بسم الله الرحمن الرحيم

Machine Learning Approach to Predict the Probability of Recurrence of Renal Cell Carcinoma After Surgery: Prediction Model Development Study

روش یادگیری ماشین برای پیش بینی احتمال بازگشت کارسینوم سلول کلیوی پس از جراحی : پیش بینی مطالعه توسعه مدل

فاطمه ذبیحی

ژورنال کلاب

۱۴۰۲/۰۸/۲۲

مشخصات ژورنال

1
Name :
JMIR mHealth and uHealth

2
Indexing:
ISI , SCOPUS , PubMed , DOAJ

3
Category : Q1

4
Impact Factor : 5

فهرست مطالب

- مقدمه و بیان مسئله
- هدف مطالعه
- روش اجرا
- نتایج
- بحث و نتیجه گیری
- دیدگاه من

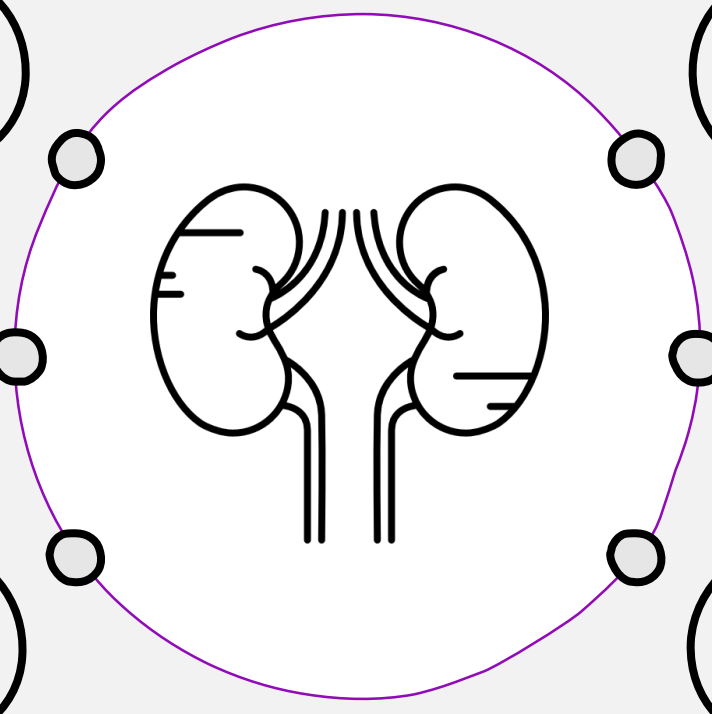
فهرست اختصارات

RCC : Renal Cell Carcinoma	سرطان سلول کلیه
NB : Naive Bayes	مدل ساده بیز
KORCC : Korean Renal Cell Carcinoma	پایگاه داده مبتنی بر وب سلول سرطانی کره
ML : Machine Learning	یادگیری ماشین
MSKCC : Memorial Sloan Kettering Cancer Center NewYork	مرکز سرطان اسلون کیورینگ مموریال

۱

مقدمه و بیان مسئله

مقدمه و بیان مسئله



1. کارسینوم سلول کلیه (RCC)
2. اقدام درمانی نفرکتومی
3. نرخ عود بالا بین ۲۰ تا ۳۰ درصد
4. مرگ بیش از ۴۰ درصد افراد مبتلا به RCC
5. اهمیت نظارت و مدیریت مداوم
6. تفکری که به دنبال این اهمیت ایجاد شد

منو

مقدمه و بیان مسئله

۹۰ درصد از سرطان های بدخیم کلیه

تعداد موارد مبتلا به سرطان RCC درکره

درمان اصلی RCC

در بین مردان ۲برابر از بقیه سرطان های بدخیم رایج تر

کشنده ترین انواع تومور

میزان عود

منو

۱ ۲ ۳ ۴ ۵

مقدمه و بیان مسئله

مطالعات گذشته

مرکز سرطان MSKCC

ایجاد نوموگرام برای پیش بینی احتمال عود سطان RCC

استفاده از داده های سانسور شده

کلیات مطالعات مقاله

جمع آوری داده از ۸ بیمارستان مراقبتی

حذف داده های سانسور شده

استفاده از داده های قابل مشاهده

هدف مطالعه

۲

هدف مطالعه

- توسعه یک الگوریتم برای پیش بینی احتمال عود RCC در ۵ و ۱۰ سال پس از عمل جراحی
- کمک گرفتن از هشت الگوریتم ML
- تهیه داد ه از یک گروه بزرگ از بیماران کره ای مبتلا به RCC در پایگاه داده مبتنی بر وب سلول سرطانی کره (KORCC)

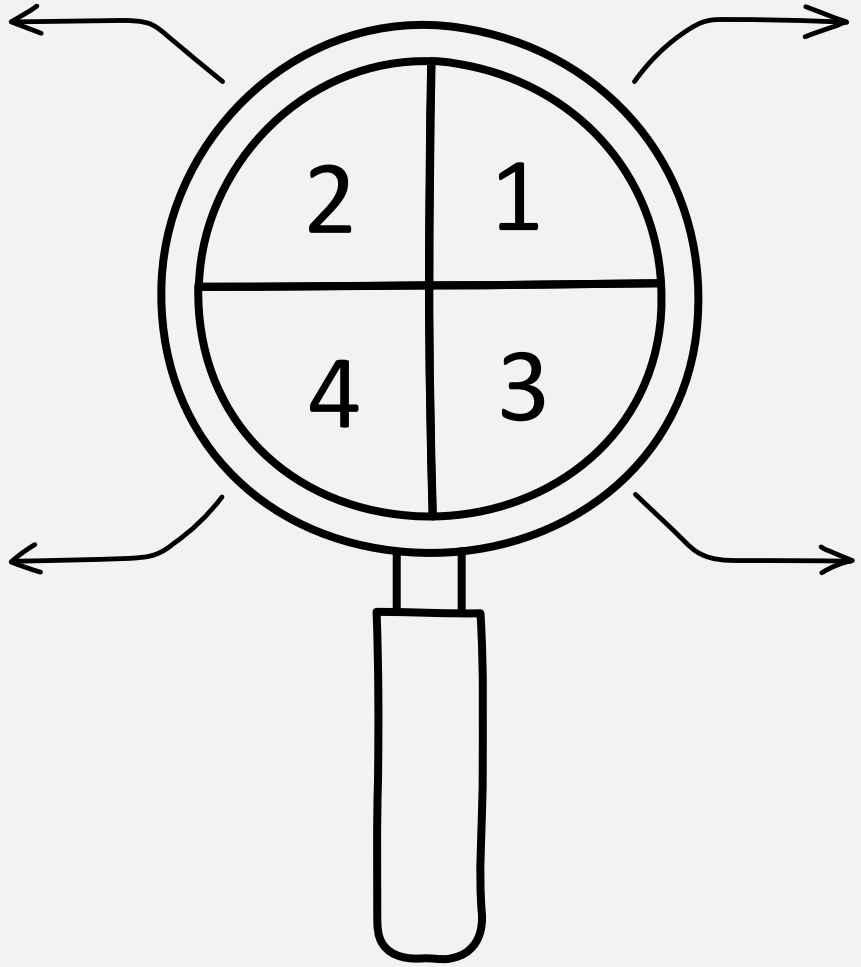
۳

روش اجرا

روش اجرا

جمع آوری داده ها از یک گروه بزرگ بیماران کره ای مبتلا به RCC در پایگاه داده مبتنی بر وب سلول سرطانی کره (KORCC)

مطالعه مورد نظر شامل ۲۰۶متغیر ، اطلاعات جمعیت شناختی ، اطلاعات آسیب شناسی و توسط هیات بازبینی سازمان دانشگاه کاتولیک کره تایید شده است

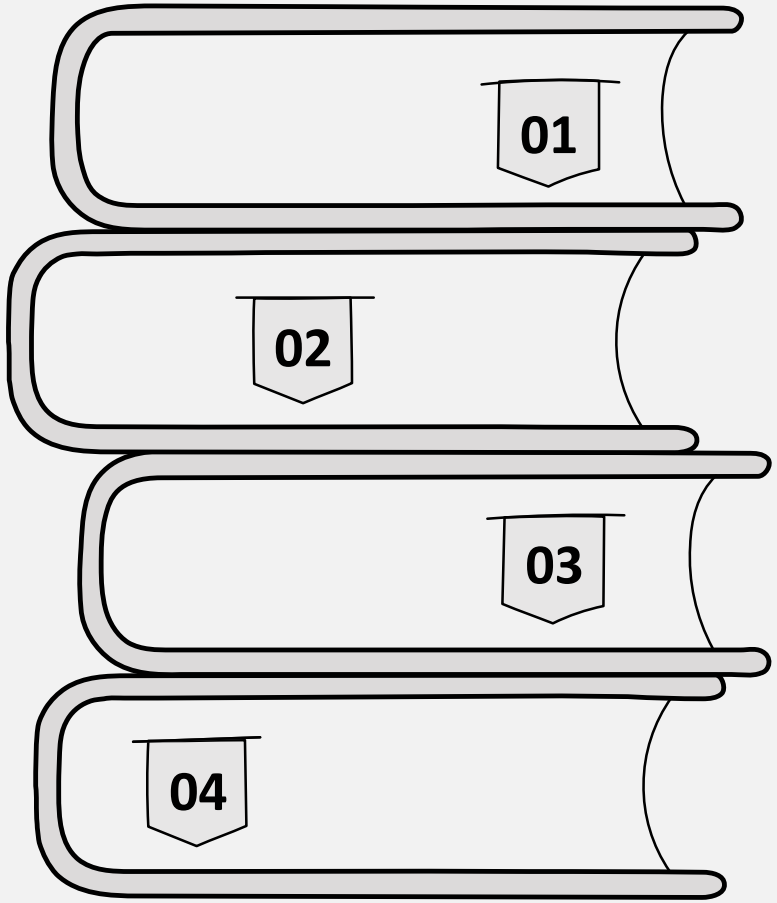


جمع آوری ۶۸۴۹داده از ۸ بیمارستان ثالث

بعد از جمع آوری داده ها متغیرها از طریق دو آزمون آماری t و آزمون مجذور کای دو بررسی شدند

منو

روش اجرا



ده متغیر انتخابی در مطالعه
براساس توصیه پزشکان

درنظر نگرفتن یک سری از
نتایج مطالعات قبل

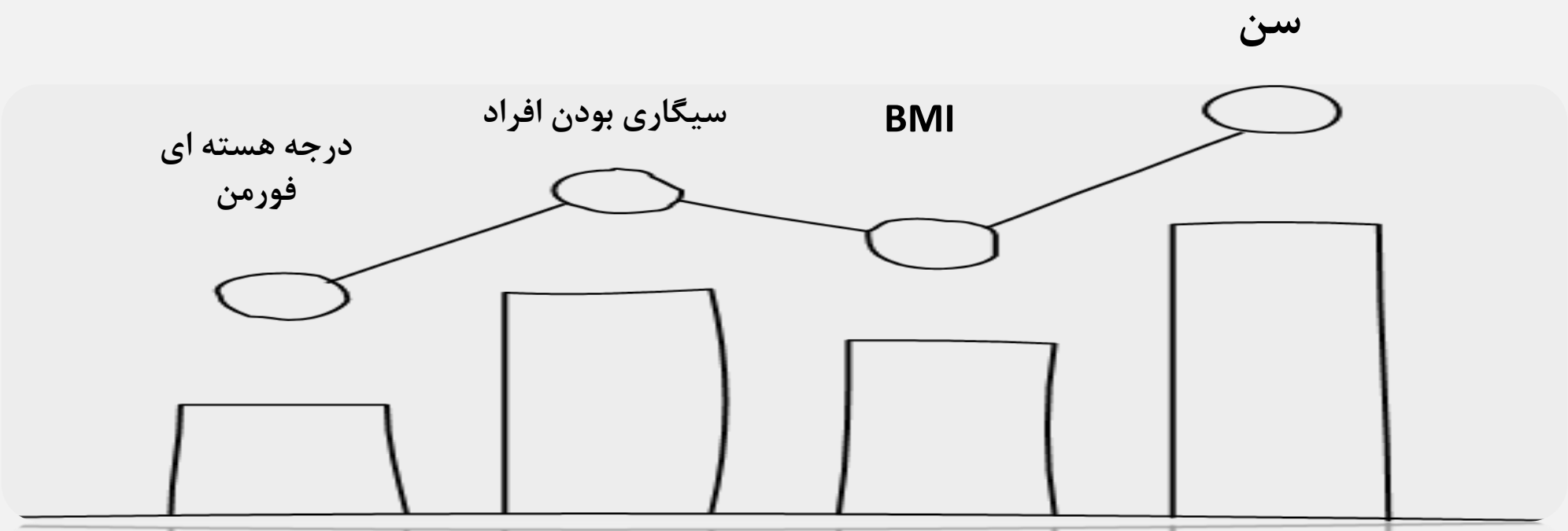
برخورد با مجموعه داده های
نا متعادل و استفاده از
تکنیک نمونه گیری بیش از
حد اقلیت ترکیبی
(SMOTE)

تحلیل آماری و توسعه مدل ML

۴

نتایج

ویژگی ها و توزیع بیماران



مقایسه متغیرها و مشخصات بیمار در دو گروه عود و غیر عود

نتایج

Table 2. Baseline characteristics of patients (N=2814).

Variable	Recurrence group (n=278)	Nonrecurrence group (n=2536)
Age (years), mean (SD)	58.4 (11.9)	55.4 (12.7)
BMI (kg/m^2), mean (SD)	23.6 (3.2)	24.7 (3.3)
Gender, n (%)		
Male	212 (76.3)	1811 (71.4)
Female	66 (23.7)	725 (28.6)
Smoking, n (%)		
Nonsmoker	207 (74.5)	2026 (79.9)
Current smoker	71 (25.5)	510 (20.1)
Necrosis, n (%)		
No	143 (51.4)	2272 (89.6)
Microscopic	30 (10.8)	126 (5.0)
Macroscopic	105 (37.8)	138 (5.4)
Lymphovascular invasion, n (%)		
No	200 (71.9)	2436 (96.1)
Yes	78 (28.1)	100 (3.9)
Capsular invasion, n (%)		
No	148 (53.2)	2114 (83.4)
Yes	130 (46.8)	422 (16.6)
Fuhrman nuclear grade, n (%)		
1	5 (1.8)	108 (4.3)
2	57 (20.5)	1293 (51.0)
3	141 (50.7)	1008 (39.7)
4	75 (27.0)	127 (5.0)

منو
۱
۲
۳
۴
۵

نتایج

عملکرد مدل
پیش بینی

استفاده از هشت الگوریتم ML

محاسبه داده های آزمون با استفاده از مقادیر :
حساسیت ، ویژگی ، دقت و AUROC

بالا بودن عملکرد الگوریتم NB نسبت به
دیگر الگوریتم ها

JMIR MEDICAL INFORMATICS Kim et a

Table 3. Diagnostic performance of machine learning algorithms for the prediction of renal cell carcinoma recurrence.

Algorithm (parameter name) and parameter value (in 5 years, in 10 years)	Sensitivity		Specificity		Accuracy		AUROC[a]	
	5-year	10-year	5-year	10-year	5-year	10-year	5-year	10-year
Kernel SVM[b,c]	0.733	0.673	0.805	0.853	0.800	0.837	0.769	0.763
Logistic regression[c]	0.644	0.692	0.839	0.816	0.823	0.805	0.741	0.754
Decision tree[c]	0.533	0.442	0.866	0.869	0.839	0.829	0.700	0.656
KNN[d] (n-neighbors)								
(100, 100)[c]	0.556	0.519	0.905	0.898	0.877	0.863	0.730	0.709
(10, 10)	0.467	0.426	0.947	0.928	0.909	0.881	0.707	0.675
(50, 50)	0.511	0.461	0.931	0.922	0.898	0.879	0.722	0.692
(200, 200)	0.556	0.481	0.899	0.902	0.871	0.863	0.727	0.691
NB[e] (alpha)								
(10, 100)[c]	0.822	0.731	0.850	0.828	0.848	0.819	0.836	0.784

منو

۵

بحث و نتیجه گیری

بحث و نتیجه گیری

منو

۱
۲
۳
۴
۵

اول

ایرادات الگوریتم پیشنهادی توسط MSKCC

دوم

دلایل برتری الگوریتم پیشنهادی جدید

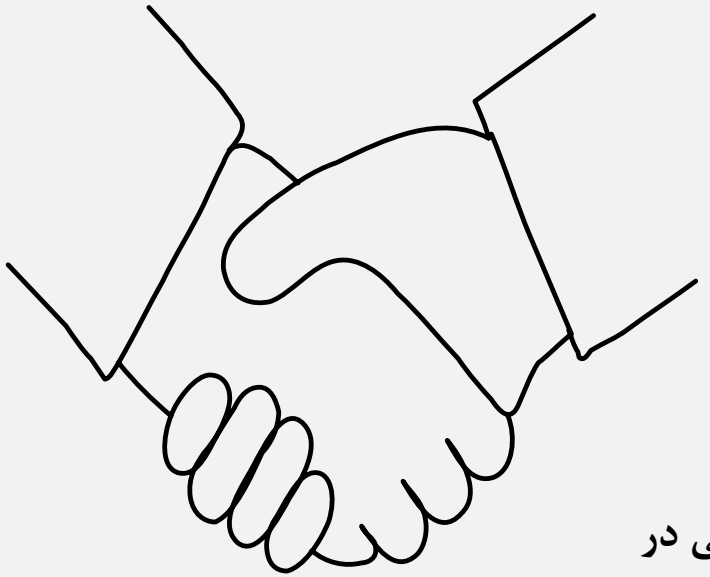
سوم

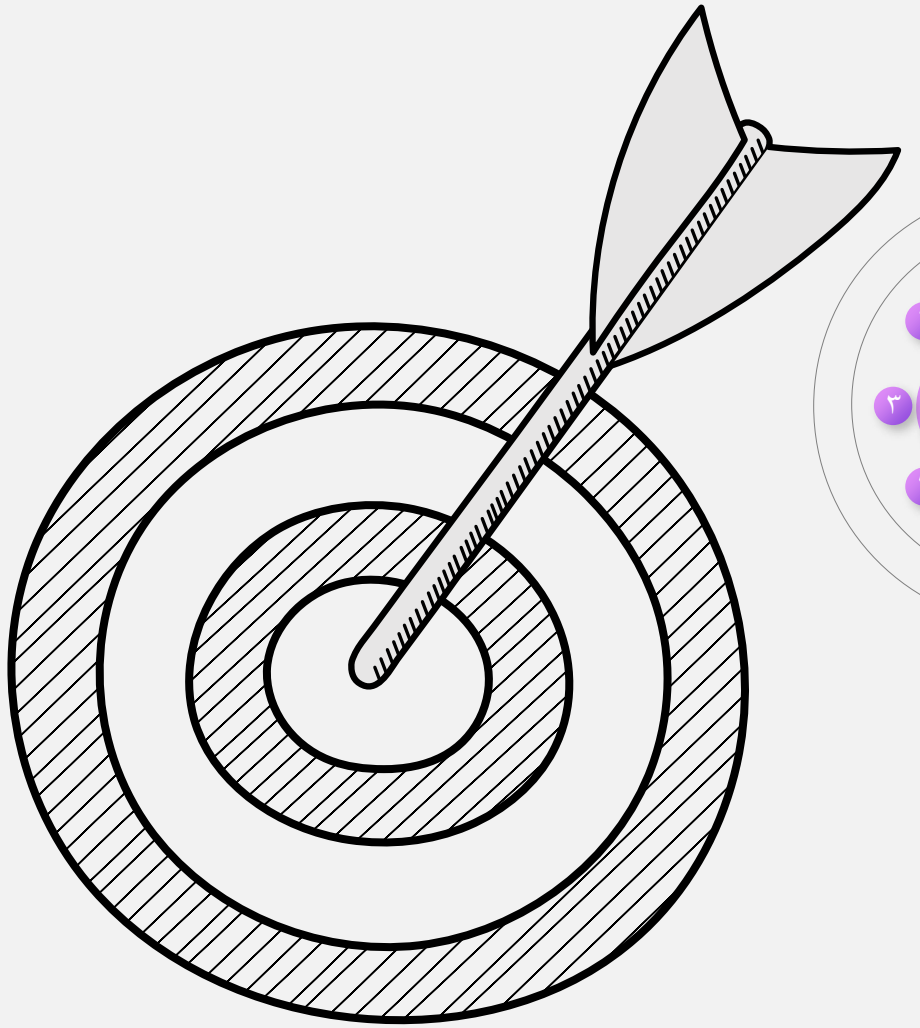
توسعه و ایجاد الگوریتم جدید

چهارم

بیان نتیجه نهایی

دیدگاه من





منو

۱ ۲ ۳ ۴ ۵

از آنجایی که متغیرهای انتخابی در مطالعه

تاثیرهای متفاوتی در عود داشتند یعنی بعضی از متغیرها باعث عود زودهنگام و بعضی باعث عود دیر هنگام میشدند

بهتر بود تا دو مدل جداگانه برای عود ۵ ساله و عود ۱۰ ساله در نظر گرفته شود

با تشکر از توجه شما