

Предикција петогодишњег преживљавања услед детекције Евинговог саркома код педијатријских пацијената

Тања Станић,
Факултет техничких наука
Универзитет у Новом Саду
Трг Доситеја Обрадовића 6
21000 Нови Сад
stanic.e247.2020@uns.ac.rs

Апстракт:

Развој прогностичког модела преживљавања може помоћи при саветовању пацијента са раком о избору лечења и за оптимизацију терапијских метода у сарадњи са онколозима и хирурзима. Циљ овог рада јесте да пружи преглед прогностичких фактора за петогодишње преживљавање. Анализирани су подаци о 2581 пацијенту, преузети из СЕЕР базе података, при чему су селектовани пацијенти са дијагностикованим Евинговим саркомом (ЕС), млађи од 20 година. Крајњи скуп података садржио је 11 независних и једну зависну променљиву, која се односи на статус преживљавања пацијента у првих пет година, након детекције Евинговог саркома. У одређивању значајних прогностичких фактора преживљавања, направљени су модели предвиђања користећи: Support Vector Machines (SVM), Decision Tree (DT, стабло одлучивања), Random Forest (RF, случајне шуме), K-Nearest Neighbor (KNN), Naïve Bayes (NB), Logistic Regression (LR), Multilayer Perceptron (MLP), Adaptive Boosting (AdaBoost) и Extreme Gradient Boosting (XGBoost) алгоритама. Важне променљиве су рангиране методама SelectKBest и RFECV. На крају, изграђена су стабла одлучивања и валидације је извршена помоћу анализе петогодишњег преживљавања. У погледу тачности модела, ф-мере и мере калибрације сви алгоритми су дали блиске резултате, где су најмање вриједности добијене моделом Naïve Bayes, а највеће AdaBoost. Важне променљиве идентификоване у овом истраживању биле су стадијум карцинома, величина тумора, врсте лечења и година детекције.

Кључне ријечи: Евингов сарком, Машинско учење, прогноза преживљавања

I. Увод

Евингов сарком (ЕС) је други најчешћи малигни сарком код деце и адолесцената и чини 1% свих тумора код педијатријских пацијената. [1] Са напретком третмана са више модела, укључујући хемотерапију, хирургију и радиотерапију, укупна петогодишња стопа преживљавања болести детектоване у локалном стадијуму се побољшала са приближно 10% на 55-65%. [2]

Лечење пацијената са метастатским и регионалним стадијумом Евингова саркома је изазовније, јер само 38% преживљава са метастатским, а 67% са регионалним стадијумом у првих пет година. [3] Рано идентификовање високо ризичних пацијената може помоћи у пружању помоћних терапија или опција испитивања. Методе које побољшавају стратификацију пацијената са ЕС могле би, према томе, резултирати побољшаним терапијским исходима.

У скороје вријеме методе машинског учења, у медицинским истраживањима, замењују традиционалне статистичке. XGBoost је способан да приведе ефикасну тачност предвиђања, једноставан за употребу и надмашио је друге алгоритме у неколико такмичења у машинском учењу. Логистичка регресија прати Гаусову дистрибуцију и обрађује све врсте променљивих, континуалне, дискретне или дихотомне, за шта није потребна нормализација.

SVM користи се за надгледану класификацију. Стабло одлучивања је надгледани алгоритам учења који илуструје резултате у лако разумљивој структури стабла где је визуализација важан фактор у анализи података. Случајна шума је извод стабла одлучивања способан је да ради и у надгледаном и у ненадгледаном режиму, рукујући континуираним и категоријским подацима у задацима регресије или класификације. [4] [5]

II. Сродна истраживања

Приликом анализе проблема пронађени су радови, засновани на истраживању детекције фактора који утичу на преживљавање услед детекције карцинома или рачунање стопе преживљавања у одређеном временском периоду од једну, двије, три, пет или десет година. Претходне студије су откриле да су клинички релевантни прогностички фактори: старост пацијента у тренутку лечења [6] [7], место тумора [8], величина тумора [9], извршена операција, примењена хемотерапија и радиотерапија [7].

Раније су клиничари користили основне софтверске програме у анализи фактора који утичу на преживљавање карцинома. Овакве конвенционалне статистичке методе нису прилагодљиве у идентификовању нових променљивих, као и у стварању креативних и интегративних визуализација. [10] [11]

Неопходно је изградити моделе како би се утврдили фактори који утичу на петогодишње преживљавање пацијената који болују од ЕС. Такође, корисно извршити избор променљивих користећи методе машинског учења у медицинском домену, где су традиционалне статистичке методе префериране међу клиничарима.

III. Скуп података:

Подаци су преузети помоћу СЕЕР програма који прикупља демографске, дијагностичке и информативне податке о лечењу свих пацијената са раком, који бораве унутар одређених Америчких региона. СЕЕР база података представља 28% САД популације. Подаци о пацијентима се достављају Националном Институту за рак без личних идентификатора. Сви подаци су јавно доступни у сврху истраживања. Аутор је потписао уговор о

коришћењу података и тиме добио приступ свим подацима.

Користећи базу података СЕЕР регистара, издвојени су сви пацијенти са детектованим Евинговим саркомом, старосне доби до 20 година. Из девет регистара прикупљени су подаци о 1231 пацијента, затим из тринаест регистара, овом претрагом, 1207 пацијената, а из осамнаест регистара на услов ове претраге одговор је био 1854 забиљежена случаја. Одбацивањем пацијената са истим идентификационим бројем, добио се коначан број посматраних случајева који износи 2581.

IV. Анализа података

СЕЕР програм подржава упите са преко 180 различитих варијабли, међутим атрибути се често односе на појединачне или групе сродних тумора, па у датој претрази таква поља остају празна. Почетни скуп података садржавао је укупно 28 атрибута. Анализа и визуализација података је вршена помоћу Python3 програмског језика и библиотека Pandas, Seaborn, Numpy, Matplotlib.

A. Стадијум болести

Анализом домена проблема утврђено је да статистијум болести у тренутку детектовања представља битан фактор у прогнози петогодишњег преживљавања. Стадијум је категорисан као: локални, регионални и дистанциран/метастазиран, кориштењем кодног система СЕЕР програма. [12] Међутим, у почетном скупу података чак за 430 пацијената статистијум болести је био непознат или недостајућ. Користећи вредности колона *Sequence number* и *Total number of malignant tumors* сваком пацијенту је додељен одговарајући статистијум. Ако је вриједност атрибута *Sequence number* била *One primary only* пацијенту је детектован само један тумор и статистијум је локални. Стадијум је регионални ако је број малигних тумора два, док је дефиниција статистијума дистанциран/метастазиран, ако је број малигних тумора већи од два. Крајњи исход је позната вриједност статистијума Евинговог саркома код сваког пацијента. Након креирања нових вриједности, које замењују непознате у колони *Stage*, колоне: број малигних тумора, број бенгних тумора и редни број у потпуности су избрисане из скупа података, како би крајња предикција била објективнија и релевантнија.

B. Величина тумора

Подаци о величини тумора су распоређени у четири колоне, у зависности од године детекције болести: 1) EOD 4 Size (1975-1987) 2) EOD 10 Size (1988-2003) 3) CS Tumor size (2004-2015) 4) Tumor Size Summary (2016+) Све четири колоне су спојене у једну, тако да је сваки пацијент имао податак о величини тумора у тренутку детекције Евинговог саркома. Величина тумора је класификована у три групе, изражена у милиметрима, < 85, 85-115 и > 115. Оваква категоризација групе тумора је уобичајена код решавања проблема сличних овом. [2] [13]

C. Раса / етничка припадност

Сви пацијенти су подељени у три расе *White*, *Black* или *Others*. У почетном скупу добијени су подаци о пацијентима из више етничких група, међутим, ради прецизније предикције, свим пацијентима из етничких група којима припада мали број пацијената, смјештени су у групу *Others*. [13]

D. Место тумора

Евингов сарком обично настаје на костима, зглобовима или меком ткиву. [2] Постојањем изузетака и проширивањем тумора на остале делове тела, у скупу података детектован је сарком на преко 30 позиција на телу. Због тачније предикције креирана је категорија *Others* у коју су уврстани пацијенти који имају сарком на некарактеристичним местима.

E. Број месеци

Почетни скуп података садржио је податак о броју месеци колико је пацијент преживио. Ови подаци послужили су за креирање циљне лабеле – колоне *Survive*.

F. Узрок смрти

Број пацијената који су живи/преживели износи 1595. Први најчешћи узрок смрти је сарком на коситма и зглобовима, затим сарком меког ткива укључујући срце. У почетном скупу налазило се и 15 пацијената који су преминули, али узрок томе није био Евингов сарком. Ови подаци послужили су за креирање циљне лабеле.

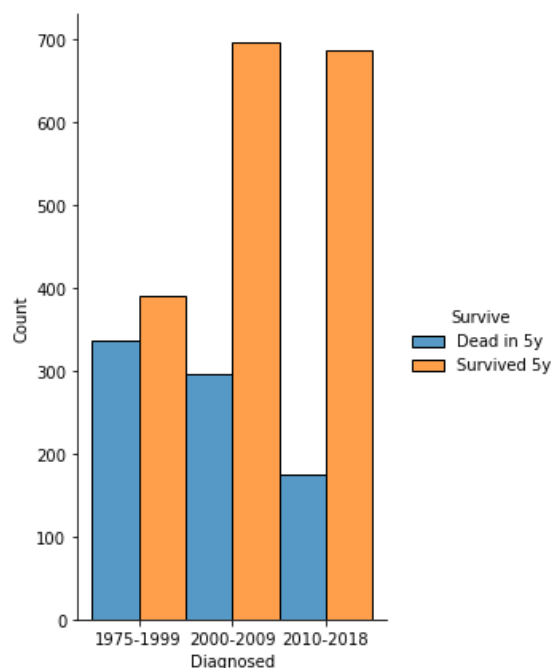
G. Survive / циљна лабела

Користећи се подацима о броју преживелих месеци, као и подацима о статусу пацијента, креирана је циљна лабела. Уколико је пацијент преминуо а број месеци је испод 60 (< 5 година) вредност циљне лабеле је *Dead in 5y*, у супротном, пацијент је преживио првих 5 година од детекције болести и вредност колоне сурвиве постаје *Survived 5y*. Ова лабела укључује

и случајеве када је пацијент преживио и када је преминуо након пет година од дијагнозе Евинговог саркома.

H. Година дијагнозе

Анализирају се подаци пацијената код којих је дијагностикован Евингов сарком од 1975 до 2018 године. Године су категорисане у три групе: 1975-1999, 2000-2009 и 2010-2018. [14] Анализом је утврђено да на петогодишње преживљавање година дијагнозе утиче директно. На слици број 1 се може уочити да се у трећој групи ('10-'18) повећао број преживелих пацијената, а смањио број преминулих у првих 5 година, за разлику од прве и друге скупине година дијагнозе. Анализом доменског проблема, може се закључити такав исход из више разлога. У каснијим годинама технологије и методе лијечења су више истражене и унапријеђене.



Слика бр. 1 Однос године дијагнозе и преживљавања пацијента

I. Операција

Приликом лијечења Евинговог саркома, „инфицирани“ део тела се је пожељно odstraniti хирушким захватом, уколико је то могуће и ако је пацијент сагласан с тим. Према томе, и информације о операцији су уврштене као фактор у предикцији преживљавања. Почетни скуп података садржио је идуће информације: 1) операција извршена, 2) операција није препоручена, 3) операција препоручена, а није извршена, 4) операција препоручена, пацијент одбио 4) операција препоручена, непознато да ли је и извршена и 5)

непознато. За овај вид истраживања довољан је само податак да ли је операција извршена или не, те су на основу вредности за дату колону, креиране су нове: да, не и непознато.

J. Хемотерапија и радиотерапија

Доменским истраживањем, изучавањем релевантних радова, а накнадно и експлоративном анализом овог скупа података, утврђено је да хемотерапија и радиотерапију директно утичу на преживљавање пацијента. [7]

K. Остали атрибути

Године пацијента у трентку дијагнозе су класификоване као 0-4, 5-9, 10-14, 15-19 година. [15]

Будући да скуп података садржи атрибуте са великим бројем недостајућих вриједности, такви атрибути су у креирању модела избачени из скупа података. Важно је нагласити да су одабрани и кориштени у анализи података за појединачне пацијенте, како би се јасније схватио модел и како би се дошло до одређених крајњих закључака. Податак о броју испитаних и захваћених лимфних чворова били су познати само за 187 пацијената (7.2%), те су те колоне у потпуности избачене из крајњег модела. Колона са информацијама о врсти извора извештаја даје информацију о начину на који су се подаци прикупили у регистре. Приликом анализе зашто код појединачних пацијената недостају вриједности у више колона, дошло се до идућег закључка: пацијенти чији су се подаци прикупљени искључиво помоћу извода из матичке књиге умрлих или лабораторија (болничких и приватних) имали су непознате вриједности атрибута *Stage* и *Surgery*. Тип извора извештаја не доприноси предикцији, па се према томе није налазио у коначном скупу података. Анализом података, имовинско стање пацијента није утицало на петогодишње преживљавање, али није избачено у моделу, циљ је био то доказати одабиром најутицајнијих атрибута за ову претрагу. [13]

Крајњи модел садржи 11 варијабли. Процентуално и бројчано, крајњи атрибути кориштени за креирање модела су приказани у табели број 1.

Промјенљива	Број пацијената	Проценат(%)
Укупно	2581	100
Age, y		
00-04	261	10.11
05-09	455	17.63
10-14	883	34.21
15-19	982	38.05
Race		
White	2275	88.14
Black	102	3.95
Other	204	7.9
Sex		
Male	1591	58.85
Female	1062	41.15
Diagnosed, y		
1975-1999	727	28.17
2000-2009	992	38.43
2010-2018	662	33.40
Stage		
Localized	1097	42.50
Regional	834	32.31
Distant	650	25.18
Site		
Bones and Joints	1879	72.80
Soft Tissue including Hart	558	21.62
Other	144	5.58
Property, \$		
< 60K	660	25.57
60K - 74K	1079	41.80
> 74K	842	32.62
Tumor size, mm		
< 85		38.20
85-115	986	41.73
> 115	1077	20.07
	518	
Surgery		
Yes	1571	60.87
No	949	36.77
Unknown	61	2.36
Chemotherapy		
Yes	2310	89.5
No	271	10.5
Radiotherapy		
Yes	913	64.63
No	1668	35.37

Табела 1. Демографске и клиничке карактеристике пацијената

V. Методологија

За предвиђање петогодишњег преживљавања пацијента кориштено је девет метода машинског учења, које се популарно користе у анализи медицинских података. Сви подаци су изражени у категоријама. Свака категорија има 2-4 могуће вриједности, из тих разлога кориштен је *One hot* метод кодирања. Применом *One Hot* кодирања

результат су 32 колоне, на основу којих се предвиђа вриједност атрибута *Survive*.

Скуп података је рандом подељен на 80% тренинг и 20% тест података. Оваква подела најчешћа је кориштена приликом медицинских истраживања, а поред тога, тестирањем је показала најбоље резултате у смислу највиших вредности мера перформанси. За оптимизацију параметара није кориштен исти скуп података, као за тренинг.

Оптимизација хиперпараметара за сваки модел машинског учења је извршена користећи *GridSearchCV* библиотеку. Сваком алгоритму прослеђено је неколико вредности за хиперпараметаре. Испробавањем вредности које су уобичајене код предикције у медицинске сврхе, добијене су почетне „најбоље“ вриједности тих параметара. [16] Након добијених вриједности тестирана је оптимизација за нешто мање и нешто веће вриједности, при чему степен величине зависи од опсега вриједности појединачних хиперпараметара. Циљ овакве оптимизације јесте да се пронађу параметри који ће извршити најбоље предвиђање.

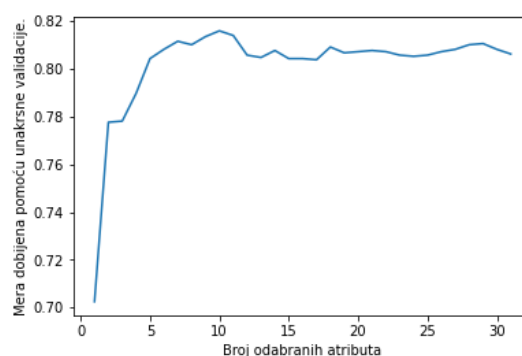
У случају SVM, најбољи параметри су: *kernal* = 'rbf', *Coefficient* = 1100, *degree* = 1, *gamma* = 0.001. Кориштена је *k* = 2 унакрсна валидација. За алгоритам стабла одлуке изабрани су идући параметри: *maximum depth* = 20, *minimum samples for split* = 2, *maximum leaf nodes* = 20. Кориштена је *k*=2 унакрсна валидација. У случају алгоритм случајне шуме кориштена је *k*=5 унакрсна валидација. Модел је дао најбољи резултат за идуће хиперпараметре: *maximum depth*=50, *number of estimators* = 200, *minimum samples leaf* = 4. Подешени параметри користећи унакрсну валидацију за *k*=2 и KNN алгоритам су: *number of neighbors*=19, *leaf size*=9. Оптимизацијом хиперпараметара за *k*=2 и

алгоритам логистичке регресије, добијене су идуће вриједности: *c*=10, *maximum iterations* = 15, *tolerance*=1e-6 За AdaBoost класификатор и

k=5 унакрсну валидацију, најбоље вредности биле су: *learning rate* = 0.01, *number of estimators* =250, *base estimator maximum depth*=6 и *base estimator minimum samples leaf*=10.

У табели број 2 приказане су мере перформанси: тачност, прецизност, поврат и ф-мера, за сваки алгоритам појединачно.

Следећи корак био је одабир променљивих за даље моделовање и визуализацију. Одговарајућ избор је важан поступак у предвиђању и анализи одлука, посебно када се ради са клиничким подацима. Избор променљивих извршен је на основу два пакета: *Recursive feature elimination with cross-validation (RFECV)* и *SelectKBest*, где се за подешено к прорачунава к колона које су највише утицале на предикцију.



Слика 2. Број одабраних обележја користећи RFECV

RFECV врши одабир најбољих обележја помоћу алгоритма стабла одлучивања, рекурзивним уклањањем обележја и унакрсним потврђеним избором најбољег броја обележја. Параметри стабла одлучивања су изабрани приликом оптимизације. За одабир најбољих атрибута

Бр.	Алгоритми	Мере евалуације			
		Тачност	Прецизност	Поврат	Ф-мера
1.	Support Vector Machines	0.85	0.85	0.94	0.89
2.	Decision Tree	0.84	0.85	0.93	0.89
3.	Random Forest	0.86	0.86	0.94	0.90
4.	K-Nearest Neighbor	0.83	0.82	0.95	0.88
5.	Naive Bayes	0.79	0.85	0.84	0.85
6.	Logistic Regression	0.81	0.82	0.92	0.87
7.	Multilayer Perceptron	0.85	0.86	0.93	0.89
8.	Adaptive Boosting	0.87	0.87	0.94	0.90
9.	Extreme Gradient Boosting	0.84	0.86	0.91	0.89

Табела 2. Мере евалуације модела

кориштена је $k=2$ унакрсна валидација. Крајњи резултат јесте одабраних 10 колона, визуелно приказано на слици број 3.

SelectKBest одабере обележја на основу k највећег скорa. Тестиран је модел за $k=10, 15$ и 20 . Уочено је да ова два алгоритма не дијагностикују потпуно исте колоне као најважније. Крајњи закључак је да на предикцију највише утичу: величина тумора, стадијум, година дијагнозе, операција, хемотерапија и радиотерапија, док је на последњем месту материјални статус пацијента. Овакав резултат избора обележја је синхронизован са експлоративном анализом података.

VI. Резултати:

Измерене су просечне перформансе различитих алгоритама у смислу броја случајева који су тачно класификовани као тражени истински позитивни (ТП) и броја случајева који су погрешно класификовани као тражени лажни позитивни (ФП). Бројеви случајева који су правилно класификовани као непотребни су истински негативни (ТН), а број случајева који су погрешно класификовани као непотребни су лажни негативни (ФН). Коришћењем ТП, ФП, ТН и ФН добијене су различите вредности прецизности, поврата, тачности и Ф-мере, користењем идућих једначина:

$$\text{Прецизност} = \text{ТП} / (\text{ТП} + \text{ФП})$$

$$\text{Поврат} = \text{ТП} / (\text{ТП} + \text{ФН})$$

$$\text{Тачност} = \text{ТП} + \text{ТН} / (\text{ТП} + \text{ФП} + \text{ТН} + \text{ФН})$$

$$\text{Ф-мера} = \frac{2(\text{Прецизност} \times \text{Поврат})}{\text{Прецизност} + \text{Поврат}}$$

Где је ф-мера хармонијска средина прецизности и поврата која представља укупне перформансе модела.

У погледу тачности модела сви алгоритми су дали блиске резултате, намање добијена тачност је применом *Naïve Bayes* алгоритма (тачност = 0.79), док је највећа применом *AdaBoost* (тачност = 0.87). Најбоља ф-мера постигнута је *AdaBoost* и алгоритмом случајне шуме. Важно је напоменути да је ф-мера свих алгоритама приближно једнака, највеће одступање је дао *Naïve Bayes*, за свега 0.05 од најбољих, већ поменутих алгоритама.

I. Поређење калибрације класификатора

Добро калибрисани класификатори су пробабилистички класификатори за које метода *predict_proba* излаз може директно протумачити као ниво поузданости. На пример, добро калибрисани класификатор треба да класификује узорке тако да међу узорцима којима је дао вредност *predict_proba* близу 0.8, тј приближно 80% заправо припада позитивној класи. На слици број 3. исцртане су калибрацијске криве свих алгоритама. Стабло одлучивања и *MLP* су генерисани је близу калибрације, што одговара мери тачности модела. На слици се може уочити највеће одступање *Naïve Bayes* алгоритма, што је опет резултат најмање тачности овог модела при предикцији у овом истраживању.

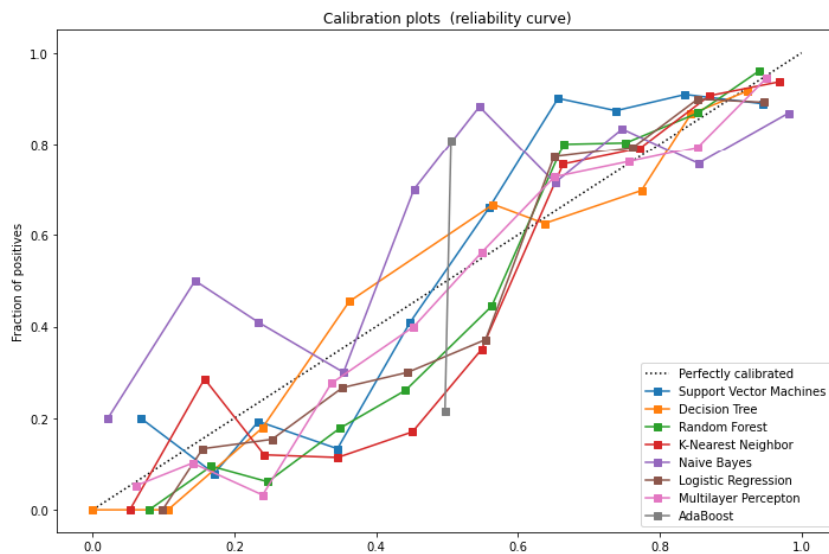
VII. Закључак

У овом истраживању направљени су модели машинског учења користећи податке о пацијентима код којих је дијагностификован Евингов сарком, преузетих из СЕЕР базе података, да би се идентификовали важни прогностички фактори за петогодишње преживљавање. Осим тога циљ овог истраживања је предвиђање петогодишњег преживљавања код ових пацијената. Ограничавајући аспект овог истраживања је то што је вршено само на популацији САД-а. У скупу података било преко 88% пацијената бијеле расе.

Током истраживања примећено је да су добијени резултати имали лошу тачност када параметри нису ефикасно иницијализовани и хиперпараметри нису правилно декларисани. Оптимизација хиперпараметара помоћу пакета *GridSearchCV* допринела је повећавању мера перформанси модела. Сви алгоритми дали су врло блиске тачности и ф-мере, с тим да се издваја *AdaBoost* као најбољи алгоритам у овом случају.

Мере изабрана за евалуацију модела је ф-мера, али приказани су и резултати прецизности, поврата и тачности модела. Такође резултати су приказани и помоћу калибрационих кривих.

Иако је велики део истраживања углавном добро конструисан и разумно добро потврђен, чини се да је потребна већа пажња приликом дизајнирања модела и примене алгоритама машинског учења, посебно у погледу количине и квалитета биолошких података. Побољшања у



Слика 3. Калибрационе криве свих алгоритама

дизајну заједно са побољшаном биолошком валидацијом несумњиво би побољшали укупан квалитет, општост и поновљивост многих класификатора заснованих на машинском учењу.

Планови за даљи развој овог решења обухватају:

- Прикупљање података о пацијентима из других извора података
- Прикупљање прецизнијих података о лечењу пацијената (дозе зрачење, циклуси хемотерапија...)
- Употреба метода полунадгледаног учења, приликом предикције
- Предикција помоћу кластеровања

Литература

- [1] Cancer.Net, Ewing sarcoma - Childhood and Adolescence:Statistics
<https://www.cancer.net/cancer-types/ewing-sarcoma-childhood-and-adolescence/statistics>, приступљено 12.05.2021. год.
- [2] Qiang Zhoua,Zong-yi Wub,Zhong-qin Lin, A nomogram to predict prognosis in Ewing sarcoma of bone, 2019.
- [3] Health NIo. National Cancer Institute, Survival Rates for Ewing Tumors,
<https://www.cancer.org/cancer/ewing-tumor/treatment/surgery/surgery-statistics>, приступљено 19.05.2021. год.
- [4] Pilaftsis A, Rubio J. The Higgs Machine Learning Challenge. Journal of Physics: Conference Series. 2015;664(7):072015.
- [5] Mogana Darshini Ganggayah, Nur Aishah Taib, Yip Cheng Har,Pietro Lio,Sarinder Kaur Dhillon, Predicting factors for survival of breast cancer patients using machine learning techniques, *BMC Medical Informatics and Decision Making*.
- [6] J. Lee, B.H. Hoang, A. Ziogas, J.A. Zell Analysis of prognostic factors in Ewing sarcoma using a population-based cancer registry, *Cancer* 116 (8) (2010) 1964–1973.
- [7] B.J. Miller, Y. Gao, K.R. Duchman, Does surgery or radiation provide the best overall survival in Ewing's sarcoma? A review of the national cancer database, *J. Surg. Oncol.* 116 (3) (2017) 384–390..
- [8] B. Biswas, S. Rastogi, S. Khan, N. Shukla, S. Deo, S. Agarwala, Developing a prognostic model for localized Ewing sarcoma family of tumors: a single institutional experience of 224 cases treated with uniform chemotherapy protocol, *J. Surg. Oncol.* 111 (6) (2015) 683–689.
- [9] G. Bacci, A. Longhi, S. Ferrari, M. Mercuri, M. Versari, F. Bertoni Prognostic factors in non-metastatic Ewing's sarcoma tumor of bone: an analysis of 579 patients treated at a single institution with adjuvant or neoadjuvant chemotherapy between 1972 and 1998, *Acta Oncol.* 45 (4) (2006) 469–475.
- [10] Eftekhari B, Mohammad K, Ardebili HE, Ghodsi M, Ketabchi E. , Comparison of artificial neural

network and logistic regression models for prediction of mortality in head trauma based on initial clinical data. *BMC Med Inform Decis Mak.* 2005;5(1):3..

- [11] Verplancke T, Van Looy S, Benoit D, Vansteelandt S, Depuydt P, Decruyenaere J, Support vector machine versus logistic regression modeling for prediction of hospital mortality in critically ill patients with haematological malignancies. *BMC Med Inform Decis Mak.* 2008;8(1):56..
- [12] A.G. Fritz, L.A.G. Ries, A.A. Hurlbut, J. Young Jr, S. Roffers, SEER Summary Staging Manual-2000 Codes and Coding Instructions, National Cancer Institute Cancer Statistics Branch, 2001.
- [13] Jun Zhang,Zhenyu Pan, Jin Yang, Jin Yang,Yuanjie Li,Jun Lyu, A nomogram for determining the disease-specific survival in Ewing sarcoma: a population study, *BMC Cancer*.
- [14] Muhammad U. Jawad, MD, Michael C. Cheung, MD, Elijah S. Min, Michaela M. Schneiderbauer, MD, Leonidas G. Koniaris, MD,Sean P. Scully, MD, PhD , Ewing Sarcoma Demonstrates Racial Disparities in Incidence-related and Sex-related Differences in Outcome.
- [15] Limin Yang, Tetsuya Takimoto and Junichiro Fujimoto, Prognostic model for predicting overall survival in children and adolescents with rhabdomyosarcoma Yang et al. *BMC Cancer* 2014, 14:654, *BMC Cancer*.
- [16] Pushpanjali Gupta, Sum-Fu Chiang, Prasan Kumar Sahoo, Suvendu Kumar Mohapatra, Suvendu Kumar Mohapatra, Djeane Debora Onthoni, Hsin-Yuan Hung, Jy-Ming Chiang , Yenlin Huang and Wen-Sy Tsai , Prediction of Colon Cancer Stages and Survival Period with Machine Learning Approach, *MDPI*.