PESQUISAS IC

Google Scholar:

Google Scholar: Machine Learning for colorectal cancer survival prediction

1: ⭐⭐⭐
Machine learning as a new horizon for colorectal cancer risk prediction? A systematic review

2: ⭐⭐⭐
Robust Machine Learning for Colorectal Cancer Risk Prediction and Stratification

3: ⭐⭐⭐
Machine learning for predicting survival of colorectal cancer patients | Scientific Reports

4:
Early Colorectal Cancer Detected by Machine Learning Model Using Gender, Age, and Complete Blood Count Data

5:
Prediction of Colon Cancer Stages and Survival Period with Machine Learning Approach

6:
Machine Learning Model for Predicting Postoperative Survival of Patients with Colorectal Cancer

7:
Predicting Colorectal Cancer Using Machine and Deep Learning Algorithms: Challenges and Opportunities

8:
Machine Learning in Colorectal Cancer Risk Prediction from Routinely Collected Data: A Review
9:

A multi-omics machine learning framework in predicting the survival of colorectal cancer patients - ScienceDirect

10:

PubMed:

Pesquisas Iniciais:

PubMed: Colorectal cancer prediction

PubMed: Prediction cancer survival

PubMed: Machine Learning in survival prediction

1:
Inflammation-Related Biomarkers for the Prediction of Prognosis in Colorectal Cancer Patients

2:
LASSO-Based Machine Learning Algorithm for Prediction of Lymph Node Metastasis in T1 Colorectal Cancer

3:
Development of a novel combined nomogram model integrating deep learning-pathomics, radiomics and immunoscore to predict postoperative outcome of colorectal cancer lung metastasis patients

4:
Patient-Derived Organoids from Colorectal Cancer with Paired Liver Metastasis Reveal Tumor Heterogeneity and Predict Response to Chemotherapy

5:
Deep learning model for the prediction of microsatellite instability in colorectal cancer: a diagnostic study

6: ⭐⭐⭐
[Prediction models applying machine learning to oral cavity cancer outcomes: A systematic review]

- Of the five studies that compared the performance of different classifiers (with ensemble machine learning models – decision tree or random forest featuring in all studies), decision tree was the algorithm of choice (based on predictive accuracy) in two studies [24,45] (page 5)

- In all, the accuracy of machine learning models for predicting oral cavity cancer recurrence status was variable ranging across all cadre of the rating scale from 64 to 100%. Two models based on decision tree showed satisfactory to excellent prediction accuracy [24,45] compared to other individual models reported. Nonetheless, these machine learning models performed better than other tumor recurrence classification or prediction systems in studies reporting on the latter [24,25,43,45].

- Prediction of locoregional tumor relapse as a discrete variable was the most evaluated outcome in the majority of the studies (page 5)

- While for those incorporating time-to-event data, the concordance indices of the machine learning models were higher than those of the Cox proportional hazard mode (page 8)

- Some of the most popular machine learning algorithms used in studies included in this review include Logistic Regression, Support Vector Machines, Decision Trees, and Bayesian Networks (page 11)

- Logistic Regression is a type of supervised machine learning algorithm used for classification of discrete outcomes using a statistical approach where prediction of a dependent variable is based on observations given in the training set. Advantages are that the predicted parameters give inference about the importance of each feature and therefore the scientist/ clinician looking at the results can decide on a scientific basis whether that particular feature is an anomaly or clinically plausible. In addition to this, Logistic Regression outputs well-calibrated probabilities along with the classification results. The disadvantages to using the Logistic Regression is that is it very difficult to capture complex relationships, and therefore this may require the researchers themselves to manually decide which feature they want to enter into the model to prevent degradation of the model's predictive value. (page 11)

- Support Vector Machines (SVM) classifies data using a hyperplane which is similar to a decision boundary between different classes. The extreme data point from each class are called Support Vectors. Since there are regularisation capabilities, Support Vector Machines has good generalisation capabilities which prevent over-fitting new datasets, and small changes to the dataset will not greatly affect the hyperplane delineation. Problems with SVM include a long training time for large datasets as well as difficulty to understand and interpret findings.

- Decision Tree models are also commonly used in medicine to predict outcomes. Decision Trees present a decision-making algorithm in a flowchart and can be understood by dividing the source dataset depending on a quality worth testing. The advantages to using a Decision Tree model is that it can handle high-dimensional or multidimensional data with great precision and normalisation of the data pre-analysis is not required. The model can handle both continuous and categorical variables as well as missing data points. Disadvantages include having poor reproducibility as the Decision Tree is very sensitive to slight changes in data which can result in large differences in the tree structure. It is also a poor model for continuous outcomes. In Tseng's study, the Decision Tree was the best performing model when using a discrete outcome of the number of recurrences [45]

- Bayesian Networks are probabilistic models that depict the conditional dependence of different variables in a graphical form. The graphical form in which information is presented allows visualisation of probabilities, understanding and analysis of the relationship between random variables. Disadvantages to the Bayesian Network is that it performs poorly on high-dimensional data. For example, in the study done by Exarchos and colleagues [40], the Bayesian Network achieved excellent validity measures when data such as clinical imaging, and molecular features were inputted into the model. Machine learning and artificial intelligent systems have been increasingly used in medicine to provide solutions to automated decision support systems for treatment personalization, as well as other tasks that improve the efficiency of the healthcare system [54,55]. Specifically in oncology research, machine learning has demonstrated success in cell type classification and treatment outcome prediction in cancers such as breast and prostatic carcinoma [56]. In oral cavity squamous cell carcinoma, in addition to prediction of disease-free survival, overall survival, recurrence, and metastasis, machine learning has been used to differentiate non-cancerous and malignant tissues for diagnosis, primarily using routinely available clinical information [32,57-61]. All these applications provide health-workers with information to plan better patient care.

7:
Development of a Machine Learning Model for Survival Risk Stratification of Patients With Advanced Oral Cancer | Oncology | JAMA Network Open

8:
DeepProg: an ensemble of deep-learning and machine-learning models for prognosis prediction using multi-omics data

9:
Prediction of driver variants in the cancer genome via machine learning methodologies

10:
A review of machine learning approaches for drug synergy prediction in cancer

11:
From patterns to patients: Advances in clinical machine learning for cancer diagnosis, prognosis, and treatment

12:
Predicting factors for survival of breast cancer patients using machine learning techniques

13:
Machine Learning and Deep Learning Approaches in Breast Cancer Survival Prediction Using Clinical Data

14:
Prediction of lung cancer patient survival via supervised machine learning classification techniques

15: (Pago)
Prediction of Cancer Treatment Using Advancements in Machine Learning

16:
Machine learning for genetics-based classification and treatment response prediction in cancer of unknown primary

17: ⭐⭐⭐
Predicting breast cancer 5-year survival using machine learning: A systematic review

- Thirty-one studies that met the inclusion criteria were included, most of which were published after 2013. The most frequently used ML methods were decision trees (19 studies, 61.3%), artificial neural networks (18 studies, 58.1%), support vector machines (16 studies, 51.6%), and ensemble learning (10 studies, 32.3%). The median sample size was 37256 (range 200 to 659820) patients, and the median predictor was 16 (range 3 to 625). The accuracy of 29 studies ranged from 0.510 to 0.971. The sensitivity of 25 studies ranged from 0.037 to 1. The specificity of 24 studies ranged from 0.008 to 0.993. The AUC of 20 studies ranged from 0.500 to 0.972. The precision of 6 studies ranged from 0.549 to 1. All of the models were internally validated, and only one was externally validated. (page 1)

Types of ML Algorithms (page 13)

Type of ML algorithms	DT	19	61.3
	ANN	18	58.1
	SVM	16	51.6
	LR	12	38.7
	Bayesian classification algorithms	6	19.4
	KNN	3	9.7
	Semi-supervised learning	3	9.7
	Ensemble learning	10	32.3
	DNN	3	9.7

Type of ML algorithms	Decision Tree (DT)	19	61,3
	Artificial Neural Network (ANN)	18	58,1
	Support Vector Machine (SVM)	16	51,6
	Logistic Regression (LR) ou Linear Regression (LR)	12	38,7
	Bayesian classification algorithms	6	19,4
	K-Nearest-Neighbors (KNN)	3	9,7
	Semi-supervised learning	3	9,7
	Ensemble learning	10	32,2
	Deep neural networks (DNN)	3	9,7

Model Evaluation Metrics (page 13)

Model evaluation metrics	Accuracy	29	93.5
	Sensitivity/Recall	25	80.6
	Specificity	24	77.4
	AUC	20	64.5
	Precision/Positive predictive value	6	19.4
	F1 score	5	16.1
	Mcc	5	16.1
	NPV	2	6.5
	G-mean	2	6.5
	C-index	1	3.2
	Cutoff	1	3.2
	Youden index	1	3.2
	Retaining time	1	3.2
	FPR	1	3.2
	FDR	1	3.2
	FNR	1	3.2

In studies that compared of two or more algorithms, ANN had the best performance in 6 studies [27, 34, 39, 46, 49, 52], DT had the best performance in 4 studies [2, 23, 26, 34], the ensemble learning algorithm had the best performance in 4 studies [42, 43, 48, 51], semisupervised learning had the best performance in 3 studies [30–32], DNN had the best performance in 3 studies [40, 41, 45], SVM had the best

performance in 2 studies [35, 37], LR had the best performance in 2 studies [24, 29], KNN had the best performance in 1 study [36], and Naive Bayes had the best performance in 1 study [47] (see S5 Table). (page 14)

However, overall, compared with LR/Cox regression model, the performance of the ML algorithm does not necessarily improve, similar to the results of previous studies (page 16)

Calibration compares the observed probability and predicted probability of the occurrence of results, which is the key to model development [67]. Only 1 study performed model calibration, and the actual availability of uncalibrated models is limited [68]. Therefore, it is recommended that researchers consider this step and report modeling information in detail. (page 16)

Model Presentation (page 13)

Model presentation	Formula	6	19.4
	Graph	5	16.1
	Formula and graph	16	51.6
	No presentation	4	12.9

18: (Pago)
Use of machine learning to predict bladder cancer survival outcomes: a systematic literature review

19:
Predicting lung cancer survival based on clinical data using machine learning: A review

- RF was the most used ML method among all studies.

- DT, RF, bagging, and XGBoost are tree-based models [66]. A tree-based model can use qualitative predictors in its prediction process without creating dummy variables hence, tree-based models are considered more effective for detecting non-monotonic or non-linear relationships between dependent variables and predictors. They are also capable of handling many high-order interactions and moderate dataset sizes more effectively than regression models [67]. Although tree-based models have many advantages, they also have some disadvantages. When a small dataset is used to train tree-based models based on the high correlation among predictors, the detection of interactions between predictors is impeded, which may lead to overfitting. However, this effect can be reduced by employing RF

- Best models in each study based on RMSE accuracy

Table A1. Best models in each study based on RMSE accuracy

Model	Accuracy	Predicted period of survival time	Publication size	Reference
RF	10.52	6 months	10,442	[12]
Custom Ensemble	15.30	5 years	10,442	[16]
Self-Ordering Maps	15.59	5 years	10,442	[17]
GRNN	0.60	5 years	683	[20]
Ridge Regression	2.70	3 years	291	[27]
ANN for male	2.32	5 years	38,262	[31]
DNN	0.314	5 years	10,001	[36]

- Best models in each study based on AUC accuracy

Table A2. Best models in each study based on AUC accuracy

Model	Accuracy	Predicted period of survival	Publication size	Reference
ADTree	93.8	5 years	57,254	[11]
MNN	78.30	2 years	239	[13]
ANN	92.0	5 years	Not reported	[14]
CNN	92.0	5 years	Not reported	[14]
RF	90.1	3 years	50,687	[10]
XGBoost	78.6	1 year	5973	[15]
RF	80.2	5 years	5123	[18]
IRF	98.0	2 years	509	[19]
Gaussian K-base NB	88.1	5 years	321	[21]
AdaBoost	71.3	2 years	809	[23]
Logistic regression	76.0	5 years	585	[24]
DL	74.4	5 years	16,613	[25]
SVM	71.0	3 years	371	[26]
NB	81.0	3 years	291	[27]
SORG ML	71.4	1 year	150	[28]
LASSO	75.3	1 year	1563	[29]
DL	81.3	5 years	704	[30]
RF	95.1	5 years	17,484	[33]
RF	68.0	5 years	739	[35]
Cox Regression	72.2	5 years	2166	[38]
XGBoost	Female 93.0	1 year	28,458	[39]

- Best model based on C-statistic accuracy

Table A3. Best model based on C-statistic accuracy

Model	Accuracy	Predicted period time of survival	Publication size	Reference
DL	73.9	5 years	17,322	[22]
DL	63.6	5 years	1137	[32]
RF	67.2	5 years	506	[34]
DL	83.4	5 years	4617	[37]

- Models in each study based on RMSE accuracy for 6 months survival time

Table B1. Models in each study based on RMSE accuracy for 6 months survival time

ML algorithms	Accuracy for each reference
RF	10.52 [12]
LR	10.63 [12]
GBT	10.65 [12]
Custom Ensemble	10.84 [12]

- Models in each study based on RMSE accuracy for 2 years survival time

Table B2. Models in each study based on RMSE accuracy for 2 years survival time

ML algorithms	Accuracy for each reference
RF	15.70 [12]
LR	15.77 [12]
GBT	15.65 [12]
Custom Ensemble	16.26 [12]

- Models in each study based on RMSE accuracy for 3 years survival time

Table B3. Models in each study based on RMSE accuracy for 3 years survival time

ML algorithms	Accuracy for each reference
LR	3.01 [27]
Ridge Regression	2.70 [27]
Line SVR	2.78 [27]
Poly SVM	2.81 [27]

- Models in each study based on RMSE accuracy for 5 years survival time

Table B4. Models in each study based on RMSE accuracy for 5 years survival time

ML algorithms	Accuracy for each reference
RF	20.51 [12], 15.63 [16]
LR	21.37 [12], 15.38 [16]
GBT	21.14 [12], 15.32 [16]
Custom Ensemble	21.18 [12], 15.30 [16]
SVM	15.82 [16]
DT	15.81 [16]
HC	16.202 [17]
MBC	16.250 [17]
K-Means Clustering	16.193 [17]
SOMs	15.591 [17]
Non-negative Matrix Factorization	16.589 [17]
GRNN	0.60 [20]
ANN	Male 2.32 [31], female 2.52 [31]
Bayes Net	0.315 [36]
DNN	0.314 [36]
Logistic Regression	0.314 [36]
Lazy Classifier LWL	0.318 [36]
J48 DT	0.32 [36]
Meta-Classifier (ASC)	0.316 [36]
Rule Learner (OneR)	0.339 [36]

- Models in each study based on AUC accuracy for 6 months survival time

Table B5. Models in each study based on AUC accuracy for 6 months survival time

ML algorithms	Accuracy for each reference
SVM	60.5 [11], 79.0 [14]
ANN	74.0 [11], 82.0 [14]
J48 DT	78.1 [11]
RF	78.1 [11], 80.0 [14]
LogitBoost	80.5 [11]
Random Subspace	80.8 [11]
ADTree	80.8 [11]
CNN	83.0 [14]
RNN	81.0 [14]
NB	77.0 [14]

- Models in each study based on AUC accuracy for 1 year survival time

Table B6. Models in each study based on AUC accuracy for 1 year survival time

ML algorithms	Accuracy for each reference
SVM	63.4 [11], 73.0 [15], male 86.4 [39], 91.6 [39]
ANN	80.2 [11]
J48 DT	78.6 [11]
RF	81.1 [11], 85.2 [10], 73.6 [15], male 86.0 [39], female 91.2 [39]
MTLR	82.1 [10]
Logit Boost	82.1 [11]
Random Subspace	82.4 [11]
ADTree	83.0 [11]
Logistic Algorithms	71.0 [15], male 85.5 [39], female 92.1 [39]
DT	Male 84.3 [39], female 90.4 [39]
XGBoost	78.6 [15], male 87.8 [39], female 93.0 [39]
SORG	71.4 [28]
LASSO	75.3 [29]
KNN	Male 83.6 [39], female 89.8 [39]

- Models in each study based on AUC accuracy for 2 years survival time

Table B7. Models in each study based on AUC accuracy for 2 years survival time

ML algorithms	Accuracy for each reference
SVM	57.0 [11], 64.0 [14], 51.6 [19]
ANN	81.6 [11], 86.0 [14]
J48 DT	81.8 [11]
RF	86.7 [11], 66.0 [14], 96.8 [19], 55.0 [24]
Logit Boost	87.7 [11]
Random Subspace	87.8 [11]
ADTree	88.4 [11]
CNN	86.0 [14]
RNN	86.0 [14]
NB	63.0 [14], 54.0 [19]
Logistic Regression	64.2 [13], 49.0 [24]
SPNN	65.0 [13]
MNN	78.3 [13]
DT	47.4 [19]
IRF	98.0 [19]
MLP	65.0 [24]
XGBoost	58.0 [24]
Gaussian NB	52.0 [24]
Light GBT	54.0 [24]
support vector clustering SVC	55.0 [24]

- Models in each study based on AUC accuracy for 3 years survival time

Table B8. Models in each study based on AUC accuracy for 3 years survival time

ML algorithms	Accuracy for each reference
SVM	71.0 [26], 73.0 [27], male 70.6 [39], female 81.1 [39]
RF	90.1 [10], 76.0 [27], male 70.2 [39], female 80.7 [39]
NB	81.0 [27]
MTLR	86.4 [10]
LR	74.0 [27]
LASSO	73.6 [29]
Logistic regression	Male 69.9 [39], female 81.8 [39]
DT	Male [39], female 80.1 [39]
XGBoost	Male 72.4 [39], female 82.9 [39]
KNN	Male 67.2 [39], female 79.5 [39]

- Models in each study based on AUC accuracy for 5 years survival time

Table B9. Models in each study based on AUC accuracy for 5 years survival time

ML algorithms	Accuracy for each reference
SVM	56.4 [11], 84.0 [14], male 72.0 [39], female 82.3 [39]
ANN	92.3 [11], 92.0 [14]
J48 DT	84.7 [11], 94.4 [33]
RF	90.5 [11], 86.0 [14], 89.9 [10], 69.2 [23], 68.8 [25], 80.259 [18], 95.1 [33], 68.0 [35], male 71.6 [39], female 82.0 [39]
Logit Boost	93.0 [11]
Random Subspace	93.2 [11], 56.0 [24]
MTLR	87.0 [10]
ADTree	93.7 [11]
CNN	92.0 [14]
RNN	91.0 [14]
NB	83.0 [14], 59.7 [21]
Logistic Regression	63.2 [23], 76.0 [24], 92.7 [33], male 71.0 [39], female 83.0 [39]
SPNN	71.0 [24]
DT	69.2 [23], 78.166 [18], male 69.4 [39], female 81.4 [39]
MLP	71.0 [24]
XGBoost	54.0 [24], male 73.5 [39], female 84.2 [39]
Gaussian NB	73.0 [24]
Light GBT	67.0 [24]
MTLR	87.0 [10]
LR	62.2 [21]
Gaussian K base NB	88.1 [21]
Bagging	70.6 [23]
AdaBoost	71.3 [23]
support vector clustering SVC	67.0 [24]
DL	74.4 [25], 81.6 [30]
LASSO	65.6 [29]
Bayes Net	94.2 [33]
Cox Regression	72.2 [38]
KNN	Male 68.6 [39], female 80.7 [39]

- Model based on C-statistic accuracy for 5 years survival time

Table B10. Model based on C-statistic accuracy for 5 years survival time

ML model	Accuracy for each reference
DL	73.9 [22], 63.6 [32], 83.4 [37]
Cox Regression	56.1 [32], GSE72094 cohort 63.0 [34], TCGA cohort 64.5 [34], 64.0 [37]
Cox NN	56.2 [32]
RF	GSE72094 cohort 67.2 [34], TCGA cohort 65.6 [34], 67.8 [37]

20:
Cervical cancer survival prediction by machine learning algorithms: a systematic review

21:
A systematic review on machine learning and deep learning techniques in cancer survival prediction

22:
Predicting breast cancer survivability: a comparison of three data mining methods - ScienceDirect

Regressão de COX:

Descrição:

Qualidades do modelo de regressão de Cox:

O modelo de regressão de Cox é frequentemente utilizado para prever a sobrevida de pacientes com câncer, especialmente quando há dados pré-processados e relevantes disponíveis. Razões pelas quais o modelo de regressão de Cox é uma escolha adequada:

1. Modelo Estatístico Robusto: A regressão de Cox é um modelo estatístico robusto amplamente utilizado em análises de sobrevida. Ele considera a influência de diversas variáveis independentes na sobrevida do paciente, permitindo a análise de múltiplos fatores preditivos.

2. Capacidade de Lidar com Dados Censurados: Em estudos de sobrevida, é comum que alguns pacientes não tenham alcançado o evento de interesse (por exemplo, morte) no momento em que os dados são analisados. O modelo de regressão de Cox pode lidar com

esses dados censurados, tornando-se útil quando há informações incompletas sobre a duração da sobrevida.

3. Interpretabilidade: A interpretação dos resultados do modelo de regressão de Cox é relativamente direta. Os coeficientes estimados para cada variável independente indicam a direção e a magnitude do efeito da variável na sobrevida, tornando-o útil para entender os fatores que influenciam os resultados.

4. Flexibilidade na Modelagem de Variáveis: O modelo de regressão de Cox permite a inclusão de variáveis contínuas e categóricas, bem como variáveis interativas, o que permite a modelagem de uma variedade de fatores que podem afetar a sobrevida do paciente.

Limitações do modelo de regressão de Cox:

1. Proporções Constantes de Risco (Pressuposto de Proporcionalidade): O modelo de regressão de Cox pressupõe que as proporções de risco entre grupos de pacientes permaneçam constantes ao longo do tempo. No entanto, essa suposição pode não ser verdadeira em todas as situações. Por exemplo, os efeitos de certos fatores de risco podem mudar ao longo do tempo, o que pode violar o pressuposto de proporcionalidade e afetar a precisão das estimativas do modelo.

2. Modelo Linear de Efeitos: O modelo de regressão de Cox assume que o efeito de cada variável independente sobre a sobrevida é linear ao longo do tempo. No entanto, em alguns casos, as relações entre as variáveis independentes e a sobrevida podem ser não lineares, o que pode resultar em uma modelagem imprecisa se essa suposição não for atendida.

3. Interpretação Causativa Limitada: O modelo de regressão de Cox pode identificar associações entre variáveis independentes e sobrevida, mas não pode determinar relações causais diretas. Outros fatores não incluídos no modelo podem influenciar os resultados, e a interpretação dos coeficientes estimados deve ser feita com cautela para evitar inferências causais incorretas.

4. Limitações na Lida com Dados Censurados: Embora o modelo de regressão de Cox seja projetado para lidar com dados censurados, certas formas de censura, como censura informativa, podem afetar a precisão das estimativas. Além disso, a presença de uma grande proporção de dados censurados pode exigir métodos de análise adicionais para garantir resultados confiáveis.

5. Dependência de Suposições Não Testadas: O modelo de regressão de Cox depende de várias suposições não testadas, como a independência dos tempos de sobrevivência e a ausência de viés de seleção. Se essas suposições forem violadas, os resultados do modelo podem ser distorcidos e não confiáveis.

Outros Modelos:

Para prever o tempo de sobrevida de um paciente com câncer baseado em informações pré-processadas, o modelo de regressão de Cox é frequentemente considerado o padrão ouro na análise de sobrevida. No entanto, se estivermos excluindo modelos de inteligência artificial e procurando por alternativas puramente estatísticas, existem algumas opções que podem ser consideradas. Aqui estão alguns modelos estatísticos que podem ser úteis:

1. Modelo de Regressão Paramétrica: Além do modelo de regressão de Cox, os modelos de regressão paramétrica são uma alternativa. Esses modelos assumem uma distribuição específica para o tempo de sobrevida, como a distribuição exponencial, Weibull ou log-normal. Eles podem ser úteis quando a suposição de proporcionalidade de risco do modelo de Cox não é atendida e quando a distribuição do tempo de sobrevida é conhecida ou pode ser razoavelmente estimada.

2. Modelo de Regressão de Aceleração de Falhas: Este modelo é uma extensão do modelo de regressão de Cox que permite a análise de dados de sobrevivência com uma taxa de falhas que muda com o tempo. Ele pode ser útil quando as proporções de risco não são constantes ao longo do tempo, permitindo que o efeito das variáveis independentes na sobrevida varie de acordo com a função de aceleração de falhas.

3. Modelo de Risco Proporcional Competitivo: Este modelo é utilizado quando há um evento de interesse competitivo que pode ocorrer além do evento de interesse principal (por exemplo, morte por outras causas em estudos de sobrevida relacionados ao câncer). Ele permite a análise da sobrevida considerando a competição entre os eventos de interesse, o que pode ser importante em certos contextos clínicos.

4. Modelo de Sobrevivência Frailty: Este modelo incorpora a heterogeneidade não observada entre os indivíduos em um estudo de sobrevida, assumindo que os efeitos das variáveis independentes podem variar devido a fatores individuais não observados (frailty). Ele pode ser útil quando há variação não explicada entre os indivíduos que pode afetar a sobrevida.

Cada um desses modelos tem suas próprias vantagens e limitações, e a escolha do modelo mais apropriado depende das características específicas do conjunto de dados, da natureza do evento de interesse e das suposições subjacentes que podem ser justificadas.

Artigos:

1: **Comparison of the cox regression to machine learning in predicting the survival of anaplastic thyroid carcinoma**

Texto:

Patients diagnosed with ATC were extracted from the Surveillance, Epidemiology, and End Results database. The outcomes were overall survival (OS) and cancer-specific survival (CSS), divided into: (1) binary data: survival or not at 6 months and 1 year; (2): time-to-event data. The Cox regression method and machine learnings were used to construct models. Model performance was evaluated using the concordance index (C-index), brier score and calibration curves. The SHapley Additive exPlanations (SHAP) method was deployed to interpret the results of machine learning models.

For binary outcomes, the Logistic algorithm performed best in the prediction of 6-month OS, 12-month OS, 6-month CSS, and 12-month CSS (C-index = 0.790, 0.811, 0.775, 0.768). For time-event outcomes, traditional Cox regression exhibited good performances (OS: C-index = 0.713; CSS: C-index = 0.712). The DeepSurv algorithm performed the best in the training set (OS: C-index = 0.945; CSS: C-index = 0.834) but performs poorly in the verification set (OS: C-index = 0.658; CSS: C-index = 0.676).

Cox regression and machine learning models combined with the SHAP method can predict the prognosis of ATC patients in clinical practice. However, due to the small sample size and lack of external validation, our findings should be interpreted with caution.

We collected all relevant data including Age, Sex, Race, Marital status, Insurance, No high school diploma, Families below poverty, AJCC TMN stage, Tumor size, Multifocality, Regional lymph node surgery, Thyroid surgery, Radiotherapy, Chemotherapy.

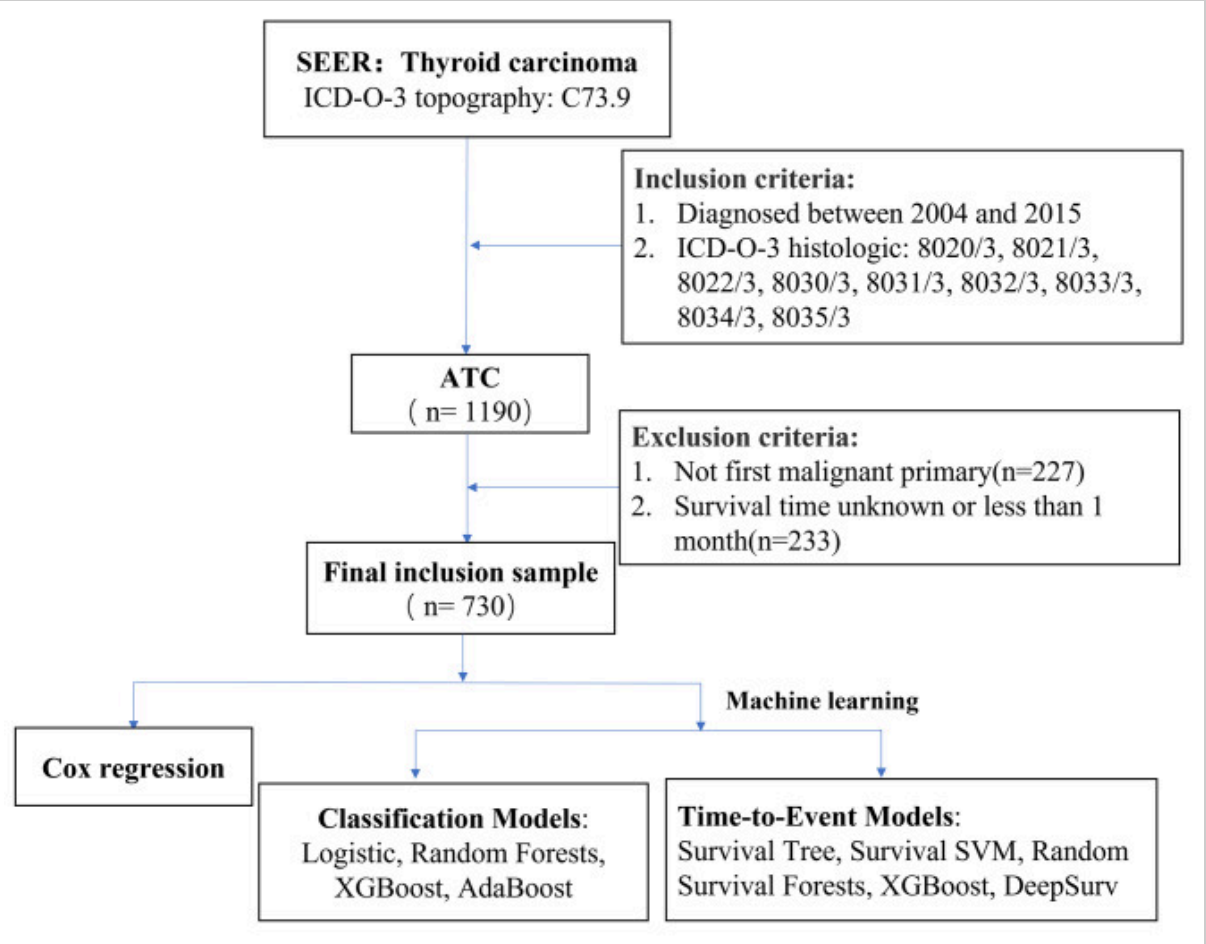
We counted the missing rates of all predictors, and retained factors with a missing rate of less than 30%. K-Nearest Neighbor (KNN) algorithms were used to fill missing values. Multicollinearity is explained by the variance inflation factor (VIF). VIF < 10 indicates that there is no multicollinearity among the variables. Correlation was determined by Spearman correlation analysis. A correlation coefficient greater than 0.5 indicates a significant correlation between variables.

For binary outcomes, we used four machine learning algorithms, Logistic, Random Forests, Extreme Gradient Boosting (XGBoost) and Adaptive Boosting (AdaBoost), to construct models and compared the pros and cons of these models

In order to avoid overfitting, the evaluation of the model comprehensively considers the results of the training set and the validation set, but mainly the results of the validation set. We used the C-index to describe the discriminativeness of the model. The C-index value can generally judge the generalization ability of the model: 0.5–0.7 means that the model has a weak generalization ability, 0.7–0.85 moderate, and 0.85-1.0 strong. In addition, we also use multiple evaluation indicators such as accuracy, sensitivity, and specificity to comprehensively evaluate the discriminative ability of the machine learning model. We used the calibration curve and brier-score to evaluate the calibration of the model. In the calibration plot, the X-axis represents the predicted survival time and the Y-axis the actual survival time with the predicted rate falling on the 45° diagonal in a perfect prediction model. The lower the Brier-score value, the better the calibration. We assessed the net benefit of the model for clinical decision making through the DCA curve. Kaplan-Meier analysis and log-rank test were used to explore differences in survival between risk subgroups.

The SHapley Additive exPlanation (SHAP) is a unified framework for interpreting the results of

machine learning models [25]. We utilized SHAP to provide explanations for the final model, including associated risk factors causing death in patients with ATC and the importance of sorting features. Our study was reported following the TRIPOD (Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis) statement [26]. All statistical analyses in this study were performed using R software (version 4.0.2) and Python software (version 3.7.6). P value of < 0.05 was considered statistically significant.



Demographic characteristics of patients with ATC

In the validation set, the C-index values of random forests algorithm were significantly lower than that in the training set, due to possible overfitting. Combining the results of the training set and the validation set, we found that the logistic algorithm presented best performance.

The results of survival analysis on time-to-event shown that the DeepSurv algorithm presented best performance in the training set and was obviously superior to Cox regression algorithm

No algoritmo DeepSurv, os valores do índice C do conjunto de treinamento e do conjunto de validação foram muito diferentes, e o overfitting foi considerado.

Summary of model performance of C-index, Accuracy, Sensitivity, Specificity, Brier-score

To evaluate the practicability of each model, we plotted the DCA curve (Supplemental Fig. 5 and Supplemental Fig. 6).This shown that Cox regression model and logistic algorithm had good clinical applicability in predicting the 6-month and 12-month survival rates of ATC and had high net benefits.

By comparing the prediction performance of different ML algorithms to the reference method (Cox regression), our findings suggested that Cox regression performed well as a conventional method for ATC survival prediction.

Among ML algorithms, Logistic algorithm demonstrated the best performance. Combining SHAP values, Logistic algorithm illustrated key predictive factors and established a high-accuracy survival prediction model.

Traditional Cox regression is the most convenient way to solve most survival prediction problems because its results are easy to interpret. However, Cox regression models should be used with a minimum of 10 outcome events per predictor variable (EPV)

As a substitute of Cox regression, the Logistic algorithm combined with SHAP values performed superiority in clinical applications. However, it is important to note that the predictive efficacy of Cox regression in predicting the survival of ATC patients were comparable with ML algorithms, suggesting that the superiority of ML was not always seen but was seen only in situations when the conventional methods meet their limits.

ATC	anaplastic thyroid carcinoma
OS	overall survival
CSS	cancer-specific survival
C-index	concordance index
SHAP	SHapley Additive exPlanations

SEER Program	Surveillance, Epidemiology, and End Results database
AJCC	American Joint Committee on Cancer
TNM	Tumor Node Metastasis
ML	machine learning
KNN	K-Nearest Neighbor
VIF	variance inflation factor
XGBoost	extreme gradient boosting
AdaBoost	adaptive boosting
SVM	Survival Support Vector Machine
AIC	Akaike information criterion

DCA curve	Decision Curve Analysis
K-M curve	Kaplan-Meier analysis

Resumo:

Este estudo fornece informações valiosas sobre a previsão de sobrevivência entre pacientes com carcinoma anaplásico de tireoide (ATC) usando algoritmos convencionais de regressão de Cox e aprendizado de máquina (ML)

Principais conclusões:
Comparação de métodos : O estudo comparou o desempenho da regressão de Cox e dos algoritmos de ML na previsão da sobrevida entre pacientes ATC.

Resultados binários e de tempo até o evento : Eles examinaram tanto os resultados binários (sobrevivência em 6 meses e 1 ano) quanto os resultados de tempo até o evento.

Métricas de desempenho : As métricas de avaliação incluíram índice de concordância (índice C), pontuação de Brier, curvas de calibração e explicações aditivas SHapley (SHAP) para interpretabilidade.

Algoritmos de melhor desempenho : a regressão logística teve melhor desempenho para resultados binários, enquanto a regressão de Cox mostrou bom desempenho para resultados de tempo até o evento. DeepSurv teve um bom desempenho no conjunto de treinamento, mas apresentou desempenho inferior no conjunto de validação, indicando potencial overfitting.

Interpretação do modelo : os valores SHAP foram usados para interpretar os resultados dos modelos de ML, identificando os principais fatores preditivos.

Metodologias:
Fonte de dados : Os dados foram extraídos do banco de dados de Vigilância, Epidemiologia e Resultados Finais (SEER), garantindo uma amostra grande e diversificada.

Variáveis Preditoras : Várias variáveis preditoras, incluindo características demográficas, características do tumor e modalidades de tratamento, foram consideradas.

Pré-processamento de dados : Os valores faltantes foram tratados usando o algoritmo K-Nearest Neighbor (KNN), e a multicolinearidade foi avaliada usando o fator de inflação de variância (VIF).

Desenvolvimento de modelo : Regressão de Cox e algoritmos de ML, incluindo regressão logística, Random Forests, XGBoost e DeepSurv foram usados para construir modelos. A seleção das variáveis foi realizada por meio de regressão stepwise bidirecional e método XGBoost.

Avaliação do modelo : O desempenho dos modelos foi avaliado usando índice C, precisão, sensibilidade, especificidade, curvas de calibração e pontuação de Brier. A discriminação e a calibração foram avaliadas para garantir a confiabilidade do modelo.

Interpretação do modelo : os valores SHAP foram utilizados para fornecer explicações para modelos de ML, identificando fatores preditivos importantes e seu impacto nos resultados.

Conclusão:
O estudo conclui que tanto a regressão de Cox quanto os algoritmos de ML, particularmente a regressão logística, combinados com valores SHAP, podem prever efetivamente a sobrevivência de pacientes ATC. No entanto, os resultados devem ser interpretados com cautela devido às limitações do pequeno tamanho da amostra e à falta de validação externa.

No geral, este estudo contribui para a compreensão dos fatores prognósticos para a sobrevivência do ATC e destaca o potencial dos algoritmos de ML para melhorar a precisão preditiva e a interpretabilidade na prática clínica.

INFOS:

Modelos Utilizados:

Logistic Regression:
    Utilizado para prever resultados binários, como sobrevivência em 6 meses e 12 meses (overall survival - OS e cancer-specific survival - CSS).
Random Forests:
    Outro modelo utilizado para prever resultados binários.
Extreme Gradient Boosting (XGBoost):
    Utilizado para previsões binárias e/ou modelos de tempo de evento.
Adaptive Boosting (AdaBoost):
    Utilizado para construir modelos de previsão de sobrevivência.
Cox Regression:
    Utilizado para análise de sobrevivência de tempo de evento.
DeepSurv:
    Modelo de aprendizado profundo usado para prever sobrevivência em dados de tempo de evento.

Dados de Entrada:

Idade, sexo, raça, estado civil, cobertura de seguro, educação (grau de escolaridade), condição socioeconômica (famílias abaixo da linha de pobreza), estágio AJCC TNM, tamanho do tumor, multifocalidade, cirurgia de linfonodos regionais, cirurgia de tireoide, radioterapia e quimioterapia.

Métrica de Avaliação:
Concordance Index (C-index):
Utilizado para avaliar a capacidade discriminativa dos modelos.
Acurácia, Sensibilidade e Especificidade:
Medidas adicionais para avaliar o desempenho dos modelos de classificação.
Brier Score e Calibration Curves:
Utilizados para avaliar a calibração e precisão dos modelos de previsão.
Decision Curve Analysis (DCA):
Usada para avaliar o benefício clínico e utilidade prática dos modelos.
Kaplan-Meier analysis e log-rank:
Usados para explorar diferenças na sobrevivência entre subgrupos de risco.

Resultados Destacados:
O modelo Logistic Regression demonstrou o melhor desempenho para prever a sobrevivência em 6 e 12 meses.
Para modelos de tempo de evento, Cox Regression exibiu um bom desempenho.
DeepSurv apresentou um desempenho excepcional no conjunto de treinamento, mas teve problemas de overfitting no conjunto de validação.

Número de Pacientes:
Amostra Final = 730

2: Survival prediction and prognostic factors in colorectal cancer after curative surgery: insights from cox regression and neural networks

Texto:

The data was used to train both Cox regression and neural network models, and their predictive accuracy was compared using diagnostic measures such as sensitivity, specificity, positive predictive value, accuracy, negative predictive value, and area under the receiver operating characteristic curve. The analyses were performed using STATA 17 and R4.0.4 software.

In conclusion, the study found that both Cox regression and neural network models are effective in predicting early recurrence and death in patients with colorectal cancer after curative surgery. The neural network model showed slightly better sensitivity and negative predictive value for death, while the Cox model had better specificity and positive predictive value for recurrence. Overall, both models demonstrated high accuracy and AUC, indicating their usefulness in predicting these outcomes.

While the Cox regression model is a widely accepted and commonly employed method in survival analysis, recent studies have shown that neural network-based analytical methods have become increasingly popular due to their ability to capture complex model relationships and improve learning performance, resulting in better diagnostic and prediction accuracy.

284 colorectal cancer patients

The studied variables include age at the time of diagnosis (years), gender (female:1; male:2), Body Mass Index (BMI: kg/m2), and clinical/pathological variables such as metastasis to other sites (no:0; yes:1), cancer site (colon:1; rectum:2), surgery (no:0; yes:1), radiotherapy (no:0; yes:1), chemotherapy (no:0; yes:1), number of chemotherapy (0:no; 1: < 6; 2:6 +), morphology (0:no adeno; 1:adeno), grade (differentiation level) (1:well; 2:moderate; 3:poor), tumor size (1: < 4; 2: > = 4 < 7; 3: = > 7), disease stage(1:B; 2:C;3:D), PT-stage(1:T2; 2:T3; 3:T4; 4:Tx), and PN-stage(1:N2; 2:N3; 3:N4; 4:Nx).

The data were analyzed using Stata (ver. 17, StataCorp, LLC, College station, Texas, USA), and R4.0.4 (https://cran.r-project.org) softwares. A significance level was considered 5%. The survival time of patients with colorectal cancer had calculated in months, and the data for categorical variables was reported as frequency and percentage. The median survival times of patients, both overall and specifically for those with recurrence, along with the median durations from the non-terminal event to the terminal event, were computed utilizing the Kaplan–Meier estimator. Besides we used this estimator to estimate the 1-, 3-, 5- and 10-year survival probabilities for the terminal event, the non-terminal events. The Cox regression model was employed to determine the factors affecting the survival time of patients. The assumption of proportional hazards was tested using Schoenfeld residual method8.

In the field of predicting postoperative mortality in healthcare, artificial neural network (ANN) and Cox regression models are used as the most common prediction models10.

One of the advantages of this method compared to the neural network is the simplicity and comprehensibility of the results8,11.

The Cox model provides researchers with the capacity to assess the impact of multiple covariates, commonly referred to as predictor variables, on the hazard rate of an event. This is achieved while accounting for censoring12. To elaborate, in the context of a medical investigation exploring the lifespan of patients exposed to different therapeutic regimens, the Cox model readily supplies hazard ratios for each unique treatment. This enables medical professionals and researchers to comprehensively comprehend the relative effects of the interventions on patient prognosis13.

Além disso, devido ao uso de distribuições estatísticas, este método pode ser utilizado para investigar o efeito simultâneo de diversas variáveis no tempo de sobrevivência 8

Furthermore, owing to the utilization of statistical distributions, this approach can be employed to examine the concurrent impact of multiple variables on survival duration8. This

prominently facilitates the discernment of the individual effects of each variable on the hazard of the event while controlling for additional factors14. Within the framework of the Cox regression model, the identification of time-dependent variables is also feasible, enabling the capture of longitudinal fluctuations over time that may influence the likelihood of the focal event. Consequently, this model provides the capability to assess the concomitant effects of these time-dependent variables on the magnitude of risk associated with distinct variable constellations, even as temporal progression unfolds15.

Neural networks possess the ability to capture intricate and non-linear patterns, making them a superior choice in such contexts16,17.

Comparison of the Cox model and neural network model.

Nonetheless, the Cox regression model outperformed the neural network model in terms of matching observed and predicted values.

The results showed that Cox regression models have a better predictive performance than more complex machine learning models25. Similarly, a 2021 study found that neural networks for predicting survival using clinical data may perform similarly to statistical methods such as the Cox model, but are often less adjusted26. In simpler cases, it may be better to use traditional statistical methods instead of neural networks26. Another study conducted in the field of urology to check the accuracy of predicting censored data using machine learning models and comparing it with the Cox regression model found that in all three datasets, the Cox regression model had acceptable predictions compared to machine learning models, including neural networks27.

Resumo:

Introdução:
A regressão de Cox é amplamente utilizada na análise de sobrevivência em pacientes com câncer.
Objetivo do estudo: Comparar a eficácia da regressão de Cox e modelos de redes neurais na predição de resultados de sobrevivência em pacientes com câncer colorretal submetidos a tratamento cirúrgico.

Material e Métodos:
Estudo retrospectivo com 284 pacientes com câncer colorretal submetidos a cirurgia entre 2001 e 2017.
Variáveis clínicas e demográficas foram analisadas.
Utilização de regressão de Cox e redes neurais para predição de recorrência e morte.

Resultados:
Ambos os modelos, Cox e redes neurais, demonstraram eficácia na predição de recorrência e morte.

Cox fornece interpretações diretas de riscos proporcionais, enquanto redes neurais capturam padrões complexos e não lineares.
Cox apresentou resultados ligeiramente melhores em comparação com redes neurais, especialmente na correspondência entre valores observados e previstos.

Discussão:
A regressão de Cox é uma ferramenta valiosa na análise de dados de sobrevivência em pacientes com câncer colorretal.
Redes neurais podem oferecer vantagens em casos de padrões complexos e não lineares, mas a interpretabilidade pode ser um desafio.

Conclusão:
A regressão de Cox e modelos de redes neurais são úteis na predição de resultados de sobrevivência em pacientes com câncer colorretal.
A escolha entre os modelos depende da natureza dos dados e da complexidade do padrão a ser modelado.

sensibilidade, especificidade, valor preditivo positivo (PPV), valor preditivo negativo (NPV), acurácia e área sob a curva ROC (Receiver Operating Characteristic)

INFOS:

Modelos:
Modelo de Regressão de Cox
Rede Neural (Multi-Layer Perceptron - MLP)

Dados de Entrada:
Idade no momento do diagnóstico, Gênero, Índice de Massa Corporal (BMI), Metástase para outros locais, Localização do câncer (colon vs. reto), Realização de cirurgia, Recebimento de radioterapia e quimioterapia, Número de sessões de quimioterapia, Grau de diferenciação do tumor, Tamanho do tumor, Estágio da doença (staging), Estágio PT (tamanho primário do tumor), Estágio PN (envolvimento dos linfonodos)

Métricas de Avaliação:
Sensibilidade
Especificidade
Valor Preditivo Positivo (PPV)
Valor Preditivo Negativo (NPV)
Área sob a Curva ROC (AUC)
Acurácia

Resultados:
O Modelo de Regressão de Cox apresentou resultados comparáveis com a Rede Neural em termos de sensibilidade, especificidade, PPV, NPV, AUC e acurácia para previsão de recorrência e mortalidade após cirurgia curativa para câncer colorretal.

Os melhores resultados para a Rede Neural foram uma sensibilidade de 88.1% e uma especificidade de 83.7% para previsão de recorrência, e uma sensibilidade de 74.5% e uma especificidade de 83.3% para previsão de mortalidade.
Os resultados do Modelo de Regressão de Cox incluíram uma sensibilidade de 73.6% e uma especificidade de 89.6% para previsão de mortalidade, e uma sensibilidade de 85.5% e uma especificidade de 98.0% para previsão de recorrência.
O modelo de Cox teve resultados ligeiramente superiores em algumas métricas, enquanto a rede neural teve melhor sensibilidade em algumas previsões.

Número de Pacientes:
284 pacientes com câncer colorretal que foram submetidos à cirurgia curativa. Destes pacientes, 131 apresentaram recorrência da doença e 121 faleceram durante o período de acompanhamento.

3: https://pubmed.ncbi.nlm.nih.gov/19622426/

4: Survival of patients with colorectal cancer in a cancer center

Análise Descritiva: Inicialmente, foram realizadas análises descritivas para caracterizar a amostra de pacientes, incluindo distribuição por idade, sexo, período de diagnóstico, localização do câncer, estágio clínico e tempo de espera para o início do tratamento.

Análise de Sobrevivência: A sobrevida global (OS) foi calculada utilizando o estimador de Kaplan-Meier, que é uma técnica para estimar a probabilidade de sobrevivência ao longo do tempo. Curvas de sobrevida foram geradas para diferentes subgrupos de pacientes com base em características como idade, sexo, período de diagnóstico e localização do câncer.

Teste de Log-Rank: Para comparar as curvas de sobrevida entre os grupos, foi utilizado o teste de Log-Rank, que avalia se há diferenças significativas nas curvas de sobrevida entre os grupos.

Modelo de Regressão de Cox: Para identificar os fatores prognósticos independentes associados à sobrevida, foi empregado o modelo de regressão de Cox. Esse modelo permite avaliar o efeito simultâneo de várias variáveis explicativas na taxa de risco de mortalidade. Foram incluídas variáveis como idade, sexo, período de diagnóstico, localização do câncer, estágio clínico e tempo de espera para o início do tratamento.

Razões de Risco (Hazard Ratios): As razões de risco (HR) foram calculadas para cada variável, juntamente com os intervalos de confiança de 95% (IC 95%). As HRs representam a magnitude do efeito de cada variável na taxa de risco de mortalidade. Valores de HR superiores a 1 indicam um aumento no risco de morte, enquanto valores inferiores a 1 indicam uma redução no risco de morte.

Resultados do Modelo de Cox: Os resultados do modelo de Cox indicaram que a idade avançada, o diagnóstico de câncer retal, e o atraso no início do tratamento estavam associados a um aumento do risco de morte. Por outro lado, o período de diagnóstico mais recente foi associado a uma redução do risco de morte. Esses resultados foram ajustados para outros fatores prognósticos, como sexo e estágio clínico.

5: https://pubmed.ncbi.nlm.nih.gov/34455119/

LIFE LINES PYTHON:

Documentação do Life Lines

Motivos do uso:

A biblioteca "lifelines" em Python é uma ferramenta poderosa para análise de sobrevivência e modelagem de riscos. Ela oferece uma série de funcionalidades para trabalhar com dados de eventos de sobrevivência, como tempo até a ocorrência de um evento (como falha de um equipamento, morte de um paciente, etc.) e variáveis explicativas que podem influenciar a ocorrência desse evento.

Modelagem de Sobrevivência:
A principal funcionalidade da biblioteca é a modelagem de sobrevivência, que envolve a análise de dados de sobrevivência para entender a distribuição do tempo até a ocorrência de um evento de interesse. A biblioteca oferece uma variedade de modelos de sobrevivência, incluindo o modelo de riscos proporcionais de Cox, modelos de riscos competitivos e modelos de riscos acelerados.

Estimadores de Kaplan-Meier:
O estimador de Kaplan-Meier é uma técnica não paramétrica usada para estimar a função de sobrevivência a partir de dados de sobrevivência censurados. Ele é comumente usado quando não há necessidade de considerar covariáveis explicativas. Em vez disso, o foco está na estimativa da função de sobrevivência ao longo do tempo.
Essa técnica é especialmente útil quando os dados de sobrevivência são observados em intervalos discretos de tempo e podem conter observações censuradas, ou seja, eventos que não foram observados até o final do estudo.
O método de Kaplan-Meier calcula a estimativa da função de sobrevivência como o produto das probabilidades de sobrevivência em cada intervalo de tempo, levando em consideração as observações censuradas. Isso resulta em uma curva de sobrevivência que mostra a probabilidade de um evento não ter ocorrido até um determinado ponto no tempo.
A função de sobrevivência estimada pelo método de Kaplan-Meier é uma estimativa não paramétrica da verdadeira função de sobrevivência da população, e é representada graficamente pela curva de Kaplan-Meier.

Modelos de Riscos Proporcionais de Cox:

Os modelos de riscos proporcionais de Cox são uma ferramenta estatística para modelar a relação entre covariáveis explicativas e o tempo até a ocorrência de um evento de interesse, enquanto mantêm a suposição de que a razão entre os riscos (hazard ratios) é constante ao longo do tempo.
Esses modelos são amplamente utilizados na análise de sobrevivência porque permitem que os pesquisadores examinam como diferentes variáveis explicativas afetam o risco de um evento ocorrer, controlando simultaneamente outros fatores.
A função de risco (ou taxa de falha) é modelada como o produto de uma função de base, que é uma função do tempo, e um termo exponencial que captura o efeito das covariáveis.
A estimativa dos parâmetros do modelo de Cox é feita por meio do método da verossimilhança parcial, e os coeficientes resultantes do modelo são interpretados como o efeito das covariáveis sobre o risco relativo de ocorrência do evento.
Uma das principais vantagens dos modelos de riscos proporcionais de Cox é que eles são semiparamétricos, o que significa que não é necessário especificar uma distribuição paramétrica para o tempo de sobrevivência. Isso os torna muito flexíveis e amplamente aplicáveis em uma variedade de cenários.
A biblioteca "lifelines" facilita a construção, ajuste e interpretação de modelos de riscos proporcionais de Cox, permitindo aos usuários examinar e quantificar o impacto das covariáveis explicativas sobre a sobrevivência.

Modelos de Riscos Competitivos:
Além dos modelos de Cox, a biblioteca também suporta a modelagem de riscos competitivos, que é usada quando há mais de um tipo de evento de interesse e a ocorrência de um evento pode afetar a probabilidade de ocorrência de outros eventos.

Modelos de Riscos Acelerados:
A biblioteca também oferece suporte para modelos de riscos acelerados, que são usados quando a taxa de falha é afetada por covariáveis, mas a forma funcional dessa relação não é especificada. Esses modelos são úteis quando os efeitos das covariáveis sobre a taxa de falha podem mudar ao longo do tempo.

Censura e Truncamento:
A biblioteca lida automaticamente com dados censurados e truncados, que são comuns em análise de sobrevivência. A censura ocorre quando o tempo de acompanhamento termina antes que o evento de interesse ocorra, enquanto o truncamento ocorre quando apenas uma amostra de dados é observada.

Visualização de Resultados:
A lifelines inclui funcionalidades para visualizar os resultados da análise de sobrevivência, como gráficos de Kaplan-Meier, curvas de sobrevida ajustadas e gráficos de risco relativo ao longo do tempo.

Flexibilidade e Facilidade de Uso:
A biblioteca é projetada para ser flexível e fácil de usar, permitindo aos usuários ajustar modelos de sobrevivência e realizar análises avançadas com apenas algumas linhas de código. Além disso, a documentação extensa e exemplos práticos tornam mais fácil para os usuários aprenderem a utilizar a biblioteca.