



## Review article

## Machine learning in colorectal cancer prediction and diagnosis: a systematic review of models' performance

Mohammad-Ali Abuzuhri<sup>a,b,\*</sup>, Shahenaz Najjar<sup>c,d,e</sup>, Mohammed Awad<sup>f</sup>,  
Ricardo Cruz-Correia<sup>g</sup>, Hiba Falna<sup>h</sup>, Emine Aktas<sup>i</sup>, Basem Mohammed Abu Al Rob<sup>j</sup>,  
Miguel Oliveira<sup>b</sup>, Ibrahim Mughnamin<sup>k</sup>, Sara Novo Esteves<sup>b</sup>, Yousef Awlad Mohammad<sup>l</sup>,  
Pedro Vieira-Marques<sup>b</sup>

<sup>a</sup> Faculty of Nursing, Arab American University, Palestine

<sup>b</sup> RISE-Health, Department of Community Medicine, Information and Health Decision Sciences, Faculty of Medicine, University of Porto, Portugal

<sup>c</sup> Institute for Healthcare Policy, Department of Public Health and Primary Care, KU Leuven University, Leuven, Belgium

<sup>d</sup> UBI Business School, Brussels, Belgium

<sup>e</sup> Department of Health Sciences, Faculty of Graduate Studies, Arab American University, Jenin, Palestine

<sup>f</sup> Computer Systems Engineering Department, Faculty of Engineering, Arab American University, Palestine

<sup>g</sup> Faculty of Medicine, University of Porto, Portugal

<sup>h</sup> Department of Biomedical Sciences, Marshall University, Huntington, WV, USA

<sup>i</sup> Department of Public Health Primary Care Environment and Health Unit, KU Leuven University, Leuven, Belgium

<sup>j</sup> Quality & Patient Safety for Hospitals & Emergencies, Ministry of Health, Palestine

<sup>k</sup> Yatta Governmental Hospital, Yatta, Hebron, Palestine

<sup>l</sup> Optometry Department, Faculty of Modern Sciences, Arab American University, Ramallah, Palestine

## ARTICLE INFO

## Keywords:

Colorectal cancer  
Machine learning  
Prediction  
Diagnosis  
Performance  
Feature selection  
Data preprocessing  
Ensemble learning

## ABSTRACT

**Introduction:** Colorectal cancer (CRC) poses a significant global health burden, demanding early and accurate detection strategies. However, Machine Learning (ML) models are increasingly being applied for CRC prediction; yet their performance requires systematic evaluation to guide adoption.

**Purpose:** This review evaluates the performance of ML models in predicting and diagnosing CRC, focusing on studies published between 2019 and 2024. It intends to identify the most frequently used ML models, determine their performance, and analyze the impact of different models' settings on their performance.

**Methods:** A comprehensive search was conducted using SCOPUS, PubMed, and Web of Science databases. Following PRISMA guidelines, studies evaluating ML models for CRC prediction were selected and reviewed. Study selection, data extraction, and risk-of-bias assessment were performed independently by multiple reviewers. Extracted data included study characteristics, model specifications, validation methods, and performance metrics.

**Results:** Thirty studies met the inclusion criteria. Random Forest (RF) was the most frequently evaluated model. At the same time, Ensemble Learning (EML), Neural Networks (ANN/DNN), and Support Vector Machines (SVM) consistently demonstrated the highest performance across multiple metrics. Most studies employed molecular datasets, and feature selection methods varied widely, significantly influencing model performance.

**Conclusions:** ML models, particularly EML, ANN, DNN, and SVM, demonstrate high diagnostic performance in CRC prediction and diagnosis, suggesting substantial diagnostic potential; any effect on decision-making and outcomes through improved accuracy and early detection requires external and prospective validation. However, variability in datasets, methodologies, and reporting quality highlights significant research gaps, including the lack of standardized validation procedures and consistent performance reporting, which are crucial for facilitating clinical adoption and informing healthcare policy decisions.

\* Corresponding author at: Faculty of Nursing, Arab American University, Palestine.

E-mail address: [mohammad.zedan@aaup.edu](mailto:mohammad.zedan@aaup.edu) (M.-A. Abuzuhri).

<https://doi.org/10.1016/j.ijmedinf.2025.106170>

Received 29 September 2025; Received in revised form 25 October 2025; Accepted 27 October 2025

Available online 28 October 2025

1386-5056/© 2025 Elsevier B.V. All rights reserved, including those for text and data mining, AI training, and similar technologies.