

(12) МЕЖДУНАРОДНАЯ ЗАЯВКА, ОПУБЛИКОВАННАЯ В  
СООТВЕТСТВИИ С ДОГОВОРОМ О ПАТЕНТНОЙ КООПЕРАЦИИ (РСТ)

(19) Всемирная Организация  
Интеллектуальной Собственности  
Международное бюро

(43) Дата международной публикации  
06 июля 2023 (06.07.2023)

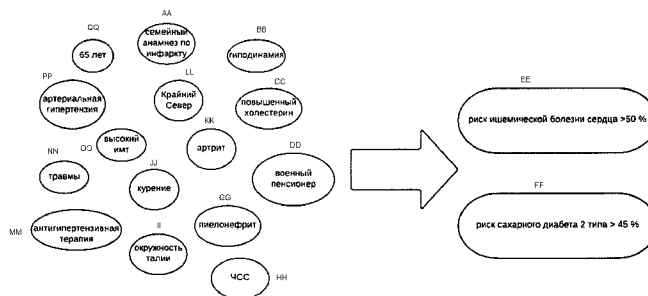


(10) Номер международной публикации  
**WO 2023/128780 A1**

- (51) Международная патентная классификация:  
*G16H 50/70* (2018.01) *G16H 50/30* (2018.01)
- (21) Номер международной заявки: РСТ/RU2021/000605
- (22) Дата международной подачи:  
28 декабря 2021 (28.12.2021)
- (25) Язык подачи: Русский
- (26) Язык публикации: Русский
- (30) Данные о приоритете:  
2021139105 27 декабря 2021 (27.12.2021) RU
- (71) Заявитель: **ОБЩЕСТВО С ОГРАНИЧЕННОЙ  
ОТВЕТСТВЕННОСТЬЮ "К-СКАЙ" (LIMITED  
LIABILITY COMPANY "K-SKAI")** [RU/RU]; на-  
бережная Варкауса, дом 17, помещение 62, офис  
20, г. Петрозаводск, Карелия Республика, 185031,  
Petrozavodsk (RU).
- (72) Изобретатели: **НОВИЦКИЙ, Роман Эдвардович  
(NOVITSKIY, Roman Edvardovich)**; ул. Коммуналь-  
ная, д. 1, кв. 74, г. Петрозаводск, Карелия Респу-  
блика, 185035, Petrozavodsk (RU). **ГУСЕВ, Александр Влади-  
мирович (GUSEV, Aleksandr Vladimirovich)**; ул. Лет-  
няя, д. 30, г. Кондопога, р-н Кондопожский, Карелия  
Республика, 186220, Republic of Karelia, g. Kondopoga  
(RU).
- (74) Агент: **КОТЛОВ, Дмитрий Владимирович  
(KOTLOV, Dmitry Vladimirovich)**; ООО "ЦИС "Скол-  
ково", Территория инновационного центра "Сколково",  
дом 4, оф. 402.1, Москва, 143026, Moscow (RU).

(54) Title: METHOD FOR THE EARLY DIAGNOSIS OF CHRONIC DISEASES IN A PATIENT

(54) Название изобретения: СПОСОБ РАННЕЙ ДИАГНОСТИКИ ХРОНИЧЕСКИХ ЗАБОЛЕВАНИЙ ПАЦИЕНТА



Фиг. 1

- AA Family history of heart attack
- BB Hypodynamia
- CC Elevated cholesterol
- DD Military pensioner
- EE Risk of ischemic heart disease >50%
- FF Risk of type 2 diabetes >45%
- GG Pyelonephritis
- HH Heart rate
- IJ Waist circumference
- JJ Smoker
- KK Arthritis
- LL Extreme North
- MM Antihypertensive therapy
- NN Injuries
- OO High BMI
- PP Arterial hypertension
- QQ Age 65

(57) Abstract: A method for the early diagnosis of chronic diseases in a patient, implemented by a computer containing a processor and a memory, includes the following steps of: receiving, on a computer, anonymized electronic medical record data from a medical information system; extracting medical and social features of a patient's state of health, as well as risk factor features, using natural language processing, wherein the extracted features are sent to the input of a trained suite of classifiers for assigning the features to a probability class for the presence of chronic diseases; dividing the extracted medical and social features about the patient's state of health into groups, and dividing the resulting groups into clusters according to diseases diagnosed, wherein groups are combined into clusters with common diagnoses; obtaining values for the medical and social features of the patient's state of health and for the risk factor



WO 2023/128780 A1

- (81) **Указанные государства** (если не указано иначе, для каждого вида национальной охраны): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.
- (84) **Указанные государства** (если не указано иначе, для каждого вида региональной охраны): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), евразийский (AM, AZ, BY, KG, KZ, RU, TJ, TM), европейский патент (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

**Опубликована:**

— с отчётом о международном поиске (статья 21.3)

features in each cluster; evaluating the risk of the development of complications of chronic diseases; obtaining, at the output of the suite of classifiers, the probability class for the presence of chronic diseases and an evaluation of the risk of the development of complications.

(57) **Реферат:** Способ ранней диагностики хронических заболеваний пациента, выполняющийся на вычислительном устройстве, содержащем процессор и память, включает следующие этапы: получают деперсонифицированные медицинские данные электронной медицинской карты от медицинской информационной системы на вычислительное устройство. Осуществляют извлечение медицинских и социальных признаков о состоянии здоровья пациента, а также признаков факторов риска посредством Natural language processing, извлеченные признаки поступают на вход обученного ансамбля классификаторов для отнесения признаков к классу вероятности наличия хронических заболеваний. Осуществляют деление извлеченных медицинских и социальных признаков о состоянии здоровья пациента на группы, деление полученных групп на кластеры по диагностированным у них заболеваниям, причем объединяют группы в кластеры с общими диагнозами. Получают значения медицинских и социальных признаков о состоянии здоровья пациента и признаков факторов риска в каждом кластере. Осуществляют оценку риска развития осложнений хронических заболеваний. На выходе ансамбля классификаторов получают класс вероятности наличия хронических заболеваний и оценку риска развития осложнений.

## СПОСОБ РАННЕЙ ДИАГНОСТИКИ ХРОНИЧЕСКИХ ЗАБОЛЕВАНИЙ ПАЦИЕНТА

### 5 ОБЛАСТЬ ТЕХНИКИ

Изобретение относится к области медицины, а также к области информационных и коммуникационных технологий для обработки медицинских данных, в частности к способу ранней диагностики хронических заболеваний пациента, основанный на кластерном анализе больших данных.

10 Представленное решение может быть использовано, по меньшей мере, в клинической практике врачами и другими медицинскими специалистами, которые занимаются диагностикой, лечением и профилактикой заболеваний, при прогнозе наступления различных медицинских событий для пациента.

### 15 УРОВЕНЬ ТЕХНИКИ

Из уровня техники известен источник информации RU 2 698 007 С1, опубликованный 21.08.2019 года и раскрывающий автоматическую систему поддержки медицинских решений при сочетанной патологии. Система содержит блок интерфейса, блок входных данных, выполненный в формате электронной истории болезней, блок хранения решений для каждой из терапевтических областей, вычислительный блок, блок проверки на противоречивость вводимых данных, блок выдачи решений, блок рекомендаций по лечению, при этом блок интерфейса выполнен с возможностью получения данных от клинициста, или от баз данных с электронными историями болезней, или из хранилищ «больших данных», при этом блок интерфейса выполнен с  
20 возможностью обмена данными с блоком входных данных, с блоком выдачи решений и с блоком рекомендаций по лечению, при этом блок входных данных выполнен в формате электронной истории болезни с возможностью передачи вводимой информации в вычислительный блок и с возможностью хранения эталонных моделей различных заболеваний, соответствующих разным органам и системам человека, при  
25 этом блок входных данных выполнен с возможностью передачи данных в блок проверки на противоречивость вводимых данных, а вычислительный блок выполнен с возможностью пересчета весовых коэффициентов, приписанных к симптомам или признакам для каждого заболевания конкретного органа или системы организма человека, и баллов в пользу конкретных заболеваний, при этом блок интерфейса, блок  
30 входных данных и блок проверки на противоречивость, блок хранения решений, вычислительный блок, блок выдачи решений, блок рекомендаций по лечению

выполнены с возможностью работы на вызовах, и/или в удаленных районах, и/или в зонах чрезвычайных ситуаций без устойчивого доступа в интернет, блок выдачи решений выполнен с возможностью вывода информации обо всех эталонных историях заболеваний с выделением признаков или симптомов, выявленных при обследовании

5 пациента, или вывода информации о тех эталонных заболеваниях, в которых есть общие признаки или симптомы с данными, полученными в ходе обследования, и с возможностью сопоставления каждой эталонной модели заболевания с введенными данными обследования пациента; анализа достаточности данных обследования для определения заболеваний; запроса дополнительных обследований в случае

10 невозможности определения заболевания или назначения плана лечения; группирования заболеваний по органам или системам организма человека, при этом блок рекомендаций по лечению выполнен с возможностью получения данных из блока входных данных и блока выдачи решений и вывода оптимального плана лечения на блок интерфейса.

15 Недостатком предлагаемого решения является то, что диагностика заболеваний осуществляется на основе симптомов заболевания. В предлагаемом решении используются только сведения о здоровье пациента и факторы риска, извлеченные из электронной медицинской карты пациента. Также отличием предлагаемого решения является определение диагностирования хронических заболеваний посредством

20 ансамбля классификаторов.

### СУЩНОСТЬ ИЗОБРЕТЕНИЯ

Технической задачей, на решение которой направлено заявленное решение, является необходимость разработки способа ранней диагностики хронических

25 заболеваний пациента, а также создании набора классификаторов для определения хронических заболеваний. Предлагаемое решение решает заявленную техническую проблему, путем разработки способа ранней диагностики хронических заболеваний пациента на основе использования набора классификаторов, которое охарактеризовано в независимом пункте формулы. Дополнительные варианты

30 реализации настоящего изобретения представлены в зависимых пунктах изобретения.

Технический результат заключается в повышении точности раннего диагностирования хронических заболеваний пациента и определения риска развития осложнений хронических заболеваний, за счет использования набора классификаторов, так как по полученным данным будет происходить проверка по всем

35 хроническим заболеваниям, на которые обучен классификатор, и оценивать риски развития хронических болезней у пациента. Дополнительным техническим

результатом является увеличение производительности серверной инфраструктуры, на которой реализуется способ при решении поставленной задачи (т.е. за счет реализации описываемого способа появляется возможность производить обработку данных с получением результата за меньшее количество времени), тем самым снижая  
5 нагрузку на центральные процессоры вычислительных устройств/серверов, за счет уменьшения количества обрабатываемых запросов.

Заявленный результат достигается за счет применения способа ранней диагностики хронических заболеваний пациента, выполняющийся на вычислительном устройстве, содержащем процессор и память, хранящую инструкции, исполняемые  
10 процессором и содержащие следующие этапы:

получают деперсонифицированные медицинские данные электронной медицинской карты от медицинской информационной системы на вычислительное устройство;

на вычислительном устройстве осуществляют извлечение медицинских и  
15 социальных признаков о состоянии здоровья пациента, а также признаки факторов риска посредством Natural language processing;

извлеченные признаки поступают на вход обученного набора классификаторов для отнесения признаков к классу вероятности наличия хронических заболеваний, при этом обучение набора классификатора заключается в обучении по меньшей мере одного  
20 классификатора на по меньшей мере одно хроническое заболевание и содержит следующие этапы:

осуществляют деление извлеченных медицинских и социальных признаков о состоянии здоровья пациента на группы;

осуществляют деление полученных групп на кластеры по диагностированным у  
25 них заболеваниям, согласно данным электронной медицинской карты и медицинских и социальных признаков о состоянии здоровья пациента, причем объединяют группы в кластеры с общими диагнозами;

получают значения медицинских и социальных признаков о состоянии здоровья пациента и признаков факторов риска в каждом кластере, по которым будет происходить  
30 отнесение извлеченных признаков к классам вероятности наличия хронических заболеваний;

осуществляют оценку риска развития осложнений хронических заболеваний, отнесенных к классу вероятности наличия хронических заболеваний по классу МКБ.;

на выходе ансамбля классификаторов получают класс вероятности наличия  
35 хронических заболеваний, по классу МКБ, и оценку риска развития осложнений хронических заболеваний.

В частном варианте реализации предлагаемого решения, медицинские и социальные признаки о состоянии здоровья пациента включают в себя: пол, возраст, социальный статус, регион проживания, физиологические параметры, лабораторные показатели.

5 В другом частном варианте реализации предлагаемого решения, признаки фактора риска характеризуют признаки, негативно влияющие на здоровье пациента.

В другом частном варианте реализации предлагаемого решения, объединение групп в кластеры осуществляется посредством метода k-средних и/или метода с-средних и/или послойная кластеризация и/или выделение связанных компонент и/или метод минимального покрывающего дерева.

### ОПИСАНИЕ ЧЕРТЕЖЕЙ

Реализация изобретения будет описана в дальнейшем в соответствии с прилагаемыми чертежами, которые представлены для пояснения сути изобретения и никоим образом не ограничивают область изобретения. К заявке прилагаются следующие чертежи:

Фиг. 1, иллюстрирует пример диагностирования сердечно-сосудистых заболеваний

### ДЕТАЛЬНОЕ ОПИСАНИЕ ИЗОБРЕТЕНИЯ

В приведенном ниже подробном описании реализации изобретения приведены многочисленные детали реализации, призванные обеспечить отчетливое понимание настоящего изобретения. Однако, квалифицированному в предметной области специалисту будет очевидно, каким образом можно использовать настоящее изобретение, как с данными деталями реализации, так и без них. В других случаях хорошо известные методы, процедуры и компоненты не были описаны подробно, чтобы не затруднять излишне понимание особенностей настоящего изобретения.

Кроме того, из приведенного изложения будет ясно, что изобретение не ограничивается приведенной реализацией. Многочисленные возможные модификации, изменения, вариации и замены, сохраняющие суть и форму настоящего изобретения, будут очевидными для квалифицированных в предметной области специалистов.

В медицинской информационной системе, используемой в медицинской организации, в процессе работы аккумулируются различные данные о пациентах, включая общую и медицинскую информацию: рост, вес, цифры артериального давления и т.д., зарегистрированные заболевания и обращения в медицинские

организации, протоколы обследований, данные врачебных осмотров, хирургических вмешательств и др.

Предлагаемое решение интегрируется с медицинской информационной системой медицинской организации посредством открытых API систем.

5 От медицинской информационной системы на вычислительное устройства, по запросу, автоматически формируется и отправляется пакет деперсонифицированных медицинских данных пациента из электронной медицинской карты пациента.

10 Электронная медицинская карта, (электронный паспорт пациента) - ЭМК; англ. electronic health record - EHR) - база данных, содержащая сведения о пациенте: физиологические параметры пациента, анамнез, истории болезней и их лечение (методы и ход лечения, прописанные препараты и т.п.), которая создается в медицинском учреждении. В том числе электронная медицинская карта пациентов содержит записи пациентов, включающих, по меньшей мере, следующие данные: дату добавления записи, коды диагнозов, симптомов, процедур и лекарств, текстовое  
15 описание истории болезни на естественном языке, ассоциированные с историей болезни биомедицинские изображения, результаты исследований и анализов пациентов.

20 Посредством методов Natural language processing, на вычислительном устройстве происходит извлечение медицинских и социальных признаков о состоянии здоровья пациента, а также признаков факторов риска из полученного пакета деперсонифицированных медицинских данных пациента из электронной медицинской карты пациента.

25 Под медицинскими признаками о состоянии здоровья пациента понимаются признаки, характеризующие индекс массы тела, артериальное давление, пульсовое давление и др., лабораторные показатели (общий и клинический анализ крови, биохимия крови и т.д.).

Под социальными признаками о состоянии здоровья пациента понимаются признаки, характеризующие пол, возраст, социальный статус, регион проживания.

30 Под факторами риска понимаются признаки, характеризующие негативное влияние на здоровье пациента, например, но не ограничиваясь курение, личный и семейный анамнез, гиподинамия, ожирение и др.

35 Извлеченные признаки поступают на вход обученному набору классификаторов для отнесения признаков к классу вероятности наличия хронических заболеваний, таких как сердечно-сосудистых заболеваний, эндокринологических заболеваний, заболеваний почек, заболеваний органов дыхания. Каждый классификатор из набора обучен на определение одного хронического заболевания. У каждого классификатора

есть свой набор данных, который он определяет для отнесения признаков к классу вероятности наличия хронических заболеваний. В случае, если в извлеченных признаках отсутствует один параметр из набора данных у классификатора, то данный классификатор не используется при оценке. Например, классификатор на определение хронической сердечной недостаточности не будет использован при оценке, если в извлеченных признаках будет отсутствовать уровень гемоглобина.

Обучение набора классификаторов заключается в обучении по меньшей мере одного классификатора на определение по меньшей мере одного хронического заболевания (определение риска наличия атеросклероза, хронической болезни легких и хронической сердечной недостаточности, хроническая болезнь почек, диабет). Этапы обучения классификатора указаны ниже.

Обучение классификаторов осуществляется с помощью библиотек scikit-learn, tensorflow, catboost, xgboost, и др. На одном и том же наборе данных проводится обучение классификаторов, после чего анализируются полученные метрики точности и матрицы ошибок, что позволяет определить оптимальный для каждой из задач алгоритм. Набор анализируемых метрик точности включает в себя:

Для алгоритмов классификации:

- accuracy (доля правильных ответов);
- точность (precision);
- полнота (recall);
- f-мера (гармоническое среднее полноты и точности);
- auc –roc (площадь под кривой ошибок);
- логистическая функция потерь.

Данные для прогнозирования заболеваний включают в себя:

- демографические данные (пол, возраст, регион проживания);
- анамнез пациентов (коды хронических неинфекционных заболеваний);
- медицинская история пациента (частота обращений и госпитализаций, диагностических манипуляций etc);
- извлеченные из ЭМК пациентов признаки в их динамической интерпретации - определенные в виде функций тенденции изменчивости физиологических и лабораторных параметров.

Реализация настоящего технического решения позволяет проводить первичный этап определения подозрений на заболевания без задействования точных медицинских сведений. На основе принадлежности пациента к одному из выделенных кластеров определяется вероятность наличия у него хронических заболеваний. Интерпретация выходных данных классификаторов происходит с использованием



порога достоверности, выше которого вероятность наличия у пациента хронических заболеваний определяется как высокая, ниже – как низкая. Например, для представителей кластера «мужчины 45-50 лет, страдающие ожирением и артериальной гипертензией, с малой историей обращений в ЛПУ» вероятность наличия сердечно-сосудистых заболеваний определяется как высокая.

Для развертывания модели прогнозирования хронических заболеваний на основе кластерного анализа больших данных необходимо выполнение следующих минимальных технических требований.

Минимальные характеристики вычислительных элементов:

Процессор: не менее 6 ядер частотой от 2.0 ГГц с поддержкой инструкций AVX;

Оперативная память: не менее 24 ГБ;

Дисковая подсистема: 100 ГБ свободного дискового пространства.

Операционная система одна из следующих:

- Ubuntu 18.04 LTS;
- Ubuntu 20.04 LTS;
- Astra Linux Common Edition 2.12.29;
- CentOS 7.7.

Необходимое ПО:

- docker 19.03 и выше
- docker-compose 1.25.5 и выше

1. Осуществляют деление извлеченных медицинских и социальных признаков о состоянии здоровья пациента на группы посредством алгоритма иерархической кластеризации. Группы состояются по ограниченному количеству признаков: пол, возрастная категория, территория проживания. Выделяются возрастные группы (согласно классификации Всемирной организации здравоохранения, а также на основе опубликованных клинических исследований, в зависимости от заболевания, например, для сердечно-сосудистых заболеваний используется шаг в 5 лет внутри возрастной группы 40+ или, например, для исследования гинекологических синдромов, принципиальной отсечкой является возраст менопаузы), после выделения возрастных группы на данных производится совместная разбивка по полу (мужчины и женщины), после разбивки по полу добавляется 3-й признак – территория проживания (согласно данным ЭМК). В результате работ алгоритма получают группы такого типа «мужчины 40-50 лет, проживающие в условиях Крайнего Севера», или «женщины 60-65 лет, проживающие в Южном Федеральном округе».

2. Осуществляют деление полученных групп на кластеры по диагностированным у них заболеваниям, согласно данным электронной медицинской карты, медицинским и социальным признакам о состоянии здоровья пациента, причем объединяют группы в кластеры с общими диагнозами. Объединение групп в кластеры с общим диагнозом происходит посредством алгоритмов: метод k-средних или метод с-средних, или послойная кластеризация, или выделение связанных компонент, или метод минимального покрывающего дерева. Методы кластеризации используют меры связи, полученные в результате сравнения диагнозов между собой. Соответственно, внутри выделенного кластера будет сформирована общность, характеризующаяся наличием общих диагнозов. Например, для группы «мужчины 40-55 лет, жители крупных городов» будет выделен подкластер «с наличием атеросклероза» или «перенесшие инфаркт». Код подкластера передается в модель в виде категориального значения (например, 0 и 1 для инфаркта, 1 – есть инфаркт в анамнезе, 0 – не было инфаркта).

Метод k-средних: - определяется число кластеров  $k$  ; - из исходного набора данных случайным образом выбирается  $k$  рядов, которые выполняют роль начальных центров кластеров; - для каждого ряда данных определяется ближайший к нему центр кластера; - вычисляются центроиды (центры тяжести кластеров); - центр кластера смещается в его центроид; - шаги от определения центра тяжести до смещения его в центроид итеративно повторяются, что обеспечивает рост межкластерных расстояний, чем меньше межкластерное расстояние, тем больше вероятность отнести пациента к нескольким кластерам и если межкластерное расстояние большое, то относят пациента к кластеру с общим диагнозом.

Метод с-средних: формируется матрица принадлежности для разбиения объектов на  $k$  кластеров; - определяются значения критерия ошибки; - все объекты перегруппируются для уменьшения значения критерия ошибки; - последние две процедуры повторяются пока изменения матрицы при перегруппировке не станут незначительными, которые свидетельствуют об определении общего диагноза.

Алгоритм минимального покрывающего дерева: - строится минимальное покрывающее дерево, а затем последовательно убираются ребра с наибольшим весом. Критерием принадлежности к общему диагнозу кластера будут - минимальные веса.

Послойная кластеризация: выделяются связанные компоненты графа, алгоритмом кластеризации формирую последовательность подграфов, которые отражают связи между кластерами. Задается порог расстояния, с помощью изменения которого можно контролировать глубину иерархии кластеров. По значению этого порога (он

рассчитывается на основе выделения связанных компонент на некотором уровне расстояний между объектами) происходит выделение древовидной структуры.

3. Получают значения медицинских и социальных признаков о состоянии здоровья пациента и признаков факторов риска в каждом кластере, по которым будет происходит отнесение извлеченных признаков к классам вероятности наличия хронических заболеваний. Данные значения являются обучающей выборкой. После того как пациент отнесен к одному из кластеров, для него получают категорийное значение, например, 5, будет значить что пациент отнесен группе «мужчины 40-55 лет, проживающие в крупных городах» Именно это значение становится входным параметром для моделей оценки рисков, т.е. на вход подаются не только данные по каждому пациенту (исходные данные), но и значение его кластера.

4. Осуществляют обучение каждого классификатора из набора на основе многослойной (не менее 3 слоев) нейронной сети прямого распространения с нормализацией во входном слое и/или градиентного бустинга, которые относят пациентов каждого кластера к положительному или отрицательному классу (положительный класс по вероятности атеросклероза, положительный класс по вероятности инсульта, отрицательный класс по вероятности хронической болезни почек и т.д. ).

5. В результате получают класс (1 или 0, где 0 – вероятность заболевания отсутствует, 1 – заболевание разовьется) вероятности наличия хронических заболеваний, по классу МКБ, каждого кластера. Например, для атеросклероза брахиоцефальных артерий риск вероятности развития диагноза у представителей возрастной группы 20-30 лет любого пола при отсутствии избыточной массы тела = 0.1, что интерпретируется как низкая вероятность.

6. Осуществляют оценку риска развития осложнений хронических заболеваний пациентов, отнесенных к классу вероятности наличия хронических заболеваний по классу МКБ. Данная оценка риска развития осложнений хронических заболеваний осуществляется на основе алгоритмов машинного обучения (нейронная сеть, градиентный бустинг или алгоритм случайного леса). Под оценкой риска подразумевается численный выход модели и его интерпретация. Для каждого алгоритма рассчитывается определенный порог, выше которого риск наступления события (развития осложнения) считают высоким, ниже – низким. Возможно введение множественных порогов для интерпретации, по которым степень риска оценивается как низкая, умеренная, высокая, очень высокая и т.д. При тренировке моделей алгоритмы учатся на наборах больших данных пациентов, у которых данные осложнения развились, и у которых таковых не было. Для численной оценки риска

развития осложнения применяются алгоритмы регрессии, позволяющие не просто классифицировать данные пациента, как, к примеру «есть риск инсульта» или «нет риска», а дать количественную оценку вероятности развития осложнения, которую платформа интерпретирует в зависимости от степени риска, например, оценка вероятности развития у данного пациента инсульта выдается в виде «высокий риск», что относит пациента к группе повышенного внимания.

Качественная интерпретация разрабатывается с участием медицинских экспертов и в реализации выглядит как цветовая схема с пояснениями (красный – высокий риск, зеленый – низкие риски или риски отсутствуют). Например, вероятность наличия атеросклеротических бляшек брахиоцефальных артерий у мужчин с ожирением (индекс массы тела  $\geq 25$ ) в возрасте 45-50 лет, без диагностированных хронических сердечно-сосудистых заболеваний, при наличии факторов риска (курение, гиподинамия, семейный анамнез, дислипидемия) модель оценила, как 0.6. Порог отсечения для этой модели, рассчитанный по критерию Юдена, составляет 0.55. Таким образом, у пациента из описанной группы высок риск наличия атеросклеротических бляшек.

На выходе набора классификаторов получают класс вероятности наличия хронических заболеваний, по классу МКБ, и оценку риска развития осложнений хронических заболеваний.

На фиг. 1 проиллюстрирован пример, где на основе медицинских и социальных признаков здоровья пациента (таких как: возраст-65 лет, социальный статус – военный пенсионер, территория проживания – крайний север, у которого присутствует – гиподинамия, повышенный холестерин, артрит, артериальная гипертензия, был инфаркт, проводилась антигипертензивная терапия, наличие травм), а также на основе признаков факторов риска (, высокий индекс массы тела, курит, указана частота сердечных сокращений, окружность талии) была определена вероятность ишемической болезни сердца более 50%, а также вероятность развития сахарного диабета 2 типа более 45%.

Степень риска сердечно-сосудистых заболеваний сцеплена с такими факторами, как пол, возраст, регион проживания и социальный статус. Следующим этапом сегментирования пациентов будет выделение внутри половозрастных групп категорий пациентов с отягощенным анамнезом (например, пациентов с диабетом 2 типа, пациентов с почечными заболеваниями и т.д) и выделение факторов риска (семейный анамнез, ожирение, курение, гиподинамия). С помощью комбинированного метода кластеризации, используя иерархическую кластеризацию, а также другие методы кластеризации, описанные выше, присваивают каждой группе пациентов с

определенным набором признаков номер группы со сходными факторами  
отягощённого анамнеза и факторами риска (например, мужчина 68 лет, пенсионер,  
индекс массы тела = 34, курит, в анамнезе имеются заболевания опорно-двигательного  
аппарата (артрозы, которые приводят к гиподинамии). Данный номер группы и данные  
5 о пациенте поступают на вход обученного набора классификаторов. Набор  
классификатор содержит классификаторы для определения следующих заболеваний:  
определение риска наличия атеросклероза, хронической болезни легких и хронической  
сердечной недостаточности, хроническая болезнь почек, наличие атеросклероза  
брахиоцефальных артерий, диабет. Так как вход набора классификатор попадают  
10 данные здоровья пациента (повышенный холестерин, артрит, артериальная  
гипертензия, был инфаркт, проводилась антигипертензивная терапия, наличие травм)  
и факторы риска (семейный анамнез, ожирение, курение, гиподинамия), то по этим  
данным начинает осуществлять работу классификаторы: определение риска  
сахарного диабета и определение риска ишемической болезни сердца, остальные  
15 классификаторы не участвуют в работе, так как для них отсутствуют данные для  
работы. На выходе получают класс вероятности наличия у пациента хронических  
заболеваний по классу МКБ.

Далее осуществляют оценку риска развития осложнений хронических  
заболеваний пациентов, пороговое значение определяется для каждого алгоритма с  
20 помощью статистики Юдена.

Если необходимо определить оценку на наличие атеросклероза  
брахиоцефальных артерий, то на вход набора классификатора подаются следующие  
данные:

- 25 общие сведения о пациенте (возраст, курение);
- общемедицинские сведения (вес, рост, окружность талии, индекс массы тела);
- данные медицинских осмотров (артериальное давление, частота сердечных  
сокращений, частота дыхания);
- сведения об анамнезе (COVID-19, подагра, сахарный диабет, псориаз,  
ревматоидный артрит, фибрилляция предсердий);
- 30 лабораторные показатели (холестерин, ЛПНП, ЛПВП, глюкоза крови, креатинин,  
АСТ, АЛТ, белок крови, триглицериды);
- данные инструментальных измерений (масса миокарда левого желудочка).

Набор классификаторов основывается на алгоритме решающих деревьев. Для  
определения атеросклероза брахиоцефальных артерий, запускается только  
35 классификатор для атеросклероза брахиоцефальных артерий, остальные  
классификаторы из набора не участвуют в анализе.

При анализе решений посредством «дерева решений» используют визуальный и аналитический инструмент поддержки принятия решений для расчёта ожидаемых значений (или ожидаемой пользы) конкурирующих альтернатив. Структура дерева представляет собой «листья» и «ветки». На рёбрах («ветках») дерева решения записаны признаки, от которых зависит целевая функция, в «листьях» записаны значения целевой функции, а в остальных узлах — признаки, по которым различаются случаи. Чтобы классифицировать новый случай, надо спуститься по дереву до листа и выдать соответствующее значение.

Дерево решений в общем виде может быть описано следующей формулой:

$$(x, Y) = (x_1, x_2, x_3, \dots, x_k, Y)$$

Зависимая переменная  $Y$  является целевой переменной, которую необходимо проанализировать, классифицировать и обобщить. Вектор  $x$  состоит из входных переменных  $x_1, x_2, x_3$  и т.д., которые используются для выполнения этой задачи.

Дерево решений состоит из трёх типов узлов:

Узлы решения

Вероятностные узлы

Замыкающие узлы

В результате работы ансамбля решающих деревьев представляется выход модели:

Вероятность наличия бляшек брахиоцефальных артерий в виде числа от 0 до 1. Интерпретация выхода осуществляется с применением порогового значения (0.55), выходное значение больше которого оценивается как высокий риск, меньше — как низкий риск наличия бляшек брахиоцефальных артерий.

В настоящих материалах заявки было представлено предпочтительное раскрытие осуществление заявленного технического решения, которое не должно использоваться как ограничивающее иные, частные воплощения его реализации, которые не выходят за рамки испрашиваемого объема правовой охраны и являются очевидными для специалистов в соответствующей области техники.

## Формула

1. Компьютерно-реализуемый способ ранней диагностики хронических заболеваний пациента, выполняющийся на вычислительном устройстве, содержащем процессор и память, хранящую инструкции, исполняемые процессором и содержащие следующие этапы:

получают деперсонифицированные медицинские данные электронной медицинской карты от медицинской информационной системы на вычислительное устройство;

на вычислительном устройстве осуществляют извлечение медицинских и социальных признаков о состоянии здоровья пациента, а также признаки факторов риска посредством Natural language processing;

осуществляют отнесение извлеченных медицинских и социальных признаков о состоянии здоровья пациента к группе пациентов;

осуществляют отнесение группы пациентов к определенному кластеру по диагностированным у пациента заболеваниям, согласно данным электронной медицинской карты и медицинских и социальных признаков о состоянии здоровья пациента, причем объединяют группу пациентов в кластеры с общими диагнозами;

получают значения медицинских и социальных признаков о состоянии здоровья пациента и признаков факторов риска в каждом кластере, по которым будет происходить отнесение извлеченных признаков к классам вероятности наличия хронических заболеваний;

кластеры с извлеченными признаками поступают на вход обученного набора классификаторов для отнесения признаков к классу вероятности наличия хронических заболеваний, при этом обучение набора классификатора заключается в обучении по меньшей мере одного классификатора на по меньшей мере одно хроническое заболевание, причем классификатор из набора не участвует в классификации, если извлеченных признаках отсутствуют входные признаки для данного классификатора;

осуществляют оценку риска развития осложнений хронических заболеваний;

на выходе набора классификаторов получают класс вероятности наличия хронических заболеваний, по классу МКБ, и оценку риска развития осложнений хронических заболеваний.

2. Способ по п.1, отличающийся тем, что медицинские и социальные признаки о состоянии здоровья пациента включают в себя: пол, возраст, социальный статус, регион проживания, физиологические параметры, лабораторные показатели.

3. Способ по п.1, отличающийся тем, что признаки фактора риска характеризуют признаки, негативно влияющие на здоровье пациента.

4. Способ по п.1, отличающийся тем, что отнесение групп в кластеры осуществляется посредством метода k-средних и/или метода с-средних и/или послойная кластеризация и/или выделение связных компонент и/или метод минимального покрывающего дерева.

5

10

15

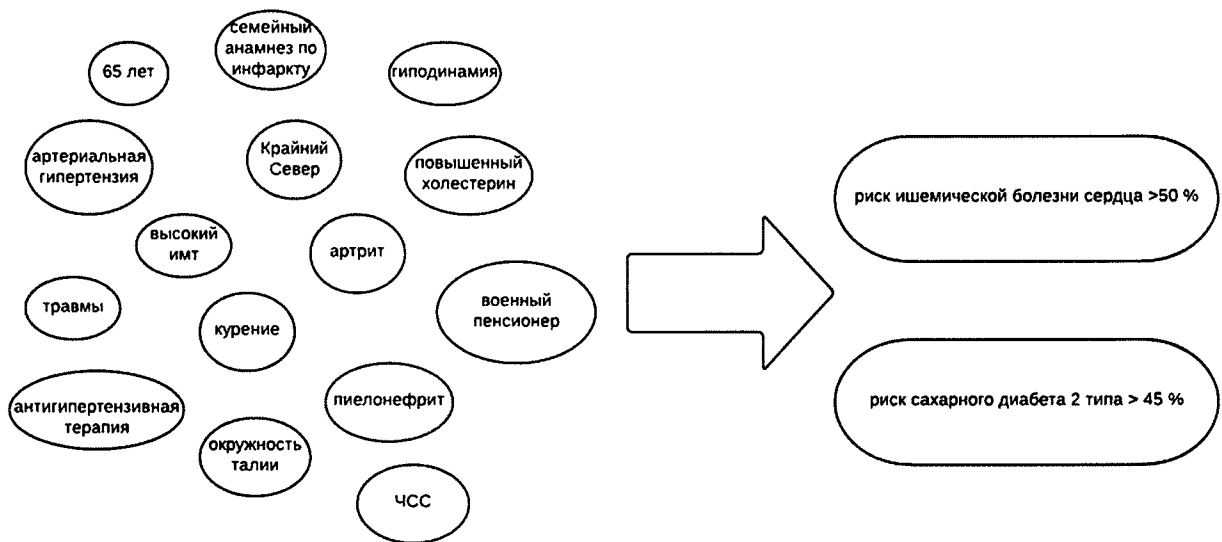
20

25

30

35





Фиг. 1

## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/RU 2021/000605

A. CLASSIFICATION OF SUBJECT MATTER		
G16H 50/70 (2018.01) G16H 50/30 (2018.01)		
According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED		
Minimum documentation searched (classification system followed by classification symbols) A61B 5/00, G16H 10/00, 10/60, 50/00, 50/20, 50/30, 50/70		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) PatSearch (RUPTO Internal), USPTO, PAJ, Espacenet, Information Retrieval System of FIPS		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2020/0335219 A1 (POINTRIGHT INC) 22.10.2020, [0008], [0009], [000], [0013], [0014], [0025], [0044], [0052], [0063], [0070], [0073], [0095], [0119], [0120], [0134], [0158], [0160], [0162], [0166], [0178], [0198], [0220], [0221], [0224], [0231]	1-4
Y	US 2009/0287837 A1 (DAVID PAUL FELSHER) 19.11.2009, [0022], [0185], [0193]	1-4
A	US 2018/0330828 A1 (HAYTER II ROBERT G) 15.11.2018	1-4
A	US 2021/0106288 A1 (HOWARD NEWTON) 15.04.2021	1-4
A	WO 2015/142708 A1 (3M INNOVATIVE PROPERTIES CO) 24.09.2015	1-4
<input type="checkbox"/> Further documents are listed in the continuation of Box C. <input type="checkbox"/> See patent family annex.		
<p>* Special categories of cited documents:</p> <p>“A” document defining the general state of the art which is not considered to be of particular relevance</p> <p>“E” earlier application or patent but published on or after the international filing date</p> <p>“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)</p> <p>“O” document referring to an oral disclosure, use, exhibition or other means</p> <p>“P” document published prior to the international filing date but later than the priority date claimed</p> <p>“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</p> <p>“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone</p> <p>“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art</p> <p>“&amp;” document member of the same patent family</p>		
Date of the actual completion of the international search 26 July 2022 (26.07.2022)		Date of mailing of the international search report 15 September 2022 (15.09.2022)
Name and mailing address of the ISA/RU		Authorized officer
Facsimile No.		Telephone No.

**ОТЧЕТ О МЕЖДУНАРОДНОМ ПОИСКЕ**

Номер международной заявки

PCT/RU 2021/000605

<p><b>A. КЛАССИФИКАЦИЯ ПРЕДМЕТА ИЗОБРЕТЕНИЯ</b>  <i>G16H 50/70</i> (2018.01)  <i>G16H 50/30</i> (2018.01)</p> <p>Согласно Международной патентной классификации МПК</p>																				
<p><b>B. ОБЛАСТЬ ПОИСКА</b>                  Проверенный минимум документации (система классификации с индексами классификации)                  A61B 5/00, G16H 10/00, 10/60, 50/00, 50/20, 50/30, 50/70</p> <p>Другая проверенная документация в той мере, в какой она включена в поисковые подборки</p> <p>Электронная база данных, использовавшаяся при поиске (название базы и, если, возможно, используемые поисковые термины)                  PatSearch (RUPTO Internal), USPTO, PAJ, Espacenet, Information Retrieval System of FIPS</p>																				
<p><b>C. ДОКУМЕНТЫ, СЧИТАЮЩИЕСЯ РЕЛЕВАНТНЫМИ:</b></p> <table border="1"> <thead> <tr> <th>Категория*</th> <th>Цитируемые документы с указанием, где это возможно, релевантных частей</th> <th>Относится к пункту №</th> </tr> </thead> <tbody> <tr> <td>Y</td> <td>US 2020/0335219 A1 (POINTRIGHT INC) 22.10.2020, [0008], [0009], [0011], [0013], [0014], [0025], [0044], [0052], [0063], [0070], [0073], [0095], [0119], [0120], [0134], [0158], [0160], [0162], [0166], [0178], [0198], [0220], [0221], [0224], [0231]</td> <td>1-4</td> </tr> <tr> <td>Y</td> <td>US 2009/0287837 A1 (DAVID PAUL FELSHER) 19.11.2009, [0022], [0185], [0193]</td> <td>1-4</td> </tr> <tr> <td>A</td> <td>US 2018/0330828 A1 (HAYTER II ROBERT G) 15.11.2018</td> <td>1-4</td> </tr> <tr> <td>A</td> <td>US 2021/0106288 A1 (HOWARD NEWTON) 15.04.2021</td> <td>1-4</td> </tr> <tr> <td>A</td> <td>WO 2015/142708 A1 (3M INNOVATIVE PROPERTIES CO) 24.09.2015</td> <td>1-4</td> </tr> </tbody> </table>			Категория*	Цитируемые документы с указанием, где это возможно, релевантных частей	Относится к пункту №	Y	US 2020/0335219 A1 (POINTRIGHT INC) 22.10.2020, [0008], [0009], [0011], [0013], [0014], [0025], [0044], [0052], [0063], [0070], [0073], [0095], [0119], [0120], [0134], [0158], [0160], [0162], [0166], [0178], [0198], [0220], [0221], [0224], [0231]	1-4	Y	US 2009/0287837 A1 (DAVID PAUL FELSHER) 19.11.2009, [0022], [0185], [0193]	1-4	A	US 2018/0330828 A1 (HAYTER II ROBERT G) 15.11.2018	1-4	A	US 2021/0106288 A1 (HOWARD NEWTON) 15.04.2021	1-4	A	WO 2015/142708 A1 (3M INNOVATIVE PROPERTIES CO) 24.09.2015	1-4
Категория*	Цитируемые документы с указанием, где это возможно, релевантных частей	Относится к пункту №																		
Y	US 2020/0335219 A1 (POINTRIGHT INC) 22.10.2020, [0008], [0009], [0011], [0013], [0014], [0025], [0044], [0052], [0063], [0070], [0073], [0095], [0119], [0120], [0134], [0158], [0160], [0162], [0166], [0178], [0198], [0220], [0221], [0224], [0231]	1-4																		
Y	US 2009/0287837 A1 (DAVID PAUL FELSHER) 19.11.2009, [0022], [0185], [0193]	1-4																		
A	US 2018/0330828 A1 (HAYTER II ROBERT G) 15.11.2018	1-4																		
A	US 2021/0106288 A1 (HOWARD NEWTON) 15.04.2021	1-4																		
A	WO 2015/142708 A1 (3M INNOVATIVE PROPERTIES CO) 24.09.2015	1-4																		
<p><input type="checkbox"/> последующие документы указаны в продолжении графы C.      <input type="checkbox"/> данные о патентах-аналогах указаны в приложении</p>																				
<p>* Особые категории ссылок документов:</p> <table border="0"> <tr> <td>“А” документ, определяющий общий уровень техники и не считающийся особо релевантным</td> <td>“Т” более поздний документ, опубликованный после даты международной подачи или приоритета, но приведенный для понимания принципа или теории, на которых основывается изобретение</td> </tr> <tr> <td>“D” документ, цитируемый заявителем в международной заявке</td> <td>“Х” документ, имеющий наиболее близкое отношение к предмету поиска; заявленное изобретение не обладает новизной или изобретательским уровнем, в сравнении с документом, взятым в отдельности</td> </tr> <tr> <td>“E” более ранняя заявка или патент, но опубликованная на дату международной подачи или после нее</td> <td>“У” документ, имеющий наиболее близкое отношение к предмету поиска; заявленное изобретение не обладает изобретательским уровнем, когда документ взят в сочетании с одним или несколькими документами той же категории, такая комбинация документов очевидна для специалиста</td> </tr> <tr> <td>“L” документ, подвергающий сомнению притязание(я) на приоритет, или который приводится с целью установления даты публикации другого ссылочного документа, а также в других целях (как указано)</td> <td>“&amp;” документ, являющийся патентом-аналогом</td> </tr> <tr> <td>“O” документ, относящийся к устному раскрытию, использованию, экспонированию и т.д.</td> <td></td> </tr> <tr> <td>“P” документ, опубликованный до даты международной подачи, но после даты испрашиваемого приоритета</td> <td></td> </tr> </table>			“А” документ, определяющий общий уровень техники и не считающийся особо релевантным	“Т” более поздний документ, опубликованный после даты международной подачи или приоритета, но приведенный для понимания принципа или теории, на которых основывается изобретение	“D” документ, цитируемый заявителем в международной заявке	“Х” документ, имеющий наиболее близкое отношение к предмету поиска; заявленное изобретение не обладает новизной или изобретательским уровнем, в сравнении с документом, взятым в отдельности	“E” более ранняя заявка или патент, но опубликованная на дату международной подачи или после нее	“У” документ, имеющий наиболее близкое отношение к предмету поиска; заявленное изобретение не обладает изобретательским уровнем, когда документ взят в сочетании с одним или несколькими документами той же категории, такая комбинация документов очевидна для специалиста	“L” документ, подвергающий сомнению притязание(я) на приоритет, или который приводится с целью установления даты публикации другого ссылочного документа, а также в других целях (как указано)	“&” документ, являющийся патентом-аналогом	“O” документ, относящийся к устному раскрытию, использованию, экспонированию и т.д.		“P” документ, опубликованный до даты международной подачи, но после даты испрашиваемого приоритета							
“А” документ, определяющий общий уровень техники и не считающийся особо релевантным	“Т” более поздний документ, опубликованный после даты международной подачи или приоритета, но приведенный для понимания принципа или теории, на которых основывается изобретение																			
“D” документ, цитируемый заявителем в международной заявке	“Х” документ, имеющий наиболее близкое отношение к предмету поиска; заявленное изобретение не обладает новизной или изобретательским уровнем, в сравнении с документом, взятым в отдельности																			
“E” более ранняя заявка или патент, но опубликованная на дату международной подачи или после нее	“У” документ, имеющий наиболее близкое отношение к предмету поиска; заявленное изобретение не обладает изобретательским уровнем, когда документ взят в сочетании с одним или несколькими документами той же категории, такая комбинация документов очевидна для специалиста																			
“L” документ, подвергающий сомнению притязание(я) на приоритет, или который приводится с целью установления даты публикации другого ссылочного документа, а также в других целях (как указано)	“&” документ, являющийся патентом-аналогом																			
“O” документ, относящийся к устному раскрытию, использованию, экспонированию и т.д.																				
“P” документ, опубликованный до даты международной подачи, но после даты испрашиваемого приоритета																				
<p>Дата действительного завершения международного поиска                  26 июля 2022 (26.07.2022)</p>		<p>Дата отправки настоящего отчета о международном поиске                  15 сентября 2022 (15.09.2022)</p>																		
<p>Наименование и адрес ISA/RU:                  Федеральный институт промышленной собственности,                  Бережковская наб., д. 30, корп. 1, Москва, Г-59, ГСП-3, 125993,                  Российская Федерация                  тел. +7(499)240-60-15, факс +7(495)531-63-18</p>		<p>Уполномоченное лицо:                  Лясковский С.                  Телефон № 8(499)240-25-91</p>																		