CD45, BCR-ABL, CD30, CD15, CD31, CD19, CD63, CD99, EpCam, MUC1, PSMA, PanCK). These biomarkers are commonly used in the identification of primary tumor in a patient and to provide guidance about disease progression and future prognosis. Particularly, for 35 healthy individuals and 39 CC patients, blood samples were analyzed to identify the presence, concentration and protein expression of CTCs. Then, the performance of 5 hyper-optimized ML classifiers was tested, namely for classification trees, Support Vector Machines (SVM), K-Nearest Neighbor (KNN), Ensemble classifiers and neural networks. The results correspond to a (5X-) 10-fold cross-validation estimation in a validation set (67 samples) and the resulting models were tested in a test set (7 samples).

ML Classification Models	Validation Accuracy	Validation Sensitivity (TPR)	Validation Specificity (TNR)
Trees	80.00 ± 1.61	86.66 ± 2.34	83.12 ± 1.75
SVM	90.89 ± 2.18	91.70 ± 0.00	90.00 ± 4.64
KNN	93.53 ± 2.23	92.78 ± 1.48	94.38 ± 5.15
Ensemble Classifiers	88.11 ± 3.35	92.22 ± 2.29	83.74 ± 6.04
Neural Networks	90.00 ± 1.23	92.78 ± 1.48	86.86 ± 4.08
Average	89.51 ± 3.19	91.11 ± 2.98	87.62 ± 4.67

Table 1: 5X-Cross-Validation=10-fold, Validation-Set = 67

ML Classification Models	Test Accuracy	Test Sensitivity (TPR)	Test Specificity (TNR)
Trees	88.56 ± 11.96	90.00 ± 13.69	86.68 ± 18.24
SVM	91.42 ± 7.83	100.00 ± 0.00	80.02 ± 18.24
KNN	88.57 ± 6.39	95.00 ± 11.18	80.02 ± 18.24
Ensemble Classifiers	82.68 ± 7.83	100.00 ± 0.00	86.68 ± 18.24
Neural Networks	91.42 ± 7.83	100.00 ± 0.00	80.02 ± 18.24
Average	90.85 ± 2.39	96.25 ± 4.79	82.68 ± 3.65

Table 2: Test-Set = 10% = 7

Results: For all ML models and for the test sets, the mean accuracy was found to be 90.85 ± 2.39, the mean sensitivity (True Positive Rate) was found to be 96.25 ± 4.79, while the mean specificity (True Negative Rate) equal to 82.68 ± 3.65.

Conclusions: The present study reports preliminary results concerning the development of ML models which exhibit notable performance in distinguishing CC samples from healthy ones. These findings indicate that ML models, using CTCs' enumeration and their protein expression data, could be included in clinical practice to assist pathologists in increasing the accuracy of diagnosis. However, albeit the results seem promising, more experiments are needed based on larger datasets to verify and extend the results of this study.

2024 ESMO GASTROINTESTINAL CANCERS Annual Congress
MUNICH GERMANY
26-29 JUNE 2024

Selected References:
• Steven J. Cohen et al., Isolation and Characterization of Circulating Tumor Cells in Patients with Metastatic Colorectal Cancer, *Clinical Colorectal Cancer* 6(2), 125-132, 2006.
• A. Tabari et al., Role of Machine Learning in Precision Oncology: Applications in Gastrointestinal Cancers, *Cancers* 15, 63, 2023.
• Zhu S-L et al., Application of machine learning in the diagnosis of gastric cancer based on noninvasive characteristics. *PLoS ONE* 15(12):e0244869, 2020.

RGCC