

Süni İntellekt əsaslı Sağqalma Təhlili əsasında Pasiyentin Yaşam Müddətinin Proqnozlaşdırılması

Məsumə Məmmədova¹, Fərqanə Abdullayeva²

ETN İnformasiya Texnologiyaları İnstitutu, Bakı, Azərbaycan

¹mmg51@mail.ru, ²a_farqana@mail.ru

Xülasə— Əhalinin sağlamlığının qorunması və uzunömürlülüyün təmin edilməsi daim inkişaf edən səhiyyə və tibbin başlıca məqsədlərindəndir. Bu məqsədə nail olmaq üçün müasir tibbi praktikada pasiyentin müayinə nəticələrinin təhlili, xəstəliyin diaqnozunun dəqiq proqnozlaşdırılması və müalicə strategiyalarının əvvəlcədən qiymətləndirilməsi vacib amillər hesab olunur. Həkimlər bu proseslərdə geniş spektrli metod və texnologiyalardan istifadə edirlər. Onlar arasında sağqalma təhlili xüsusi yer tutur və bu sahədə effektiv metodologiyalardan biri hesab olunur. Səhiyyənin və tibbin sürətlə rəqəmsallaşması pasiyentin yaşam müddətinin daha dəqiq proqnozlaşdırılmasında innovativ yanaşmaların və metodların tətbiqi üçün yeni imkanlar yaratmışdır. Bu məqalədə tibbi verilənlər əsasında pasiyentin yaşam müddətini proqnozlaşdırmaq məqsədilə süni intellekt əsaslı sağqalma təhlili üçün yanaşma işlənib hazırlanmışdır. Təklif olunan yanaşmanın effektivliyi *Concordance Index* və *AUC* metrikaları əsasında qiymətləndirilmişdir. Alqoritmlərin müqayisəli təhlili göstərmişdir ki, pasiyentin yaşam müddətini ən yüksək dəqiqliklə proqnozlaşdırın model *Fast Survival SVM* olmuşdur.

Açar sözlər— sağqalma təhlili; süni intellekt; Kaplan-Meier ayrısı; *Random Survival Forest*; *Concordance index*, *time-to-event data*

I. GİRİŞ

Səhiyyə və tibb elmlərində xəstəliklərin erkən diaqnostikası, xəstənin (pasiyentin) sağlamlıq vəziyyətinin və müalicənin effektivliyinin qiymətləndirilməsi, pasiyentin xəstəxanaya daxil olduğu vaxtdan ölüm hadisəsinin baş verə biləcəyi vaxta qədər yaşadığı müddətin proqnozlaşdırılması və bu müddətə təsir edən amillərin müəyyən edilməsi kritik əhəmiyyət daşıyan tədqiqat sahəsi hesab olunur [1]. Bu müddətin proqnozlaşdırılması məsələlərinin tədqiqi və həlli tibbin mühüm statistik alətlərindən biri olan sağqalma təhlili (survival analysis) vasitəsilə aparılır. Sağqalma təhlili hadisənin baş verməsinə qədər qalan müddəti (time-to-event) proqnozlaşdırmaq üçün istifadə edilən statistik metodlar toplusudur [2].

Sağqalma təhlili xüsusilə klinik onkoloji tədqiqatlarda mühüm yer tutur. Tibbin bu sahəsində aparılan əksər tədqiqatlarda sağqalma təhlilinin tətbiqi xərçəng xəstəliyinin diaqnozu qoyulandan sonra təyin edilən müalicə üsullarının təsirini müqayisə etməyə və hansı müalicənin daha məqbul olduğunu müəyyənləşdirməyə imkan verir, əməliyyat, kimyaterapiya və ya şüa terapiyasından sonra xəstələrin yaşam müddətini proqnozlaşdırır [3].

Hadisənin baş verməsinə qədər qalan müddətin proqnozlaşdırılması üçün Cox Proportional Hazards, Kaplan Meier, log-rank tests adlı ənənəvi statistik sağqalma təhlili model və üsullarından istifadə edilir [4]. Lakin bu ənənəvi statistik sağqalma təhlili üsullarının miqyaslaşma imkanlarının olmaması, mürəkkəb qeyri-xətti əlaqələri idarə edə bilməməsi, çoxsaylı əlamətlərlə, o cümlədən böyük ölçülü verilənlərlə işləmək imkanının olmaması onların müasir səhiyyədə tətbiqini məhdudlaşdırır. Sadalanan problemlərin aradan qaldırılması üçün süni intellekt əsaslı sağqalma təhlili yanaşmalarından istifadə olunması zəruri hesab olunur [5, 6, 7].

Məqalədə tibbi verilənlər əsasında pasiyentin yaşam müddətinin proqnozlaşdırılması üçün süni intellekt əsaslı sağqalma təhlili alqoritmləri işlənmişdir.

II. SAĞQALMA TƏHLİLİNİN TARIXI

Sağqalma təhlili verilənlərin analizini həyata keçirmək üçün statistik üsullardan ibarət elm sahəsidir. Burada əsas çıxış parametri olaraq pasiyentin müəyyən bir hadisədən, məsələn, müalicəyə başlama və ya xəstəliyin diaqnoz qoyulmasından sonra nə qədər yaşayacağı müddət qəbul edilir. [8]. Sağqalma təhlili müşahidənin başlanma anından hadisənin baş vermə anına qədər olan vaxtı öyrənir. Sağqalma təhlili ilk dəfə ölüm hadisələrinin baş verməsi səbəblərinin öyrənilməsi məsələlərində tətbiq edilmişdir. Hazırda sağqalma təhlili xəstəliyin başlanmasına qədər qalan vaxtın, birjanın çökməsinə qədər qalan vaxtın, avadanlıqların sıradan çıxmasına qədər qalan vaxtın, zəlzələnin baş verməsinə qədər qalan vaxtın öyrənilməsi kimi geniş sahələrdə də tətbiq edilir [9]. Sağqalma təhlili metodları leykoz xəstələri və remissiyada olan vaxtın, normal fərdlər üçün ürək xəstəliyinin inkişaf vaxtının, yaşlı əhali və ölümə qədər qalan vaxtın, ürək transplantasiyası və ölümə qədər qalan vaxtın proqnozlaşdırılması məsələlərində də tətbiq edilir.

Sağqalma təhlilinin yaranması tarixi İkinci Dünya müharibəsindən sonra hərbi texnikanın etibarlılığına artan maraqla əlaqələndirilir.

Ölüm haqqında verilənləri, o cümlədən imtinaların vaxtını tədqiq edən statistik metodlardan ibarət sağqalma təhlili aparatı, müştərilərin daha təhlükəsiz və etibarlı məhsullara tələbatı artdığından xüsusi ilə sənayedə sürətlə yayılmağa başlamışdır.

III. SAĞQALMA TƏHLİLİNİN BƏZİ ANLAYIŞLARI

Sağqalma funksiyası (*Survival Function*). Sağqalma funksiyası adətən Kaplan-Meier əyrisi şəklində təsvir edilir. Kaplan-Meier əyrisində şaquli enmə baş verən hadisəni göstərir. Bu funksiya fərdin müəyyən bir gün üçün sağqalma ehtimalını hesablayır. Tibbi tədqiqatlarda adətən müalicədən və ya diaqnozdan sonra müəyyən bir müddət yaşayan xəstələrin hissəsini ölçmək üçün istifadə edilir. Kaplan-Meier funksiyasının düsturu aşağıdakı kimi ifadə edilir:

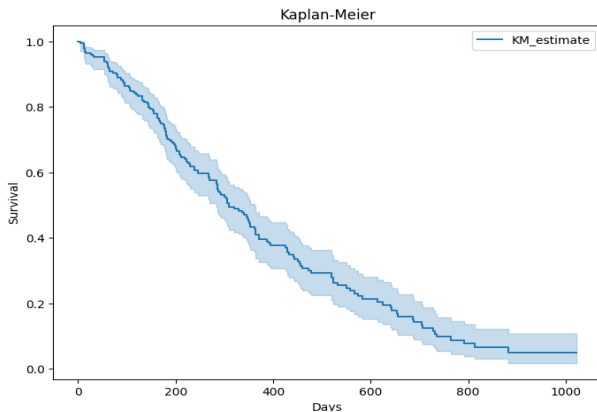
$$S(t_i) = S(t_{i-1}) \times \left(1 - \frac{d_i}{n_i}\right)$$

burada $S(t_i)$ - t_i -ci anda sağqalma ehtimalıdır, d_i - i -ci anda baş verən hadisələrin sayıdır, n_i - i -ci anda sağ qalan pasiyentlərin sayıdır.

Kaplan-Meier düsturu hər hansı şəxsin t_i -ci anda sağqalma ehtimalının t_{i-1} -ci anda sağqalma ehtimalından asılılığını göstərir.

Tibbi tədqiqatlarda real ssenarilərin əksəriyyətində sağqalma şansı adətən zamandan asılı olaraq aşağı düşür. Məsələn, insan nə qədər uzun müddət xəstə olarsa, onun sağqalma ehtimalı müalicə olunmadıqda bir o qədər aşağı olar. Kaplan-Meier düsturu onu göstərir ki, sağqalma ehtimalı zamandan asılı olaraq kiçilməyə doğru hərəkət edir. Yanaşmada fərz edilir ki, əgər şəxs x -ci günə qədər sağ qalıbsa, deməli o əvvəlki günlərin hər birində də sağ qalmalı idi. Bugünün sağqalma ehtimalı dünyanın sağqalma ehtimalından kiçik və ya bərabərdir. Bütün bu gündəlik ehtimalları bir-birinə vurduqda ümumi sağqalma əyrisi alınır.

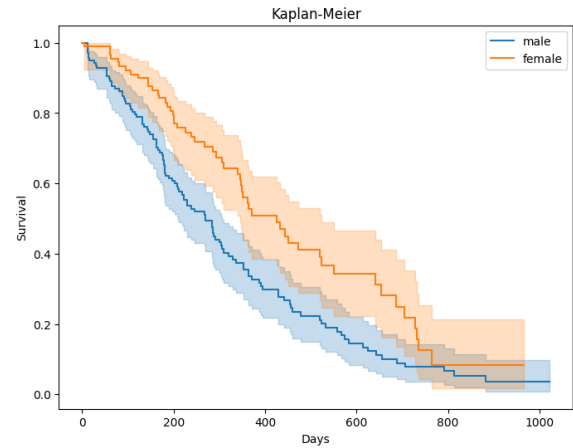
Ağciyər xərçəngi olan insanlar haqqında toplanmış verilənlər üzərində Kaplan-Meier funksiyasını tətbiq etdikdə aşağıdakı qrafik alınmışdır [10].



Şəkil 1. Kaplan-Meier əyrisi [10].

Bu əyri sağqalma ehtimalının zamana görə necə dəyişdiyini göstərir. İlk mərhələdə sağqalma ehtimalı yüksəkdir, daha sonra sürətlə aşağı düşməyə cəhd edir. Yəni daha çox insanın ölməsi sağqalma ehtimalını azaldır. Bu əyridən göründüyü kimi 5% insan müəyyən zaman periodundan sonra sağ qala bilmişdir. Həmin oblasta qədər olan insanlar isə dünyadan köçmüşdür, ya da sağala bilməmişdir.

Ağciyər xərçəngi olan insanlar haqqında toplanmış verilənlər üzərində Kaplan-Meier funksiyası qadın və kişilər üçün də fərqli nəticələr göstərmişdir (Şəkil 2).



Şəkil 2. Qadın və kişilərə nəzərən çəkilmiş Kaplan-Meier əyrisi [10].

Qeyd edək ki, sağqalma təhlilinin Kaplan-Meier üsulundan başqa loqarifmik sıralama testləri (log-rank tests) və Koks nisbətlili risklər (Cox proportional hazards) adlı modelləri də geniş istifadə edilir [11].

Senzura (Censoring). Hadisənin baş vermə vaxtının dəqiq məlum olmadığı situasiyaları əks etdirir.

Risk Funksiyası (Hazard function). Vaxtdan asılı olaraq ölüm riskini müəyyən edən funksiyadır.

IV. SÜNI INTELEKT ƏSASLI SAĞQALMA TƏHLİLİ ÜSULLARI

Müasir səhiyyə və tibbi tədqiqatların aparılmasında əhəmiyyətli rol oynayan süni intellekt əsaslı sağqalma təhlili üsulları ənənəvi statistik modellərlə müqayisədə daha yüksək dəqiqliklə nəticə əldə etməyə imkan verir.

Bu modellər xüsusən böyük verilənlər üzərində işləyərkən əlverişli hesab edilir. Bu növ üsullar, xəstəliklərin inkişafı, müalicə cavabları və xəstələrin sağqalma ehtimalları barədə daha dəqiq məlumatlar əldə etməyə imkan verir. Süni intellekt əsaslı sağqalma təhlili üsullarına aşağıdakıları aid etmək olar [11]:

Random Survival Forest (RSF) Təsadüfi Meşənin xüsusi olaraq sağqalma analizi üçün hazırlanmış bir növüdür. RSF-də hər bir qərar ağacı sağqalma ağacı qurur və ağacda hər bir qovşadakı nümunələrin sağqalma ehtimalını Kaplan-Meier modeli qiymətləndirir.

Gradient Boosting Survival Analysis (GBSA). Sağqalma məlumatlarını modelləşdirmək üçün gradient boosting modellərindən istifadə edən sağqalma təhlili əsaslı maşın təlimi yanaşmasıdır. GBSA modelində itki funksiyası olaraq Koks modelinin loqarifmik ehtimalı istifadə edilir.

Fast Survival SVM. SVM alqoritminin modifikasiyasıdır, xüsusi olaraq sağqalma təhlili üçün modelləşdirilmişdir.

FastSurvivalSVM modelində itki funksiyası olaraq Cox tipli loqarifmik ehtimal istifadə edilir.

V. EKSPERIMENTLƏR

Sağqalma təhlilinin aparılması adətən tədqiqat məsələsində müddətlə bağlı komponent olduqda mümkün olur. Sağqalma təhlili hər hansı hadisə ilə onun baş verdiyi vaxt arasında qarşılıqlı əlaqələrin müəyyən edilməsi tələb olunan məsələlərdə tətbiq olunur.

Ekspperimentlərin aparılması üçün [12] verilənlər bazası istifadə edilmişdir. Verilənlər bazasında 2 nəticə çıxarma parametrləri (əlamətləri) istifadə edilmişdir: 1) event (hadisə - öldü/sağ qaldı), vaxt (nə qədər vaxt qaldı). İstifadə edilmiş verilənlər bazasının strukturu Şəkil 3-də göstərilmişdir.

Unnamed: 0	inst	time	status	age	sex	ph.ecog	ph.karno	pat.karno	meal.cal	wt.loss	
0	1	3.0	306	1	74	1	1.0	90.0	100.0	1175.0	NaN
1	2	3.0	455	1	68	1	0.0	90.0	90.0	1225.0	15.0
2	3	3.0	1010	0	56	1	0.0	90.0	90.0	NaN	15.0
3	4	5.0	210	1	57	1	1.0	90.0	60.0	1150.0	11.0
4	5	1.0	883	1	60	1	0.0	100.0	90.0	NaN	0.0

Şəkil 3. NCCTG verilənlər bazasının strukturu

Verilənlər bazasının əlamətləri haqqında məlumat aşağıda verilmişdir.

CƏDVƏL 1. NCCTG VERİLƏNLƏR BAZASININ ƏLAMƏTLƏRİ

Əlamətlər	Şərh
inst	Xəstənin müşahidə edildiyi müəssisənin kodu
time	Xəstənin sağ qaldığı və ya senzura məruz qaldığı günlər
status	Senzura statusu (0 = senzura, 1 = öldü)
age	Xəstənin illərlə ifadə olunan yaşı
sex	1 = Kişi, 2 = Qadın
ph.ecog	ECOG performans (0 - 4; 0 = asimptomatik, 4 = yatağa bağlı)
ph.karno	Həkim tərəfindən qiymətləndirilmiş performans balı (Karnofski) (0 = pis, 100 = yaxşı)
pat.karno	Yuxarıdakı kimi, lakin xəstə tərəfindən qiymətləndirilib
meal.cal	Yemək zamanı istehlak edilən kalorilər
wt.loss	Son altı ayda itirilən çəki (lbs.) miqdarı

Sağqalma təhlilinin aparılması üçün ənənəvi maşın təlimi modellərindən fərqli olaraq 2 hədəf dəyişənlər istifadə edilir. Burada y hədəf dəyişəni elə bölünməlidir ki, hadisəni əks etdirən indikatorun qiyməti bu ədəd, vaxtı göstərən indikatorun qiyməti sürüşkənvergüllü ədəd olmalıdır. Bizim yanaşmada bu iki dəyişən qismində **status** və **time** parametrləri götürülmüşdür. Verilənlər bazasını 228 nümunə təşkil edir. Bu nümunələrdən 182 nümunə təlim üçün, 46 nümunə test üçün istifadə edilmişdir.

Tibbi verilənlər üzərində sağqalma təhlilinin aparılması üçün Random Survival Forest, Gradient Boosting Survival Analysis, Fast Survival SVM, Fast Kernel Survival SVM alqoritmlərindən istifadə edilmişdir. Bu alqoritmlərin effektivliyinin qiymətləndirilməsi üçün C-indeksindən (concordance index) və Əyri Altındakı Sahə (ing. Area Under the Curve, AUC) metrikasından istifadə edilmişdir. Qurulmuş alqoritmlərin NCCTG Lung Cancer Data verilənlər bazası üzərində C-index və AUC metrikası əsasında test nəticələri Cədvəl 2-ə əlavə edilmişdir.

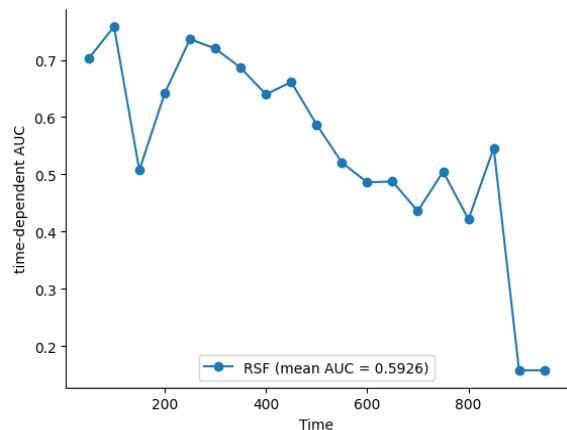
CƏDVƏL 2. ALQORITMLƏRİN TEST NƏTİCƏLƏRİ

Alqoritmlər	C-index score	Mean AUC score
Random Survival Forest	0.6248	0.5926
Gradient Boosting Survival Analysis	0.6251	0.5891
Fast Survival SVM	0.6452	0.5988
Fast Kernel Survival SVM	0.5352	0.4608

Cədvəldən göründüyü kimi Fast Survival SVM modeli digər modellərlə müqayisədə daha yüksək nəticə göstərmişdir. Belə ki, Fast Survival SVM modelinin C-indeks qiyməti 0.6452 olduğu halda Random Survival Forest, Gradient Boosting Survival Analysis və Fast Kernel Survival SVM modellərinin qiymətləri uyğun olaraq 0.6248, 0.6251 və 0.5352 olmuşdur.

C-indeksinin qiymətinin 0.5-dən yuxarı olması proqnozlaşdırma modelinin yaxşı olduğunu göstərir. C-index modelin ümumi effektivliyini qiymətləndirir. Müxtəlif vaxt nöqtələrində modelin effektivliyinin qiymətləndirilməsi üçün isə *cumulative dynamic AUC* metrikasından istifadə edilir.

RSF modelinin effektivliyinin *cumulative dynamic AUC* metrikası əsasında qiymətləndirilməsinin vizual təsviri aşağıdakı şəkildə verilmişdir.



Şəkil 4. RSF modelinin AUC əyrisi.

Şəkildən göründüyü kimi modelin orta AUC qiyməti 0.5926 olmuşdur.

Qeyd edək ki, Random Survival Forest modelinin illər üzrə dəyişkən qiymətlər almasına verilənlər bazasının ayrı-ayrı əlamətləri təsir edir. Bu əlamətləri müəyyən etmək üçün

Permutation Importance funksiyasından istifadə edərək əlamətlərə çəki əmsalları hesablayaq. Hesablanmış çəki əmsallarının qiymətləri Cədvəl 3-də verilmişdir.

CƏDVƏL 3. RANDOM SURVIVAL FOREST ÇƏKİLƏRİ

Weight	Feature
0.0430 ± 0.0427	sex
0.0364 ± 0.0489	pat.karno
0.0149 ± 0.0443	ph.ecog
0.0061 ± 0.0381	age
0.0014 ± 0.0152	Unnamed: 0
-0.0050 ± 0.0177	inst
-0.0061 ± 0.0291	wt.loss
-0.0070 ± 0.0162	meal.cal
-0.0182 ± 0.0307	ph.karno

Cədvəldən göründüyü kimi sex, pat.karno, ph.ecog Random Survival Forest modelinin nəticələrinə təsir edən əsas əlamətlərdir.

Gradient Boosting Survival Analysis modelinin nəticələrinə sex, pat.karno, wt.loss, ph.ecog əlamətləri təsir etmişdir.

CƏDVƏL 4. GRADIENT BOOSTING SURVIVAL ANALYSIS ÇƏKİLƏRİ

Weight	Feature
0.0626 ± 0.0560	sex
0.0451 ± 0.0460	pat.karno
0.0416 ± 0.0526	Unnamed: 0
0.0260 ± 0.0368	wt.loss
0.0101 ± 0.0280	ph.ecog
0.0053 ± 0.0417	age
-0.0022 ± 0.0331	inst
-0.0092 ± 0.0183	meal.cal
-0.0116 ± 0.0289	ph.karno

NƏTİCƏ

Məqalədə ağciyər xərçəngi xəstələrinin yaşam müddətinin proqnozlaşdırılması üçün süni intellekt əsaslı sağqalma təhlili üsulları tədqiq edilmişdir. Qurulmuş alqoritmlərin real verilənlər bazası üzərində test nəticələrinin müqayisəli analizi aparılmışdır. Aparılmış eksperimentlər təklif edilən sağqalma modelləri və alqoritmlər ansamblının ənənəvi sağqalma təhlili üsulları ilə müqayisədə daha yüksək dəqiqliyə malik olduğunu

göstərir və bu istiqamətdə tədqiqatların davam etdirilməsini stimullaşdırır.

MINNƏTDARLIQ

Bu iş Azərbaycan Elm Fondunun maliyyə dəstəyi ilə yerinə yetirilmişdir – Qrant № AEF-MCG-2023- 1(43)-13/04/1-M-04.

ƏDƏBİYYAT

[1] E.T. Lee, O.T. Go, “Survival analysis in public health research,” Annual Review of Public Health, vol 18., pp. 105-134, 1997.

[2] C. Kartsonaki, “Survival analysis,” Diagnostic Histopathology, vol. 22, no. 7, pp. 263–270, 2016.

[3] P. Nathan, J.C. Hassel, P. Rutkowski, “Overall Survival Benefit with Tebentafusp in Metastatic Uveal Melanoma,” N Engl J Med. Vol. 385, no. 13, pp. 1196–1206, 2021.

[4] T.G. Clark, M.J. Bradburn, S.B. Love, D.G. Altman, “Survival Analysis, Part I: Basic Concepts and first analyses,” Br J Cancer, vol. 89, no. 2, pp. 232–238, 2003.

[5] S. Gao, V.D. Calhoun, J. Sui, Machine learning in major depression: From classification to treatment outcome prediction, CNS Neurosci Ther, vol. 24, no. 11, pp. 1037–1052, 2018.

[6] V. Mhasawade, Y. Zhao, R. Chunara, Machine learning and algorithmic fairness in public and population health, Nat Mach Intell, vol. 3, no. 8, pp. 659–666, 2021.

[7] P. Doupe, J. Faghmous, S. Basu, “Machine Learning for Health Services Researchers,” Value in Health, vol. 22, no. 7, pp. 808–15, 2019.

[8] R. Singh, K. Mukhopadhyay, “Survival analysis in clinical trials: Basics and must know areas,” Perspect Clin Res, vol. 2, no. 4, pp. 145–148, 2011.

[9] T. Smith, B. Smith, “Survival analysis and the application of Cox’s proportional hazards modeling using SAS,” Proceedings of the Twenty Sixth Annual SAS Users Group International Conference, 2001, pp. 22-25.

[10] R. Lyu, Survival Analysis of Lung Cancer Patients from TCGA Cohort, Advances in Lung Cancer, vol. 9, pp. 1-15, 2020.

[11] R. Sonabend, A. Bender, “Machine Learning in Survival Analysis,” 2025, p. 229.

[12] NCCTG lung cancer data set, <https://monolixsuite.slp-software.com/monolix/2024R1/ncctg-lung-cancer-data-set>