ERASMUS UNIVERSITY ROTTERDAM

ERASMUS SCHOOL OF ECONOMICS

Simulating Haemoglobin values for MISCAN-Colon using machine learning: XGBoost, NN, and SVM[1]

Author

Yoëlle Kilsdonk (513530)

Supervisors

E. P. O'Neill (EUR), dr. I. Lansdorp-Vogelaar (EMC), R. van den Puttelaar (EMC), D. van den Berg (EMC)

May 19, 2022

Abstract

Keywords— MISCAN, Machine Learning

[1]The views stated in this thesis are those of the author and not necessarily those of the supervisors, second assessor, Erasmus School of Economics, Erasmus University Rotterdam or Erasmus Medical Centre.

Table of contents

List of Figures

List of Tables

1 Introduction

The MISCAN-Colon (MIcrosimulaten SCreening ANalysis) model is a microsimulation model, developed by the Erasmus University Medical Center (EMC) of Rotterdam, for the evaluation of colorectal cancer (CRC) screening. It evaluates different CRC screening policies by comparing their costs and effectiveness (Loeve et al., 1999).

Before, the MISCAN-Colon model did not simulate a haemoglobin (Hb) value for every faecal immunochemical test (FIT), instead, it simulated a positive or negative FIT result. Recently, the Public Health department of Erasmus Medical Center explored the extension of the MISCAN-Colon model with a simulation model for the Hb values. Thus, with this new extension, in order to evaluate the benefits of personalised screening strategies based on the previous Hb concentration found in a person's stool, MISCAN-Colon needs an accurate simulation model for the Hb values. Consequently, the aim of this thesis will be to extend the MISCAN-Colon model with a simulation model for the Hb concentration found in a person's stool.

So far, the proposed method to simulate the Hb values is a mixed-effect zero-inflated negative binomial model (ZINB). However, previously conducted research conducted by EMC finds that mixed-effect machine learning (MEml) models outperform the mixed-effect ZINB model significantly for this purpose. The optimal MEml model was chosen to be a decision tree, due to its interpretability. This research investigates the gains in accuracy due to the prioritization of model performance over interpretability. To this end, we incorporate three types of black-box machine learning methods: a Neural Network (NN), XGBoost as an Ensemble method (XGBoost) and a support vector machine (SVM) as Kernel based method.

Due to correlation within observations (since patients with positive FITs have multiple outcomes for each test) the assumption of iid observations is often violated. Consequently, this study follows the approach by Ngufor et al. (2019), who propose a MEml model that incorporates random-effects into machine learning algorithms for efficient analysis of longitudinal data, using the aforementioned methods.

This research consists of two phases, the first being outside of MISCAN to predict Hb concentrations and phase two is the calibration and implementation of the models[2] in MISCAN, with as main goal to obtain a Hb simulation model for which the simulated Hb concentrations resemble the observed concentrations of real-life Dutch population screening data.

RQ1 How can we combine the structure of mixed-effect models with black box machine learning algorithms?

[2]The models which will be included depend on time constraints and model performance.

RQ2 Which MEml model – NN, XGBoost, or SVM – is best suited for predicting the Hb concentration in CRC screening?

RQ2a Can we – and is it worth it to – improve the existing HB simulation model in MISCAN colon using black box machine learning techniques, at the cost of interpretability?

2 Literature

2.1 Colorectal cancer

Colorectal cancer (CRC) is the development of cancer from the colon or rectum, which usually starts as a benign adenoma. CRC is one of the most commonly diagnosed and most deadly cancers worldwide (Torre et al., 2015; Sung et al., 2021). According to the Dutch Rijksinstituut voor Volkgezondheid en Milieu, 5% of people will develop CRC in the Netherlands. Nearly nine in ten cases occur in people older than 55. Risk factors for CRC include age, gender, genetics, environment, diet, physical activity, and smoking (Botteri et al., 2008; Thanikachalam and Khan, 2019). Moreover, the worldwide burden of CRC is expected to further increase due to, *inter alia*, the growth and aging of the population (Jiang et al., 2022).

Figure 1 shows the progression of CRC in five stages. In stage 0, the adenoma is *in situ* and has not grown beyond the mucosa (i.e., the inner lining) of the colon or rectum. Stage I is when the adenoma has grown beyond the mucosa, without spreading to the lymphatic system or distant organs. In stage II the adenoma has invaded the colonic or rectal wall, with possible infection of nearby organs. Finally, in stages III and IV, the metastatic adenocarcinoma has spread to lymph nodes and distant organs.

Adenomas of the colon are estimated to be present in 20-53% of the U.S. population older than 50 years of age, with a prevalence of 3.4-7.6% for advanced histopathological features and 0.2-0.6% for adenocarcinomas (Strum, 2016). Hence, given that only a small percentage of adenomas becomes malignant, we distinguish between progressive and non-progressive adenomas, where non-progressive adenomas do not develop into an adenocarcinoma (see Figure 1). We also distinguish between clinical and preclinical stages, where preclinical indicates that the cancer is not yet diagnosed. Preclinical cancer can then progress from stage I to stage IV, where symptoms may develop in each stage, which in turn may lead to disease diagnosis (Compton and Greene, 2004). Once the cancer has been diagnosed, the cancer is referred to as clinical.

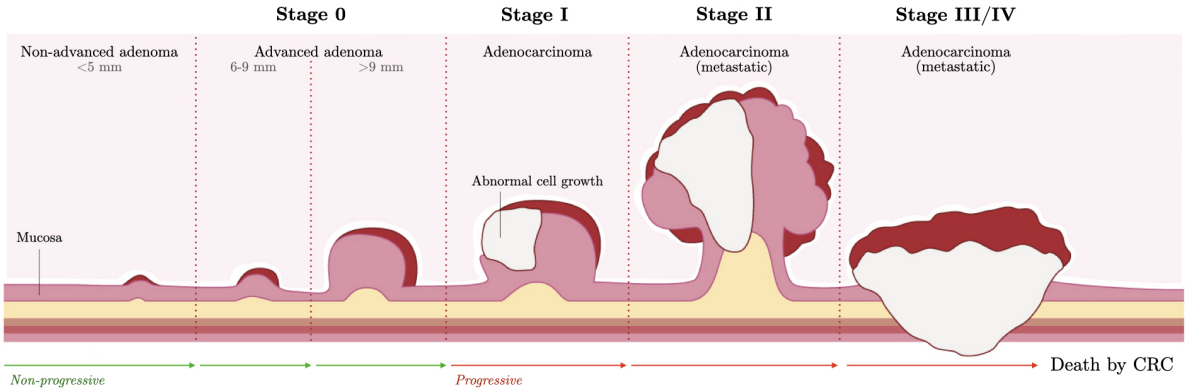

Figure 1: Progression of colorectal cancer in stages

2.1.1 Screening

The effect of screening is twofold. First, existing evidence indicates that most colorectal cancers eventually develop from adenomas. Hence, early detection and removal might prevent CRC (Loeve et al., 1999). Second, early detection of an (a)symptomatic cancer may result in an improvement in prognosis. More specifically, a large body of literature finds that screening results in a reduction in mortality, as cancers can be detected at an early and curable stage (Jiang et al., 2022; Levin et al., 2008; Whitlock et al., 2012; Toribara and Sleisenger, 1995).

Screening tests can be subdivided into two categories: stool-based tests and visual exams. The guaiac-based fecal occult blood test (gFOBT) and fecal immunochemical test (FIT) belong to the first category, in which the stool is tested for cancer. The most common visual exams are (flexible) sigmoidoscopy, and colonoscopy, which investigate the structure of the colon and rectum for abnormal tissue. According to the review by Ding et al. (2022), a colonoscopy is most effective in reducing CRC-related deaths, at an approximate 68% decrease (Brenner et al., 2014). The FIT test is on average 7% more effective in reducing death numbers compared to the gFOBT test (14-16%) (Hewitson et al., 2008; Zorzi et al., 2015). The FIT test also has a higher participation rate and positivity rate compared to gFOBT in the CRC screening programs, while reporting fewer false negatives (Mousavinezhad et al., 2016). Moreover, the FIT test is relatively close in effectiveness compared to flexible sigmoidoscopies with reported reduction of approximately 28%, while being considerably less invasive (Holme et al., 2013). The general advice when you choose to be screened with a test other than colonoscopy, is that any abnormal test result should be followed up with a timely colonoscopy (Ding et al., 2022).

In the Netherlands, each person between the age of 55-75 is asked to participate in the biennial population screening for bowel cancer once every two years since January of 2014[3]. The participants receive a faecal immunochemical test (FIT), which is sent back to the hospital after

[3]For more information see: https://www.rivm.nl/darmkanker.

taking a stool sample. In the event of an aberrant result, a referral is made for a colonoscopy and, if necessary, treatment. During the colonoscopy, if any abnormalities are present, small amounts of tissue can be removed for analysis (i.e., a biopsy), and abnormal growths, or adenomas, can be identified and removed. This way, colon cancer can be detected at an early stage. According to the Integraal Kankercentrum Nederland, patients diagnosed with colorectal cancer through the population screening had a more favorable stage distribution than patients without screening (see Figure 2). Also, patients who were diagnosed through population screening were more likely to receive less invasive treatments.

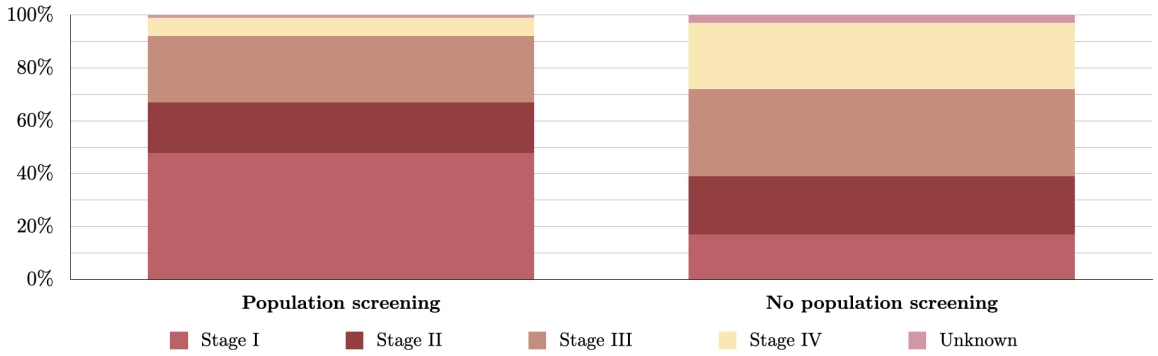

Figure 2: Distribution of diagnosed cancers in patients with, and without screening (source: https://iknl.nl/)

2.2 MISCAN-Colon

Welch and Black (2010) provide a summary of current evidence that early detection leads to overdiagnosis in breast, lung, and prostate cancer, where overdiagnosis is defined as the diagnosis of a medical condition or disease that would not cause symptoms or death during a patient's lifetime. Overdiagnosis is associated with long-term psychosocial harm, lower quality of life, and unwanted/unnecessary usage of (follow-up) tests, treatment, and healthcare facilities (Barton et al., 2001; Brodersen and Siersma, 2013; Jenniskens et al., 2017; Van der Steeg et al., 2011). In contrast, Brasso et al. (2010) and Wardle et al. (2003) find no adverse psychological effects due to cancer screening, although they do not specifically investigate the effects of overdiagnosis.

Overdiagnosis could be particularly harmful if it leads to unnecessary treatments, each of which comes with their specific risk. For an assessment of operative risk in CRC surgery, we refer to Fazio et al. (2004).

Given the previously stated disadvantages to screening, it is of high importance to identify the most cost-efficient and the most effective screening policy. To this end, the MISCAN-Colon (MIcrosimulaten SCreening ANalysis) microsimulation model was developed by the Erasmus

University Medical Center of Rotterdam. After modeling CRC in a population, it can be used to evaluate different screening policies, and surveillance strategies following polyp removal (Loeve et al., 1999). This model is an adapted version of Habbema et al. (1985)'s MISCAN microsimulation model for the evaluation of screening.

The model takes demographic characteristics, the epidemiology and natural history of the disease, and the characteristics of screening as input, and subsequently simulates a large number of individual life histories in which several colorectal lesions can emerge. By comparing the simulated life histories with and without screening, MISCAN-Colon can evaluate the costs and benefits of a specific screening strategy.

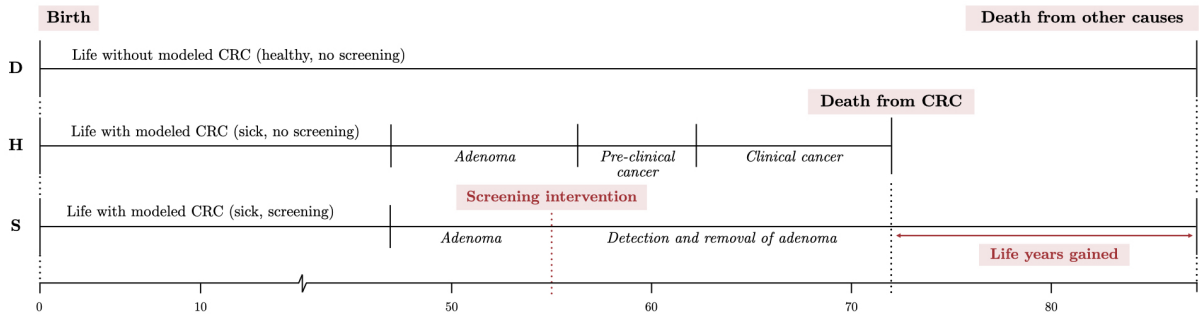

Figure 3: Simulations from the MISCAN-Colon model, where the upper bar shows the demography part (D), the middle bar adds the natural history (H) to D, and the lower bar adds both H and screening (S)

Figure 3 shows an exemplified version of the three parts of MISCAN-Colon, using a fictive individual named Robin. The upper line simulates the life of Robin without cancer, referred to as the demography part, who dies at 87 years old of other causes than CRC. The middle line simulates Robin's life *with* cancer, but without screening, which adds a natural history to the demography part. In this scenario, Robin dies at 72 due to CRC. The bottom line simulates Robin's life when screening is overlayed. As a result, Robin gains 15 life years.

Three remarks on Figure 3. First, the survival of a lesion after diagnosis depends on the stage of the cancer (and other risk factors). Thus, screening does not ensure that an individual no longer dies from CRC. The possible prognostic consequences after a positive test result for cancer screening are: total cure, delay in moment of death, no change in moment of death, or premature death by complications of treatment. Second, the figure only shows examples of individuals with *one* lesion for simplicity, but the MISCAN-Colon model allows for the modelling of multiple lesions. New lesions that appear after clinical diagnosis of cancer are accounted for in the simulated survival of the clinically diagnosed cancer. Third, this figure only shows adenomas that progress to cancer, but it is also possible that an individual develops non-lethal adenomas which would never result in death of an individual, and it is possible that a lesion is invasive

from the beginning, i.e., a cancer without a preceding polyp.

3 Data

All data for this research is obtained from the Dutch CRC Screening Program during the period 2014-2021. Four rounds of data are available from the biennial screening with the FIT test. Only persons who consistently responded to the invitations for screening and any follow-up examination with colonoscopy were included. Persons who did not respond to one of the invitations were excluded. Persons for whom the results of any follow-up examination was missing are also excluded. Table 1 shows the original variables included in the data set.

Table 1: Original variables in the data set provided by the Erasmus Medical Centre

Variable	**Description**	**Range**
`Age`	Age of respondent at time of screening	55 − 78
`Bloodtest result`	Indicator for result of screening bloodtest	0 (Favourable), 1 (Unfavourable)
`Haemoglobin current`	Hb value	0 − 306
`Haemoglobin threshold`	Threshold value used to determine bloodtest result	275, 88
`ID`	Personal identification	1 − 2,493,999
`Round`	Indicator for presence of individual per round	0 (Participated), 1 (Non-respondent, non-participant)
`Stage current`[1]	Stage of cancer at time of screening	1 (Healthy), 2 (Non-advanced adenoma), 3 (Advanced adenoma), 4 (Cancer stage 1), 5 (Cancer stage 2), 6 (Cancer stage 3), 7 (Cancer stage 4)
`Sex`	Gender of respondent	0 (Male), 1 (Female)

Notes: [1]`Stage current` is one-hot encoded, such that the resulting dummy variables are equal to one for the current stage of cancer, and zero otherwise.

Given that mixed effects models are virtually unexplored for NN and SVM, we must make an active effort to minimally violate the iid assumption. To this end, we introduce the additional variables described in Table 2, to allow for as much individual variation as possible.

Table 2: Variables which are added to the original data set

Variable	Description	Range
`FIT number`[1]	Indicator for sequence number of the FIT test	$1-3$
`Haemoglobin difference`	Difference between current and previously obtained Hb value at time of screening	-306 − 306
`Haemoglobin max`	Maximum obtained Hb value over all tests at time of screening	$0-306$
`Haemoglobin previous`	Previously obtained Hb value at time of screening	$0-306$

Notes: [1]`Fit number` is one-hot encoded, such that the resulting dummy variables are equal to one for the current FIT test, and zero otherwise. [2]`Stage previous` is one-hot encoded in the same way as `Stage current` (see Table 1).

4 Methodology

The currently employed ZINB model assumes that there are two distinct data generation processes. With probability p, the Hb concentration is zero and with probability of $(1-p)$ the Hb concentration is drawn from a mixed-effects negative binomial distribution. That is, conditional on an (assumed iid normal) subject-specific risk factor γ, the model assumes that the responses y_{it} for a single subject i are independent and follow a distribution from the exponential family with mean: $E(y_{it}|\gamma_i) = \mu_{it} = \exp(\eta_{it})$, where $\eta_{it} = \beta' x_{it} + \gamma_i$, using $g(\cdot) = h^{-1}(\cdot) = exp^{-1}(\cdot)$ as link function. Here β is the vector of population fixed-effect coefficients.

Ngufor et al. (2019) propose a mixed-effect machine learning (MEml) framework. The MEml model estimates the fixed-effects component $(\beta' x_{it})$ using machine learning algorithms. Thus η_{it} is now defined as $\eta_{it} = f(x_{it}) + \gamma_i$. They estimate $f(\cdot)$ using only tree based algorithms for interpretability, however they demonstrate that any supervised learning algorithm can be used. Therefore, I wish to contribute to the existing literature by testing the performance of other machine learning algorithms. The machine learning algorithms I want to investigate are: XGBoost, Support Vector Regression and Neural Network. The reason being that these machine learning algorithms are known for their high accuracy (and consequently low interpretability).

The first method I intend to use the XGBoost (XGB) algorithm developed by Chen and Guestrin (2016) as a high-performing prediction model. XGB is a gradient boosting algorithm that classifies through majority voting of sequentially built shallow decision trees, while updating the importance weights of observations at every tree. The XGB algorithm minimizes a negative log-likelihood loss function that measures the difference between the prediction $\hat{y}_i$ and the true outcome y_i for each individual, using $\Omega(f_s)$ as a regularization term that penalizes the complexity of each regression tree $f_s \in (f_1, ..., f_q)$. Clearly, the majority voting and re-weighting make for a

highly opaque algorithm. Although the decisions of the algorithm can be presented in a simpler form through approximation, the real working of the algorithm cannot be easily interpreted, due to the high complexity of the probability predictions.

The second algorithm is Support Vector Machine (SVM). The SVM aims to maximize the distance between the hyperplane and the support vectors, which are the datapoints closest to both sides of the hyperplane. It is also able to handle not perfectly separable through the introduction of soft margins. Moreover, SVMs can handle highly non-linear data by using kernels. It implicitly maps the input vector to higher dimensional feature spaces by the transformation which rearranges the data set in such a way that it is linearly solvable. In the higher dimensional feature space the problem becomes a linear surface that fits the data.

The last algorithm is Neural Networks (NN). NNs have a layered structure with different nodes. This structure consists of an input layer, possibly hidden layers and an output layer, which are all connected. Each node, or artificial neuron, connects to another and has an associated weight and threshold. If the output of any individual node is above the specified threshold value, that node is activated, sending data to the next layer of the network. Otherwise, no data is passed along to the next layer of the network. These weights can be seen as the coefficients that determine the influence of the regressors. During the learning process of the NN these weights adjusts. The input layer of the NN consists of n nodes, where n is equal to the number of explanatory variables. Since we model the Hb concentration, we only have one node in the output layer. For the hidden layers and output function we need activation functions. For both for the hidden layers, and the output layer, I intend use the widely used ReLU activation function, however further investigation will have to show if this is indeed the most appropriate choice.

Moreover, since we have a large dataset, and all of the aforementioned techniques require proper tuning, which in turn would take up a lot of time. I intend to use the method proposed by Bergstra et al. (2013): Randomized Hyperopt, which is a Bayesian search among candidate values of hyperparameters. The Hyperopt method can be seen as an exploration/exploitation strategy, that starts by exploring the performance across the candidate hyperparameter space, and subsequently randomly exploits the most promising subspace of hyperparameters. For the same number of iterations, this method can lead to better hyperparameter settings than the ones of random search.

Furthermore, since the data is zero-inflated, and therefore highly unbalanced, I wish to explore state-of-the-art rebalancing techniques, either using cost functions (assuming expert knowledge is available during my internship), or using a combination of SMOTE-NC (Chawla et al., 2002),

along with either ENN (Wilson, 1972), Tomek Links (Tomek, 1976), or NearMiss, depending on the actual size of the dataset.

Also, besides using a mixed-effect machine learning framework to account for dependencies within the data, I am planning on introducing additional variables: some version of a lagged dependent variable (previous FIT values, also because Grobbee et al. (2017) find that an undetectable Hb concentration two years ago decreases the current risk of having CRC), and the difference between the current and previous test to allow for 'directional trends' so to speak (increase since previous test, or decrease since previous test).

5 Results

6 Conclusion

Further research optimization of screening This study demonstrates a positive correlation between FIT value and risk of interval cancer even for very low values. It further suggests that an increase in the screening interval could be reasonable in the low FIT categories.

Our findings could help to identify the areas that could be improved and finally optimize the CRC screening programs.

References

Barton, M. B., Moore, S., Polk, S., Shtatland, E., Elmore, J. G., and Fletcher, S. W. (2001). Increased patient concern after false-positive mammograms. *Journal of General Internal Medicine*, 16(3):150–156.

Bergstra, J., Yamins, D., and Cox, D. D. (2013). Hyperopt: A Python Library for Optimizing the Hyperparameters of Machine Learning Algorithms. In *Proceedings of the 12th Python in Science Conference*, volume 13, page 20. Citeseer.

Botteri, E., Iodice, S., Bagnardi, V., Raimondi, S., Lowenfels, A. B., and Maisonneuve, P. (2008). Smoking and Colorectal Cancer: A Meta-analysis. *Jama*, 300(23):2765–2778.

Brasso, K., Ladelund, S., Frederiksen, B. L., and Jørgensen, T. (2010). Psychological distress following fecal occult blood test in colorectal cancer screening–a population-based study. *Scandinavian Journal of Gastroenterology*, 45(10):1211–1216.

Brenner, H., Stock, C., and Hoffmeister, M. (2014). Effect of screening sigmoidoscopy and screening colonoscopy on colorectal cancer incidence and mortality: systematic review and meta-analysis of randomised controlled trials and observational studies. *British Medical Journal*, 348.

Brodersen, J. and Siersma, V. D. (2013). Long-Term Psychosocial Consequences of False-Positive Screening Mammography. *The Annals of Family Medicine*, 11(2):106–115.

Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, 16:321–357.

Chen, T. and Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 785–794. ACM.

Compton, C. C. and Greene, F. L. (2004). The Staging of Colorectal Cancer: 2004 and Beyond. *CA: A Cancer Journal for Clinicians*, 54(6):295–308.

Ding, H., Lin, J., Xu, Z., Chen, X., Wang, H. H., Huang, L., Huang, J., Zheng, Z., and Wong, M. C. (2022). A Global Evaluation of the Performance Indicators of Colorectal Cancer Screening with Fecal Immunochemical Tests and Colonoscopy: A Systematic Review and Meta-Analysis. *Cancers*, 14(4):1073.

Fazio, V. W., Tekkis, P. P., Remzi, F., and Lavery, I. C. (2004). Assessment of operative risk in colorectal cancer surgery: the Cleveland Clinic Foundation colorectal cancer model. *Diseases of the colon & rectum*, 47(12):2015–2024.

Grobbee, E. J., Schreuders, E. H., Hansen, B. E., Bruno, M. J., Lansdorp-Vogelaar, I., Spaander, M. C., and Kuipers, E. J. (2017). Association Between Concentrations of Hemoglobin Determined by Fecal Immunochemical Tests and Long-term Development of Advanced Colorectal Neoplasia. *Gastroenterology*, 153(5):1251–1259.

Habbema, J., Van Oortmarssen, G., Lubbe, J. T. N., and Van der Maas, P. (1985). The MISCAN simulation program for the evaluation of screening for disease. *Computer Methods and Programs in Biomedicine*, 20(1):79–93.

Hewitson, P., Glasziou, P., Watson, E., Towler, B., and Irwig, L. (2008). Cochrane Systematic Review of Colorectal Cancer Screening Using the Fecal Occult Blood Test (Hemoccult): An Update. *Official journal of the American College of Gastroenterology| ACG*, 103(6):1541–1549.

Holme, Ø., Bretthauer, M., Fretheim, A., Odgaard-Jensen, J., and Hoff, G. (2013). Flexible sigmoidoscopy versus faecal occult blood testing for colorectal cancer screening in asymptomatic individuals (Review). *Cochrane Database of Systematic Reviews*, (9).

Jenniskens, K., De Groot, J. A., Reitsma, J. B., Moons, K. G., Hooft, L., and Naaktgeboren, C. A. (2017). Overdiagnosis across medical disciplines: a scoping review. *BMJ open*, 7(12):e018448.

Jiang, Y., Yuan, H., Li, Z., Ji, X., Shen, Q., Tuo, J., Bi, J., Li, H., and Xiang, Y. (2022). Global pattern and trends of colorectal cancer survival: a systematic review of population-based registration data. *Cancer biology & medicine*, 19(2):175.

Levin, B., Lieberman, D. A., McFarland, B., Andrews, K. S., Brooks, D., Bond, J., Dash, C., Giardiello, F. M., Glick, S., Johnson, D., et al. (2008). Screening and Surveillance for the Early Detection of Colorectal Cancer and Adenomatous Polyps, 2008: A Joint Guideline From the American Cancer Society, the US Multi-Society Task Force on Colorectal Cancer, and the American College of Radiology. *Gastroenterology*, 134(5):1570–1595.

Loeve, F., Boer, R., van Oortmarssen, G. J., van Ballegooijen, M., and Habbema, J. D. F. (1999). The MISCAN-COLON Simulation Model for the Evaluation of Colorectal Cancer Screening. *Computers and Biomedical Research*, 32(1):13–33.

Mousavinezhad, M., Majdzadeh, R., Sari, A. A., Delavari, A., and Mohtasham, F. (2016). The effectiveness of FOBT vs. FIT: A meta-analysis on colorectal cancer screening test. *Medical Journal of the Islamic Republic of Iran*, 30:366.

Ngufor, C., Van Houten, H., Caffo, B. S., Shah, N. D., and McCoy, R. G. (2019). Mixed Effect Machine Learning: A framework for predicting longitudinal change in hemoglobin A1c. *Journal of Biomedical Informatics*, 89:56–67.

Strum, W. B. (2016). Colorectal Adenomas. *New England Journal of Medicine*, 374(11):1065–1075.

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., and Bray, F. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians*, 71(3):209–249.

Thanikachalam, K. and Khan, G. (2019). Colorectal Cancer and Nutrition. *Nutrients*, 11(1):164.

Tomek, I. (1976). Two modifications of CNN. *IEEE Transactions Systems, Man and Cybernetics*, 6:769–772.

Toribara, N. W. and Sleisenger, M. H. (1995). Screening for Colorectal Cancer. *New England Journal of Medicine*, 332(13):861–867.

Torre, L. A., Bray, F., Siegel, R. L., Ferlay, J., Lortet-Tieulent, J., and Jemal, A. (2015). Global Cancer Statistics, 2012. *CA: a cancer journal for clinicians*, 65(2):87–108.

Van der Steeg, A., Keyzer-Dekker, C., De Vries, J., and Roukema, J. (2011). Effect of abnormal screening mammogram on quality of life. *Journal of British Surgery*, 98(4):537–542.

Wardle, J., Williamson, S., Sutton, S., Biran, A., McCaffery, K., Cuzick, J., and Atkin, W. (2003). Psychological Impact of Colorectal Cancer Screening. *Health Psychology*, 22(1):54.

Welch, H. G. and Black, W. C. (2010). Overdiagnosis in cancer. *Journal of the National Cancer Institute*, 102(9):605–613.

Whitlock, E. P., Lin, J. S., Liles, E., Beil, T. L., and Fu, R. (2012). Screening for Colorectal Cancer: A Targeted, Updated Systematic Review for the U.S. Preventive Services Task Force. *Annals of Internal Medicine*, 157(2):120–134.

Wilson, D. L. (1972). Asymptotic Properties of Nearest Neighbor Rules Using Edited Data. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-2(3):408–421.

Zorzi, M., Fedeli, U., Schievano, E., Bovo, E., Guzzinati, S., Baracco, S., Fedato, C., Saugo, M., and Dei Tos, A. P. (2015). Impact on colorectal cancer mortality of screening programmes based on the faecal immunochemical test. *Gut*, 64(5):784–790.