

УДК 618.19

Технічні науки

Носовець Олена Костянтинівна

кандидат технічних наук,

доцент кафедри біомедичної кібернетики

Національний технічний університет України

«Київський політехнічний інститут імені Ігоря Сікорського»

Носовец Елена Константиновна

кандидат технических наук,

доцент кафедры биомедицинской кибернетики

Национальный технический университет Украины

«Киевский политехнический институт имени Игоря Сикорского»

Nosovets Olena

Candidate of Technical Sciences,

Associate Professor of the Department of Biomedical Cybernetics

National Technical University of Ukraine

"Igor Sikorsky Kyiv Polytechnic Institute"

Новохатня Олександра Сергіївна

студентка магістратури

Національного технічного університету України

«Київський політехнічний інститут імені Ігоря Сікорського»

Новохатня Александра Сергеевна

студентка магистратуры

Национального технического университета Украины

«Киевский политехнический институт имени Игоря Сикорского»

Novokhatnia Oleksandra

Master's Student of the

National Technical University of Ukraine

"Igor Sikorsky Kyiv Polytechnic Institute"

**ІНФОРМАЦІЙНА СИСТЕМА ОЦІНКИ НАСЛІДКІВ ПІСЛЯ
ЛІКУВАННЯ РАКУ ПЕРЕДМІХУРОВОЇ ЗАЛОЗИ
ІНФОРМАЦІЙНА СИСТЕМА ОЦІНКИ НАСЛІДКІВ ПІСЛЯ
ЛІКУВАННЯ РАКУ ПЕРЕДМІХУРОВОЇ ЗАЛОЗИ
INFORMATION SYSTEM OF EVALUATION OF CONSEQUENCES
AFTER TREATMENT OF PROSTATE CANCER**

***Анотація.** В роботі визначено значущі параметри для моделей прогнозування. Створення інформаційної системи оцінки наслідків після лікування раку передміхурової залози на основі розроблених математичних моделей для передбачення рецидиву та виживаності.*

***Ключові слова:** рак передміхурової залози, прогнозування, регресія, класифікація, логістична регресія, інформаційна система, МГУА.*

***Аннотация.** В работе определены значимые параметры для моделей прогнозирования. Создание информационной системы оценки последствий после лечения рака предстательной железы на основе разработанных математических моделей для предсказания рецидива и выживаемости.*

***Ключевые слова:** рак предстательной железы, прогнозирования, регрессия, классификация, логистическая регрессия, информационная система, МГУА.*

***Summary.** The work defines significant parameters for forecasting models. Creation of an information system for assessing the effects of post-treatment prostate cancer on the basis of developed mathematical models for prediction of relapse and survival.*

***Key words:** prostate cancer, prediction, regression, classification, logistic regression, information system, GMDH.*

Актуальність дослідження. Згідно з даними Національного інституту онкології США в Америці більш ніж в 160 тис чоловіків щороку діагностується рак простати [1, с. 20].

Смертність від раку передміхурової залози зростає протягом останніх 10 років. Втім, частка чоловіків, які вмирають від цієї хвороби, у період з 2010 по 2015 рік скоротились на 6%.

На сьогодні до найбільшої кількості смертей призводять такі види раку, як рак легенів та шлунка, а рак простати - на третьому місці.

В останні два десятиліття минулого століття і на початку нинішнього істотно підвищилася частота виявлення раку простати. Відбувається це як за рахунок більш раннього розпізнавання, так і внаслідок природного приросту, природа якого залишається малозрозумілою.

Необхідно вчасно не тільки діагностувати хворобу, а й запобігти виникненню рецидиву. Оцінивши ризики після лікування, можна уникнути виникнення ускладнень.

Матеріали дослідження. У роботі представлені результати обстеження та лікування 135 хворих на рак передміхурової залози, які проходили лікування в Київській міській онкологічній лікарні та ДУ «Інститут урології АМН України» протягом 2012-2016 рр. В дослідженні брали участь 405 хворих віком 51-75 років (середній вік - 65,2±9,8 років).

Клінічний діагноз встановлювали на підставі визначення загального початкового рівня простат специфічного антигену (ПСА) в сироватці крові, пальцевого ректального дослідження простати, комп'ютерної томографії, трансректального ультразвукового дослідження передміхурової залози, остеосцинтиграфії, рентгенографії органів грудної порожнини. Верифікація діагнозу – за даними трансректальної мультифокальної біопсії простати, а також біопсії пухлини проведеної трансуретральної резекції [6].

Методи дослідження. В якості статистичного пакету для обробки даних та відбору прогностично цінних ознак використано пакет IBM SPSS Statistics 21.0. Для побудови прогностичної класифікаційної моделі для визначення настання рецидиву використано програмний продукт GMDH Shell [5]. Для побудови класифікаційної прогностичної моделі для визначення ймовірності настання летального наслідку використано мову програмування Python 3 та бібліотеку Scikit-learn.

Результати та обговорення.

Відбір прогностично цінних ознак, які в подальшому розглядалися як фактори ризику настання рецидиву захворювання, проводилися шляхом розрахунку коефіцієнтів ризику (Hazard Ratio, HR), при побудові регресійної моделі Кокса [4].

Результат оцінки показників дозволив відібрати такі предиктори для математичного моделювання рецидиву як:

- вік (HR 0,956 (0,917-0,996% ДІ));
- статус хірургічного краю резекції (HR 4,019 (1,981-8,155% ДІ));
- розповсюдженість первинної пухлини (HR 1,425 (1,18-1,722% ДІ));
- дисеміновані пухлинні клітини (HR 0,284 (0,140-0,577% ДІ));
- біомаркер для виявлення ракових стоволовбурових клітин CD44 (HR 0,299 (0,121-0,740% ДІ));
- експресія CD44 (HR 2,619 (1,801-3,808% ДІ));
- проведення променевої терапії (HR 4,284 (1,989-9,071% ДІ));
- проведення орхектомії (HR 8,691 (3,434-21,994% ДІ));
- неoad'ювантне лікування (HR 3,635 (1,725-7,663% ДІ)).

Результати оцінки прогностичної цінності показників для математичної моделі виживаності, дозволили відібрати такі предиктори як:

- статус хірургічного краю резекції (HR 7,967 (3,212-19,762% ДІ));
- поширення на лімфатичні вузли (HR 22,614 (7,425-68,868% ДІ));

- розповсюдженість первинної пухлини (HR 1,552 (1,272-1,893% ДІ));
- дисеміновані пухлинні клітини (HR 0,090 (0,033-0,241% ДІ));
- експресія біомаркера для виявлення ракових стоволовбурових клітин CD44 (HR 2,873 (2,138-3,982% ДІ));
- проведення променевої терапії (HR 4,389 (1,938-9,257% ДІ));
- проведення орхектомії HR 7,991 (3,683-22,344% ДІ));
- неoad'ювантне лікування (HR 5,388 (2,184-13,295% ДІ)).

Для побудова математичної моделі прогнозування настання рецидиву використані методи групового обліку аргументів [6] (покрокове включення, комбінаторний алгоритм, змішаний і нейросетевой), бінарна логістична регресія, RandomForest, Support Vector Machine (метод опорних векторів) [2; 3]. Аналіз побудованих моделей проведено шляхом розрахунку показників чутливості, специфічності, а також загальної прогностичної цінності моделі. Найбільший позитивну прогностичну точність має модель отримана методом покрокового включення МГУА = 98.4%.

Загальний вигляд побудованої моделі:

$$Y = -3.38759 \times 10^{-6} + 0.54145 \times \sqrt[3]{X1} + 0.817054 \times \sqrt[3]{X2} \times \sqrt[3]{X3} - \\ - 0.708652 \times \sqrt[3]{X2} \times \sqrt[3]{X4} - 0.665543 \times \sqrt[3]{X5} \times \sqrt[3]{X1} + 0.149379 \times \sqrt[3]{X6} \times \\ \times \sqrt[3]{X5} - 0.0687171 \times \sqrt[3]{X3} \times \sqrt[3]{X1} + 0.546218 \times (\sqrt[3]{X1})^2$$

X1– проведення орхектомії

X2– дисеміновані пухлинні клітини

X3– біомаркер для виявлення ракових стоволовбурових клітин CD44

X4- експресія CD44

X5– проведення променевої терапії

X6– вік

Для побудови математичної моделі прогнозування настання летального результату використані методи групового обліку аргументів

(покрокове включення, комбінаторний алгоритм, змішаний і нейросетевой), бінарна логістична регресія, RandomForest, Support Vector Machine. Аналіз побудованих моделей проведено шляхом розрахунку показників чутливості, специфічності, а також загальної прогностичним оцінці моделі. Найбільший позитивну прогностичну точність має модель отримана методом логістичної регресії = 95.3%.

Загальний вигляд побудованої моделі:

$$Y = \frac{1}{1 + e^{-(0,361 \times X_1 + 3,856 \times X_2 + 1,49 \times X_4 + 1,709 \times X_5 + 0,388 \times X_6 + -3,187 \times X_3 - 0,056 \times X_7 - 11,955)}}$$

X1– проведення орхектомії

X2– дисеміновані пухлинні клітини

X3– неoad'ювантне лікування

X4- експресія CD44

X5– проведення променевої терапії

X6– вік

X7– біомаркер для виявлення ракових стоволовбурових клітин CD44

Розробка інформаційної системи, яка реалізує створені математичні моделі була здійснена на мові програмування Java і за допомогою фреймворків Spring IoC, Spring Security, Spring Data JPA. Система складається з двох сервісів, один з яких відповідає за збереження даних і взаємодія з користувачем, а інший містить натреновані моделі і виконує обчислення за запитом першого сервера.

Програмне забезпечення дозволяє системі оцінити ризиків у хворих після лікування раку передміхурової залози. Розроблений інтерфейс користувача є досить простим і інтуїтивно зрозумілим. Основні можливості системи:

- Реєстрація користувача;
- Захист інформації шляхом авторизації користувача;
- Можливість зворотного зв'язку з базою даних;

- Накопичення даних про стан пацієнта в різний період часу;
- Аналіз ефективності лікувального процесу;
- Можливість використання системи як в умовах лікувальних установ, так і в центрах математичної статистики.

Заключення. Оцінка ризику настання рецидиву дасть можливість правильно призначати лікування та запобігати розвитку хвороби. За результатами проведеної регресії Кокса були виявлені ряд інформативних показників, що можуть бути використані в діагностичних цілях. Побудована математична класифікаційні моделі визначення ймовірності настання рецидиву та летального випадку у хворих раком передміхурової залози.

Література

1. Siegel R., Naishadham D., Jemal A. (2015) Cancer statistics, 2012. *CA Cancer J. Clin.*, 62(1): 10–29.
2. Sklearn.ensemble.RandomForestClassifier [Електронний ресурс]. – Режим доступу:
<http://scikitlearn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html>.
3. Sklearn.linear_model.LogisticRegression [Електронний ресурс]. – Режим доступу: http://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html.
4. Введение в машинное обучение с помощью Python и Scikit-Learn [Електронний ресурс]. – Режим доступу: <https://habrahabr.ru/company/mlclass/blog/247751>
5. Есипов Б.А., Губанов Е.С., Боряев Е.А. Математическая модель и программа прогнозирования успешности лечения на основе регрессионного анализа. - Известия Самарского научного центра Российской академии наук – 5 с.

6. Метод группового учёта аргументов [Электронный ресурс]. – Режим доступа:http://www.machinelearning.ru/wiki/index.php?title=Метод_группового_учета_аргументов