



Research article

Predicting mortality and recurrence in colorectal cancer: Comparative assessment of predictive models

Shayeste Alinia^a, Mohammad Asghari-Jafarabadi^b, Leila Mahmoudi^{a,*},
Ghodratollah Roshanaei^c, Maliheh Safari^d

^a Department of Statistics and Epidemiology, School of Medicine, Zanjan University of Medical Sciences, Zanjan, Iran

^b Road Traffic Injury Research Center, Tabriz University of Medical Sciences, Tabriz, Iran

^c Modeling of Non-communicable Diseases Research Center, Department of Biostatistics, School of Public Health, Hamadan University of Medical Sciences, Hamadan, Iran

^d Department of Biostatistics, School of Medicine, Arak University of Medical Sciences, Arak, Iran

ARTICLE INFO

Keywords:

Cox regression
Deep learning neural network (DLNN)
Machine learning
Survival analysis
Colorectal cancer
Curative surgery
Early recurrence

ABSTRACT

Introduction: Colorectal cancer (CRC), also known as colorectal cancer, is a significant disease marked by high fatality rates, ranking as the third leading cause of global mortality. The main objective of this study was to assess the accuracy of predictive models in predicting both mortality events and the probability of disease recurrence.

Method: A retrospective analysis was conducted on a cohort of 284 individuals diagnosed with colorectal cancer between 2001 and 2017. Demographic and clinical data, including gender, disease stage, age at diagnosis, recurrence status, and treatment details, were meticulously recorded. We rigorously evaluated various predictive models, including Decision Trees, Random Forests, Random Survival Forests (RSF), Gradient Boosting, mboost, Deep Learning Neural Network (DLNN), and Cox regression. Performance metrics, such as sensitivity, positive predictive value (PPV), specificity, area under the receiver operating characteristic curve (ROC area), and overall accuracy, were calculated for each model to predict mortality and disease recurrence. The analysis was performed using R version 4.1.3 software and the Python programming language.

Results: For mortality prediction, the mboost model demonstrated the highest sensitivity at 96.9% (95% CI: 0.83–0.99) and an ROC area of 0.88. It also exhibited high specificity at 80% (95% CI: 0.59–0.93), a positive predictive value of 86.1% (95% CI: 0.70–0.95), and an overall accuracy of 89% (95% CI: 0.78–0.96). Random Forests showed perfect sensitivity of 100% (95% CI: 0.85–1) but had low specificity at 0% (95% CI: 0–0.52) and poor overall accuracy (50%). On the other hand, DLNN had the lowest performance metrics for mortality prediction, with a sensitivity of 24% (95% CI: 0.222–0.268), specificity of 75% (95% CI: 0.73–0.77), and a lower positive predictive value of 42% (95% CI: 0.38–0.45). The Gradient Boosting model showed the best performance in predicting recurrence, achieving perfect sensitivity of 100% (95% CI: 0.87–1) and high specificity at 92.9% (95% CI: 0.76–0.99). It also had a high positive predictive value of 93.3% (95% CI: 0.77–0.99). Gradient Boosting, with an ROC area of 96.4%, and mboost, with an ROC area of 75%, demonstrated remarkable performance. DLNN had the lowest performance

* Corresponding author. Department of Statistics and Epidemiology, School of Medicine, Zanjan University of Medical Sciences, Mahdavi Blvd, Zanjan, Postal Code: 4513956111, Iran.

E-mail address: leilamahmoudi7171@gmail.com (L. Mahmoudi).

<https://doi.org/10.1016/j.heliyon.2024.e27854>

Received 4 August 2023; Received in revised form 5 March 2024; Accepted 7 March 2024

Available online 12 March 2024

2405-8440/© 2024 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

metrics for recurrence prediction, with sensitivity at 1.75% (95% CI: 0.01–0.02), specificity at 98% (95% CI: 0.97–0.98), and a lower positive predictive value at 52.6% (95% CI: 0.39–0.65). *Conclusion:* In summary, the mboost model demonstrated outstanding performance in predicting mortality, achieving exceptional results across various evaluation metrics. Random Forests exhibited perfect sensitivity but showed poor specificity and overall accuracy. The DLNN model displayed the lowest performance metrics for mortality prediction. In terms of recurrence prediction, the Gradient Boosting model outperformed other models with perfect sensitivity, high specificity, and positive predictive value. The DLNN model had the lowest performance metrics for recurrence prediction. Overall, the results emphasize the effectiveness of the mboost and Gradient Boosting models in predicting mortality and recurrence in colorectal cancer patients.

Abbreviations	
CRC	Colorectal cancer
SE	Sensitivity
SP	Specificity
PPV	Positive predictive value
CI	Confidence interval
SD	Standard deviation
ROC	Receiver operating characteristic
AI	Artificial intelligence
CPH	Cox proportional hazards
ML	Machine learning
DL	deep learning
DNN	Deep Neural Network
DLNN	Deep Learning Neural Network
RSF	Random Survival Forests

1. Introduction

CRC poses a significant global health challenge, marked by high mortality rates and associated complexities. On a global scale, it stands as the third most common cancer and ranks as the fourth leading cause of cancer-related deaths [1]. According to the American Cancer Society, the percentage of colorectal cancer cases occurring in individuals under 55 years old doubled from 1995 to 2019, increasing from 11% to 20% [2]. In 2020, 12% of colorectal cancer diagnoses will involve individuals under 50 [3]. Rates have been rising since the mid-1980s in adults aged 20–39 and since the mid-1990s in adults aged 40–54. Individuals born in 1990 face twice the risk of colon cancer and four times the risk of rectal cancer compared to those born in 1950 [4].

The findings indicate a growing burden of CRC in Iran, attributed to factors like population aging and a more prevalent westernized lifestyle. The estimated number of CRC patients within 1, 2–3, and 4–5 years in 2015 was high, with a substantial increase projected for 2020. The study highlights the highest 5-year prevalence in specific age groups, with adenocarcinoma being the most prevalent histologic subtype. Conclusively, the study suggests a significant growth in the prevalence of CRC survivors in Iran in the coming years [5,6]. Risk factors for CRC encompass insufficient exercise, a high BMI, a diet rich in fats, alcohol consumption, limited fruit and vegetable intake, tobacco use, a family history of CRC, and the use of certain pharmaceutical agents, such as oral contraceptives and non-steroidal anti-inflammatory drugs (NSAIDs) [7]. An encouraging aspect of CRC is its high potential for successful treatment when detected in its early stages [7]. To improve survival rates and provide patients with a better chance of recovery, post-CRC treatment follow-up programs concentrate on early identification of disease recurrence, facilitating more effective surgical interventions [8].

In the context of this topic, ML can be formally defined as a specialized field within AI that employs software algorithms to enhance data analysis. Its primary function is to identify and comprehend patterns within extensive datasets [9]. These algorithms leverage large datasets containing input and output information to recognize patterns and progressively ‘learn,’ allowing the machine to autonomously generate recommendations or make decisions. Through iterative repetitions and algorithm refinements, the machine becomes proficient in processing input and predicting corresponding outputs [10]. Consequently, the integration of ML holds significant potential in developing a reliable risk assessment tool for young adults affected by CRC. It is noteworthy that ML techniques, when incorporating clinical risk factors, have already demonstrated successful implementation in breast cancer risk prediction, leading to a substantial improvement in predictive accuracy from 60% to 90% [11].

The Cox model, widely acknowledged and frequently used in survival analysis, is employed to explore the correlation between explanatory variables and the response variable [12]. Numerous studies consistently suggest that ML-based approaches exhibit comparable, if not superior, performance in predicting patient survival compared to conventional CPH analysis [13–15].

Deep neural networks are complex models that utilize hierarchical layers to process and manage output information. These

networks have the capability to learn and refine themselves through multiple tiers, capturing diverse aspects of input data [16]. This transformative process, known as “deep learning,” allows machines to autonomously develop models, resembling the functioning of the human brain, without explicit programming [9,17]. In a thorough systematic review, the application of DL in CRC was meticulously examined through digital image analysis of histopathological images. The study duly acknowledged the limitations associated with this approach and emphasized the importance of researchers proposing innovative solutions [18]. Another study conducted a comprehensive evaluation of ML algorithms in predicting CRC, highlighting the remarkable efficacy of ML techniques in cancer prediction [19].

The primary objective of this study was to assess the effectiveness of predictive models in accurately estimating mortality rates and disease recurrence. This topic has garnered significant interest within the medical community, particularly in the context of applying DL and ML techniques in cancer diagnosis.

2. Materials and methods

2.1. Study population and data collection

In this retrospective cohort study, 284 patients diagnosed with CRC who underwent surgical resection at the Imam Khomeini clinic in Hamadan, Iran, between 2001 and 2017 were recruited. Demographic and clinical (pathological) data were extracted from patients' medical records, with a specific emphasis on primary outcome measures, including recurrence, the time until recurrence, and mortality events.

Patient data were extracted from medical records, encompassing various demographic parameters such as gender (categorized as female or male), BMI (measured in kg/m^2), and age at diagnosis (expressed in years). Clinical factors related to surgical intervention, including chemotherapy and radiotherapy, along with morphological characteristics, were categorized into binary groups (0 for absence, 1 for presence) for analysis. Disease recurrence evaluation was based on the duration between surgery and the subsequent cancer occurrence. The number of chemotherapy sessions was classified into three categories (0 for no sessions, 1 for fewer than 6 sessions, and 2 for 6 or more sessions). Cancer site was documented as colon (coded as 1) or rectum (coded as 2), and tumor grade (differentiation level) was categorized as well (coded as 1), moderate (coded as 2), or poor (coded as 3). Disease stage was categorized as (1: B, 2: C, 3: D), and tumor size was stratified as (1: <4 , 2: $\geq 4 < 7$, 3: ≥ 7). For deceased patients during the study, contact details were collected, and a follow-up investigation recorded their current status in the records.

2.2. Statistical analysis

The statistical analysis for this study utilized the R software (version 4.1.3; available at <https://www.r-project.org/>) and Python (version 3.12; available at <https://www.python.org/>). Descriptive statistics were employed to present an overview of the characteristics of study participants. A significance level of 5% was applied for hypothesis testing. The survival duration of individuals diagnosed with CRC was measured in months. Various predictive models, including Decision Trees, Random Forests, RSF, Gradient Boosting, mboost, DLNN, and Cox regression, were assessed for predicting death and recurrence. Performance metrics, such as accuracy, sensitivity, specificity, and ROC area, were examined to evaluate the effectiveness of the predictive models. In this study, Python packages such as Keras and TensorFlow, as well as R software packages including rpart, ranger, survival, caret, gbm, mboost, and RandomForest, were employed [20–22].

3. Models

3.1. Cox proportional hazards

The CPH regression analysis is a specialized statistical method used to assess the influence of various variables, referred to as covariates, on the likelihood of an event, such as mortality. This model consists of two main components: a baseline hazard that changes over time and individual-specific characteristics (e.g., age or gender) multiplied by their respective coefficients. In the CPH model, it is crucial to note that no specific form is assumed for the hazard [23]. Furthermore, regression models employed in survival analysis often rely on assumptions that are not universally satisfied. These conventional models may inadequately capture intricate relationships and higher-order interactions [24]. We utilized the CPHs model for predicting death and disease recurrence.

3.2. Machine learning models

In recent years, ML techniques have shown significant effectiveness in various practical domains, highlighting their ability to capture complex relationships and make accurate predictions. However, in the context of survival analysis, ML methods face challenges in properly handling censored data and accurately estimating time [25].

In the realm of survival analysis, Decision Trees serve as hierarchical decision models that are created based on decision rules derived from input features. They provide a clear interpretation and explanation of rules, making them suitable for smaller-scale problems [25]. Conversely, Random Forests consist of a collection of decision trees constructed randomly, with their outcomes combined. Each tree is constructed independently, resulting in enhanced diversity and prediction accuracy. Random Forests exhibit high resistance to overfitting, making them well-suited for large-scale problems [24,25]. When it comes to analyzing survival and

censored data, RSF represent a specific type of Random Forests. RSF employs a collection of decision trees and is particularly well-suited for complex survival analysis tasks. It brings improvements to the analysis of censored data and yields more accurate results in survival analysis problems [26].

Both mboost and Gradient Boosting fall under the category of 'Boosting Algorithms,' sharing the core concept of combining weak models to form a stronger one [27]. Despite this commonality, they possess distinct characteristics. mboost is a Boosting algorithm that places significant emphasis on the distribution of responses. It constructs a potent model by amalgamating base functions with varying importance weights. An advantage of mboost lies in its ability to regulate the model's complexity by adjusting the number and type of base functions based on the specific problem and data [27]. In contrast, Gradient Boosting is a Boosting method that iteratively optimizes the objective function and its gradient. It systematically enhances the gradient of the objective function to build a new model. Typically, Gradient Boosting utilizes simple base functions like decision trees and incorporates shrinkage techniques to mitigate errors [24,25,28].

3.3. Deep Learning Neural Network

DLNNs have received considerable recognition for their impressive capability to autonomously acquire hierarchical data representations, enabling the identification of complex patterns and correlations. A key characteristic of DLNNs is their ability to efficiently learn representations, eliminating the need for labor-intensive manual feature engineering. Through the extraction of relevant features and understanding intricate data interdependencies, DLNNs demonstrate exceptional effectiveness across various tasks [18,29].

Nevertheless, DLNNs also have certain limitations that need to be considered. Firstly, they typically demand a substantial amount of labeled training data to achieve effective generalization. Insufficient data can lead to overfitting and suboptimal performance of DLNNs. Therefore, obtaining a large dataset is often crucial for successful training [30,31]. Another challenge associated with DLNNs is their computational complexity. Training DLNNs can be computationally intensive, especially when dealing with deep architectures that involve numerous parameters. This complexity often requires significant computational resources, such as powerful graphics processing units (GPUs), to efficiently manage the training process [32].

3.4. Comparison of machine learning and Deep Learning Neural Networks

ML presents clear advantages over DLNN. Firstly, ML algorithms exhibit outstanding performance on small datasets, often outperforming DLNN [33]. Secondly, ML models commonly integrate interpretable parameters, enabling researchers and users to enhance their understanding of the results and the influential factors within the model [34]. In contrast, DLNN is frequently considered a complex model with non-interpretable parameters. Lastly, ML provides greater control over training and model behavior. For example, in a random forest model, the number of trees can be easily controlled, whereas determining the number of layers and neurons in DLNN typically requires manual specification [33,34]. These advantages establish ML as a valuable approach for a wide range of applications.

To ensure the robustness of our model training, we systematically divided our dataset into two distinct subsets: the training set and the testing set. This division was randomized to mitigate potential biases. The primary goal of the training dataset was to enable our ML models to acquire intrinsic data patterns, allowing for accurate predictions. Conversely, the testing dataset was assigned for assessing model performance and evaluating their predictive capability on novel, unobserved data.

3.5. Model evaluation and performance metrics

We assigned 80% of our dataset for training purposes, reserving the remaining 20% exclusively for testing. Following the model fitting process mentioned earlier, we evaluated the model's classification performance using a confusion matrix. To create this matrix, our model was applied to the test set, and its predictions were compared with the actual labels. Our analysis involved quantifying the number of correctly classified samples in each class (true positives), samples mistakenly assigned to other classes (false positives), and samples inaccurately attributed to the specific target class (false negatives). Additionally, we calculated the count of accurately classified samples in the negative class (true negatives). These statistics were then organized into a two-dimensional matrix known as the confusion matrix, providing a comprehensive summary of the model's classification performance, taking into account both the actual classifications and the corresponding predictions.

In assessing the effectiveness and suitability of each model, we employed various diagnostic metrics, such as PPV, SP, SE, AC, and calculated the area under the ROC curve. To ensure reliability, we also incorporated a 95% confidence interval with these metrics.

3.6. Ethical considerations

In adherence to ethical guidelines and participant consent, this research protocol obtained approval from the Institutional Review Board (IRB) at Tabriz University of Medical Sciences, under the ethics code IR.TBZMED.REC.1400.457.

4. Result

4.1. Participants

In the provided study, a total of 284 participants were enrolled, with 134 (47.1%) being women. Among these participants, 121

(42.6%) experienced mortality, and 131 (46.1%) faced disease recurrence. It is noteworthy that only 16 (19.8%) of the patients who experienced recurrence survived until the study's conclusion. Additionally, 16 individuals (10.5%) encountered a terminal event without a preceding non-terminal event. The mean age at the time of diagnosis was 55.6 years (SD 13.1), ranging from 21 to 84 years. Patients with recurrence had an average age at diagnosis of 56.7 years (SD 13.4), while patients without recurrence had an average age at diagnosis of 54.7 years (SD 12.8).

In terms of survival, the analysis indicated that the median survival for the entire patient cohort was 0.61 months (95% CI: 42.2–79.8). It's important to highlight that individuals who experienced recurrence had a median survival of 47 months (CI: 21–73). The average and median durations from the non-terminal event to the terminal event were 26.2 months (CI: 19.1–33.2) and 10 months (CI: 12.2–33.2), respectively.

Moreover, it is crucial to note that within the group of patients who faced disease recurrence, a substantial number (110, representing 84%) showed the spread of the disease to other anatomical sites, highlighting the inherently aggressive nature of the condition. In this specific subgroup, it is important to mention that 76 individuals (58%) underwent more than six cycles of chemotherapy, suggesting a potential link between treatment intensity and disease progression. It is noteworthy that among patients who passed away during the study period, 94 individuals (77.7%) experienced metastases to other sites, with 61 individuals (50.4%) having undergone more than six courses of chemotherapy. Additionally, 23 patients, constituting 19% of the study cohort, were classified as stage B, while a considerably larger subset, comprising 57 patients (47.1%), were categorized as stage D.

4.2. Comparison of the models

Table 1 presents the performance metrics of various predictive models for death and recurrence in the context of the study. The models evaluated include Decision Trees, Random Forests, Survival forest with ranger, Gradient Boosting, mboost, DLNN, and Cox regression.

For the prediction of death, the mboost model demonstrated the highest sensitivity (96.9%) and area under the receiver operating characteristic curve (ROC area) of 0.88. It also exhibited high specificity (80%), positive predictive value) ppv ((86.1%), and overall accuracy (89%). Random Forests showed perfect sensitivity (100%), but low specificity (0%) and poor overall accuracy (50%). The Cox regression model demonstrated an overall accuracy of 49.1% and an ROC area of 0.543, indicating average performance in terms of accuracy and discrimination. On the other hand, DLNN showed the lowest performance metrics for death prediction, with sensitivity of 24%, specificity of 75%, and a lower positive predictive value (42%). The variable importance was assessed using the mboost model for predicting patient mortality. Among the variables under investigation, Disease_stage exhibited the highest selection frequency, indicating its significant role in mortality prediction. The remaining important variables, in descending order, were surgery, n_chemo_cat (number of chemotherapy sessions), and cancer_site. The graphical representation of these results can be observed in Fig. 1.

Table 1
Performance metrics of predictive models for death and recurrence.

	Model	SE	SP	PPV	AC	ROC area
Death	Decision Trees	78.3% (0.56–0.92)	60.6% (0.42–0.77)	58.1% (0.39–0.75)	67% (0.54–0.79)	69% (0.57–0.81)
	Random Forests	100% (0.85–1)	0% (0 - 0.52)	82.1% (0.63–0.93)	96 % (0.81–0.99)	50% (0.001–0.99)
	Random Survival Forests (RSF)	77.5% (0.66–0.86)	63.4% (0.51–0.74)	67.9% (0.56–0.77)	70.4% (0.62–0.77)	71% (0.64–0.79)
	Gradient Boosting	75.9% (0.56–0.89)	66.7% (0.46–0.83)	71% (0.52–0.85)	71.4% (0.57–0.83)	71.3% (0.59–0.83)
	mboost	96.9% (0.83–0.99)	80% (0.59–0.93)	86.1% (0.70–0.95)	89% (0.78–0.96)	88% (0.75–0.99)
	DLNN	24% (0.222–0.268)	75% (0.73–0.77)	42% (0.38–0.45 (54% (0.52–0.557 (30.5% (0.288–0.32)
	Cox regression	12.5% (0.035–0.29)	96% (0.79–0.99)	80% (0.28–0.99)	49.1% (0.35–0.62)	54.3% (0.47–0.61)
Recurrence	Decision Trees	56.1% (0.29–0.79)	47% (0.31–0.63)	43.3% (0.25–0.62)	53% (0.39–0.67)	51% (0.37–0.66)
	Random Forests	100% (0.82–1)	0% (0–0.28)	67.9% (0.47–0.84)	67.8% (0.47–0.84)	50% (0.49–0.51)
	Random Survival Forests (RSF)	77.8% (0.66–0.86 (71% (0.58–0.81)	73.7% (0.62–0.83)	74% (0.66–0.81)	74.4% (0.65–0.82)
	Gradient Boosting	100% (0.87–1)	92.9% (0.76–0.99)	93.3% (0.77–0.99)	96% (0.87–0.99)	96.4% (0.91–0.99)
	mboost	96.8% (0.83–0.99)	53.8% (0.33–0.73)	71.4% (0.55–0.84)	76 % (0.63–0.87)	75% (0.62–0.87)
	DLNN	1.75% (0.01–0.02 (98% (0.97–0.98 (52.6% (0.39–0.65 (47.4% (0.45–0.49 (26.8% (0.24–0.29)
	Cox regression	64.5% (0.45–0.80)	76.9% (0.56–0.91)	76.9% (0.56–0.91)	70% (0.56–0.81)	70.7% (0.58–0.82)

SE=Sensitivity; SP=Specificity; PPV=Positive predictive value; ROC=Receiver operating characteristics; AC= Accuracy.

The Gradient Boosting model stood out as the top performer for predicting recurrence, achieving perfect sensitivity (100%) and a high specificity of 92.9%, indicating its accurate identification of positive and negative cases. It also exhibited a high positive predictive value of 93.3%. Similarly, the mboost model demonstrated a high sensitivity of 96.8%, effectively capturing positive cases. However, its specificity was lower (53.8%) compared to Gradient Boosting. Both models showcased strong discriminatory ability and achieved high overall accuracy. Gradient Boosting, with an ROC area of 96.4%, and mboost, with an ROC area of 75%, exhibited remarkable performance. Additionally, the Cox regression model exhibited a general accuracy of 70% and an ROC area of 0.707, indicating average performance. On the other hand, DLNN showed the lowest performance metrics for recurrence prediction, with a sensitivity of 1.75%, specificity of 98%, and a lower ppv (52.6%). Based on the outcomes of the analysis of the importance of variables in the Gradient Boosting model for predicting recurrence, it can be concluded that the variables Disease_stage, BMI, tumor_size_4_7, age_cat and gender are more important in predicting recurrence among the investigated variables) [Fig. 2](#)).

5. Discussion

In our thorough investigation, we evaluated the predictive accuracy of a DLNN model and ML techniques for forecasting early recurrence and mortality subsequent to curative surgical procedures in a cohort of 284 patients undergoing resection for colorectal cancer. The study involved 47.1% female participants and presented significant findings related to mortality and disease recurrence. Precisely, 42.6% encountered mortality, and only 19.8% of those with recurrence survived until the study's completion. The median survival for the entire cohort was 0.61 months, whereas individuals with recurrence demonstrated a median survival of 47 months.

The findings of the study revealed that the mboost model achieved an impressive 89% accuracy and a notable sensitivity of 96.9%

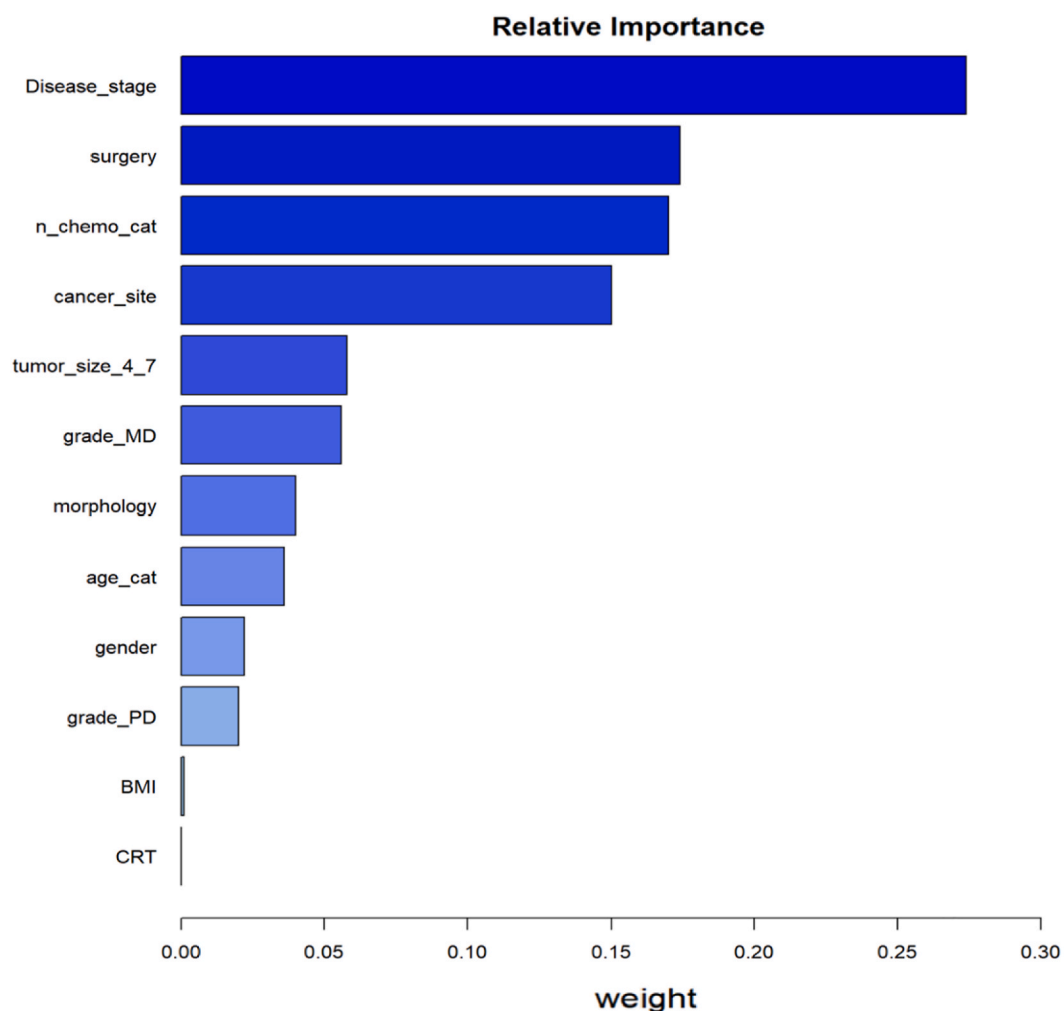


Fig. 1. The significance of the variables incorporated into the mboost machine learning model for the prediction of early mortality in colorectal cancer following tumor resection. N – chemo –cat: Number of chemotherapies; Grade_PD: Grade poorly differentiation level; Grade_MD: Grade moderately differentiation level; CRT: Radiotherapy.

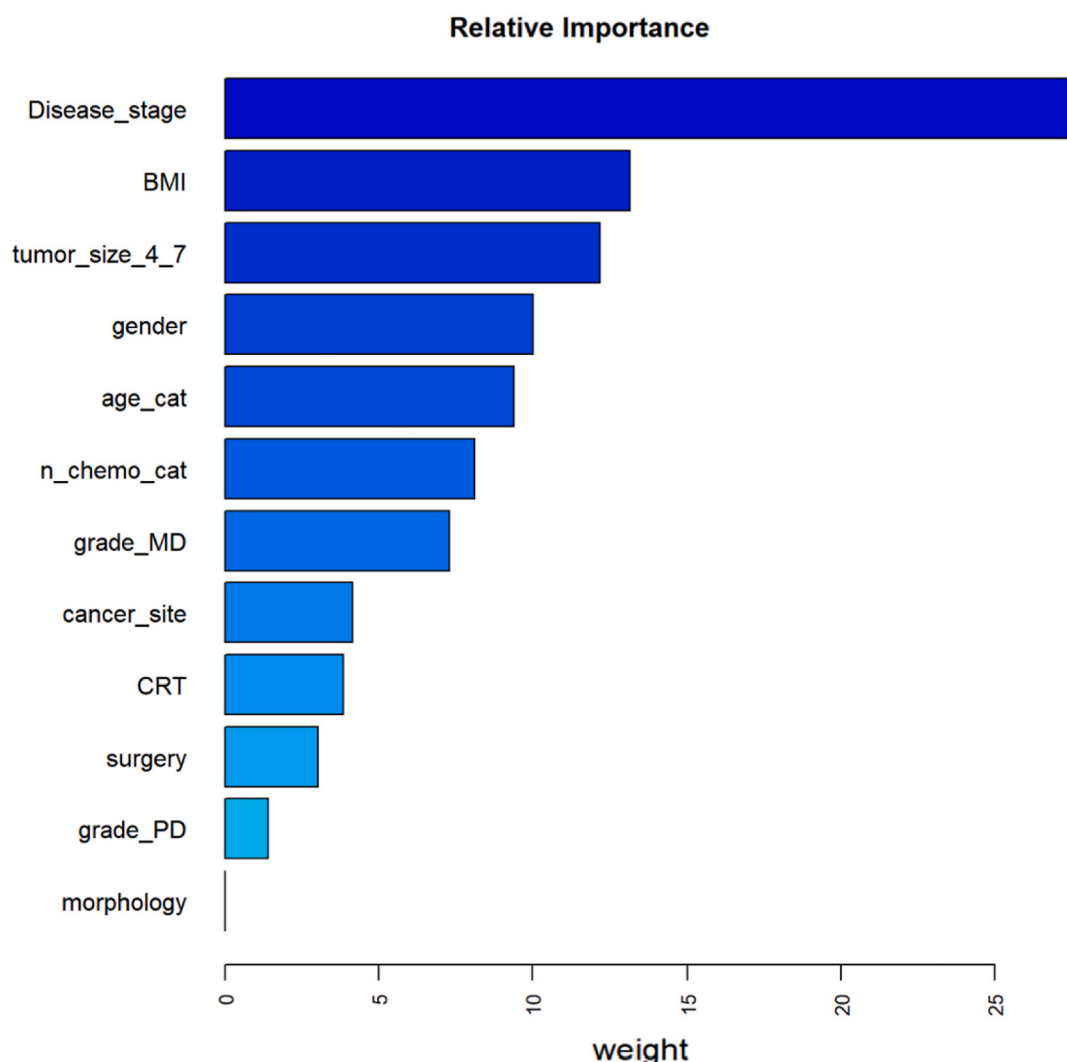


Fig. 2. The significance of individual features within the Gradient Boosting machine learning algorithm for predicting colorectal cancer recurrence following tumor resection. N – chemo –cat: Number of chemotherapies; Grade_PD: Grade poorly differentiation level; Grade_MD: Grade moderately differentiation level; CRT: Radiotherapy.

in predicting mortality. Similarly, the Gradient Boosting model demonstrated exceptional performance in forecasting recurrence, achieving a high accuracy of 96.4% (ROC area) and a perfect sensitivity of 100%. It is important to note, however, that the DLNN model exhibited comparatively lower performance in predicting both mortality (SE: 24%) and recurrence (SE: 1.75%). Nonetheless, the ML models surpassed the DLNN model by demonstrating better alignment between observed and predicted values.

The results of this investigation are consistent with previous studies in the field of predicting patient survival after colon cancer surgery [8,35–37]. For example, in 2021, Wang et al. developed a nomogram to predict overall survival following curative surgery for colon cancer. Their study highlighted the independent predictive significance of TNM stage, chemotherapy, age, and tumor size in overall survival [36]. In 2022, Høydaal Ø et al. examined the long-term survival rates of elderly individuals undergoing substantial resection with curative intent. The researchers found that the long-term survival rates among older patients were similar to those observed in their younger counterparts [38]. In a 2019 study by Ju et al., there was an emphasis on the importance of improving individual recurrence risk prediction for stage II/III colorectal cancer [39].

In a 2023 study by Khene ZE, it was found that employing ML approaches for predicting recurrence after surgical resection of Renal Cell Carcinoma (RCC) resulted in more accurate predictions compared to the validated models commonly used in clinical practice [40]. Similarly, Sun H et al., 's 2023 study highlighted that ML algorithms offer improved alternatives for clinically estimating the survival probability of patients with Laryngeal Squamous Cell Carcinoma [41]. Another investigation by Mehedi Hassan et al. (2023) explored the effectiveness of ML algorithms in breast cancer detection, comparing them with the LASSO method. The Random Forest algorithm demonstrated the highest accuracy at 90.68%, indicating its potential for precise detection [42].

In a recent review conducted by Dabiah Alboaneen and colleagues (2023), 42 studies on the utilization of ML and DL in the

detection and diagnosis of CRC were examined. The review highlights the significant impact of artificial intelligence in healthcare, especially in enhancing predictive capabilities for CRC and other malignancies through ML and DL algorithms [43]. Xuhai Zhao et al. (2022) conducted research comparing ML models and a nomogram for predicting distant metastasis in male breast cancer patients. The XGB model outperformed other models and the nomogram, demonstrating superior AUC values and accurate predictions of distant, bone, and lung metastasis [44].

In a thorough investigation, the XGB model demonstrated robust predictive capabilities for bone and lung metastasis [45,46]. Moreover, its predictive efficacy exceeded that of the nomogram in predicting lymph node metastasis among patients with invasive micropapillary breast cancer [47]. De Bin et al. (2023) introduced innovative strategies for analyzing high-dimensional data with time-to-event outcomes. They applied a gradient boosting algorithm to fit the FHT model, presenting a novel framework that overcomes limitations of conventional methods and enhances data analysis [48]. In a separate study, Sameera Senanayake et al. (2019) explored ML methods for predicting transplant outcomes in kidney transplantation. Their approach incorporated artificial neural networks, decision trees, and Bayesian belief networks. However, the performance of these ML methods varied compared to traditional regression methods [49].

The research presented in this study is part of a larger series of investigations aimed at providing a comprehensive understanding of various research topics. This integrative approach is applicable to other sets of research studies as well. The combined results of these studies reveal important predictive variables for recurrence, including disease stage, BMI, tumor size, age, and gender. Additionally, significant variables for predicting mortality, such as surgical interventions, the number of chemotherapy sessions, and cancer site as determined by the optimized model, are emphasized. The synthesized information derived from these findings has the potential to inform healthcare policy-making and shape medical and health recommendations.

5.1. Strengths and limitations

This study has several notable strengths. Firstly, it employed appropriate statistical methods, such as confidence intervals and ROC curve analysis, to assess and compare model performance. Secondly, machine learning techniques, including mboost and Gradient Boosting, were utilized to predict mortality and recurrence, demonstrating exceptional accuracy and predictive power. Additionally, the analysis of features and their significance in predicting mortality and recurrence enhances our understanding of influential factors in these events, providing insights for improving prediction models. Moreover, the use of ML models facilitates the examination and comparison of various variables and their impact on predicting mortality and recurrence in cancer. The research also evaluated the effectiveness of different models in forecasting both mortality and recurrence, contributing to the identification of the most proficient models or those exhibiting comparative superiority.

Nevertheless, there are several constraints and limitations to consider in this study. Firstly, the study included a limited number of participants, with a sample size of only 284 individuals. As a result, this restricted sample size may not offer a comprehensive and representative understanding of mortality and recurrence prediction in colorectal cancer. Moreover, this study focused on specific samples; therefore, the findings may not be universally applicable to the entire population of individuals diagnosed with colorectal cancer. Consequently, additional studies with larger and more diverse populations are necessary to validate the findings. Furthermore, the study design was retrospective, which may introduce biases and limit the ability to establish causal relationships. Prospective studies with longitudinal follow-up would provide more robust evidence.

6. Conclusion

In conclusion, this study has demonstrated that the utilization of ML methods, specifically mboost and Gradient Boosting models, provides a high level of accuracy and discriminative ability in predicting mortality and recurrence in colorectal cancer. Notably, influential factors including disease staging, surgery, the number of chemotherapy sessions, and patient characteristics significantly contribute to predicting these events. This study underscores the importance of thorough investigation and comparison of various machine learning models in this field, as it can lead to substantial improvements in the diagnosis and prediction of mortality and recurrence in colorectal cancer. Therefore, incorporating these models and the identified influential variables from this study can be beneficial in enhancing patient management and improving healthcare quality in colorectal cancer.

In alignment with the findings of this inquiry, several suggestions are proposed for future research. Firstly, we recommend conducting research encompassing larger sample sizes and more heterogeneous populations to enhance the generalizability of the findings. This approach would allow for a more comprehensive and accurate assessment of the factors influencing the prediction of death and recurrence in cancer. Secondly, investigating various variables and their associations with death and recurrence in cancer is crucial for improving prognostic capabilities. Future studies could focus on factors such as age, gender, disease stages, treatment-related side effects, and other cancer-specific variables. Furthermore, the incorporation of genetic data should be considered. Genetic information can offer valuable insights into an individual's susceptibility to cancer and their response to medical interventions. The integration of genetic data into statistical models shows potential for improving predictive accuracy and enabling personalized treatment strategies. Additionally, it is wise to explore alternative survival models. While the Cox proportional hazards model is widely employed in survival analysis, the accelerated failure time model and the competing risks model, among others, may present more suitable options in specific circumstances. A comprehensive exploration of these models could yield novel perspectives on cancer-related outcomes.

Financial Support

It is crucial to emphasize that external funding or support was not provided for this study.

Accessibility to data and materials

The data that support the findings of this study are available from LM, but restrictions are applied to the availability of these data, which were used under license for the current study, and are not publicly available. Data are, however, available from the authors upon reasonable request by LM.

CRediT authorship contribution statement

Shayeste Alinia: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Mohammad Asghari-Jafarabadi:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Leila Mahmoudi:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Ghodratollah Roshanaei:** Data curation. **Maliheh Safari:** Data curation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- [1] E. Dekker, P.J. Tanis, J.L.A. Vleugels, P.M. Kasi, M.B. Wallace, Colorectal cancer, *Lancet* 394 (10207) (2019) 1467–1480.
- [2] American Cancer Society, Colorectal Cancer Facts & Figures 2023–2025, vol. 66, American Cancer Society, Atlanta, 2020, pp. 1–41.
- [3] R.L. Siegel, K.D. Miller, A. Goding Sauer, S.A. Fedewa, L.F. Butterly, J.C. Anderson, et al., Colorectal cancer statistics, 2020, *CA Cancer J Clin* 70 (3) (2020) 145–164.
- [4] A.N. Willauer, Y. Liu, A.A.L. Pereira, M. Lam, J.S. Morris, K.P.S. Raghav, et al., Clinical and molecular characterization of early-onset colorectal cancer, *Cancer* 125 (12) (2019) 2002–2010.
- [5] H. Vardanjani, A. Haghdoost, K. Bagheri-Lankarani, M. Hadipour, Estimation and projection of prevalence of colorectal cancer in Iran, 2015–2020, *Adv. Biomed. Res.* 7 (1) (2018) 20.
- [6] N. Izadi, F. Koohi, M. Safarpour, P. Naseri, S. Rahimi, S. Khodakarim, Estimating the cure proportion of colorectal cancer and related factors after surgery in patients using parametric cure models, *Gastroenterol Hepatol from Bed to Bench* 13 (2) (2020) 125–132.
- [7] M. Lazzaroni, F. Bellerba, M. Calvello, F. Macrae, A.K. Win, M. Jenkins, et al., A meta-analysis of obesity and risk of colorectal cancer in patients with lynch syndrome: the impact of sex and genetics, *Nutrients* 13 (5) (2021 May).
- [8] S. Alinia, M. Asghari-Jafarabadi, L. Mahmoudi, S. Norouzi, M. Safari, G. Roshanaei, Survival prediction and prognostic factors in colorectal cancer after curative surgery: insights from cox regression and neural networks, *Sci. Rep.* 13 (1) (2023 Sep) 15675.
- [9] J.M. Helm, A.M. Swiergosz, H.S. Haerberle, J.M. Karnuta, J.L. Schaffer, V.E. Krebs, et al., Machine learning and artificial intelligence: definitions, applications, and future directions, *Curr Rev Musculoskelet Med* 13 (1) (2020 Feb) 69–76.
- [10] S.A. Bini, Artificial intelligence, machine learning, deep learning, and cognitive computing: what do these terms mean and how will they impact health care? *J. Arthroplasty* 33 (8) (2018 Aug) 2358–2361.
- [11] C. Ming, V. Viassolo, N. Probst-Hensch, P.O. Chappuis, I.D. Dinov, M.C. Katapodi, Machine learning techniques for personalized breast cancer risk prediction: comparison with the BCRAT and BOADICEA models, *Breast Cancer Res.* 21 (1) (2019 Jun) 75.
- [12] S.V. Deo, V. Deo, V. Sundaram, Survival analysis—part 2: cox proportional hazards model, *Indian J. Thorac. Cardiovasc. Surg.* 37 (2) (2021) 229–233.
- [13] J. Peng, Y. Lu, L. Chen, K. Qiu, F. Chen, J. Liu, et al., The prognostic value of machine learning techniques versus cox regression model for head and neck cancer, *Methods* 205 (2022) 123–132.
- [14] J.L. Katzman, U. Shaham, A. Cloninger, J. Bates, T. Jiang, Y. Kluger, DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network, *BMC Med. Res. Methodol.* 18 (1) (2018 Feb) 24.
- [15] D.W. Kim, S. Lee, S. Kwon, W. Nam, I.H. Cha, H.J. Kim, Deep learning-based survival prediction of oral cancer patients, *Sci. Rep.* 9 (1) (2019) 1–10.
- [16] L. Zhao, D. Feng, Deep neural networks for survival analysis using pseudo values, *IEEE J Biomed Heal Informatics* 24 (11) (2020) 3308–3314.
- [17] P.N. Ramkumar, J.M. Karnuta, S.M. Navarro, H.S. Haerberle, R. Iorio, M.A. Mont, et al., Preoperative prediction of value metrics and a patient-specific payment model for primary total hip arthroplasty: development and validation of a deep learning model, *J. Arthroplasty* 34 (10) (2019 Oct) 2228–2234.e1.
- [18] A. Davri, E. Birbas, T. Kanavos, G. Ntritsos, N. Giannakeas, A.T. Tzallas, et al., Deep learning on histopathological images for colorectal cancer diagnosis: a systematic review, *Diagnostics* 12 (4) (2022 Mar).
- [19] K. Swanson, E. Wu, A. Zhang, A.A. Alizadeh, J. Zou, From patterns to patients: advances in clinical machine learning for cancer diagnosis, prognosis, and treatment, *Cell* 186 (8) (2023 Apr) 1772–1791.
- [20] S.P. Antonio Gulli, Deep Learning with Keras. Illustrate, Packt Publishing, 2017, p. 318.
- [21] CRAN Task View, Machine Learning & Statistical Learning, 2023.
- [22] M. Abadi, others, TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems, 2016.
- [23] A. Moncada-Torres, M.C. van Maaren, M.P. Hendriks, S. Siesling, G. Geleijnse, Explainable machine learning can outperform Cox regression predictions and provide insights in breast cancer survival, *Sci. Rep.* 11 (1) (2021) 1–13.
- [24] H. Zhou, W. Fan, D. Qin, P. Liu, Z. Gao, H. Lv, et al., Development, validation and comparison of artificial neural network and logistic regression models predicting eosinophilic chronic rhinosinusitis with nasal polyps, *Allergy Asthma Immunol Res* 15 (1) (2023 Jan) 67–82.
- [25] P. Wang, Y. Li, C.K. Reddy, Machine learning for survival analysis: a survey, *ACM Comput. Surv.* 51 (6) (2019) 1–39.
- [26] Y. Chen, G. Li, W. Jiang, R.C. Nie, H. Deng, Y. Chen, et al., Prognostic risk factor of major salivary gland carcinomas and survival prediction model based on random survival forests, *Cancer Med.* 12 (9) (2023 May) 10899–10907.

- [27] N. Bansal, D. Singh, M. Kumar, ScienceDirect Computation of energy across the type-C piano key weir using gene expression programming and extreme gradient boosting (XGBoost) algorithm, *Energy Rep.* 9 (2023) 310–321.
- [28] N. MacNell, L. Feinstein, J. Wilkerson, P.M. Salo, S.A. Molsberry, M.B. Fessler, et al., Implementing machine learning methods with complex survey data: lessons learned on the impacts of accounting sampling weights in gradient boosting, *PLoS One* 18 (1) (2023) e0280387.
- [29] N. Someeh, M. Asghari Jafarabadi, S.M. Shamshirgaran, F. Farzipoor, The outcome in patients with brain stroke: a deep learning neural network modeling, *J Res Med Sci Off J Isfahan Univ Med Sci* 25 (2020) 78.
- [30] B. Lee, S.H. Chun, J.H. Hong, I.S. Woo, S. Kim, J.W. Jeong, DeepBTS : prediction of recurrence- free survival of non-small cell lung cancer using a time-binned deep neural network, *Sci. Rep.* 1–10 (2020).
- [31] R.W. Pettit, R. Fullem, C. Cheng, C.I. Amos, Artificial intelligence, machine learning, and deep learning for clinical outcome prediction, *Emerg Top Life Sci* 5 (6) (2021 Dec) 729–745.
- [32] Z.A. Alhulaybi, M.A. Martuza, Modeling the Mechanical Properties of a Polymer-Based Mixed-Matrix Membrane Using Deep Learning Neural Networks, 2023.
- [33] S. Huang, J. Yang, S. Fong, Q. Zhao, Artificial intelligence in cancer diagnosis and prognosis: opportunities and challenges, *Cancer Lett.* 471 (2020) 61–71.
- [34] D. Li, J. Hu, L. Zhang, L. Li, Q. Yin, J. Shi, et al., Deep learning and machine intelligence: new computational modeling techniques for discovery of the combination rules and pharmacodynamic characteristics of Traditional Chinese Medicine, *Eur. J. Pharmacol.* 933 (2022 Oct) 175260.
- [35] T. Hong, D. Cai, L. Jin, Y. Zhang, T. Lu, D. Hua, et al., Development and validation of a nomogram to predict survival after curative resection of nonmetastatic colorectal cancer, *Cancer Med.* 9 (12) (2020 Jun) 4126–4136.
- [36] S. Wang, Y. Liu, Y. Shi, J. Guan, M. Liu, W. Wang, Development and external validation of a nomogram predicting overall survival after curative resection of colon cancer, *J. Int. Med. Res.* 49 (5) (2021 May) 3000605211015023.
- [37] M.R. Saraiva, I. Rosa, I. Claro, Early-onset colorectal cancer: a review of current knowledge, *World J. Gastroenterol.* 29 (8) (2023 Feb) 1289–1303.
- [38] Ø. Høydahl, T.-H. Edna, A. Xanthoulis, S. Lydersen, B.H. Endreseth, The impact of age on rectal cancer treatment, complications and survival, *BMC Cancer* 22 (1) (2022 Sep) 975.
- [39] C. Xin, Y. Lai, L. Ji, Y. Wang, S. Li, L. Hao, et al., A novel 9-gene signature for the prediction of postoperative recurrence in stage II/III colorectal cancer, *Front. Genet.* 13 (2022) 1097234.
- [40] Z.-E. Khene, P. Bigot, N. Doumerc, I. Ouzaid, R. Boissier, F.-X. Nouhaud, et al., Application of machine learning models to predict recurrence after surgical resection of nonmetastatic renal cell carcinoma, *Eur Urol Oncol* 6 (3) (2023 Jun) 323–330.
- [41] H. Sun, S. Wu, S. Li, X. Jiang, Which model is better in predicting the survival of laryngeal squamous cell carcinoma?: comparison of the random survival forest based on machine learning algorithms to Cox regression: analyses based on SEER database, *Medicine (Baltim.)* 102 (10) (2023 Mar) e33144.
- [42] M.M. Hassan, M.M. Hassan, F. Yasmin, M.A.R. Khan, S. Zaman, Galibuzzaman, et al., A comparative assessment of machine learning algorithms with the Least Absolute Shrinkage and Selection Operator for breast cancer detection and prediction, *Decis. Anal. J* 7 (2023 Jun 1) 100245.
- [43] D. Alboaneen, R. Alqarni, S. Alqahtani, M. Alrashidi, R. Alhuda, Predicting Colorectal Cancer Using Machine and Deep Learning Algorithms : Challenges and Opportunities, 2023.
- [44] X. Zhao, C. Jiang, The prediction of distant metastasis risk for male breast cancer patients based on an interpretable machine learning model, *BMC Med Inform Decis Mak* 23 (1) (2023 Apr) 74.
- [45] R. Lo Gullo, S. Eskreis-Winkler, E.A. Morris, K. Pinker, Machine learning with multiparametric magnetic resonance imaging of the breast for early prediction of response to neoadjuvant chemotherapy, *Breast* 49 (2020 Feb) 115–122.
- [46] C.-M. Zhou, Q. Xue, Y. Wang, J. Tong, M. Ji, J.-J. Yang, Machine learning to predict the cancer-specific mortality of patients with primary non-metastatic invasive breast cancer, *Surg. Today* 51 (5) (2021 May) 756–763.
- [47] C. Jiang, Y. Xiu, K. Qiao, X. Yu, S. Zhang, Y. Huang, Prediction of lymph node metastasis in patients with breast invasive micropapillary carcinoma based on machine learning and SHapley Additive exPlanations framework, *Front. Oncol.* 12 (2022) 981059.
- [48] R De Bin, V.G. Stikbakke, A boosting first-hitting-time model for survival analysis in high-dimensional settings, *Lifetime Data Anal.* 29 (2) (2023) 420–440.
- [49] S. Senanayake, N. White, N. Graves, H. Healy, K. Baboolal, S. Kularatna, Machine learning in predicting graft failure following kidney transplantation: a systematic review of published predictive models, *Int. J. Med. Inf.* 130 (2019 Oct) 103957.