

(19)



**Евразийское  
патентное  
ведомство**

(21) **202090714** (13) **A2**

**(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ЕВРАЗИЙСКОЙ ЗАЯВКЕ**

(43) Дата публикации заявки  
2020.10.30

(51) Int. Cl. *G16H 50/20* (2018.01)  
*G16H 50/30* (2018.01)

(22) Дата подачи заявки  
2020.04.10

**(54) СПОСОБ СКРИНИНГОВОГО ОПРЕДЕЛЕНИЯ ВЕРОЯТНОСТИ НАЛИЧИЯ РАКА  
МОЧЕВОГО ПУЗЫРЯ**

(31) 2019111091

(32) 2019.04.12

(33) RU

(71) Заявитель:  
**ФЕДЕРАЛЬНОЕ  
ГОСУДАРСТВЕННОЕ  
АВТОНОМНОЕ  
ОБРАЗОВАТЕЛЬНОЕ  
УЧРЕЖДЕНИЕ ВЫСШЕГО  
ОБРАЗОВАНИЯ ПЕРВЫЙ  
МОСКОВСКИЙ  
ГОСУДАРСТВЕННЫЙ  
МЕДИЦИНСКИЙ УНИВЕРСИТЕТ  
ИМЕНИ И.М. СЕЧЕНОВА  
МИНИСТЕРСТВА  
ЗДРАВООХРАНЕНИЯ  
РОССИЙСКОЙ ФЕДЕРАЦИИ  
(СЕЧЕНОВСКИЙ УНИВЕРСИТЕТ)  
(ФГАОУ ВО ПЕРВЫЙ МГМУ ИМ.  
И.М. СЕЧЕНОВА МИНЗДРАВА  
РОССИИ (СЕЧЕНОВСКИЙ  
УНИВЕРСИТЕТ)) (RU)**

(72) Изобретатель:

**Глыбочко Петр Витальевич,  
Свистунов Андрей Алексеевич,  
Фомин Виктор Викторович, Копылов  
Филипп Юрьевич, Секачева  
Марина Игоревна, Еникеев Дмитрий  
Викторович, Гитель Евгений  
Павлович, Рагимов Алигейдар  
Алекперович, Поддубская Елена  
Владимировна (RU)**

(74) Представитель:  
**Куприянова О.И. (RU)**

(57) Изобретение относится к области медицины, а именно онкологии, и может быть использовано для скринингового определения вероятности наличия РМП или выявления данного онкологического заболевания на ранней стадии. Скрининговое определение вероятности наличия РМП основано на измерении уровня биомаркеров в образце биологической жидкости, полученном у субъекта: sVCAM.1, ApoA1, CA19.9, ApoA2, CYFRA.21.1, Ddimer, ApoB, hsCRP, TTR, B2M, и информации о поле пациента, с последующей обработкой совокупности полученных данных с использованием по меньшей мере одной классификационной модели, обученной для определения вероятности наличия РМП.

**A2**

**202090714**

**202090714**

**A2**

## **СПОСОБ СКРИНИНГОВОГО ОПРЕДЕЛЕНИЯ ВЕРОЯТНОСТИ НАЛИЧИЯ РАКА МОЧЕВОГО ПУЗЫРЯ**

### **Область техники, к которой относится изобретение**

Изобретение относится к области медицины, а именно онкологии, и может быть использовано для скринингового определения вероятности наличия рака мочевого пузыря (РМП) или выявления данного онкологического заболевания на ранней стадии.

### **Уровень техники**

Злокачественные опухоли представляют собой одну из самых значимых проблем здравоохранения не только в России, но и во всем мире.

Онкологические заболевания являются второй по частоте причиной смерти в России. Средний показатель заболеваемости злокачественными новообразованиями по России в 2016г. составил 408,6 чел. на 100000 населения. Средний показатель смертности – 201,6 чел. на 100000 населения. Онкологическая заболеваемость растет во всем мире. За последние 10 лет она увеличилась более чем на 20%.

В случае развития злокачественного заболевания, стадия, на которой онкологический процесс будет выявлен, является одним из определяющих факторов, обуславливающих продолжительность жизни пациента.

Ежегодно в мире раком мочевого пузыря заболевают более 300 тыс. человек. В структуре онкологической заболеваемости России РМП занимает значительную долю, при этом характерно позднее выявление РМП, что затрудняет лечение больных и вызывает частое рецидивирование.

Наиболее простым в исполнении и доступным неинвазивным методом выявления РМЖ, является цитологическое исследование мочи. Данный метод отличается высокой специфичностью, но не всегда обладает высокой чувствительностью, в частности, в случае высокой дифференцировки опухоли.

Для обнаружения рака мочевого пузыря известно также применение инструментальных методов диагностики – цистоскопии, урографии, компьютерной томографии, ультразвукового исследования, магнитно-резонансной томографии и других методов сканирования.

Основным методом диагностики опухолей мочевого пузыря на сегодняшний день является цистоскопия, которая производится с помощью эндоскопа, который вводят через мочеиспускательный канал (уретру).

Ультразвуковое исследование мочевого пузыря (трансабдоминальное, трансуретральное и др. сканирование) также используют для диагностики опухолей мочевого пузыря. Точность скринингового ультразвукового исследования возрастает с увеличением размеров опухолей, составляя 82% при опухолях более 5 мм, однако УЗИ сопряжено со сложностью определения локальной протяженности опухоли.

Компьютерная томография (КТ) является важным неинвазивным методом обследования больных раком мочевого пузыря, рекомендована для выявления рецидивов рака (возможных метастазов в печень и надпочечники), при поражении регионарных лимфатических узлов. Для уточнения степени инфильтрации стенки пузыря и перехода её на окружающие ткани иногда применяют рентгенологическое исследование в условиях двойного контрастирования, при котором в пузырь и в окружающую клетчатку вводят кислород. Метод магнитно-резонансного сканирования позволяет уточнить стадию заболевания, оценить состояние стенки мочевого пузыря и окружающих органов, определить прорастание глубокого мышечного слоя мочевого пузыря.

Тем не менее существующие методы лучевой диагностики не позволяют абсолютно точно определить распространение первичной опухоли. Чувствительность компьютерной томографии в определении экстравезикального распространения опухоли варьирует от 60 до 96%, а специфичность – от 66 до 93%.

Традиционное эндоскопическое и лучевое обследование не позволяют обнаружить мелкие опухолевые изменения слизистой оболочки мочевого пузыря на ранних стадиях развития, зачастую мелкие папиллярные опухоли (менее 0,3 см) остаются незамеченными.

В качестве альтернативы инструментальным методам исследования мочевого пузыря могут выступать методы диагностики, основанные на определении биохимических маркеров в биологических тканях и жидкостях пациента, например, цельной крови, сыворотке или плазме. В качестве таких маркеров, например, могут быть использованы различные антигены, протеины и метаболиты, секретируемые злокачественными клетками или образующиеся в процессе их гибели. На данный момент цитологическое исследование мочи является наиболее популярным неинвазивным методом диагностики РМП, однако данный метод не подходит для диагностики высокодифференцированных опухолей (Leiblich, Recent Developments in the Search for Urinary Biomarkers in Bladder Cancer. Curr Urol Rep. 2017; 18(12): 100). В качестве дополнительных скрининговых методов так же может выступать определение антигена рака мочевого пузыря (Urine Bladder Cancer Antigen (UBC)), фрагмента цитокератина 21 (CYFRA 21-1), белка ядерного матрикса NMP22, тканевого полипептидного антигена (tissue polypeptide antigen (TPA)), а так же ряда других

биомаркеров в моче (Sanchez-Carbayo et al., Comparative sensitivity of urinary CYFRA 21-1, urinary bladder cancer antigen, tissue polypeptide antigen, tissue polypeptide antigen and NMP22 to detect bladder cancer. J Urol. 1999 Dec;162(6):1951-6). Например, из патентов RU2470301, RU2463354 известны способы диагностики рака мочевого пузыря, основанные на определении концентрации белков KIFC1 и TFDP1 в образце мочи и/или крови от пациента. Также известно применение молекулярных методов (например, детекция хромосомных перестроек в эпителиальных клетках, определенных в моче) и FDP теста, предполагающего определение отношения фибрина к продуктам его распада (Сивков и др., Молекулярно-генетические маркеры рака мочевого пузыря в клинической практике. Урология. 2014. № 3. С. 48–54). Как уже было отмечено, тесты, основанные на определении единичных биомаркеров, как правило, являются менее точными и не позволяют выявить РМП на ранних стадиях его развития.

В настоящее время в диагностике онкологических заболеваний, в том числе и РМП, получают распространение мультиплексные системы, предполагающие определение риска наличия заболевания на основании данных по нескольким биомаркерам, измеренным у пациента. Такой подход позволяет увеличить чувствительность и специфичность тест-систем. Так, в исследовании (Soukup et al., Panel of Urinary Diagnostic Markers for Non-Invasive Detection of Primary and Recurrent Urothelial Urinary Bladder Carcinoma. Urol Int. 2015;95(1):56-64) было показано, что дискриминационная способность мультимаркерного теста была выше по сравнению с единичными маркерами. Основным недостатком применяемых в клинической практике мультиплексных панелей является недостаточная точность при обнаружении низкодифференцированных форм РМП (Lotan and Roehrborn, Sensitivity and specificity of commonly available bladder tumor markers versus cytology: results of a comprehensive literature review and meta-analyses. Urology. 2003 Jan;61(1):109-18). Стоит отметить, что, в то время как все вышеизложенные методы предполагают определение биомаркеров в моче, диагностический потенциал циркулирующих в крови остается недостаточно изученным (Nandagopal and Sonpavde, Circulating Biomarkers in Bladder Cancer. Bladder Cancer. 2016; 2(4): 369–379).

Из патента US9892229 (прототип) известен способ определения вероятности наличия рака мочевого пузыря у субъекта по определению в биологической жидкости (моче, крови, плазме, сыворотке) набора биомаркеров, выбранных из группы, включающей: CRP; EGF; IL-6; IL-1  $\alpha$ ; MMP9; IL-4; TM; IL-2; TNF $\alpha$ ; sTNFRI ; STNFR2; MMP9/NGAL комплекс; CEA; CK18; IL-1  $\beta$ ; IL-8; dDimer; VEGF; NGAL; BTA; FPSA; vWF; HA; NSE; MCP1; NMP22; TPSA и FAS, а также учитывающий пол субъекта, наличие камней в почках,

доброкачественной гиперплазии предстательной железы, прием антигипертензивных, антиагрегантных и противоязвенных препаратов, статус курения. Однако, даже при включении дополнительных факторов в расчет риска наличия РМП, качество теста, оцениваемое по AUROC, не является оптимальным (Литвин, А. А. Использование данных доказательной медицины в клинической практике (сообщение 3 — диагностические исследования) / А. А. Литвин, А. Л. Калинин, Н. М. Тризна // Проблемы здоровья и экологии. – 2008. - № 4 (18). – С. 12-19).

Заявляемое изобретение основано на исследовании нового комплекса маркеров, позволяющего повысить точность и достоверность определения наличия заболевания при скрининге РМП у конкретного пациента европеоидной популяции, формирование на этой основе той или иной группы риска и выявление тех пациентов, которые нуждаются в углубленном дорогостоящем обследовании для обнаружения ранней стадии РМП.

### **Раскрытие изобретения**

Технической проблемой, решаемой настоящим изобретением, является создание более точного способа определения вероятности наличия РМП в европеоидной популяции.

Достижимым техническим результатом является повышение точности скринингового выявления наличия рака у пациентов европеоидной популяции, причем уже на ранних стадиях его развития, посредством биостатистической обработки результатов анализа фракции сыворотки и плазмы крови с определением концентрации комплексной группы биомаркеров.

Технический результат достигается посредством реализации способа скринингового определения вероятности наличия РМП, включающего измерение уровня биомаркеров в образце биологической жидкости, полученной у субъекта: sVCAM.1, ApoA1, CA19.9, ApoA2, CYFRA.21.1, Ddimer, ApoB, hsCRP, TTR, B2M, а также определение пола пациента, с последующей обработкой совокупности полученных значений биомаркеров с использованием, по меньшей мере, одной классификационной модели, обученной для определения высокой или низкой вероятности наличия РМП.

В качестве классификационных моделей используют метод «случайного леса» (random forest), и/или линейный дискриминантный анализ, и/или метод опорных векторов.

Обученную классификационную модель получают посредством реализации следующих шагов:

- формируют обучающую и тестовую выборку записей субъектов с измеренными значениями биомаркеров sVCAM.1, ApoA1, CA19.9, ApoA2, CYFRA.21.1, Ddimer, ApoB, hsCRP, TTR, B2M, включающие записи о пациентах разного пола и возраста;

- обучают классификационную модель выявлению заданной патологии, используя записи обучающей и тестовой выборки;

- сохраняют связи и веса обученной классификационной модели, для последующего определения вероятности наличия РМП по итогам обработки измеренных данных биомаркеров субъекта.

При формировании обучающей и тестовой выборки, включают записи субъектов с выявленной патологией - наличие РМП и отсутствие РМП.

Технический результат достигается посредством реализации системы скринингового определения вероятности наличия РМП, включающей

- модуль ввода измеренных значений биомаркеров субъекта sVCAM.1, ApoA1, CA19.9, ApoA2, CYFRA.21.1, Ddimer, ApoB, hsCRP, TTR, B2M;

- модуль хранения данных, выполненный с возможностью хранения обучающей и тестовой выборки классификационной модели, связей и весов обученной классификационной модели, записей субъектов с измеренными значениями биомаркеров sVCAM.1, ApoA1, CA19.9, ApoA2, CYFRA.21.1, Ddimer, ApoB, hsCRP, TTR, B2M, включающие записи о пациентах разного пола и возраста;

- модуль обученной классификационной модели, выполненный с возможностью построения и обучения, по меньшей мере, одной классификационной модели для определения наличия заданной патологии по упомянутым маркерам, взятым из модуля хранения данных;

- модуль диагностики, выполненный с возможностью обработки введенных значений биомаркеров субъекта с использованием, по меньшей мере, одной обученной классификационной модели;

- модуль вывода данных, выполненный с возможностью получения данных о высокой или низкой вероятности наличия РМП.

Точность заявляемого мультиплексного способа диагностики РМП обеспечивается за счет использования комплекса из 10 биомаркеров и информации о поле пациента, а также

за счет использования нескольких классификационных моделей с последующим усреднением модельных результатов.

Кроме того, предлагаемый способ адаптирован для популяций обследуемых европеоидной расы, что исключает прогнозные ошибки, связанные с межпопуляционными различиями в молекулярных механизмах канцерогенеза. Так, были выявлены различия в экспрессии генов, характере соматических мутаций и паттернах метилирования ДНК из образцов опухолей, взятых у пациенток европеоидной и черной расы (Huo et al., Comparison of Breast Cancer Molecular Features and Survival by African and European Ancestry in The Cancer Genome Atlas. JAMA Oncol. 2017 Dec 1;3(12):1654-1662).

### **Краткое описание чертежей**

Изобретение поясняется чертежами, где:

На фиг.1 А. представлена диаграмма рассеяния «возраст пациента - концентрация биомаркеров». Точки – индивидуальные измерения, линии – предсказания линейной регрессионной модели. На графиках приведены значения корреляционных коэффициентов, рассчитанных по методу Пирсона и Р-значения, рассчитанные по тесту Стьюдента; на фиг.1 Б. представлена диаграмма размаха для оценки значимости гендерных различий в концентрациях биомаркеров. На графиках приведены Р-значения, полученные при помощи критерия Стьюдента. Серым цветом показаны данные для женщин, черным – для мужчин;

На фиг.2 - ROC-кривые для оценки предсказательной способности отдельных биомаркеров (тип линий соответствует биомаркеру);

На фиг.3. - Примеры деревьев решений, полученных в результате обучения многофакторного классификационного алгоритма random forest на экспериментальных данных по 10 биомаркерам;

На фиг. 4 - Визуализация результатов разделения пациентов на 2 класса (здоровые доноры и пациенты с РМП) при помощи линейного дискриминантного анализа по 10 биомаркерам;

На фиг.5 - Примеры 3-мерных проекций разделения объединенной популяции пациентов на 2 класса (здоровые доноры и пациенты с РМП) при помощи метода опорных векторов по 10 биомаркерам;

На фиг.6 - Доля классификаторов стратифицированная по AUROC в зависимости от количества включенных в них биомаркеров. Обучение проводилось при помощи А. Метода опорных векторов Б. Линейного дискриминантного анализа.

На фиг.7 - ROC-кривые для оценки предсказательной способности различных классификационных алгоритмов. А. Весь набор данных был использован как для обучения модели, так и для ее валидации; Б. 80% данных было использовано для обучения модели, 20% - для валидации.

На фиг.8 - Блок-схема системы, предназначенной для оценки вероятности наличия РМП на основе данных пациента.

На фиг.9 - Алгоритм оценки вероятности наличия РМП на основе данных пациента.

### **Осуществление изобретения**

Исходная группа биомаркеров, используемая в диагностическом тесте на определение вероятности наличия РМП, была получена с использованием многофакторной классификационной модели. Подобные методы позволяют находить комбинации биомаркеров, обладающих наибольшим диагностическим потенциалом. Математическая модель проходит обучение на экспериментальных измерениях заданного набора биомаркеров, полученных на смешанной выборке из здоровых добровольцев и пациентов с РМП. Обученная модель может быть использована для оценки риска наличия заболевания у пациента на основе показателей его биомаркеров.

В рамках проведенной работы на этапе разработки диагностически значимого комплекса показателей были использованы данные измерений 13 биомаркеров (AFP, CEA, CA 19-9, tPSA, B2M, hsCRP, Ddimer, CYFRA 21-1, ApoA1, ApoA2, Apo B, TTR, sVCAM-1), полученные на выборке здоровых добровольцев (n=203, 104 женщины и 99 мужчин 36-80 лет, средний возраст 53 года) и пациентов с РМП (n=59, 12 женщин и 47 мужчин 28-82 лет, средний возраст 61 год).

Статистическая обработка экспериментальных данных и разработка классификационных моделей проводилась в среде R {RDevelopmentCoreTeam (2007). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0}.

Первым этапом являлся статистический анализ и визуализация данных. Для здоровых добровольцев было оценено влияние пола и возраста на показатели биомаркеров (Фиг.1).

По итогам проведения исследования на данном этапе был сделан вывод об отсутствии значимой корреляции между возрастом пациентов и показателями большинства



биомаркеров, в то же время наблюдались значимые гендерные различия в показателях CEA, CA 19-9, Ddimer, ApoA1 и TTR.

На следующем этапе проводилась оценка значимости различий в уровнях отдельных биомаркеров между здоровыми добровольцами и пациентами с РМП при помощи критерия Стьюдента после нормализации экспериментальных данных путем log-трансформации (Таблица 1).

Таблица 1. Сравнение показателей биомаркеров в плазме крови здоровых доноров и пациентов с РМП

| Биомаркер         | Здоровые доноры | Пациенты с РМП | P-значения <sup>1</sup> |
|-------------------|-----------------|----------------|-------------------------|
| AFP, МЕ/мл        | 2.84            | 4.041          | 0.2                     |
| ApoA1, г/л        | 1.60            | 1.22           | 2e-19 ***               |
| ApoA2, г/л        | 0.30            | 0.23           | 2e-12 ***               |
| ApoB, г/л         | 1.03            | 1.46           | 4e-07 ***               |
| B2M, нг/мл        | 1477.29         | 1772.63        | 0.001 **                |
| CA19.9, МЕ/мл     | 6.60            | 35.36          | 7e-10 ***               |
| CEA, нг/мл        | 1.86            | 54.47          | 0.6                     |
| CYFRA.21.1, нг/мл | 1.37            | 5.17           | 2e-08 ***               |
| Ddimer, нг/мл     | 119.75          | 463.93         | 1e-09 ***               |
| hsCRP, мг/л       | 1.77            | 11.39          | 1e-17 ***               |
| sVCAM.1, нг/мл    | 658.37          | 1102.29        | 2e-19 ***               |
| t.PSA, нг/мл      | 1.130           | 2.18           | 0.2                     |
| TTR, мг/дл        | 25.64           | 22.14          | 3e-04 ***               |

<sup>1</sup> – сравнение по тесту Стьюдента после нормализации экспериментальных данных; \* - P-значения<0.05, \*\* - P-значения<0.01, \*\*\* - P-значения<0.001.

На основании проведенного анализа был сделан вывод об отсутствии значимых различий в концентрациях AFP, CEA, t.PSA между здоровыми добровольцами и пациентами с РМП. Отмечены значимые различия в уровнях ApoA1, ApoA2, ApoB, B2M, CA19.9, CYFRA.21.1, sVCAM.1 и TTR, не включенных в прототип.

Для оценки диагностической ценности отдельных биомаркеров использовался метод логистических регрессий. В данных статистических моделях рассматривалась взаимосвязь между концентрацией биомаркера и вероятностью наличия заболевания (уравнение 1):

$$P(Y) = 1 / (1 + e^{-(b_0 + b_1 \cdot X)}) \quad (1)$$

где P(Y) – вероятность наличия заболевания,  $b_0$  и  $b_1$  – коэффициенты, определяемые по экспериментальным данным, X – предиктор (концентрация биомаркера).

Предсказательная способность логистических моделей оценивалась при помощи ROC-анализа, предполагающего определение чувствительности, специфичности и точности метода относительно тестового или общего набора данных. Для этого значение пороговой вероятности, определяющей наличие заболевания, варьировалось в пределах от 0 до 1 с заданным шагом, для каждого шага рассчитывалась доля верно диагностированных случаев заболевания (чувствительность) ( $S_e$ ), правильно определенных случаев отсутствия заболевания (специфичность) ( $S_p$ ), а также общая доля правильно диагностированных случаев, как наличия, так и отсутствия заболевания (точность) ( $Acc$ ), (уравнения 2-4):

$$S_e = TP / (TP + FN) \cdot 100\% \quad (2)$$

$$S_p = TN / (TN + FP) \cdot 100\% \quad (3)$$

$$Acc = (TN + TP) / (TN + FP + TP + FN) \cdot 100\% \quad (4)$$

где TP – верно классифицированный положительный результат (верно диагностированное заболевание), FP – ложноположительный результат (ошибочно диагностированное заболевание), TN – верно классифицированный отрицательный результат (верно диагностированное отсутствие заболевания), FN – ложноотрицательный результат (ошибочно диагностированное отсутствие заболевания).

Полученный набор значений чувствительности и специфичности использовался для построения ROC-кривой. В качестве интегрального показателя качества моделей использовалась площадь под ROC-кривой (AUROC): предикторы с максимальной предиктивной способностью показывают наибольшие значения AUROC. Результаты ROC-анализа приведены на фиг. 2 и в таблице 2.

Таблица 2. Диагностическая ценность отдельных биомаркеров

| Биомаркер         | Специфичность, % | Чувствительность, % | Точность, % | AUROC |
|-------------------|------------------|---------------------|-------------|-------|
| <b>sVCAM.1</b>    | 93               | 78                  | 89          | 0.92  |
| <b>ApoA1</b>      | 87               | 80                  | 85          | 0.9   |
| <b>CA19.9</b>     | 72               | 100                 | 76          | 0.87  |
| <b>ApoA2</b>      | 81               | 75                  | 79          | 0.85  |
| <b>CYFRA.21.1</b> | 60               | 90                  | 67          | 0.83  |
| <b>Ddimer</b>     | 67               | 81                  | 70          | 0.81  |
| <b>ApoB</b>       | 88               | 69                  | 84          | 0.81  |
| <b>hsCRP</b>      | 51               | 100                 | 62          | 0.8   |
| <b>TTR</b>        | 91               | 36                  | 78          | 0.66  |
| <b>t.PSA</b>      | 63               | 66                  | 64          | 0.65  |
| <b>B2M</b>        | 94               | 31                  | 80          | 0.63  |
| <b>AFP</b>        | 93               | 40                  | 88          | 0.57  |
| <b>CEA</b>        | 58               | 58                  | 58          | 0.53  |

На основе результатов статистического анализа данных и оценки предсказательной способности однофакторных логистических моделей были отобраны биомаркеры, которые впоследствии были включены в многофакторные классификационные модели. Критерием включения биомаркеров являлись  $p\text{-value} < 0.005$  (Таблица 1) и  $\text{AUROC} \geq 0.6$  (Таблица 2). Таким образом, для построения классификационных моделей были отобраны экспериментальные измерения 10 биомаркеров (**sVCAM.1, ApoA1, CA19.9, ApoA2, CYFRA.21.1, Ddimer, ApoB, hsCRP, TTR, B2M**), также использовалась информация о поле пациента.

Разработка многофакторных классификационных моделей являлась завершающим этапом исследования. Различные способы машинного обучения (random forest, линейный дискриминантный анализ, метод опорных векторов) были использованы в рамках текущей задачи. Оценка параметров моделей (обучение), производилась на объединённых данных, полученных на здоровых добровольцах и пациентах с РМП, и была направлена на минимизацию предсказательных ошибок алгоритма. Детальное описание использованных методов изложено в книге (Bishop CM, Pattern recognition and machine learning. Springer. 2006).

Метод «random forest» (RF) подразумевает создание совокупности кросс-валидированных решающих деревьев. Каждое из таких деревьев проходит обучение на подвыборке данных, включающей информацию лишь по части биомаркеров и наблюдений, и валидируется на подвыборке, не использованной для его построения (бэггинг). На основании предсказаний каждого из построенных деревьев решений пациент причисляется к одной из групп (здоровые доноры или пациенты с РМП), финальное предсказание классификатора определяется большинством голосов построенных деревьев (см. фиг.3 А, Б).

Использование линейного дискриминантного анализа (LDA) предполагает поиск линейной комбинации биомаркеров - дискриминанты, обеспечивающей наилучшее разделение всей популяции обследуемых на здоровых добровольцев и пациентов с РМП. Линейная дискриминанта может быть рассчитана:  $z(x) = \beta_1 x_1 + \dots + \beta_n x_n$ , где  $x_i$  — это концентрации  $i$ -го биомаркера,  $\beta_i$  — коэффициенты модели. Данная задача решается за счет нахождения оси, проекция на которую обеспечивает максимальное отношение общей дисперсии линейной комбинации биомаркеров выборки к сумме дисперсий линейной комбинации биомаркеров внутри классов (см. фиг.4).

Таблица 3. Значения линейных коэффициентов при дискриминанте (LDAcomponent 1)

| Фактор | Коэффициент |
|--------|-------------|
|--------|-------------|

|            |           |
|------------|-----------|
| Пол (М)    | 1.32E-02  |
| СА19.9     | 2.38E-03  |
| B2M        | -9.24E-04 |
| hsCRP      | 2.38E-04  |
| Ddimer     | -4.79E-04 |
| CYFRA.21.1 | 3.15E-01  |
| ApoA1      | -1.51E+00 |
| ApoA2      | -2.46E+00 |
| ApoB       | 1.31E+00  |
| TTR        | -7.69E-03 |
| sVCAM.1    | 4.52E-03  |

Использование метода опорных векторов (SVM) предполагает нахождение  $(n-1)$ -мерной гиперплоскости, разделяющей  $n$ -мерное пространство значений биомаркеров на два класса. Пусть имеется обучающая выборка  $(x_1, y_1), \dots, (x_n, y_n)$ ,  $x_i \in R^n, y_i \in \{-1, 1\}$ , где  $x_i$  — это вектор значений биомаркеров, а  $y_i$  определяет принадлежность пациента к классу. Классифицирующая функция может быть определена как  $F(x) = \text{sign}(\langle w, x \rangle + b)$ , где  $w$  — нормальный вектор к разделяющей гиперплоскости,  $b$  — вспомогательный параметр, а функция может принимать значения 1 или -1 в зависимости от класса объекта. Обучение алгоритма подразумевает поиск такой гиперплоскости, которая обеспечивает наименьшую эмпирическую ошибку классификации и максимизирует расстояние между значениями биомаркеров пациентов, относящихся к разным классам (см. фиг.5):

На первом этапе построения многофакторных моделей проводилось изучение диагностической ценности различных комбинаций биомаркеров из приведенной выше группы. Для этого все возможные комбинации, включающие от 2 до 10 биомаркеров, были использованы для построения классификационных моделей (2026 вариантов). Для обучения использовались объединённые данные, полученные на здоровых добровольцах и пациентах с РМП, и методы линейного дискриминантного анализа и опорных векторов. Разработанные модели были ранжированы в соответствии с их предсказательным потенциалом, оцененным по показателю AUROC (фиг. 7). Как видно из фиг. 6, даже классификаторы, состоящие из 2-3 биомаркеров, обладают высокой дискриминационной способностью.

Финальной фазой построения классификаторов являлась их валидация.

Объединённые данные, полученные на здоровых добровольцах и пациентах с РМП были случайным образом разделены на обучающую и тестовую выборки. Оценка параметров моделей (обучение), производилась на обучающей выборке и была направлена на минимизацию предсказательных ошибок алгоритма. Валидация обученных моделей

заклучалась в оценке их предсказательной способности на тестовой выборке. Предсказательная способность многофакторных классификационных моделей оценивалась при помощи ROC-анализа как это было сделано ранее для отдельных биомаркеров (фиг. 7, Таблица 4).

Таблица 4. Диагностическая ценность обученных многофакторных классификационных моделей, включающих 10 биомаркеров

| Метод                                                                             | Чувствительность,<br>% | Специфичность,<br>% | Точность,<br>% | Пороговая<br>вероятность,<br>% | AUROC |
|-----------------------------------------------------------------------------------|------------------------|---------------------|----------------|--------------------------------|-------|
| Весь набор данных был использован как для обучения модели, так и для ее валидации |                        |                     |                |                                |       |
| RF                                                                                | 100                    | 100                 | 100            | 50                             | 1     |
| LDA                                                                               | 76                     | 100                 | 96             | 50                             | 0.98  |
| SVM                                                                               | 100                    | 100                 | 100            | 50                             | 1     |
| 80% данных было использовано для обучения модели, 20% - для валидации             |                        |                     |                |                                |       |
| RF                                                                                | 100                    | 100                 | 100            | 50                             | 1     |
| LDA                                                                               | 75                     | 100                 | 96             | 50                             | 0.96  |
| SVM                                                                               | 75                     | 100                 | 96             | 50                             | 0.96  |

Диагностический потенциал наиболее перспективных комбинаций из 2-3 биомаркеров представлен в таблице 5.

Таблица 5. Диагностическая ценность обученных многофакторных классификационных моделей, включающих 2-3 биомаркера, обладающие наибольшим диагностическим потенциалом (AUROC>0.95)

| Комбинация                | AUROC (SVM) | AUROC (LDA) |
|---------------------------|-------------|-------------|
| Пол + ApoA1, sVCAM.1      | 0.984       | 0.958       |
| Пол + CYFRA.21.1, sVCAM.1 | 0.964       | 0.960       |
| Пол + ApoA2, sVCAM.1      | 0.961       | 0.950       |
| Пол + CA19.9, sVCAM.1     | 0.955       | 0.946       |

Как видно из Таблицы 4 комбинации из модели, включающие все 10 биомаркеров, обладают наибольшим диагностическим потенциалом, при этом даже для комбинаций, включающих 2-3 биомаркера, отмечен более высокий показатель AUROC по сравнению с единичными биомаркерами (Таблица 5). Замечено, что sVCAM.1 является компонентом наиболее перспективных моделей. На основании рассчитанных значений AUROC можно отметить более высокую предсказательную способность предложенных многофакторных

классификаторных моделей по сравнению с рассматриваемым прототипом, где значения AUROC не достигали 0.9.

Финальные классификационные модели представляют собой обученные алгоритмы, позволяющие предсказать вероятность наличия РМП на основании экспериментальных измерений биомаркеров пациентов с учетом гендерных различий.

Финальное решение - определение вероятности наличия РМП, рассчитывается как медиана значений вероятностей РМП, рассчитанных в 3 классификационных моделях (RF, LDA SVM), обученных на всей выборке пациентов (см., например, Kittler J, Hatef M, Duin RPW et al, On Combining Classifiers. IEEE Transactions on Pattern Analysis and Machine Intelligence, VOL. 20, NO. 3, MARCH 1998 226-39.)

Для реализации заявляемого способа было разработано программное обеспечение (ПО), позволяющее на основе данных конкретного пациента (пол и результаты измерения биомаркеров) рассчитывать вероятность наличия у него РМП. Блок-схема реализации изобретения представлена на фиг. 8.

Компьютерно-реализуемая система состоит из (1) интерфейса, включающего устройство ввода данных пациента (пол и результаты измерений биомаркеров) и вывода результатов расчета (вероятность наличия РМП); (2) блока памяти, содержащего обученные классификаторные модели и программные продукты, необходимые для работы с ними (R portable, Google Chrome Portable) и (3) программного модуля, с помощью которого реализуется программный код, необходимый для обмена данными между интерфейсом и блоком памяти. Для создания графического интерфейса был использован пакет shiny (Winston Chang, Joe Cheng, JJ Allaire, Yihui Xie and Jonathan McPherson (2017). shiny: Web Application Framework for R. R package version 1.0.5. <https://CRAN.R-project.org/package=shiny>), созданный на базе среды R {RDevelopmentCoreTeam (2007). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0}. Для работы с данным пакетом необходимо наличие программных продуктов R portable и Google Chrome portable, хранящихся в блоке памяти. Для работы с предложенными моделями необходимы следующие пакеты: (1) RandomForest (A. Liaw and M. Wiener (2002). Classification and Regression by randomForest. R News 2(3), 18-22); (2) MASS (Venables, W. N. & Ripley, B. D. (2002) Modern Applied Statistics with S. Fourth Edition. Springer, New York. ISBN 0-387-95457-0); (3) e1071 (David Meyer, Evgenia Dimitriadou, Kurt Hornik, Andreas Weingessel and Friedrich Leisch (2017). e1071: Misc

Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien. R package version 1.6-8. <https://CRAN.R-project.org/package=e1071>).

Алгоритм оценки вероятности наличия РМП на основе данных пациента представлен на фиг.9.

Данные пациента вводятся через интерфейс и подаются в качестве входных переменных в разработанные модели, в каждой из которых производится расчет вероятности наличия РМП. Далее по результатам модельных предсказаний рассчитывается среднее значение, которое выводится в окно вывода.

Диагностическая мультиплексная панель для оценки риска РМП включает биомаркеры, показавшие максимальный предсказательный потенциал в рамках проводимого исследования (рис. 2, таблица 2): sVCAM.1, ApoA1, CA19.9, ApoA2, CYFRA.21.1, Ddimer, ApoB. Кроме того, в заявляемый комплекс включены дополнительные биомаркеры, обладающие меньшим предсказательным потенциалом, однако значимо различные между здоровыми добровольцами и пациентами с РМП (Таблица. 1): hsCRP, TTR и B2M в исследуемой популяции.

Ниже представлены клинические примеры применения способа.

**Пример 1.** Больной Г., 82 года.

Пациенту было предложено принять участие в программе Онкопоиска.

Пациент (823) обследован в рамках программы. Получены следующие результаты:

CA 19-9 8,17 МЕ/мл, tPSA 6 нг/мл, B2M 2878 нг/мл, hsCRP 35 нг/мл, D-dimer 889,0 нг/мл, CYFRA 21-1 5,77 нг/мл, Apo A1 1,38 г/л, Apo A2 0,151 г/л, Apo B 1,89 г/л, TTR (prealb) 23,0 мг/дл, sVCAM-1 2293 нг/мл, Rantes 49616 пг/мл, VEGFR1 98 пг/мл, LRG-1 165740 нг/мл, Apo A4 40.2 мкг/мл.

При обработке полученных результатов заявляемым способом выявлена вероятность РМП, значительно превышающая пороговое значение, составляющее 50% (Таблица 4): модель RF – 94.6%, модель LDA – 100%, модель SVM – 98.2%, усредненное итоговое значение вероятности наличия РМП по трем моделям составило 97.6%.

Пациент приглашен на обследование.

Жалоб активно не предъявляет.

Анамнез: ранее наблюдался у уролога по поводу гиперплазии простаты.

При физикальном осмотре мочевого пузыря не пальпируется. Наружные половые органы – без особенностей. При ректальном исследовании простата незначительно увеличена в размерах, эластической консистенции, безболезненная, с четкими контурами.

Выполнено УЗИ органов мочеполовой системы.

Правая почка с четкими ровными контурами, размерами 10,4х5,0 см, паренхима однородная, толщиной 20 мм. Чашечно-лоханочная система не расширена.

Левая почка с четкими ровными контурами, размерами 11,0х5,5см. Паренхима однородная, толщиной до 2,0 см. В проекции среднего сегмента лоцируется гиперэхогенная структура 0,8 см без акустической дорожки (ангиомиолипома). Чашечно-лоханочная система не расширена.

Простата с четкими ровными контурами, неоднородной эхоструктуры. В центральной зоне с обеих сторон лоцируются гиперэхогенные включения без акустической дорожки. Предстательная железа размерами 2,6х4,2х3,8см, объемом до 21,7 см куб. Семенные пузырьки не расширены, структурно однородные. Кровоток в простате распределен равномерно и симметрично, очагов патологической васкуляризации не выявлено.

МРТ органов малого таза с контрастированием.

У задней стенки мочевого пузыря от дна до устья уретры определяется экзофитнорастущее образование размером 40х30х40 мм с четкими, волнистыми контурами. В зоне образования расположены устья мочеточников. Признаков урообструкции на момент исследования не обнаружено. Мочевой пузырь наполнен умеренно. Задняя стенка пузыря выпрямлена, наружный контур четкий, внутренний не прослеживается на фоне образования, дифференцировка слоев нарушена.

Простата размерами 30х50х35 мм. Зональность сохранена, без особенностей.

При цистоскопии в области правой заднебоковой стенки определяется опухолевое образование, размерами 40х45мм, расположенное на ножке.

Выполнена операция: лазерная en-blok резекция стенки мочевого пузыря.

Гистологическое заключение: морфология папиллярного уротелиального низкодифференцированного рака, прорастающего до мышечной оболочки.

Диагноз: Рак мочевого пузыря T1N0M0.

**Пример 2.** Больной Н., 71 год.

Пациенту было предложено принять участие в программе Онкопоиска.

Пациент (802) обследован в рамках программы. Получены следующие результаты:

СА 19-9 7,94 МЕ/мл, tPSA 6.27 нг/мл, B2M 1557 нг/мл, hsCRP 127 нг/мл, D-dimer 524,0 нг/мл, CYFRA 21-1 0,77 нг/мл, Apo A1 0,8 г/л, Apo A2 0,15 г/л, Apo B 0,66 г/л, TTR (prealb) 5,0 мг/дл, sVCAM-1 1867 нг/мл, Rantes 23259 пг/мл, VEGFR1 192 пг/мл, LRG-1 155030 нг/мл, Apo A4 38 мкг/мл.



При обработке полученных результатов заявляемым способом выявлена вероятность РМП, значительно превышающая пороговое значение, составляющее 50% (Таблица 4): модель RF – 88.2%, модель LDA – 100%, модель SVM – 98.2%, усредненное итоговое значение вероятности наличия РМП по трем моделям составило 95.5%.

Пациент приглашен на обследование.

Осмотрен урологом. По данным физикального осмотра патологии мочеполовой системы не выявлено.

Выполнено УЗИ органов мочеполовой системы.

Правая почка с четкими ровными контурами, размерами 10,0х5,0 см, паренхима однородная, толщиной 1,8 см. Дилатации чашечно-лоханочной системы нет. Левая почка с четкими ровными контурами, размерами 10,0х5,0 см. Паренхима однородная, толщиной до 1,8 см. Дилатации чашечно-лоханочной системы нет. Мочевой пузырь с четкими контурами, по передней стенке визуализируется экзофитное образование 3,6х2,4х3,3см. Остаточной мочи нет. Простата с четкими ровными контурами, неоднородной эхоструктуры. Предстательная железа объемом до 56 см куб. Паренхима однородная.

МРТ органов малого таза с контрастированием.

У внутренней поверхности передней стенки мочевого пузыря определяется объемное образование неправильной формы, неоднородной структуры (с участками кровоизлияний) с неровными четкими контурами, размерами 38х30х35 мм с четкими, неровными контурами. Образование имеет широкое основание, около 30х25мм, распространяется экзо- и эндофитно, с вовлечением глубоких мышечных слоев. Признаков урообструкции на момент исследования не обнаружено. Мочевой пузырь наполнен умеренно.

При цистоскопии на передней стенке мочевого пузыря визуализируется солидное экзофитное образование до 40х30 мм, с участками кровоизлияний, переходящее справа в образование по типу тутовой ягоды. Определяется деформация левой боковой стенки мочевого пузыря с отёчной, петехиально измененной слизистой на участке до 20 мм. На переднеправой стенке мочевого пузыря определяется несколько стелющихся опухолевых образований, размерами до 10мм.

Выполнена операция: трансуретральная резекция экзофитных частей новообразований. Внутрипузырно на 120 минут введен митомицин 40мг.

Гистологическое заключение: умеренно- и низкодифференцированная папиллярная уротелиальная карцинома, прорастающая поверхностные мышцы.

Диагноз: Рак мочевого пузыря T2N0M0.

## ФОРМУЛА ИЗОБРЕТЕНИЯ

1. Способ скринингового определения вероятности наличия рака мочевого пузыря, включающий измерение уровня биомаркеров в образце биологической жидкости, полученном у субъекта: sVCAM.1, ApoA1, CA19.9, ApoA2, CYFRA.21.1, Ddimer, ApoB, hsCRP, TTR, B2M, а также определение пола пациента, с последующей обработкой совокупности полученных значений биомаркеров с использованием, по меньшей мере, одной классификационной модели, обученной для определения высокой или низкой вероятности наличия рака мочевого пузыря.

2. Способ по п.1, характеризующийся тем, что в качестве классификационных моделей используют метод «случайного леса» (random forest), и/или линейный дискриминантный анализ, и/или метод опорных векторов.

3. Способ по п.1, характеризующийся тем, что обученную классификационную модель получают посредством реализации следующих шагов:

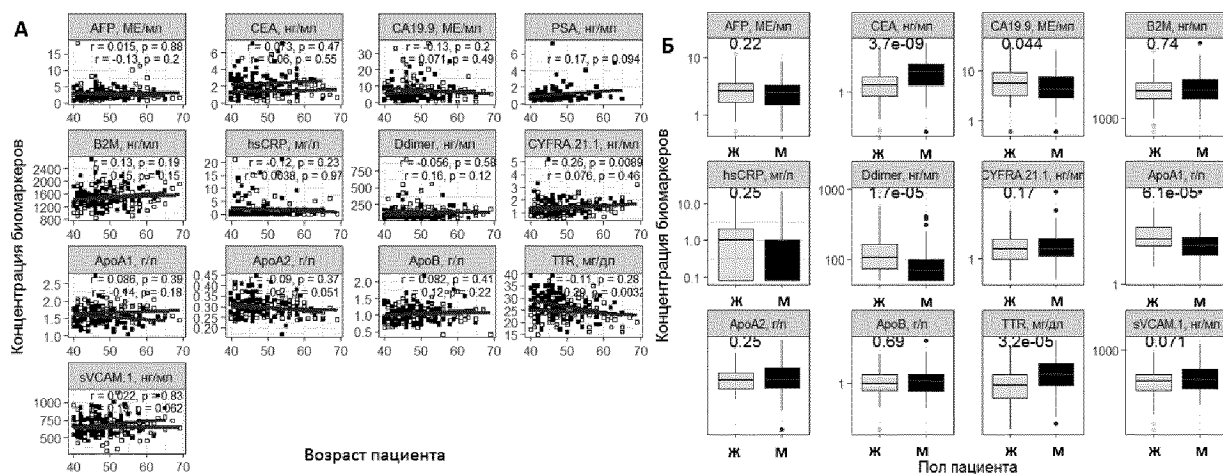
- формируют обучающую и тестовую выборку записей субъектов с измеренными значениями биомаркеров sVCAM.1, ApoA1, CA19.9, ApoA2, CYFRA.21.1, Ddimer, ApoB, hsCRP, TTR, B2M, включающие записи о пациентах разного пола и возраста;

- обучают классификационную модель выявлению заданной патологии, используя записи обучающей и тестовой выборки;

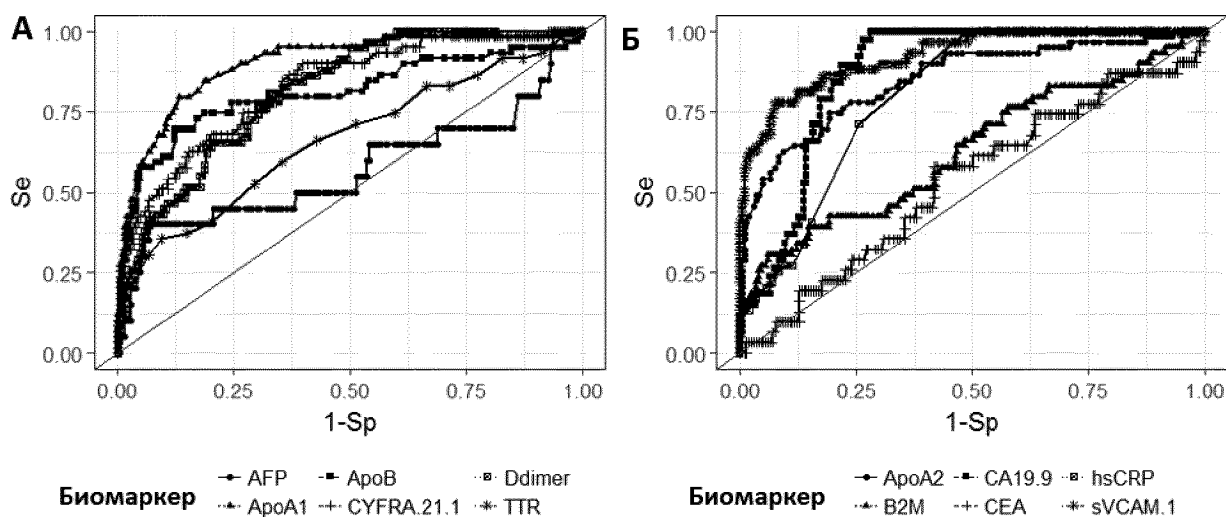
- сохраняют связи и веса обученной классификационной модели, для последующего определения вероятности наличия рака мочевого пузыря по итогам обработки измеренных данных биомаркеров субъекта.

4. Способ по п. 3, характеризующийся тем, что при формировании обучающей и тестовой выборки, включают записи субъектов с выявленной патологией - наличие рака и отсутствие рака мочевого пузыря.

# СПОСОБ СКРИНИНГОВОГО ОПРЕДЕЛЕНИЯ ВЕРОЯТНОСТИ НАЛИЧИЯ РАКА МОЧЕВОГО ПУЗЫРЯ

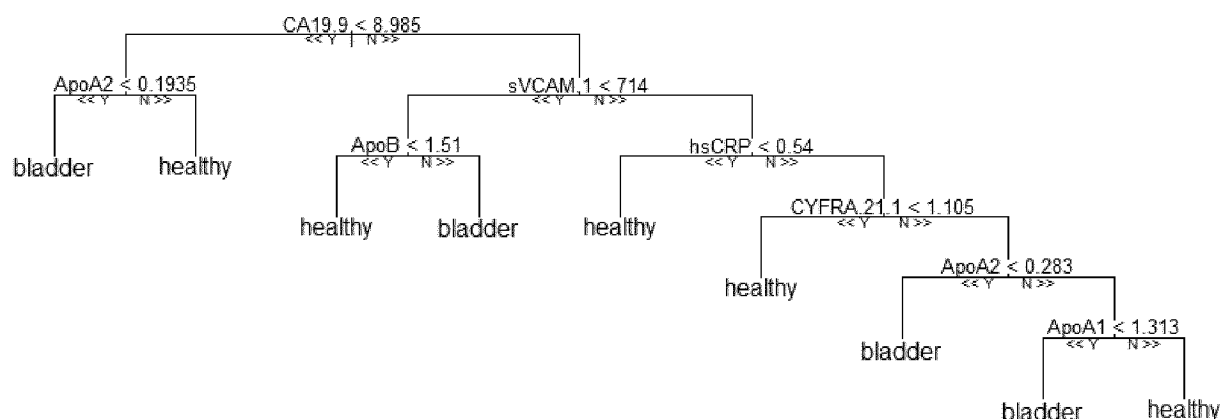


ФИГ.1

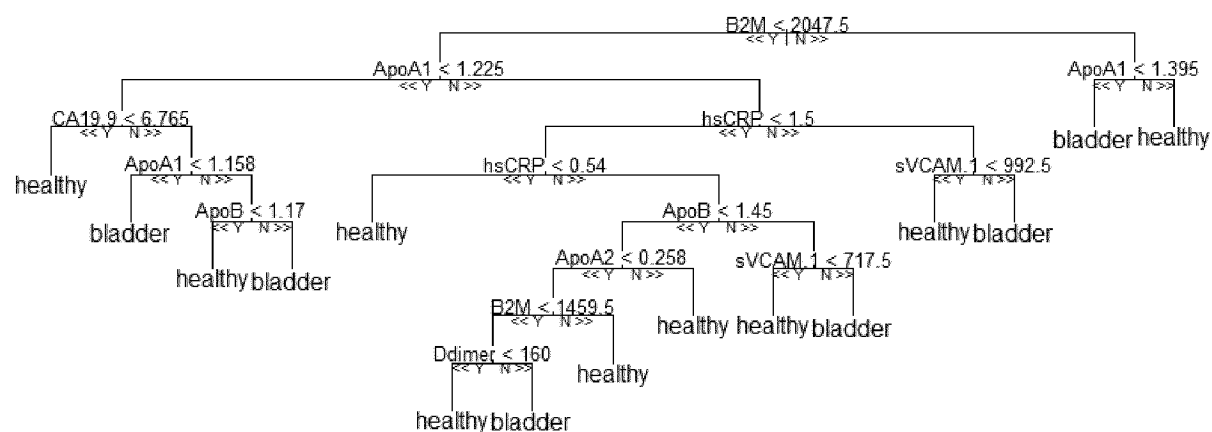


ФИГ.2

# СПОСОБ СКРИНИНГОВОГО ОПРЕДЕЛЕНИЯ ВЕРОЯТНОСТИ НАЛИЧИЯ РАКА МОЧЕВОГО ПУЗЫРЯ

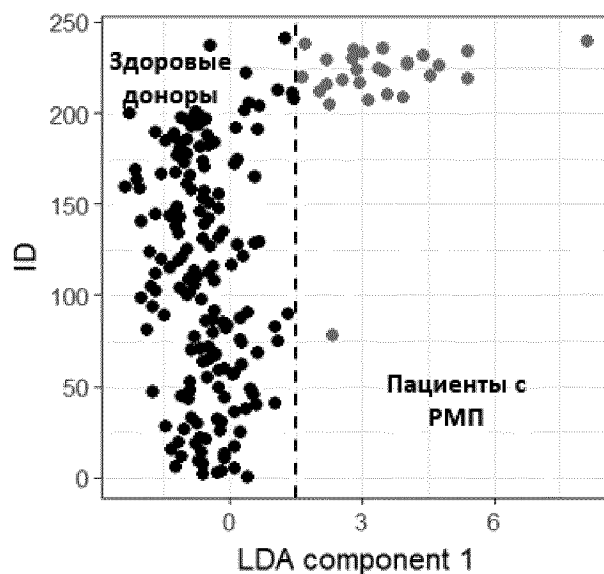


ФИГ.3 А

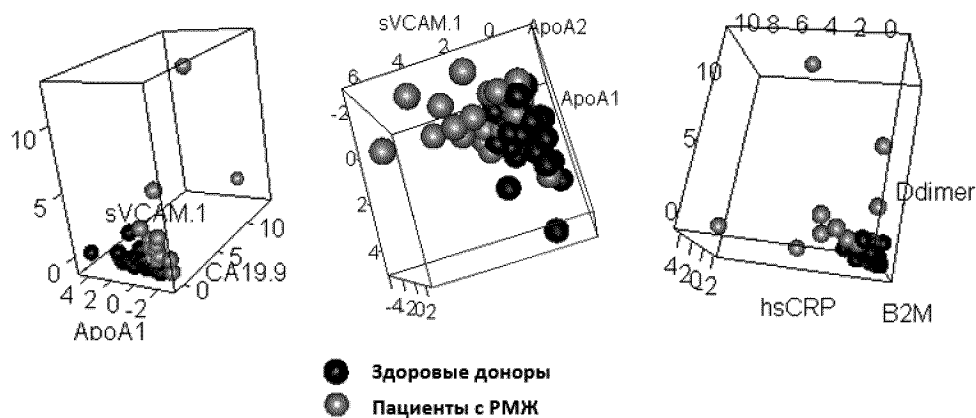


ФИГ.3 Б

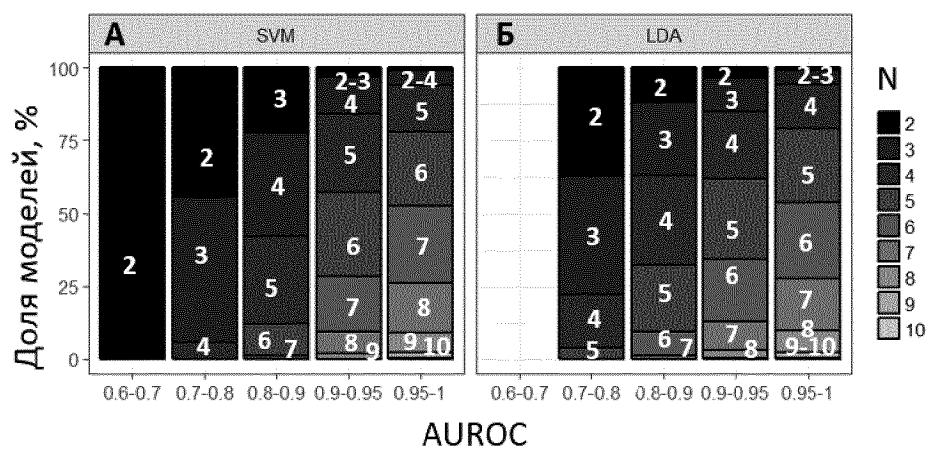
# СПОСОБ СКРИНИНГОВОГО ОПРЕДЕЛЕНИЯ ВЕРОЯТНОСТИ НАЛИЧИЯ РАКА МОЧЕВОГО ПУЗЫРЯ



ФИГ.4

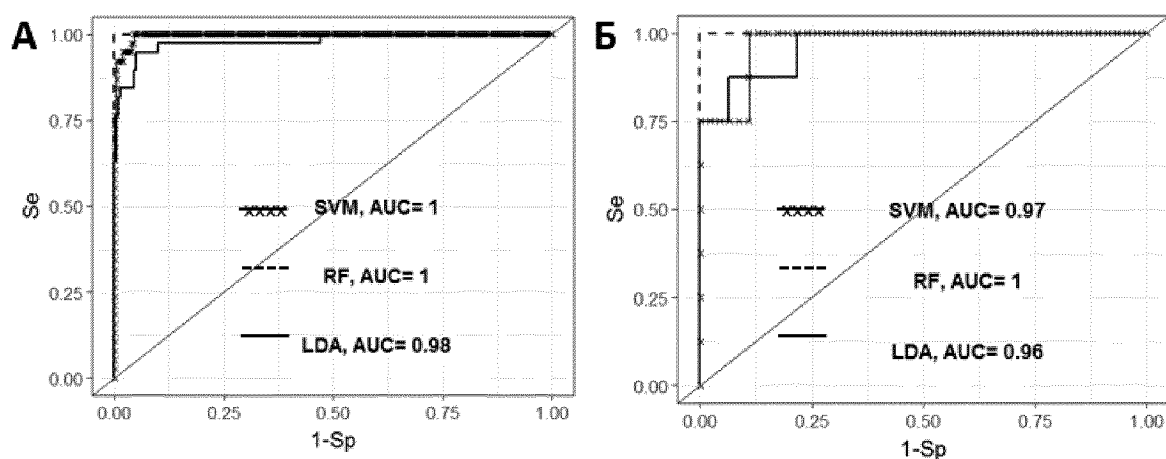


ФИГ.5

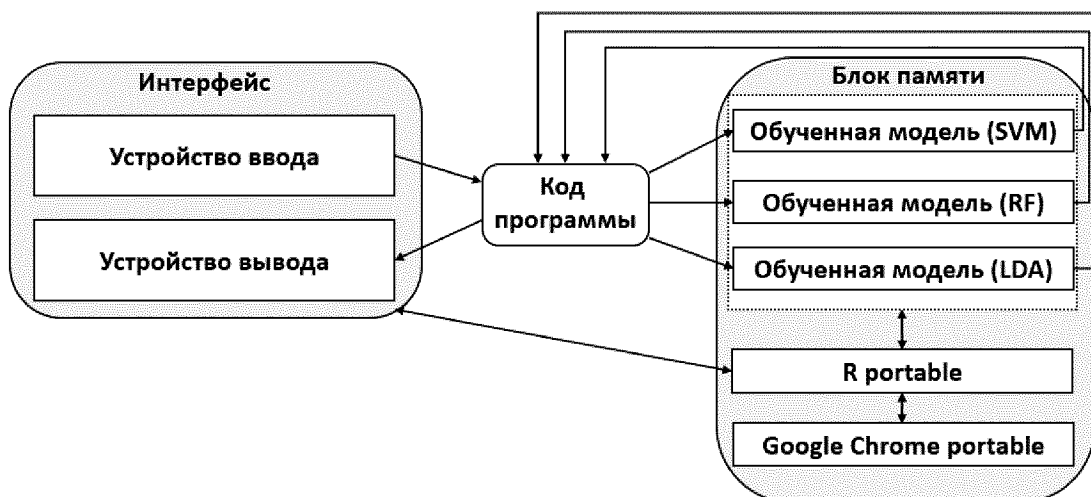


ФИГ.6

# СПОСОБ СКРИНИНГОВОГО ОПРЕДЕЛЕНИЯ ВЕРОЯТНОСТИ НАЛИЧИЯ РАКА МОЧЕВОГО ПУЗЫРЯ

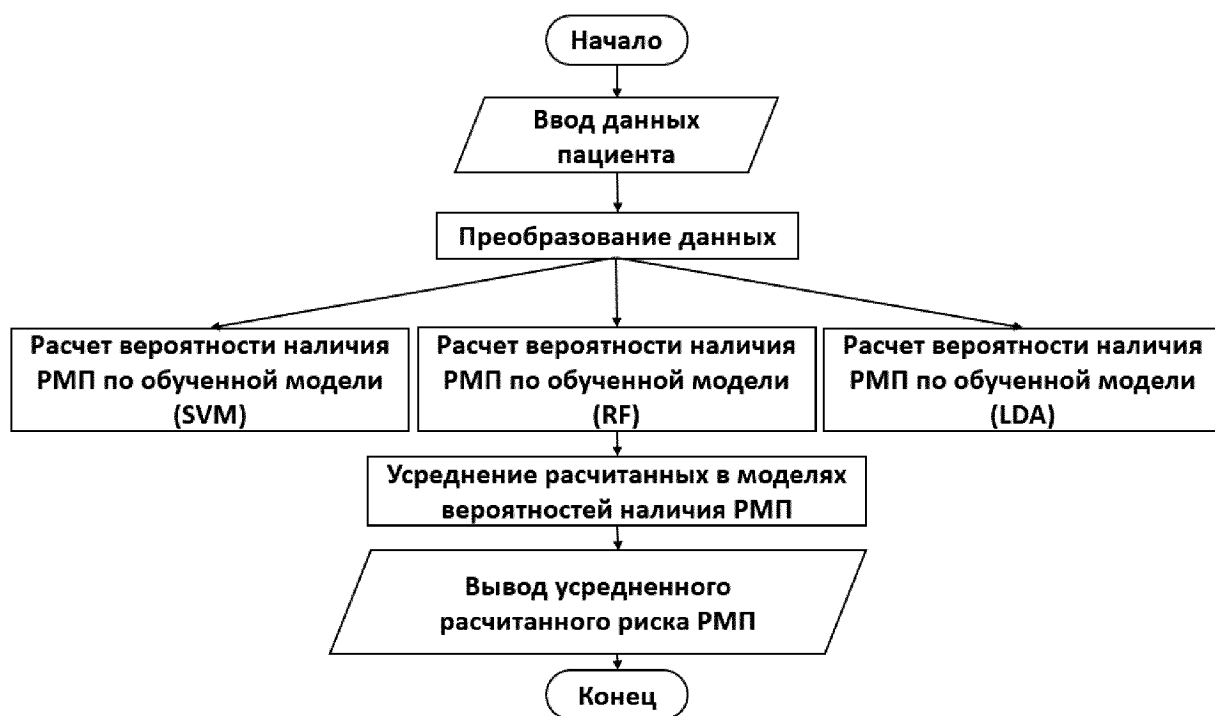


ФИГ.7



ФИГ.8

# СПОСОБ СКРИНИНГОВОГО ОПРЕДЕЛЕНИЯ ВЕРОЯТНОСТИ НАЛИЧИЯ РАКА МОЧЕВОГО ПУЗЫРЯ



ФИГ.9