

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

(19) Всемирная Организация  
Интеллектуальной Собственности  
Международное бюро

(43) Дата международной публикации  
15 июня 2023 (15.06.2023)



(10) Номер международной публикации  
**WO 2023/106960 A1**

(51) Международная патентная классификация:  
G16H 50/30 (2018.01)

(21) Номер международной заявки: РСТ/RU2021/000580

(22) Дата международной подачи:  
16 декабря 2021 (16.12.2021)

(25) Язык подачи: Русский

(26) Язык публикации: Русский

(30) Данные о приоритете:  
2021136205 08 декабря 2021 (08.12.2021) RU

(71) Заявитель: **ОБЩЕСТВО С ОГРАНИЧЕННОЙ  
ОТВЕТСТВЕННОСТЬЮ "К-СКАЙ" (LIMITED  
LIABILITY COMPANY "K-SKAI")** [RU/RU]; набе-  
режная Варкуса, 17, помещение 62, офис 20 Каре-

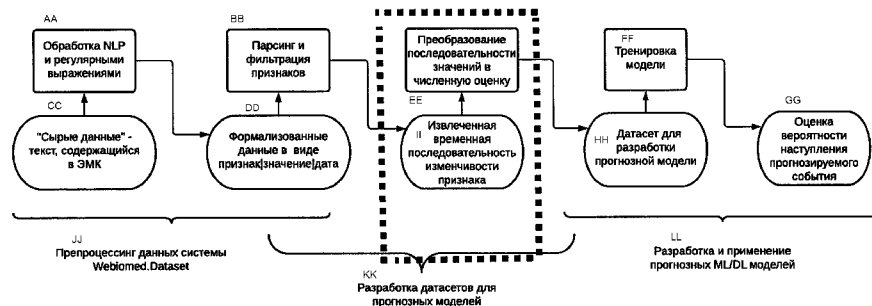
лия Республика, г. Петрозаводск, 185031, Republic of  
Karelia, g. Petrozavodsk (RU).

(72) **Изобретатели: НОВИЦКИЙ, Роман Эдвардович  
(NOVITSKIY, Roman Edvardovich);** ул. Коммуналь-  
ная, 1, кв. 74 Карелия Республика, г. Петроза-  
водск, 185035, Republic of Karelia, g. Petrozavodsk  
(RU). **ГУСЕВ, Александр Владимирович (GUSEV,  
Aleksandr Vladimirovich);** ул. Летняя, 30 Р-н Кондо-  
пожский, Карелия Республика, г. Кондопога, 186220, R-  
n Kondopozhskiy, 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 PREDICTING THE ONSET OF A MEDICAL EVENT IN A PERSON'S HEALTH

(54) Название изобретения: СПОСОБ ПРОГНОЗИРОВАНИЯ НАСТУПЛЕНИЯ МЕДИЦИНСКОГО СОБЫТИЯ В  
ЗДОРОВЬЕ ЧЕЛОВЕКА



Фиг. 1

- AA NLP processing using regular expressions
- BB Parsing and feature filtration
- CC "Raw data" - text contained in electronic medical records
- DD Formalized data in feature/value/date form
- EE Conversion of sequence of values into numerical estimate
- FF Training of model
- GG Estimation of probability of the onset of an event to be predicted
- HH Dataset for creating a predictive model
- II Extracted chronological sequence of feature variability
- JJ Preprocessing of system data using Webiomed.Dataset
- KK Creation of datasets for predictive models
- LL Creation and application of predictive ML/DL models

(57) **Abstract:** The invention relates to a method for predicting the onset of a medical event in a person's health using machine learning, taking into account estimates of the variability of personal health data over time. The technical result of the claimed invention is that of allowing more accurate and better quality prediction of the onset of a medical event in a person's health, taking the pattern of change over time of a patient's health indices into account in the prediction process. The method involves extracting personal health data containing personal health parameters by means of processing medical documents describing a person's state of health; using the extracted personal health data to generate chronological sequences of personal health data, each chronological sequence being produced for the values of one health parameter; determining a set of health parameters for predicting the onset of a medical event in the person's health, said set

[продолжение на следующей странице]



WO 2023/106960 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), ОАПИ (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)

of parameters depending on the medical event to be predicted; converting the generated chronological sequence of health parameter values for each health parameter from the set of parameters for predicting the onset of a medical event into an estimative index of the pattern of change of the values of the given health parameter; obtaining a set of estimative indices of the pattern of change over time of the health parameter values for predicting the onset of a medical event in the person's health; feeding the obtained set of estimative indices of the pattern of change over time of the health parameter values for predicting the onset of a medical event in the person's health to the input of a predictive machine learning model, and obtaining at the output thereof the probability of the onset of a medical event in the person's health.

(57) **Реферат:** Изобретение относится к способу прогнозирования наступления медицинского события в здоровье человека с применением машинного обучения с учетом оценок изменчивости данных о состоянии здоровья человека во времени. Техническим результатом заявляемого изобретения является повышение точности и качества прогнозирования наступления медицинского события в здоровье пациента, учет динамики изменения показателей здоровья пациента во времени в процессе прогнозирования. В способе извлекают данные о состоянии здоровья человека, содержащие параметры здоровья человека, путем обработки медицинских документов, характеризующих состояние здоровья человека; на основе извлеченных данных о состоянии здоровья человека формируют временные последовательности данных о состоянии здоровья человека; причем каждую временную последовательность получают для значений одного параметра здоровья; определяют набор параметров здоровья для прогнозирования наступления медицинского события в здоровье человека; причем указанный набор параметров здоровья зависит от прогнозируемого медицинского события о состоянии здоровья человека; для каждого параметра здоровья из набора параметров для прогнозирования наступления медицинского события в здоровье человека преобразуют сформированную временную последовательность значений параметра здоровья в оценочный показатель динамики изменения значений параметра здоровья; и получают набор оценочных показателей динамики изменения во времени значений параметров здоровья для прогнозирования наступления медицинского события в здоровье человека; подают на вход прогностической модели машинного обучения полученный набор оценочных показателей динамики изменения во времени значений параметров здоровья для прогнозирования наступления медицинского события в здоровье человека и на выходе прогностической модели машинного обучения получают вероятность наступления медицинского события в здоровье человека.

## СПОСОБ ПРОГНОЗИРОВАНИЯ НАСТУПЛЕНИЯ МЕДИЦИНСКОГО СОБЫТИЯ В ЗДОРОВЬЕ ЧЕЛОВЕКА

5

### ОБЛАСТЬ ТЕХНИКИ

Изобретение относится к области информационных и коммуникационных технологий для обработки медицинских данных, в частности к способу прогнозирования наступления медицинского события в здоровье человека с применением машинного обучения с учетом оценок изменчивости данных о состоянии здоровья человека во времени.

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, например госпитализации в стационар, летального исхода, развития специфического осложнения, развития сопутствующего заболевания, при планировании профилактических мероприятий, диспансеризации пациентов, а также при разработке комплекса предиктивных моделей, прогнозирующих стандартные медицинские события (госпитализация в стационар, летальный исход, развитие специфического осложнения, развитие сопутствующего заболевания и др.).

### УРОВЕНЬ ТЕХНИКИ

Большинство прогнозных моделей, задачей которых является оценка вероятности наступления какого-либо события в здоровье человека в будущем или оценка вероятности наличия какого-то состояния в организме человека в данное время, принимают «на вход» (в качестве исходного) набор показателей, представленных в том числе абсолютными числовыми значениями (например, рост, вес, холестерин, артериальное давление и т.д.). Как правило, эти значения определены на момент обследования. При выполнении машинного обучения, основанного на фиксированных значениях входных признаков, независимо от количества вносимых данных и алгоритмов, точность модели упирается «в потолок» и не достигает необходимого для реального применения значения. Вне зависимости от материальных и временных вложений применяемые сейчас в мире

модели не достигают требуемого, идеального значения точности, которое, например, в здравоохранении, должно быть не ниже 80-85%, в идеальной ситуации приближаться к 95-97%.

5 Модель, построенная на фиксированных значениях входных признаков, не позволяет проанализировать состояние здоровья организма в обратной перспективе, учитывая все индивидуальные особенности.

Например, при заданном весе человека 80 кг, артериальном давлении 140, холестерине 8, такая модель не может понять природу изменений организма этого человека. Так, исходя из этих вводных, может наблюдаться, как минимум, два  
10 противоречащих друг другу прогноза:

1) Всю жизнь у этого человека было АД 120, а сейчас 140 => динамика отрицательная. Это значит, что явно идет ухудшение состояния здоровья.

2) Всю жизнь у этого человека было АД 180, а сейчас 140 => динамика положительная. Это значит, что наметилось улучшение.

15 Таким образом, распространенная на данный момент модель прогнозирования не разделяет между собой вышеуказанные «сценарии» и выдает различным людям единый прогноз, что как минимум в 50 % случаев неточно. Или, например, модель учитывает такой параметр пациента, как «не курю». Но пациент мог бросить курить 10 лет назад, а мог бросить курить вчера – и его состояние  
20 обусловлено последствиями отказа или накопленного эффекта.

Упрощение работы когнитивных технологий приводит к снижению прогностической точности моделей. Модели не понимают динамику изменений состояния организма и поэтому, хоть и могут применяться в среде разработки (где нет необходимости учитывать уникальные или нечасто встречающиеся  
25 отклонения), в реальной клинической практике точности достигнуть не удастся, а в лаборатории невозможно учесть все возможные жизненные сценарии, которые случаются на практике.

В заявке на изобретение US2017032241A1, дата публикации 02.02.2017, описаны способы, системы и устройства, включая компьютерные программы, закодированные на компьютерных носителях, для использования рекуррентных  
30 нейронных сетей для анализа событий, связанных со здоровьем. Один из способов включает получение первой временной последовательности событий о здоровье, где первая временная последовательность включает соответствующие данные о здоровье, связанные с конкретным пациентом на каждом из множества временных шагов; обработку первой временной последовательности событий о здоровье с  
35

помощью рекуррентной нейронной сети для генерации выходных данных нейронной сети для первой временной последовательности; и генерацию из выходных данных нейронной сети для первой временной последовательности данных анализа здоровья, которые характеризуют будущие события о здоровье, которые могут произойти после последнего временного шага во временной последовательности.

В патенте US9652712B2, дата публикации 16.05.2017, раскрыты способы, системы и устройства, включая компьютерные программы, закодированные на компьютерных носителях, для использования рекуррентных нейронных сетей для анализа событий, связанных со здоровьем. Один из способов включает: обработку каждой из множества начальных временных последовательностей событий о здоровье для создания, для каждой из начальных временных последовательностей, соответствующего внутреннего состояния сети рекуррентной нейронной сети для каждого временного шага в начальной временной последовательности; хранение, для каждой из начальных временных последовательностей, одного или более внутренних состояний сети для временных шагов во временной последовательности в хранилище; получение первой временной последовательности; обработка первой временной последовательности с помощью рекуррентной нейронной сети для создания внутреннего состояния последовательности для первой временной последовательности; и выбор одной или более начальных временных последовательностей, которые, вероятно, включают события, связанные со здоровьем, которые предсказывают будущие события, связанные со здоровьем, в первой временной последовательности.

В заявке на изобретение US20200152333A1, дата публикации 14.05.2020, раскрыты способы, системы и устройства, включая компьютерные программы, закодированные на компьютерных носителях, для прогнозирования будущих неблагоприятных событий со здоровьем с помощью нейронных сетей. Один из способов включает получение данных электронной медицинской карты пациента; генерирование на основе данных электронной медицинской карты входной последовательности, включающей соответствующее представление признака на каждом из множества временных шагов временного окна, включающее, для каждого временного шага временного окна: определение, для каждого из возможных числовых признаков, произошел ли числовой признак во время временного окна; и генерирование, для каждого из возможных числовых признаков,

одного или более признаков присутствия, которые определяют, произошел ли числовой признак во время временного окна; и обработку входной последовательности с помощью нейронной сети для генерирования выходного сигнала нейронной сети, который характеризует прогнозируемую вероятность того, что неблагоприятное событие со здоровьем произойдет с пациентом.

Однако в данных решениях отсутствует обработка отдельно каждой временной последовательности значений параметров, характеризующих состояние здоровья пациента, (например, последовательности значений артериального давления, показателя сахара крови, гликированного гемоглобина и т.д.), преобразование временной последовательности значений каждого параметра здоровья в оценочный показатель (например, числовой индекс), характеризующий динамику изменения во времени параметра (например, оценка межвизитной или временной изменчивости параметра), и подача на вход прогнозной модели не единичных значений, а оценочных показателей.

15

#### СУЩНОСТЬ ИЗОБРЕТЕНИЯ

Техническим результатом заявляемого изобретения является повышение точности прогнозирования наступления медицинского события в здоровье пациента, учет динамики изменения показателей здоровья пациента во времени в процессе прогнозирования, повышение качества предоставляемого прогноза с использованием машинного обучения с учетом оценок изменчивости данных о состоянии здоровья человека во времени.

Указанный технический результат достигается за счёт того, что Компьютерно-реализуемый способ прогнозирования наступления медицинского события в здоровье человека с применением машинного обучения с учетом оценок изменчивости данных о состоянии здоровья человека во времени, в котором:

извлекают данные о состоянии здоровья человека, содержащие параметры здоровья человека, путем обработки медицинских документов, характеризующих состояние здоровья человека;

на основе извлеченных данных о состоянии здоровья человека формируют временные последовательности данных о состоянии здоровья человека; причем каждую временную последовательность получают для значений одного параметра здоровья;

определяют набор параметров здоровья для прогнозирования наступления медицинского события в здоровье человека; причем указанный набор параметров здоровья зависит от прогнозируемого медицинского события в здоровье человека;

5 для каждого параметра здоровья из набора параметров для прогнозирования наступления медицинского события в здоровье человека преобразуют сформированную временную последовательность значений параметра здоровья в оценочный показатель динамики изменения значений параметра здоровья; и получают набор оценочных показателей динамики изменения во времени значений параметров здоровья для прогнозирования наступления медицинского события в здоровье человека;

10 подают на вход прогностической модели машинного обучения полученный набор оценочных показателей динамики изменения во времени значений параметров здоровья для прогнозирования наступления медицинского события в здоровье человека и на выходе прогностической модели машинного обучения получают вероятность наступления медицинского события в здоровье человека.

В способе медицинскими событиями в здоровье человека могут являться, по меньшей мере, следующие: госпитализация в стационар, летальный исход, развитие специфического осложнения, развитие сопутствующего заболевания.

20 В способе параметрами здоровья для прогнозирования наступления медицинского события в здоровье человека могут являться, по меньшей мере, следующие: физиологические параметры, данные лабораторных исследований.

В способе оценочным показателем может являться индекс долгосрочной изменчивости параметра здоровья человека.

25 В способе каждую временную последовательность могут получать для одного параметра здоровья за заданный период времени.

### ОПИСАНИЕ ЧЕРТЕЖЕЙ

30 Реализация изобретения будет описана в дальнейшем в соответствии с прилагаемыми чертежами, которые представлены для пояснения сути изобретения и никоим образом не ограничивают область изобретения.

Заявляемое изобретение проиллюстрировано фигурами 1-4, на которых изображены:

35 Фиг. 1 – иллюстрирует вариант блок-схемы способа прогнозирования наступления медицинского события в здоровье человека с учетом оценок изменчивости данных о состоянии здоровья человека во времени.

Фиг. 2 – иллюстрирует вариант блок-схемы процесса преобразования данных в настоящем изобретении.

Фиг. 3 – иллюстрирует пример цикла работы с данными в настоящем изобретении.

5 Фиг. 4 – иллюстрирует общую схему вычислительного устройства для реализации настоящего изобретения.

#### ДЕТАЛЬНОЕ ОПИСАНИЕ ИЗОБРЕТЕНИЯ

10 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  
15 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  
20 настоящего изобретения, будут очевидными для квалифицированных в предметной области специалистов.

В настоящем изобретении на вход прогностической модели подаются не абсолютные значения признаков, характеризующих состояние здоровья человека, а аналитическая оценка динамики изменения данных признаков во времени  
25 (например, числовая характеристика, индекс). То есть для каждого признака получают не абсолютное число, а значение в зависимости от ретроспективного анализа данного признака здоровья человека во времени.

В заявленном решении ведут работу с последовательностями дискретных по времени числовых значений конкретных параметров здоровья пациента  
30 (объективных данных, лабораторных показателей и т.д.), что позволяет учитывать динамику параметра и благоприятно влияет на точность и специфичность прогноза. Заявленное решение может использоваться в прогностических моделях разного назначения и класса решаемой задачи (классификация, регрессия), при условии обучения их на реальных медицинских данных, которые отражают состояние  
35 здоровья человека, например, на данных электронных медицинских карт (ЭМК).



Настоящее решение применяется для готовых, уже сформированных последовательностей и преследует цель повышения точности прогноза, основанного на использовании данных этих последовательностей. Т.е., формирование последовательности происходит на стадии обработки медицинских документов комплексом НЛП-моделей (англ. Neuro-Linguistic Programming, сокр. NLP), маппинга, регулярных выражений, в результате чего формируется массив датированных извлеченных признаков (Фиг. 1). Решение работает отдельно с каждой временной последовательностью одного признака (например, последовательность значений такого параметра здоровья человека, как например артериального давления или показателя сахара крови или гликированного гемоглобина и т.д.), полученной для одного пациента в течение определенного периода времени. Целью преобразований временной последовательности значений в числовой индекс является подача на вход прогнозной модели не единичной записи, а оценки динамики изменения во времени значений признака (параметра). Такой подход позволяет учесть динамику показателя здоровья пациента в процессе тренировки прогнозных моделей и повысить качество предоставляемого ими прогноза (Фиг. 1).

В качестве такой оценки (оценочного показателя) может использоваться индекс долгосрочной изменчивости признака (параметра). Например, таким индексом может являться оценка межвизитной или временной изменчивости признака.

Так, например, в клинической практике широко используется межвизитная вариабельность артериального давления, определяемая как отклонения артериального давления от среднего уровня, рассчитанного за конкретный период (год или несколько). Так, возрастание межвизитной вариабельности систолического давления считается предиктором повышенного риска сердечно-сосудистых исходов при сердечной недостаточности и артериальной гипертензии. На практике в качестве показателей вариабельности используются показатели стандартного отклонения или коэффициент вариации. При этом стандартное отклонение является менее предпочтительным выбором из-за своей зависимости от среднего уровня параметра.

Кроме того, в качестве оценочного показателя могут выступать:

- средняя реальная вариабельность (ARV - Average Real Variability),
- вариабельность, не зависящая от среднего (VIM - Variability Independent of Mean).

Также выбор формул для оценочного показателя можно осуществлять экспериментальным путем. Формулы для каждого признака будут выбираться в ходе корреляционного анализа. Для каждого параметра рассчитывают набор характеристик, включающих в себя:

- 5           - стандартное отклонение от среднего (STD – STAndart Deviation);
- коэффициент вариации;
- средняя реальная вариабельность ARV;
- вариабельность, не зависящая от среднего VIM.

10           На следующем шаге проводится корреляционный анализ взаимосвязи каждого преобразованного показателя с целевым параметром (feature importance). Данный подход позволит определить оптимальный метод расчета индексов для каждого признака.

15           Также может использоваться переход к логическим переменным, таким как «положительный прирост величины» или «отрицательный прирост величины», причем подобная оценка будет даваться на разные временные промежутки, например, «положительный прирост показателя индекса массы тела в течение последних 12 месяцев» или «снижение показателя креатинина после назначения пероральных препаратов».

20           Кроме того, может быть проведен анализ параметра, для которого выявлена значительная амплитуда показателя за определенный период времени, например, выявлена значительная амплитуда параметров биохимического анализа крови за предшествующие 12 месяцев. В таком случае анализируют сопутствующие события – заболевания, медицинские рекомендации (исполняются или нет), вариабельность других показателей здоровья.

25           Ниже приведен пример модели прогнозирования летального исхода в течение 1 года и 5 лет у пациентов с сахарным диабетом 2 типа. Модели используют метод градиентного бустинга для определения численного значения вероятности наступления летального исхода у пациентов из определенной выборки (критерии включения - наличие сахарного диабета 2 типа в списке основных  
30           диагнозов пациента, возраст в пределах 18-99 лет). При использовании градиентного бустинга прогнозы создаются на основе последовательных ансамблей алгоритмов, в котором каждая последующая модель обучается с использованием данных об ошибках предыдущих моделей для дальнейшего  
35           снижения ошибок. Результатом срабатывания каждого алгоритма является результирующее значение, выдаваемое ансамблем – это сумма результатов

отдельных регрессионных моделей. В ансамбле моделей по очереди работают входящие в состав ансамбля алгоритмы, при этом каждый последующий обучается в том числе на ошибках данных предыдущего. Таким образом, выдаваемый моделью результат уже содержит в себе минимизацию ошибки. Входные параметры модели включают в себя: возраст, пол, применение антигипертензивной терапии, вес, гемоглобин крови, головная боль, диастолическое артериальное давление, систолическое артериальное давление, креатинин крови, содержание лейкоцитов в крови, общий холестерин, применение пероральных препаратов, рост, скорость оседания эритроцитов (СОЭ), температура тела, содержание тромбоцитов, утомляемость, частота дыхания, содержание эритроцитов в крови.

В Таблице 1 приведен предшествующий уровень техники - примерный вид набора данных для обучения модели прогнозирования летального исхода, для формирования которого были взяты фиксированные значения каждого параметра.

Таблица 1.

Пациент	Пол	Возраст	Вес	Общий холестерин	Гемоглобин	.....	Креатинин
111111	1	69	75	4	132		68
222222	0	55	70	3.5	140		70

В настоящем решении изменен подход к машинному обучению, в котором используют не фиксированные значения каждого параметра, а статистические оценки изменчивости значений каждого параметра во времени (индексы), при этом учитывается динамика изменения каждого признака. Схематично процесс преобразования данных представлен на Фиг. 2. При этом при использовании этого подхода Таблица 1 примет следующий вид:

Таблица 2.

Пациент	Пол	Возраст	Оценки параметра «Вес»				....	Оценки параметра «креатинин»			
			$\Delta$ за год	Среднее	Max	Min		$\Delta$ за год	Среднее	Max	Min
111111	1	69	+4	73	76	69	0	70	73	65	
222222	0	55	-7	75	79	70	+6	72	74	65	

Данный подход можно применять при прогнозировании стандартных медицинских событий (например, госпитализация в стационар, летальный исход, развитие специфического осложнения, развитие сопутствующего заболевания и др.) с помощью прогностических моделей машинного обучения.

Электронная медицинская карта пациента содержит набор медицинских документов, из которых с помощью НЛП-моделей (англ. Neuro-Linguistic Programming, сокр. NLP) или маппинга можно извлечь конкретные параметры здоровья в формализованном виде (Фиг. 1). На Фиг. 3 приведен пример цикла работы с данными в заявленном решении. Из формализованных медицинских данных, которые прошли препроцессинг, формируют новые признаки (оценки изменчивости параметров). Хранилище формализованных медицинских данных может пополняться новыми признаками, которые используются для обучения различных прогностических моделей, а также для их дообучения.

Особенность данных, содержащихся в ЭМК заключается в том, что все сведения медицинского характера собираются в момент обращения в ЛПУ, причем состав этих данных может отличаться в зависимости от причины обращения, а частота внесения в ЭМК одного и того же параметра - от плана лечения и состояния пациента. Таким образом, для каждого пациента мы можем получить набор дискретных записей, характеризующих каждый параметр здоровья. В таблице 3 приведен пример набора медицинских сведений, полученных для одного пациента в разные даты и в ходе разных обращений. Так, при вызове врача на дом или обращении в поликлинику формируется один набор сведений, включающий в себя жалобы пациента и данные объективного осмотра, после чего дается рекомендация на проведение дополнительных обследований (лабораторных или инструментальных), в результате которых формируется принципиально новый набор данных.

Таблица 3 – Пример извлеченных из электронной медицинской карты значений параметров здоровья.

Таблица 3.

	Глюкоза крови	Общий холестерин	Систолическое АД	Диастолическое АД	Частота сердечных сокращений
13.04.2015	-	-	110 мм рт.ст.	70 мм рт.ст.	80 уд в мин
2.02.2016	-	-	180 мм рт.ст.	100 мм рт.ст.	-
1.08.2015	3,5 ммоль/л	4,6 ммоль/л	-	-	-
5.02.2016	3,6 ммоль/л	7,5 ммоль/л	-	-	-

Работа с данными ЭМК дает возможность провести масштабное исследование данных и выделить для отдельных нозологий тот набор параметров, динамика которых принципиальна для развития исходов конкретного заболевания.

Этот список параметров определяется при участии медицинского эксперта отдельно для каждого заболевания (или группы заболеваний) и проверяется методом feature importance – отбором информативных признаков. То есть, для прогнозирования исходов сердечно-сосудистых заболеваний оценивается динамика одного набора параметров, для прогнозирования развития легочных болезней - другого и т.д.

Например, разница наборов входных параметров для прогноза госпитализации пациентов с ССЗ и с легочными заболеваниями в течение 12 месяцев представлена в Таблице 4.

10 Таблица 4.

Параметры для прогнозирования госпитализации пациентов с сердечно-сосудистыми заболеваниями	Параметры для прогнозирования госпитализации пациентов с легочными заболеваниями
индекс массы тела	индекс массы тела
частота сердечных сокращений	систолическое артериальное давление
систолическое артериальное давление	диастолическое артериальное давление
диастолическое артериальное давление	частота дыхания
общий холестерин	частота сердечных сокращений
липопротеиды низкой плотности	глюкоза крови
липопротеиды высокой плотности	креатинин крови
триглицериды	триглицериды
скорость оседания эритроцитов	гемоглобин крови
тромбоциты	общий холестерин
глюкоза крови	эритроциты
лейкоциты	лейкоциты крови

Таким образом, любой набор данных будет включать физиологические параметры (например, артериальное давление, пульсовое давление, индекс массы тела, частота сердечных сокращений и др. ), данные лабораторных исследований (общий анализ крови, биохимия крови, общий анализ мочи, исследования на антитела и др.) и сведения об анамнезе (например, наличие в анамнезе хронических неинфекционных заболеваний, таких как диабет, ишемическая болезнь сердца, пиелонефрит и т.д.; общее количество основных диагнозов; количество впервые установленных диагнозов за последний год; количество госпитализаций в стационар, изменение состояний, возникновение осложнений). Этот набор уточняется на стадии проведения отбора информативных признаков.

Для формализации данных, так называемых данных нерегулярных временных последовательностей (именно к этому виду относятся записи электронных медицинских карт) в общепринятой практике анализа медицинских данных наибольшей популярностью пользуются следующие подходы:

5           извлечение фиксированного количества значений из последовательности, например, длину ряда, среднее значение или амплитуду между первым и последним значением;

          сведение всей последовательности в одну точку путем взятия взвешенной суммы всех значений, где временная метка каждого значения используется для  
10           присвоения соответствующих весов;

          создание временных паттернов или временной логики по типу "возникновение клинического события А предшествует падению клинического события В";

          математические аппроксимации, представляющие исходные  
15           последовательности в упрощенном виде, наиболее часто используется Symbolic Aggregate approXimation (SAX);

          кластеризация сгенерированных последовательностей, благодаря чему последовательности, принадлежащие к одному классу, заменяются его медоидом.

          В настоящем решении комбинируют несколько наиболее  
20           зарекомендовавших себя практик для различных групп показателей:

          изменчивость морфологических данных (таких как вес, индекс массы тела) можно формализовать с помощью простых математических оценок, т.к. в этом случае достаточно понимания тенденции изменчивости. Так, в большинстве случаев изменение индекса массы тела (ИМТ) в сторону увеличения/уменьшения  
25           является достаточной информацией для прогностической оценки, уточнение конкретной величины приращения не влияет на точность прогноза. Соответственно, в обучающий датасет можно ввести бинарный показатель delta\_bmi (изменчивость ИМТ за определенный период времени), принимающий значения 0 (отрицательный прирост) или 1 (положительный прирост);

          динамика лабораторных показателей будет индексироваться с помощью  
30           пошаговых преобразований временной последовательности и их кластеризации.

          Индексация может проводиться методами оценки межвизитной  
35           вариабельности показателя. Для реализации этой задачи оптимальным подходом является простое экспоненциальное сглаживание, при котором вес значения параметра уменьшается по мере углубления в исторические данные. Так, для

прогнозирования целевого медицинского события (например, госпитализации в стационар) важны все доступные наблюдения, но определяющий вес будут иметь последние записи временной последовательности.

5 Также можно провести многофакторный анализ и выделить клинические события, прямо влияющие на динамику показателей, например, назначение антигипертензивных препаратов, влекущее за собой снижение артериального давления. Подобное определение значимости признаков для целевого события реализуется с помощью моделей множественной регрессии.

10 Например, в наборе данных для прогноза госпитализации в стационар пациентов с ССЗ, динамика индекса массы тела пациента оценивается логической переменной «growth\_bmi», где переменная принимает значение True, если разница последнего и первого извлеченного имеет положительное значение и False, если эта величина отрицательна. Оценка динамики лабораторных показателей (холестерин крови, глюкоза крови и др.) получается следующим образом –  
15 рассчитываются STD, ARV и VIM для каждого информативного параметра, этот набор используется в экспериментальном обучении с минимум 5 алгоритмами (градиентный бустинг, адаптивный бустинг, многослойная нейронная сеть, логистическая регрессия, деревья решающих правил), после чего методом feature selection определяется финальный набор входных параметров и на основе метрик  
20 выбирается итоговый алгоритм машинного обучения. Так, для прогнозирования госпитализации пациентов с ССЗ в стационар в течение года использовалась многослойная нейронная сеть прямого распространения с 3-мя слоями (MLP) с кросс-валидацией во входном слое. На выходе модель выдает вероятность наступления целевого события (госпитализации в стационар). При замене твердого  
25 значения входных параметров модели на их динамическую оценку метрики модели в тестировании будут не менее 0.87 (0.85 – 0.89) при CI = 0.95 для ROC – AUC и чувствительность модели возрастет до 0.9.

Ниже приведен пример осуществления настоящего изобретения.

30 Извлекают данные о состоянии здоровья человека, содержащие параметры здоровья человека, путем обработки медицинских документов, характеризующих состояние здоровья человека. Например, данные о состоянии здоровья человека могут быть извлечены из ЭМК человека с помощью НЛП-моделей или маппинга. Параметрами здоровья человека являются, например, физиологические  
35 параметры (например, артериальное давление, пульсовое давление, индекс массы тела, частота сердечных сокращений и др.), данные лабораторных исследований

(общий анализ крови, биохимия крови, общий анализ мочи, исследования на антитела и др.). На основе извлеченных данных о состоянии здоровья человека формируют временные последовательности данных о состоянии здоровья человека; при этом каждую временную последовательность получают для значений одного параметра здоровья. Определяют набор параметров здоровья для прогнозирования наступления медицинского события в здоровье человека. При этом указанный набор параметров здоровья зависит от прогнозируемого медицинского события в здоровье человека. Так, например, набор параметров для прогнозирования госпитализации пациентов с сердечно-сосудистыми заболеваниями будет отличаться от набора параметров для прогнозирования госпитализации пациентов с легочными заболеваниями (Таблица 4). Для каждого параметра здоровья из набора параметров сформированную временную последовательность значений параметра здоровья преобразуют в оценочный показатель динамики изменения значений параметра здоровья, например, индекс долгосрочной изменчивости параметра здоровья человека. Получают набор оценочных показателей динамики изменения во времени значений параметров здоровья для прогнозирования наступления медицинского события в здоровье человека. Подают на вход прогностической модели машинного обучения полученный набор оценочных показателей динамики изменения во времени значений параметров здоровья для прогнозирования наступления медицинского события в здоровье человека и на выходе прогностической модели машинного обучения получают вероятность наступления медицинского события в здоровье человека, например, вероятность госпитализации в стационар, вероятность летального исхода, вероятность развития специфического осложнения, вероятность развития сопутствующего заболевания.

На Фиг. 4 представлена общая схема вычислительного устройства (400), обеспечивающего обработку данных, необходимую для реализации заявленного решения.

В общем случае устройство (400) содержит такие компоненты, как: один или более процессоров (401), по меньшей мере одну память (402), средство хранения данных (403), интерфейсы ввода/вывода (404), средство В/В (405), средства сетевого взаимодействия (406).

Процессор (401) устройства выполняет основные вычислительные операции, необходимые для функционирования устройства (400) или функциональности одного или более его компонентов. Процессор (401) исполняет



необходимые машиночитаемые команды, содержащиеся в оперативной памяти (402).

Память (402), как правило, выполнена в виде ОЗУ и содержит необходимую программную логику, обеспечивающую требуемый функционал.

5 Средство хранения данных (403) может выполняться в виде HDD, SSD дисков, рейд массива, сетевого хранилища, флэш-памяти, оптических накопителей информации (CD, DVD, MD, Blue-Ray дисков) и т.п. Средство (403) позволяет выполнять долгосрочное хранение различного вида информации.

10 Интерфейсы (404) представляют собой стандартные средства для подключения и работы с серверной частью, например, USB, RS232, RJ45, LPT, COM, HDMI, PS/2, Lightning, FireWire и т.п.

Выбор интерфейсов (404) зависит от конкретного исполнения устройства (400), которое может представлять собой персональный компьютер, мейнфрейм, серверный кластер, тонкий клиент, смартфон, ноутбук и т.п.

15 В качестве средств В/В данных (405) в любом воплощении системы должна использоваться клавиатура. Аппаратное исполнение клавиатуры может быть любым известным: это может быть, как встроенная клавиатура, используемая на ноутбуке или нетбуке, так и обособленное устройство, подключенное к настольному компьютеру, серверу или иному компьютерному устройству. Подключение при этом  
20 может быть, как проводным, при котором соединительный кабель клавиатуры подключен к порту PS/2 или USB, расположенному на системном блоке настольного компьютера, так и беспроводным, при котором клавиатура осуществляет обмен данными по каналу беспроводной связи, например, радиоканалу, с базовой станцией, которая, в свою очередь, непосредственно подключена к системному  
25 блоку, например, к одному из USB-портов. Помимо клавиатуры, в составе средств В/В данных также может использоваться: джойстик, дисплей (сенсорный дисплей), проектор, тачпад, манипулятор мышь, трекбол, световое перо, динамики, микрофон и т.п.

30 Средства сетевого взаимодействия (406) выбираются из устройств, обеспечивающих сетевой прием и передачу данных, например, Ethernet карту, WLAN/Wi-Fi модуль, Bluetooth модуль, BLE модуль, NFC модуль, IrDa, RFID модуль, GSM модем и т.п. С помощью средств (405) обеспечивается организация обмена данными по проводному или беспроводному каналу передачи данных, например, WAN, PAN, ЛВС (LAN), Интранет, Интернет, WLAN, WMAN или GSM, 3G, 4G, 5G.

Компоненты устройства (400) сопряжены посредством общей шины передачи данных (407).

В настоящих материалах заявки представлено предпочтительное раскрытие осуществления 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.

Специалисту в данной области техники должно быть понятно, что различные вариации заявляемого способа и системы не изменяют сущность изобретения, а лишь определяют его конкретные воплощения и применения.

**ФОРМУЛА ИЗОБРЕТЕНИЯ**

1. Компьютерно-реализуемый способ прогнозирования наступления медицинского события в здоровье человека с применением машинного обучения с учетом оценок изменчивости данных о состоянии здоровья человека во времени, в  
5 котором:

извлекают данные о состоянии здоровья человека, содержащие параметры здоровья человека, путем обработки медицинских документов, характеризующих состояние здоровья человека;

10 на основе извлеченных данных о состоянии здоровья человека формируют временные последовательности данных о состоянии здоровья человека; причем каждую временную последовательность получают для значений одного параметра здоровья;

определяют набор параметров здоровья для прогнозирования наступления медицинского события в здоровье человека; причем указанный набор параметров  
15 здоровья зависит от прогнозируемого медицинского события в здоровье человека;

для каждого параметра здоровья из набора параметров для прогнозирования наступления медицинского события в здоровье человека преобразуют сформированную временную последовательность значений параметра здоровья в оценочный показатель динамики изменения значений  
20 параметра здоровья; и получают набор оценочных показателей динамики изменения во времени значений параметров здоровья для прогнозирования наступления медицинского события в здоровье человека;

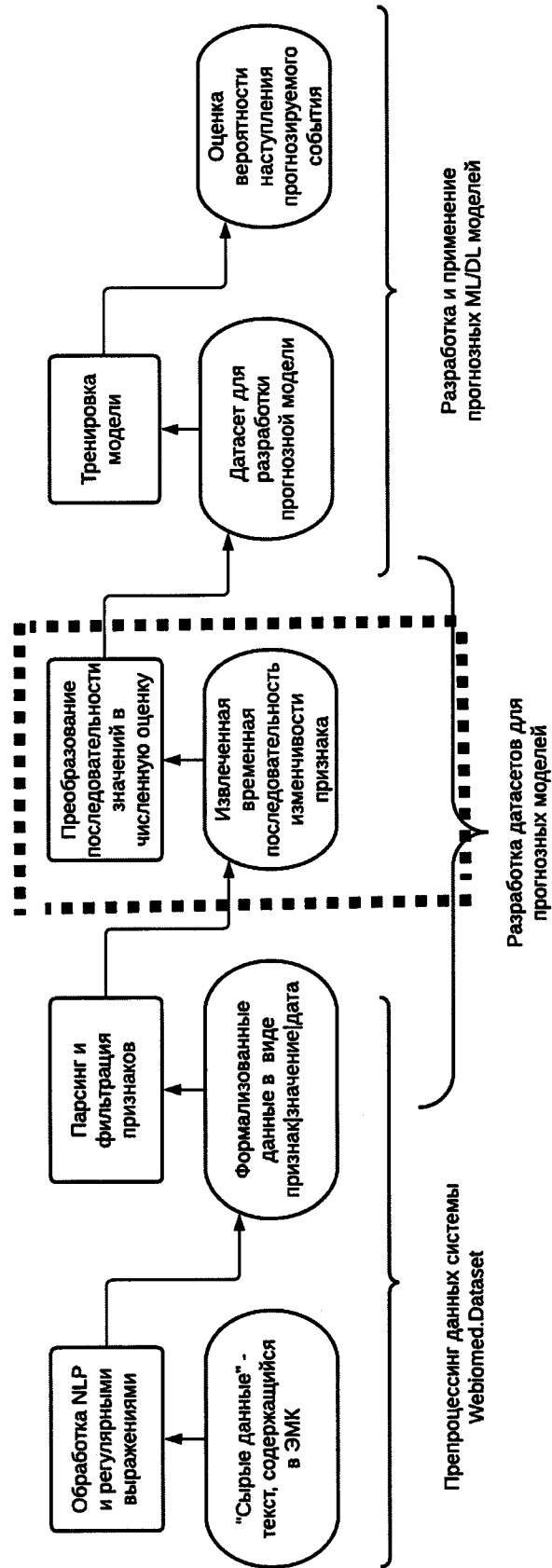
подают на вход прогностической модели машинного обучения полученный набор оценочных показателей динамики изменения во времени значений  
25 параметров здоровья для прогнозирования наступления медицинского события в здоровье человека и на выходе прогностической модели машинного обучения получают вероятность наступления медицинского события в здоровье человека.

2. Способ по п. 1, *характеризующийся тем, что* медицинскими событиями в здоровье человека являются, по меньшей мере, следующие:  
30 госпитализация в стационар, летальный исход, развитие специфического осложнения, развитие сопутствующего заболевания.

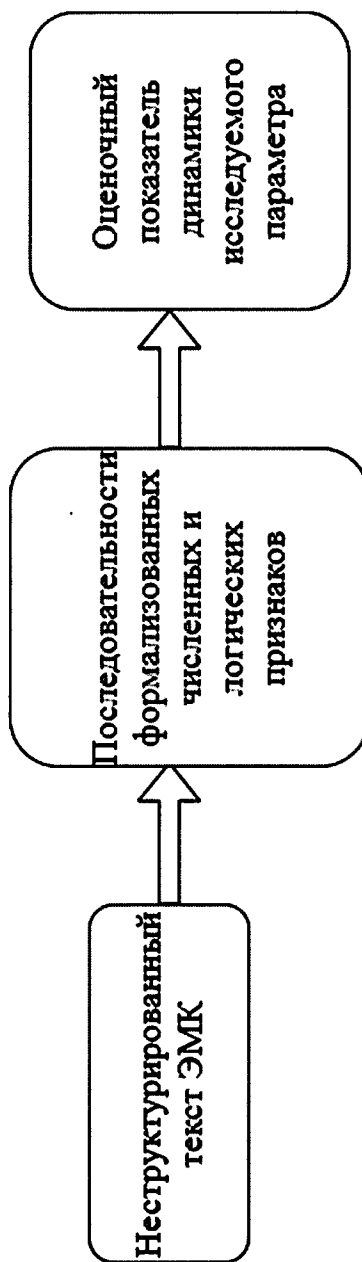
3. Способ по п. 1, *характеризующийся тем, что* параметрами здоровья для прогнозирования наступления медицинского события в здоровье человека являются, по меньшей мере, следующие: физиологические параметры, данные  
35 лабораторных исследований.

4. Способ по п. 1, *характеризующийся тем, что оценочным показателем является индекс долгосрочной изменчивости параметра здоровья человека.*

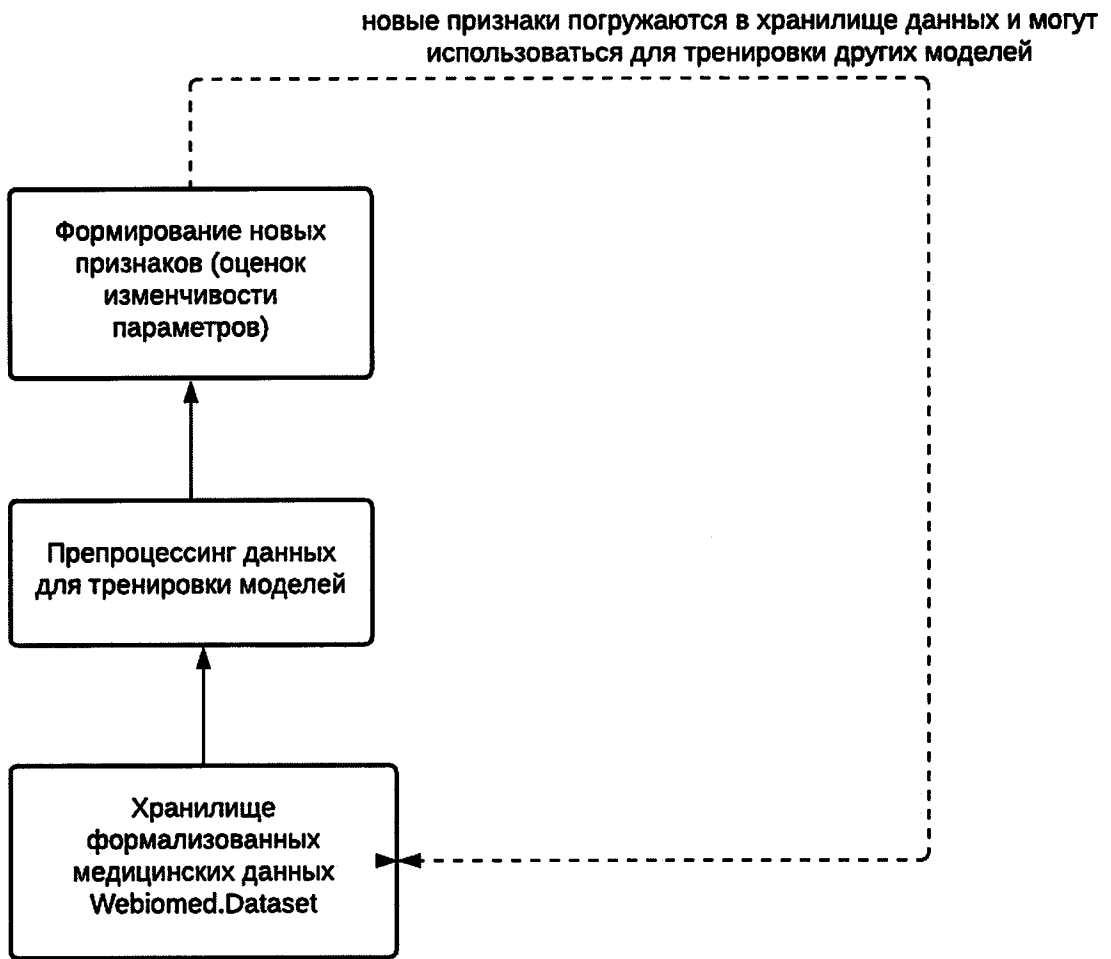
5. Способ по п. 1, *характеризующийся тем, что каждую временную последовательность получают для одного параметра здоровья за заданный период времени.*



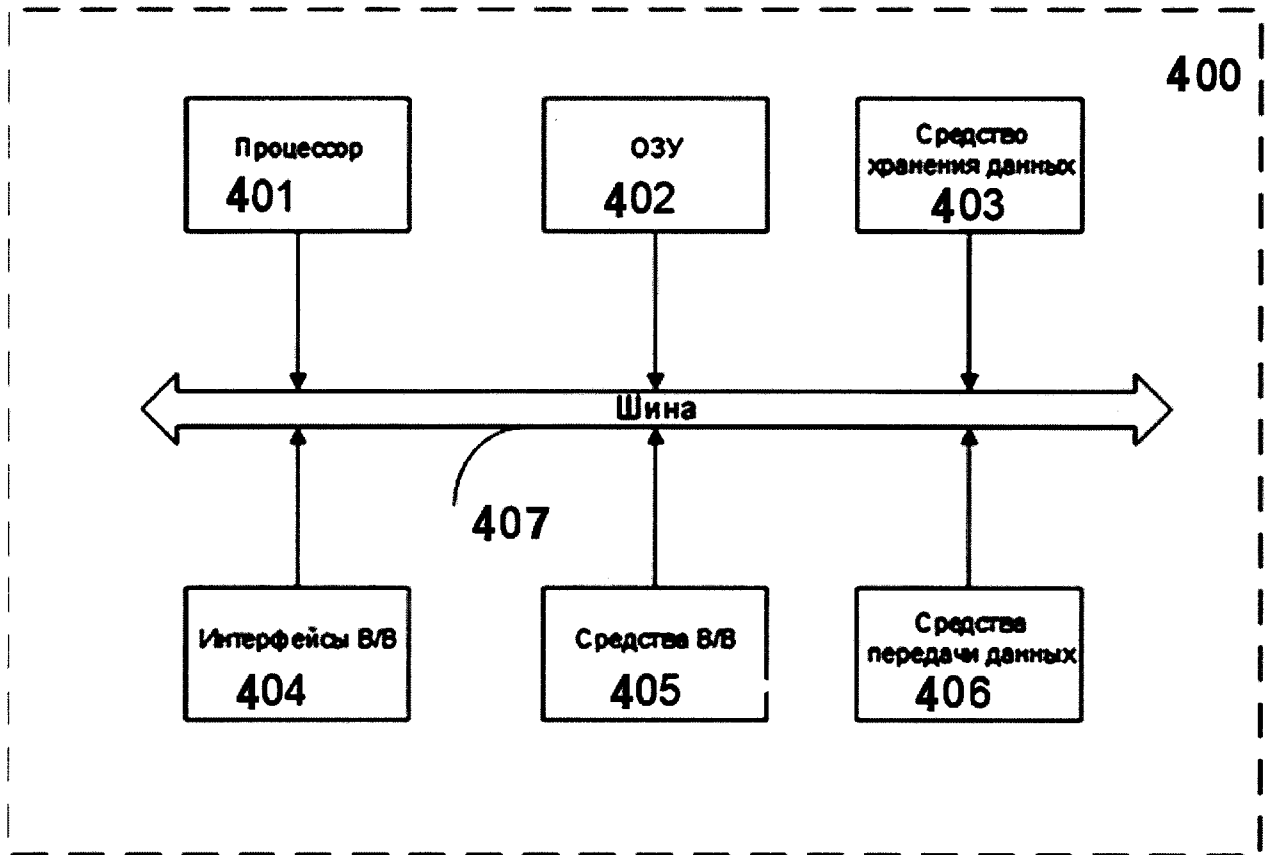
Фиг. 1



Фиг. 2



Фиг. 3



Фиг. 4



## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/RU 2021/000580

A. CLASSIFICATION OF SUBJECT MATTER 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 50/00, 50/30, G06N 20/00 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, Esp@cenet, DWPI, EAPATIS, PATENTSCOPE		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2019/216350 A1 (ZOLL MEDICAL CORP) 18.07.2019, [0005], [0009], [0021], [0029], [0032], [0049], [0111], [0125], [0131], [0135], [0209], [0213], [0258], [0264], [0281], [0302], [0303], [0324], [0327], [0331], [0334], [0355], [0362], [0366], [0373], [0420], [0422], [0512], [0518]	1-5
A	US 2018/301221 A1 (ROTHMAN STEVEN I) 18.10.2018	1-5
A	WO 2021/133935 A1 (KPN INNOVATIONS LLC) 01.07.2021	1-5
A	WO 2021/066918 A1 (DAWNLIGHT TECH INC) 08.04.2021	1-5
<input type="checkbox"/> Further documents are listed in the continuation of Box C. <input type="checkbox"/> See patent family annex.		
* Special categories of cited documents: “A” document defining the general state of the art which is not considered to be of particular relevance “E” earlier application or patent but published on or after the international filing date “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) “O” document referring to an oral disclosure, use, exhibition or other means “P” document published prior to the international filing date but later than the priority date claimed “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 “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 “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 “&” document member of the same patent family		
Date of the actual completion of the international search 24 August 2022 (24.08.2022)		Date of mailing of the international search report 01 September 2022 (01.09.2022)
Name and mailing address of the ISA/RU Facsimile No.		Authorized officer Telephone No.

**ОТЧЕТ О МЕЖДУНАРОДНОМ ПОИСКЕ**

Номер международной заявки

PCT/RU 2021/000580

<p><b>A. КЛАССИФИКАЦИЯ ПРЕДМЕТА ИЗОБРЕТЕНИЯ</b>  <b>G16H 50/30 (2018.01)</b></p> <p>Согласно Международной патентной классификации МПК</p>																	
<p><b>B. ОБЛАСТЬ ПОИСКА</b></p> <p>Проверенный минимум документации (система классификации с индексами классификации)</p> <p>A61B 5/00, G16H 50/00, 50/30, G06N 20/00</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)</p> <p>PatSearch (RUPTO internal), USPTO, PAJ, Esp@cenet, DWPI, EAPATIS, PATENTSCOPE</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>X</td> <td>US 2019/216350 A1 (ZOLL MEDICAL CORP) 18.07.2019, [0005], [0009], [0021], [0029], [0032], [0049], [0111], [0125], [0131], [0135], [0209], [0213], [0258], [0264], [0281], [0302], [0303], [0324], [0327], [0331], [0334], [0355], [0362], [0366], [0373], [0420], [0422], [0512], [0518]</td> <td>1-5</td> </tr> <tr> <td>A</td> <td>US 2018/301221 A1 (ROTHMAN STEVEN I) 18.10.2018</td> <td>1-5</td> </tr> <tr> <td>A</td> <td>WO 2021/133935 A1 (KPN INNOVATIONS LLC) 01.07.2021</td> <td>1-5</td> </tr> <tr> <td>A</td> <td>WO 2021/066918 A1 (DAWNLIGHT TECH INC) 08.04.2021</td> <td>1-5</td> </tr> </tbody> </table>			Категория*	Цитируемые документы с указанием, где это возможно, релевантных частей	Относится к пункту №	X	US 2019/216350 A1 (ZOLL MEDICAL CORP) 18.07.2019, [0005], [0009], [0021], [0029], [0032], [0049], [0111], [0125], [0131], [0135], [0209], [0213], [0258], [0264], [0281], [0302], [0303], [0324], [0327], [0331], [0334], [0355], [0362], [0366], [0373], [0420], [0422], [0512], [0518]	1-5	A	US 2018/301221 A1 (ROTHMAN STEVEN I) 18.10.2018	1-5	A	WO 2021/133935 A1 (KPN INNOVATIONS LLC) 01.07.2021	1-5	A	WO 2021/066918 A1 (DAWNLIGHT TECH INC) 08.04.2021	1-5
Категория*	Цитируемые документы с указанием, где это возможно, релевантных частей	Относится к пункту №															
X	US 2019/216350 A1 (ZOLL MEDICAL CORP) 18.07.2019, [0005], [0009], [0021], [0029], [0032], [0049], [0111], [0125], [0131], [0135], [0209], [0213], [0258], [0264], [0281], [0302], [0303], [0324], [0327], [0331], [0334], [0355], [0362], [0366], [0373], [0420], [0422], [0512], [0518]	1-5															
A	US 2018/301221 A1 (ROTHMAN STEVEN I) 18.10.2018	1-5															
A	WO 2021/133935 A1 (KPN INNOVATIONS LLC) 01.07.2021	1-5															
A	WO 2021/066918 A1 (DAWNLIGHT TECH INC) 08.04.2021	1-5															
<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 ссылочных документов:</p> <p>“А” документ, определяющий общий уровень техники и не считающийся особо релевантным</p> <p>“D” документ, цитируемый заявителем в международной заявке</p> <p>“E” более ранняя заявка или патент, но опубликованная на дату международной подачи или после нее</p> <p>“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)</p> <p>“O” документ, относящийся к устному раскрытию, использованию, экспонированию и т.д.</p> <p>“P” документ, опубликованный до даты международной подачи, но после даты испрашиваемого приоритета</p> <p>“T” более поздний документ, опубликованный после даты международной подачи или приоритета, но приведенный для понимания принципа или теории, на которых основывается изобретение</p> <p>“X” документ, имеющий наиболее близкое отношение к предмету поиска; заявленное изобретение не обладает новизной или изобретательским уровнем, в сравнении с документом, взятым в отдельности</p> <p>“Y” 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</p> <p>“&amp;” документ, являющийся патентом-аналогом</p>																	
<p>Дата действительного завершения международного поиска</p> <p>24 августа 2022 (24.08.2022)</p>		<p>Дата отправки настоящего отчета о международном поиске</p> <p>01 сентября 2022 (01.09.2022)</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:                  Лясовский С.                  Телефон № 8 499 240 25 91</p>															