



АНАЛИЗ НА АКТУАЛНОТО СЪСТОЯНИЕ В ДИАГНОСТИКАТА НА РАКА НА ГЪРДАТА В БЪЛГАРИЯ ЧРЕЗ ИЗПОЛЗВАНЕ НА СОФТУЕР ЗА АНАЛИЗ С ИЗКУСТВЕН ИНТЕЛЕКТ

Д. Димитров¹, М. Караманлиев¹, Ив. Петрова¹, М. Шошкова¹, Д. Бойчев²

¹ Клиника по онкологична хирургия, УМБАЛ „Д-р Георги Странски“, катедра „Пропедевтика на хирургическите болести“, Медицински университет – Плевен,

² Бизнес развитие, Скиллайн, София,

ANALYSIS OF THE CURRENT STATE OF BREAST CANCER DIAGNOSIS IN BULGARIA USING ARTIFICIAL INTELLIGENCE ANALYSIS SOFTWARE

D. Dimitrov¹, M. Karamanliev¹, I. Petrova¹, M. Shoshkova¹, D. Boychev²

¹ Department of surgical oncology, University Hospital “Georgi Stranski”, Department of Propedeutics of surgical diseases, Medical University – Pleven, Pleven, Bulgaria

² Business development, Sqilline, Sofia, Bulgaria

РЕЗЮМЕ

Ракът на млечната жлеза (РМЖ) е социално-значимо заболяване. Според GLOBOCAN за 2020 РМЖ е на първо място в света по честота на нови случаи при двата пола (11.7%) и на четвърто място по смъртност (6,9%). Според последните данни от Българския раков регистър от 2015 г. РМЖ заема първо място по честота на нови случаи при жени с 26,8% и е на първо място по смъртност при жени с 17,4%. Анализираха се данните от януари 2019 г. до май 2023 г. на Национално ниво. Използва се електронната информация, подавана от болниците след дехоспитализация на пациентите, достъпна в платформата Danny. Това е софтуерно решение, което обединява и хармонизира различни източници на данни от онкологични, кардиологични и други практики чрез използване на изкуствен интелект. За периода са извършени общо 12 989 оперативни интервенции по КП 193 и КП 194 с диагноза на пациенти МКБ С50. При 8 673 пациенти е извършена биопсия на формация, както следва: дебелоиглена биопсия при 4

SUMMARY

Breast cancer (BC) is a socially important disease. According to GLOBOCAN for 2020, breast cancer ranks first in the world in the incidence of new cases in both sexes (11.7%) and fourth in mortality (6.9%). According to the latest data from the Bulgarian Cancer Registry from 2015, BC ranks first in the incidence of new cases in women with 26.8% and is first in mortality in women with 17.4%. Data from January 2019 to May 2023 on a national-wide basis were analyzed. Electronic information submitted by hospitals after patient dehospitalization available on the Danny platform by Sqilline is used. It is a software solution that brings together and harmonizes diverse data sources from oncology, cardiology, and other practices using artificial intelligence. A total of 12,989 surgical interventions were performed during the period into clinical pathway (CP) 193 and clinacal pathway (CP) 194 with patients diagnosed with ICD C50. Breast biopsy was performed in 8,673 patients as follows: core-needle biopsy (CNB) in 4,915 patients (56.67%), excisional biopsy in 3,051 patients (35.18%), and

915 пациенти (56,67%), ексцизионна биопсия при 3 051 пациенти (35,18%) и инцизионна биопсия при 707 пациенти (8,15%). Геффрир е извършен при 5 982 пациенти (46,05%), открит чрез търсене на ключова дума „геффрир“ в текста на епикризите.

В България все още съществуват трудности при осъществяването на съпоставими със световните изисквания честоти на осъществяване на дебелоиглена биопсия за диагноза на РМЖ и непрепоръчително високи честоти на използването на геффрир в диагностиката на РМЖ.

КЛЮЧОВИ ДУМИ: рак на млечната жлеза, платформа Danny, изкуствен интелект, биопсия, геффрир, дебелоиглена биопсия

incisional biopsy in 707 patients (8.15%). Frozen section was performed in 5,982 patients (46.05%), detected by searching for the keyword "геффрир" in the text of the documents.

In Bulgaria, there are still difficulties in achieving comparable rates of core-needle biopsy for the diagnosis of breast cancer and unacceptably high rates of the use of frozen section in the diagnosis of breast cancer.

The aim of this study is to investigate whether the standards in breast cancer diagnosis in Bulgaria are being met using artificial intelligence analysis software.

KEY WORDS: breast cancer, platform "Danny", artificial intelligence, frozen section, biopsy, core-needle biopsy

ВЪВЕДЕНИЕ

Ракът на млечната жлеза (РМЖ) е социално значимо заболяване. Според GLOBOCAN за 2020 РМЖ е на първо място в света по честота на нови случаи при двата пола (11.7%) и на четвърто място по смъртност (6,9%) [1]. Според последните данни от Българския раков регистър от 2015 г. РМЖ заема първо място по честота на нови случаи при жени с 26,8% и е на първо място по смъртност при жени с 17,4% [2].

Ранното откриване на рака на гърдата чрез мамография, клиничен преглед на гърдата и самото изследване продължава да е от основно значение за подобряване на преживяемостта (3–6). Последните изследвания обаче хвърлиха светлина и върху нови образни методи и биомаркери, които предлагат по-голяма чувствителност и специфичност при ранната диагностика и наблюдението на отговора на лечението [4,7].

Европейските насоки за осигуряване на качество при рака на гърдата налагат въвеждане на програма за скрининг на гърдата във всички европейски страни [8]. Досега България не е въвела изцяло национална програма за скрининг на рака на гърдата.

Европейското дружество по медицинска онкология (ESMO) и Националната мрежа за борба с рака (NCCN) предоставят насоки, които очертават дебелоиглената биопсия като златен стандарт за диагностициране на рака на гърдата и че запазващата гърдата хирургия с биопсия на сентинелен лимфен възел (SLNB) е метод на избор [8].

Важно е да се отбележи, че анализът на замразени срезове (геффрир) има някои ограничения, включително възможността за грешки при вземането на проби и интерпретационни предизвикателства поради артефакти от замразяването. Насоките на Националната мрежа за борба с рака (NCCN) непрепоръчват рутинно извършване на геффрир за диагностициране на рак на гърдата. Той може да бъде полезен в определени ситуации, като например оценка на хирургичните граници или оценка на засягането на лимфните възли по време на операцията.

Целта на настоящата статия е проучване, до каква степен се спазват стандартите в диагностиката на рака на гърдата в България, чрез използване на софтуер за анализ с изкуствен интелект.

МАТЕРИАЛИ И МЕТОДИ

Анализираха се данните от януари 2019 г. до май 2023 г. на Национално ниво. Използва се електронната информация, подавана от болниците след дехоспитализиране на пациентите, дос-

тъпна в платформата Danny. Това е софтуерно решение, което обединява и хармонизира различни източници на данни от онкологични, кардиологични и други практики, чрез използване на изкуствен интелект. Осъществено от SAP HANA, водеща база данни в паметта, която позволява безпроблемна интеграция на структурирани и неструктурирани данни от многобройни източници на здравни данни.

Като използва собствени алгоритми за машинно обучение и обработка на естествен език (NLP), Danny извлича структурирани и неструктурирани здравни данни. Тя преминава през процеси на предварителна обработка и нормализация, за да се гарантира качеството на данните, преди да се извършат анализи и да се генерират прогнози.

За да защити неприкосновеността на личния живот на пациентите, платформата Danny включва няколко функции за сигурност. Това включва деидентификация на данните за защита на самоличността на пациентите, съответствие с Общия регламент за защита на данните (GDPR), прилагане на добре структурирана политика за оторизация за различни нива на достъп и анонимизиране на данните.

Освен това, платформата включва статистически инструменти, които осигуряват поглед върху реалните болнични данни, и позволяват анализ в различни категории. Тези функции увеличават възможностите на платформата да извлича ценни прозрения от данните за здравеопазването и да подпомага процесите на вземане на решения.

В този ретроспективен анализ на базата данни са използвани дигитализирани медицински досиета в България, за да се извлекат данни за пациенти, приети за операция на рак на гърдата в периода януари 2019 г. - май 2023 г.

Компанията собственик на Danny platform има частни договорни отношения с повечето Университетски болници и Онкологични центрове в страната, както и с множество по-малки Многопрофилни и Специализирани болници. За създаването на този анализ не са използвани публични данни, а медицински записи на ниво уникален пациент. Информацията е била напълно анонимизирана преди постъпване в Danny platform и нейната статистическа обработка. Изработването на текущия доклад бе осъществено чрез предоставяне на достъп на авторите и съответните научни институции до конкретни изискани параметри и информация от платформата след сключване на частни партньорски отношения.

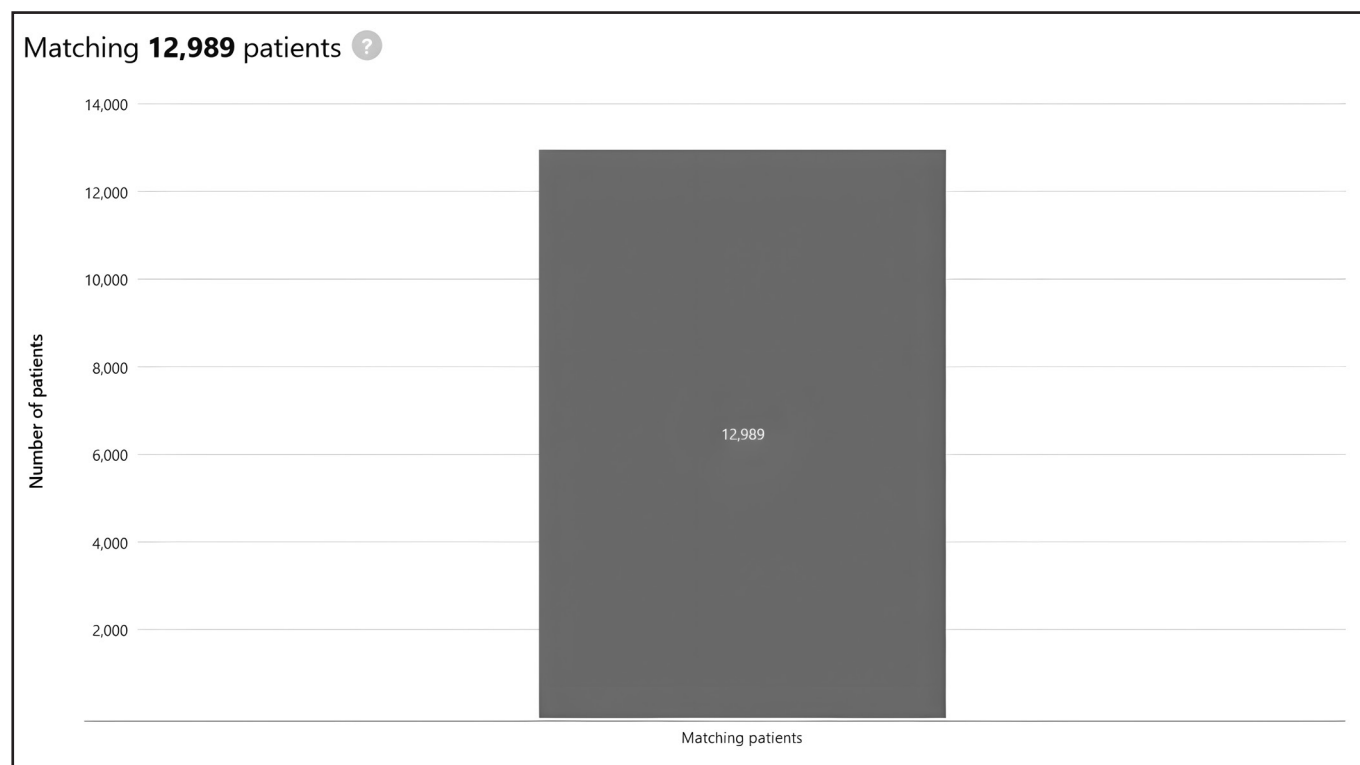
Според данни от Danny platform, работеща активно в сферата на онкологията в България през периода 01.2019 г. - 05.2023 г., е имало 16 986 ново-диагностицирани случая на рак на гърдата. В текущата публикация са анализирани оперативни интервенции при 12 989 пациенти в същия период, което представлява 76,47% от всички ново-диагностицирани пациенти в периода в България.

Структуриране на данните

В това проучване използвахме платформата Danny - система, предназначена за наблюдение на пациенти, които се движат между различни здравни заведения, като болници, амбулаторни практики или лаборатории. Платформата гарантира, че не се съхранява никаква лична информация, тъй като уникалният идентификационен номер на пациента се анонимизира чрез хаширащ алгоритъм. Този процес генерира личен номер, който остава последователен във всички източници на данни, позволявайки проследяване на пациентите, без да се разкрива тяхната идентификараща информация.

Въпреки че някои от входящите данни вече са били структурирани в специфични параметри, значителна част от тях се състоят от свободен текст, ръчно въведен от лекари и други здравни специалисти. За да направим тези неструктурирани данни използвани, извършихме допълнителни стъпки за предварителна обработка и извличане по време на фазата на импортиране на данни. Тези стъпки включваха консолидиране на данните на ниво транзакции в унифицирана структура на данните, която можеше лесно да се интерпретира, и да се подложи на техники за извличане на стойност чрез обработка на естествен език (NLP).

За извличане на конкретни стойности на параметрите от свободния текст, използвахме различни подходи, включително регулярни изрази, съвпадение на низове, сходство на думи и по-усъвършенствани методи на NLP. Изборът на най-подходящия подход зависеше от задълбочен и



Фиг. 1. Пациенти хоспитализирани по КП 193 и КП 194 с диагноза МКБ С50 за периода януари 2019 г. – май 2023 г.

непрекъснат анализ на свободния текст и свързаните с него параметри. Това щателно проучване ни позволи да изберем най-подходящата техника за всеки сценарий.

Нормализиране на данните

След извличането на стойностите на параметрите от данните, беше необходимо допълнително нормализиране, поради значителните различия в синтаксиса в различните източници на данни. За да се справим с този проблем, за нормализирането на всеки параметър беше използван подход, основан на правила. За лабораторните резултати беше използвана международната система от единици и официално признатите мерни единици на всеки лабораторен тест. Основните използвани системи за кодиране бяха МКБ-10 за кодиране на заболявания и МКБ-9-СМ за кодиране на процедури, които са официално утвърдените системи за кодиране от Националната здравноосигурителна каса в България.

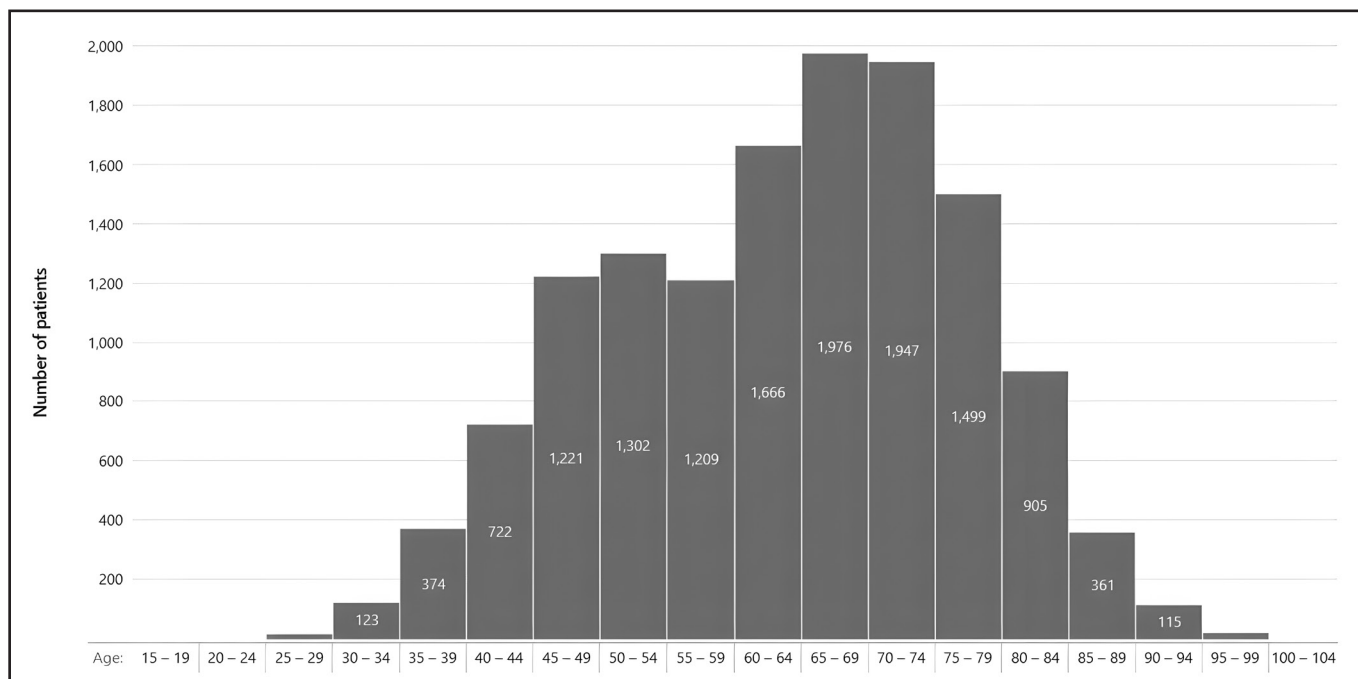
В случаите, когато системата не успяваше да намери подходящо правило за нормализация за определен параметър, той беше маркиран за допълнително разглеждане. Тогава експертите по контрол на качеството отговаряха за провеждането на допълнителни последващи процедури за създаване на нови правила за нормализация, специфични за този параметър. Този повтарящ се процес гарантираше, че базата данни поддържа актуален списък с правила, което позволяваше непрекъснатото анализиране и запазване на постъпващите данни.

Чрез прилагането на този подход, нормализирането на стойностите на параметрите беше подобро, което улесни последователността и точността при анализа и тълкуването на данните.

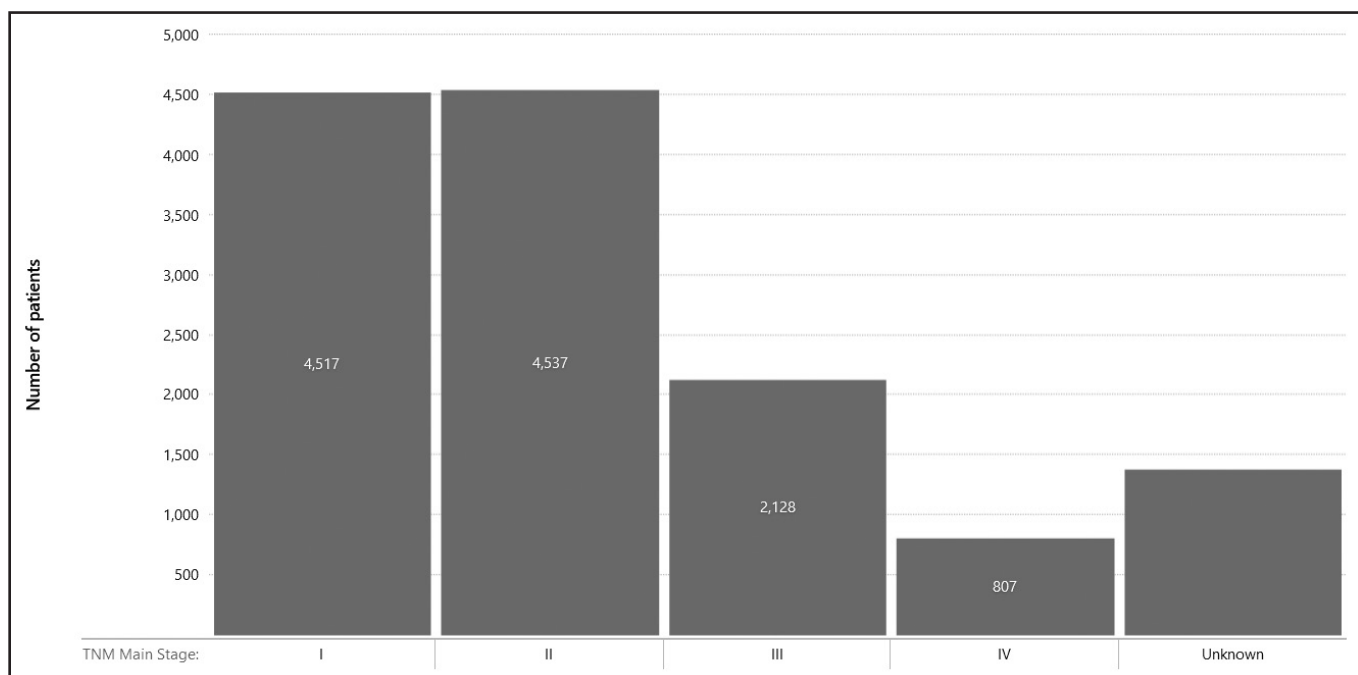
Това проучване е подкрепено от Европейския фонд за регионално развитие чрез Оперативна програма "Наука и образование за интелигентен растеж" по договор №BG05M2OP001-1.002-0010-C01(2018-2023).

РЕЗУЛТАТИ

За периода са извършени общо 12 989 оперативни интервенции по КП 193 и КП 194 с диагноза на пациенти МКБ С50.



Фиг. 2. Разпределение на пациентите по възраст



Фиг. 3. Разпределение на пациентите по TNM стадий

На фигури 2 и 3 е показано разпределението на пациентите по възраст и стадий на заболяването.

Забелязва се тенденцията за пик на рака на гърдата във възрастта 60-79 г., както и немалкият брой пациенти диагностицирани преди 40-годишна възраст. Над 9000 пациенти са в първи и втори стадий на заболяването, а пациентите диагностицирани в четвърти стадий са 807 (6.2%).

При 8 673 пациенти е извършена биопсия на формация, както следва: дебелоиглена биопсия при 4 915 пациенти (56,67%), ексцизионна биопсия при 3 051 пациенти (35,18%) и инцизионна биопсия при 707 пациенти (8,15%).

Таблица 1. Брой извършени диагностични процедури за РМЖ за периода

Общ брой оперирани пациенти	12989 пациенти
Дебелоиглена биопсия	4915 пациенти (37,84% от общия брой пациенти)
Ексцизионна биопсия	3051 пациенти
Инцизионна биопсия	707 пациенти
Геффрир	5982 пациенти (46,05% от общия брой пациенти)

Геффрир е извършен при 5 982 пациенти (46,05%), открит чрез търсене на ключова дума „геффрир“ в текста на епикризите.

Общият брой пациенти по групи в табл. 1 е по-голям от броя на уникалните пациенти, защото при 1666 от тях се среща повече от една патолого-диагностична модалност, в частност инициална биопсия плюс последващ интраоперативен геффрир.

ОБСЪЖДАНЕ

Забелязва се тенденция за по-ранно откриване на рака на гърдата в България, с 69.7% пациенти диагностицирани в първи и втори стадий на заболяването, в настоящата кохорта. Това вероятно е свързано с по-висока здравна култура на пациентите и все по-широкото навлизане извършването на мамографии. Все пак част от пациентите от втори стадии и всички пациенти в трети стадии са показани за неoadювантна терапия. Извършването на такава е най-подходящо след диагноза с дебелоиглена биопсия.

Дебелоиглената биопсия е неоспорим златен стандарт в диагностиката на РМЖ. През годините е показала множеството си предимства пред останалите типове биопсии (инцизионна, ексцизионна и тънкоиглена биопсия) [8–11]. Използването ѝ позволява минимално-инвазивно добиване на материал и информация за хистологичния вид на тумора, рецепторен статус и определяне на субтипове. Бъдейки в ерата на неoadювантната терапия в лечението на РМЖ, дебелоиглената биопсия е най-добрият начин за определяне на възможните кандидати за неoadювантна терапия. Неoadювантна терапия след ексцизионна биопсия или геффрир са несъществуващи термини и извършването на такива процедури на пациенти, които биха били подходящи за неoadювантна терапия влошава прогнозата им.

Дебелоиглената биопсия на гърда е лесен, изпълним, сигурен и надежден метод с висока чувствителност и специфичност. Най-честите варианти на изпълнението му са под мануален и под ултразвуков контрол. Ултразвуковият контрол значително подобрява успеваемостта на метода и няма минимален размер на формация, под която не може да се използва метода [12–14]. В настоящото проучване при едва 37,84% (4915/12989) от пациентите е доказан рак на гърдата чрез дебелоиглена биопсия.

Геффрирът в диагностиката на РМЖ е изключително непрепоръчителен. Той намира своето място в изследването на резекционни линии и съмнителни лимфни възли. Извършването на ексцизия на формация, последващ геффрир и вземане на решение за радикална операция в един етап е неприемливо от медицинска, научна, морално-етична и деонтологична гледна точка.

Danny platform в България, както и другите аналитични платформи в света са софтуерни системи, които обработват и анализират големи количества медицински данни, за да идентифицират модели и тенденции, които могат да бъдат използвани за подобряване на грижите за пациентите.

Развитието в съвременната онкология е изключително силно водено от изводите при анализ на големи масиви от данни (big data). Това започва още в края на предния век и началото на сегашния с откриването на първите таргетни молекули - trastuzumab при HER2+ рак на гърдата и imatinib през 2002, таргетиращ BCR-ABL при хронична миелоидна левкемия. Тези събития отприщват своеобразна вълна от проучвания и проекти, свързани със секвенирането на генома в онкологията. Един от тях е Раковият геномен атлас (The Cancer Genome ATLAS – TCGA), при които се анализират първоначално се секвенират 33 различни вида рак и общото количество мултимодална информация надхвърля 2.5 пета байта [15].

Благодарение на този и много други проекти, използващи аналитични платформи за големи обеми данни, в нашите данни имаме установени панели с генетични маркери, които да се изследват при различните видове рак.

През 2021 година, учените от Caris Life Sciences Phoenix, Arizona, публикуват свое проучване, при което разработват модел на изкуствен интелект, които на базата на молекулярния подпис на тумора, могат да кажат дали пациентът ще има по-добри резултати при стандартната първа линия на терапия или алтернативни медикаментозни терапии. В конкретния случай се касае за пациенти с метастатичен колоректален карцином. Заключениеето на авторите е, че разработката и употребата на подобни модели може да подобри резултатите при пациенти с метастатичен колоректален карцином, както и други видове рак [16].

Именно такива са и прогнозите на множество експерти в областта. Бъдещето на онкологията и медицината като цяло е базирано на обработката и анализа на големи количества данни, което ще доведе до силно персонализирана медицина и по-добри резултати.

Изкуственият интелект навлиза все повече в ежедневието и в медицинската практика. Не успяхме да открием подобни статии, които да проучват състояние на диагностика или лечение на РМЖ с подобен софтуер използващ изкуствен интелект. Това вероятно се дължи на факта, че в развитите страни има национален регистър и национална база данни, която е достъпна, и анализът от там е лесен. Изкуственият интелект се използва широко в образната диагностика на РМЖ и има обещаващо бъдеще в персонализираната гена терапия [17–21].

ЗАКЛЮЧЕНИЕ

В България все още съществуват трудности при осъществяването на съпоставими със световните изисквания честоти на осъществяване на дебелоиглена биопсия за диагноза на РМЖ и непрепоръчително високи честоти на използването на гефрир в диагностиката на РМЖ.

КНИГОПИС / REFERENCES

1. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin.* 2021;71(3):209-249. doi:10.3322/caac.21660
2. Валерианова З., Атанасов Т., Вуков М. (Редактори). *Заболяемост От Рак в България, 2014 и 2015.* Български Национален Раков Регистър. София, 2017.
3. Wang L. Early Diagnosis of Breast Cancer. *Sensors.* 2017;17(7). doi:10.3390/s17071572
4. Shieh Y, Eklund M, Madlensky L, et al. Breast Cancer Screening in the Precision Medicine Era: Risk-Based Screening in a Population-Based Trial. *J Natl Cancer Inst.* 2017;109(5). doi:10.1093/jnci/djw290
5. Cedolini C, Bertozzi S, Londero AP, et al. Type of Breast Cancer Diagnosis, Screening, and Survival. *Clin Breast Cancer.* 2014;14(4):235-240. doi:https://doi.org/10.1016/j.clbc.2014.02.004
6. Jafari SH, Saadatpour Z, Salmaninejad A, et al. Breast cancer diagnosis: Imaging techniques and biochemical markers. *J Cell Physiol.* 2018;233(7):5200-5213. doi:https://doi.org/10.1002/jcp.26379
7. Morris E, Feig SA, Drexler M, Lehman C. Implications of Overdiagnosis: Impact on Screening Mammography Practices. *Popul Health Manag.* 2015;18 Suppl 1(Suppl 1):S3-11. doi:10.1089/pop.2015.29023.mor
8. Cardoso F, Kyriakides S, Ohno S, et al. Early breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol.* 2019;30(8):1194-1220. doi:10.1093/annonc/mdz173
9. Munch-Petersen HD, Rasmussen BB, Balslev E. Reliability of histological malignancy grade, ER and HER2 status on core needle biopsy vs surgical specimen in breast cancer. *APMIS.* 2014;122(9):750-754. doi:https://doi.org/10.1111/apm.12213
10. Kwok TC, Rakha EA, Lee AHS, et al. Histological grading of breast cancer on needle core biopsy: the role of immunohistochemical assessment of proliferation. *Histopathology.* 2010;57(2):212-219. doi:https://doi.org/10.1111/j.1365-2559.2010.03620.x
11. Meattini I, Bicchierai G, Saieva C, et al. Impact of molecular subtypes classification concordance between preoperative core needle biopsy and surgical specimen on early breast cancer management: Single-institution experience and review of published literature. *Eur J Surg Oncol.* 2017;43(4):642-648. doi:https://doi.org/10.1016/j.ejso.2016.10.025
12. Shah VI, Raju U, Chitale D, Deshpande V, Gregory N, Strand V. False-negative core needle biopsies of the breast. *Cancer.* 2003;97(8):1824-1831. doi:https://doi.org/10.1002/cncr.11278
13. Osanai T, Gomi N, Wakita T, et al. Ultrasound-guided Core Needle Biopsy for Breast Cancer: Prelimi-

- nary Report. *Jpn J Clin Oncol.* 2000;30(2):65-67. doi:10.1093/jjco/hyd015
14. Schueller G, Jaromi S, Ponhold L, et al. US-guided 14-gauge Core-Needle Breast Biopsy: Results of a Validation Study in 1352 Cases. *Radiology.* 2008;248(2):406-413. doi:10.1148/radiol.2482071994
 15. Stephens ZD, Lee SY, Faghri F, et al. Big Data: Astronomical or Genomical?. *PLoS Biol.* 2015;13(7):e1002195. Published 2015 Jul 7. doi:10.1371/journal.pbio.1002195
 16. Abraham JP, Magee D, Cremolini C, et al. Clinical Validation of a Machine-learning-derived Signature Predictive of Outcomes from First-line Oxaliplatin-based Chemotherapy in Advanced Colorectal Cancer. *Clin Cancer Res.* 2021;27(4):1174-1183. doi:10.1158/1078-0432.CCR-20-3286
 17. Tan P, Chen X, Zhang H, Wei Q, Luo K. Artificial intelligence aids in development of nanomedicines for cancer management. *Semin Cancer Biol.* 2023;89:61-75. doi:https://doi.org/10.1016/j.semcancer.2023.01.005
 18. Sechopoulos I, Teuwen J, Mann R. Artificial intelligence for breast cancer detection in mammography and digital breast tomosynthesis: State of the art. *Semin Cancer Biol.* 2021;72:214-225. doi:https://doi.org/10.1016/j.semcancer.2020.06.002
 19. Becker AS, Marcon M, Ghafoor S, Wurnig MC, Frauenfelder T, Boss A. Deep Learning in Mammography: Diagnostic Accuracy of a Multipurpose Image Analysis Software in the Detection of Breast Cancer. *Invest Radiol.* 2017;52(7). https://journals.lww.com/investigativeradiology/Fulltext/2017/07000/Deep_Learning_in_Mammography__Diagnostic_Accuracy.7.aspx
 20. Syed AH, Khan T. Evolution of research trends in artificial intelligence for breast cancer diagnosis and prognosis over the past two decades: A bibliometric analysis. *Front Oncol.* 2022;12. doi:10.3389/fonc.2022.854927
 21. Urso L, Manco L, Castello A, et al. PET-Derived Radiomics and Artificial Intelligence in Breast Cancer: A Systematic Review. *Int J Mol Sci.* 2022;23(21). doi:10.3390/ijms232113409

АДРЕС ЗА КОРЕСПОНДЕНЦИЯ:

Д-р Мартин Караманлиев, дм
Клиника по онкологична хирургия,
УМБАЛ „Д-р Георги Странски“
Плевен 5809
бул. Георги Кочев 8А,
e-mail: martinkaramanliev@gmail.com

ADDRESS FOR CORRESPONDENCE:

Dr. Martin Karamanliev, MD, PhD
Clinic of Oncological Surgery,
University Hospital „Dr. Georgi Stranski“
8A, Georgi Kochev blvd.
5809 Pleven, Bulgaria
e-mail: martinkaramanliev@gmail.com