

(19)



Евразийское
патентное
ведомство

(21) 202392267

(13) A1

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ЕВРАЗИЙСКОЙ ЗАЯВКЕ

(43) Дата публикации заявки
2024.11.13

(51) Int. Cl. G10L 15/01 (2013.01)
G10L 25/00 (2006.01)
G16Y 40/20 (2020.01)

(22) Дата подачи заявки
2023.07.21

(54) СПОСОБ ДИАГНОСТИКИ БОЛЕЗНИ ПАРКИНСОНА В СЕТИ ИНТЕРНЕТ ВЕЩЕЙ, ОСНОВАННЫЙ НА АНАЛИЗЕ ГОЛОСОВЫХ МАРКЕРОВ

(96) 2023/ЕА/0037 (ВУ) 2023.07.21

(72) Изобретатель:

(71) Заявитель:
УЧРЕЖДЕНИЕ ОБРАЗОВАНИЯ
"БЕЛОРУССКИЙ
ГОСУДАРСТВЕННЫЙ
УНИВЕРСИТЕТ ИНФОРМАТИКИ И
РАДИОЭЛЕКТРОНИКИ" (ВУ)

Вишняков Владимир Анатольевич,
Ся Ивей (ВУ)

(57) Способ диагностики болезни Паркинсона в сети интернет вещей, основанный на анализе голосовых маркеров, относится к удаленной диагностике и может быть использован для обнаружения у пациента болезни Паркинсона на основе диагностики голоса при помощи обученной нейронной сети. Экспериментальные результаты показали, что использование модели GRU с 6 слоями привело к точности 86,67% на тестовом наборе данных. Задача настоящего изобретения направлена на оперативность диагностики за счет использования сети ИВ, снижение времени диагностики болезни Паркинсона по сравнению с клиникой и повышение точности классификации моделей за счет нейронной сети. Изобретение относится к способу диагностики пациентов и может быть использовано для обнаружения у пациента болезни Паркинсона при помощи сети интернета вещей на основе голосовой диагностики. Поставленная цель достигается тем, что способ идентификации болезни Паркинсона с использованием GRU нейронной сети и сети интернета вещей для оперативной удаленной диагностики включает последовательное выполнение следующих этапов: удаленная запись на смартфон голосов пациентов с установленной БП; передача этих аудиоданных на сервер сети интернета вещей, обучение на их основе нейронной сети; запись на смартфон голоса пациента с подозрением на наличие БП, передача их на сервер сети интернета вещей, который включает в себя ранее обученную нейронную сеть; предварительная обработка голоса пациента с целью выделения, шумоподавления и сегментации этих данных; распознавание аудиоданных на основе обученной нейронной сети; принятие решения о вероятности наличия БП; направление полученных результатов в приложение смартфона определенным пользователям или медицинским учреждениям.

Структура нейронной сети



A1

202392267

202392267

A1

Способ диагностики болезни Паркинсона в сети интернет вещей, основанный на анализе голосовых маркеров

Изобретение относится к способу диагностики пациентов и может быть использовано для обнаружения у пациента болезни Паркинсона помощи сети интернет вещей на основе голосовой диагностики.

Известно, что наиболее близким предложенному изобретению является способ диагностики болезни Паркинсона на основе динамических характеристик речевого контекста [1]. Он предоставляет возможность использования динамических характеристик речевого контекста для обнаружения болезни Паркинсона (БП) с целью повышения точности методов прогнозирования на основе речи. Инновацией данного способа является предварительная обработка речевых данных, извлечение особенностей произношения и их корректировка, создание модели распознавания речи на основе двунаправленной нейронной сети с использованием долгосрочной и краткосрочной памяти и быстрое обнаружение болезни Паркинсона. Однако недостатком данного изобретения является недостаточная точность, отсутствие удаленной диагностики и вероятность распознавания БП, составляющая 84,29%.

Задачей настоящего изобретения является повышение точности распознавания БП и удаленной диагностики с использованием технологии сети интернет вещей и нейронной сети.

Поставленная цель достигается тем, что способ диагностики болезни Паркинсона основан на анализе голосовых маркеров в сети интернет вещей и заключается в удаленной записи на смартфон голосов пациентов с установленной БП, обучении на их основе нейронной сети. Затем удаленной записи аудиоданных пациентов с подозрением на наличие БП, передаче их по протоколу MQTT на сервер интернета вещей, который включает уже обученную нейронную сеть, проведении на сервере предварительной обработки полученных аудиоданных с целью выделения, шумоподавления и сегментации этих данных, проведении классификации и распознавания предварительно очищенных аудиоданных на ранее обученной нейронной сети и выполнении постановки диагноза о наличии БП, далее направлении полученных результатов диагностики на приложения смартфона определенным пользователям или медицинским учреждениям.

Основанный на нейронной сети способ голосовой диагностики БП включает следующие этапы:

1. Удаленная запись на смартфон голосов пациентов с установленным диагнозом БП.

2. Передача собранных аудиоданных от пациентов с БП по протоколу MQTT как наиболее распространенного, на сервер интернета вещей и обучение на них модели нейронной сети, которая представляет собой комплекс распознавания данных. Нейронная сеть проходит обучение на полученном объеме аудиозаписей голосов пациентов с БП. В процессе обучения нейронная сеть извлекает специфические голосовые признаки, связанные с симптомами БП, такие как дрожь, нестабильность интонации и изменение ритма речи.

3. Запись на смартфон голоса пациента с подозрением на наличие БП. Во время сбора данных микрофон настраивают на частоту 44,1 КГц и затем пациент повторяет гласную / а / для получения устойчивого ее произношения.

4. Программное обеспечение передает данные по протоколу MQTT. Через это программное обеспечение смартфон отправляет голосовые данные на сервер. Далее сервер использует обученную модель нейронной сети для обработки голосовых данных, отправленных со смартфона.

5. Для эффективной обработки полученных аудиоданных применяются несколько этапов. Среди них — предварительное усиление (выделение), разбиение голосового сигнала (сегментация) с использованием скользящего фиксированного окна и удаление шума.

Процесс предусиления помогает усилить важные частотные компоненты в аудиосигнале. Техники снижения шума применяются для минимизации нежелательного фонового шума, обеспечивая более ясный и точный анализ нужного аудио контента.

Разбиение голосового сигнала с использованием скользящего фиксированного окна предполагает разделение аудиоданных на более мелкие перекрывающиеся сегменты для дальнейшего анализа. Это позволяет извлекать и выбирать конкретные особенности в каждом сегменте, облегчая более точный и детальный анализ аудиоданных.

Благодаря реализации этих шагов система может эффективно предварительно обрабатывать полученные аудиоданные, усиливая важные компоненты, снижая помехи шума и сегментируя данные для извлечения и выбора особенностей.

Для предварительного выделения речевых данных используется КИХ-цифровой фильтр (цифровой фильтр с конечной импульсной характеристикой) верхних частот (ВЧ) первого порядка. Основной целью предварительного усиления является усиление энергии высокочастотной составляющей речевого сигнала для компенсации потерь в высокочастотной

области и сохранения информации звукового канала. Потери в диапазоне ВЧ могут возникать во время передачи или записи аудиосигналов из-за характеристик передающей среды и ограничений оборудования записи. Предположим, что значение n -й точки дискретизации входного сигнала равно $x[n]$, тогда уравнение предварительного усиления может быть записано следующим образом:

$$y(n) = x(n) - ax(n - 1) \quad (1)$$

где $y(n)$ обозначает выходной сигнал, $x(n)$ обозначает входной сигнал, а n обозначает индекс дискретного времени. Параметр a управляет характеристиками фильтра. В случае цифрового КИХ-фильтра высоких частот первого порядка параметр a вычисляется следующим образом.

$$a = \frac{1 - \frac{2\pi f_c}{f_s}}{1 + \frac{2\pi f_c}{f_s}} \quad (2)$$

где f_c обозначает частоту среза, а f_s - частоту дискретизации. Подставляя частоту дискретизации 44,1 Гц и частоту среза 300 Гц в уравнение, $a=0,9932$. Передаточная функция $H(z)$ цифрового КИХ-фильтра высоких частот первого порядка имеет вид.

$$H(z) = 1 - 0.9932z^{-1} \quad (3)$$

Аудиосигнал делится на несколько коротких сегментов для обработки, каждый короткий сегмент называется кадром, а затем к каждому кадру добавляется оконная функция. Преимущество использования оконной функции заключается в том, что она обеспечивает плавный переход от кадра к кадру. Однако использование оконной функции сопряжено с некоторыми издержками, т.е. начало и конец кадра ослабляются. Поэтому процесс разделения кадров требует некоторого перекрытия между соседними кадрами.

В изобретении используется функция окна Хэмминга, размер окна составляет 1024, частота речевых данных составляет 44,1 кГц, коэффициент перекрытия окна составляет 50%, а время речи в окне составляет около 23 мс.

Для уменьшения шума речи используется алгоритм спектрального вычитания, для преобразования голосовых данных из временной области в частотную область используется кратковременное преобразование Фурье. Для каждой полосы частот вычисляется отношение спектра мощности сигнала к спектру мощности шума, и если это отношение ниже определенного порога, сигнал в этой полосе частот устанавливается равным нулю следующим образом:

$$|s_i(\omega)| = |x_i(\omega)| - \mu = \begin{cases} |x_i(\omega)| - \mu, & |x_i(\omega)| > \mu \\ 0, & |x_i(\omega)| \leq \mu \end{cases} \quad (4)$$

где $s_i(\omega)$ – представление спектра речи в частотной области на i -м кадре, ω обозначает индекс частоты;

$x_i(\omega)$ – представление спектра зашумленной речи в частотной области на i -м кадре, ω обозначает индекс частоты;

μ – спектральный порог, который используется для определения, какие части спектра считаются сигналом, а какие шумом.

Если величина $x_i(\omega)$ меньше или равна μ , она рассматривается как шум и удаляется из спектра речи с шумоподавлением.

Предположим, что исходный сигнал $x(k) = n(k) + s(k)$, где $n(k)$ – шумовой сигнал, а $s(k)$ – речевой сигнал. Выбирается исходный сигнал $x(k)$, не содержащий речевого сигнала $s(k)$, в качестве шумового сигнала $n(k)$, то есть выбирается только отрезок с шумом, не содержащий речевую информацию. Ниже приведена математическая формула для параметра μ :

– Формирование фрейма: шумовой сигнал $n(k)$ кадрируется так же, как и исходный сигнал. Шумовой сигнал $n(k)$ делится на перекрывающиеся кадры, каждый из которых имеет длину N точек выборки, с использованием оконной функции со сдвигом кадра на M точек выборки для кадрирования. Индекс начальной точки отсчета первого кадра равен iM , а индекс конечной точки отсчета равен $iM+N-1$.

$$n_{i(n)} = n(iM + n) \quad (5)$$

где $n_{i(n)}$ обозначает n -ю точку выборки i -го кадра шумового сигнала.

– добавление окна: для каждого кадра $n_{i(n)}$ умножается оконная функция $W(n)$.

$$n_{iW(n)} = n_{i(n)} * W(n) \quad (6)$$

где $n_{iw(n)}$ обозначает n -ю точку выборки i -го кадра зашумленного сигнала после применения окна.

– быстрое преобразование Фурье (БПФ): БПФ выполняется для каждого кадра $n_{iw(n)}$ после добавления окна, чтобы получить представление шума в частотной области.

$$N_{i(w)} = FFT(n_{iw(n)}) \quad (7)$$

где $N_{i(w)}$ обозначает представление i -го кадра шума в частотной области.

– расчет среднего значения: представление частотной области всех кадров шума усредняется.

$$\mu = \left(\frac{1}{I}\right) * \sum_{i=1}^{i=I} N_{i(w)} \quad (8)$$

где $N(w)$ обозначает среднее представление шума в частотной области, а I обозначает общее количество кадров шума. Поскольку окружение каждого человека, при сборе его речи, является уникальным, значение μ для каждого человека будет отличаться.

6. Извлечение особенностей голосового сигнала из каждого кадра. Признаками изобретения для выделения признаков голосовых данных, в частности, являются: 21 базовый признак, 11 частотно-временных признаков, 84 признака спектрального коэффициента Мейера, 182 признака на основе вейвлет-преобразования, 22 признака складки голоса и 432 признака вейвлет-преобразования с регулируемой добротностью [2]. Набор признаков данных был нормализован, и был использован алгоритм PCA (анализ основных компонентов). Алгоритм PCA преобразует набор данных из 752 функций в 75 путем вычисления ковариационной матрицы, разложения собственных значений и выбора основных компонентов. Это позволяет уменьшить размерность данных и сократить количество признаков, сохранив при этом наиболее важную информацию в наборе данных.

7. Для оценки и оптимизации работы модели GRU (Gated Recurrent Unit – рекуррентная сеть с воротными блоками) набор данных признаков был разделен на обучающий и тестовый в соотношении 9:1. Обучающий набор данных подавался в модель GRU в течение 1000 тренировочных сессий для оптимизации параметров и весов модели. Для более точной оценки эффективности модели использовался метод 5-кратной кросс-валидации.

После того как оптимизированная 6-слойная модель GRU была обучена, ее эффективность была проверена на тестовом наборе данных.

8. Классификация и распознавание предварительно очищенных аудиоданных ранее обученной нейронной сетью, которая классифицирует их как симптомы БП или нет, основываясь на обученных моделях и алгоритмах, и выполнение постановки диагноза о наличии БП. Как показано на рисунке 1 и в таблице 1, конкретный процесс классификации включает:

Размерность входной последовательности данных может быть представлена как «С × В × Т», где:

С – признаков, указывающих размерность признаков на каждом временном шаге входной последовательности.

В – размер пакета, то есть количество одновременно обрабатываемых примеров.

Т – количество временных шагов, то есть длина или временная размерность входной последовательности.

Последовательность входных данных имеет размерность 75 (С) × 1 (В) × 1 (Т).

В данном случае 75(С) × 1(В) × 1(Т) указывает на то, что входная последовательность имеет 75 признаков (каналов), размер пакета равен 1, а количество временных шагов – 1. Это означает, что на каждый образец приходится только один временной шаг, а размерность входной последовательности составляет 75. Это подразумевает, что каждый образец в наборе данных состоит из 75 точек каналов или признаков.

Таблица I. Активации и обучаемые свойства нейронных сетей GRU

Название	Активации	Обучаемые свойства
Вход последовательности	75(С) × 1(В) × 1(Т)	
Слой GRU	6(С) × 1(В)	Веса входа 18 × 75 Рекуррентные веса 18 × 6 Смещение 18 × 1
Слой RELU	6(С) × 1(В)	
Полносвязный слой	2(С) × 1(В)	Веса 2 × 1 Смещение 18 × 1
Слой Softmax	2(С) × 1(В)	
Выходной слой	2(С) × 1(В)	

Как показано на рисунке 2, GRU является моделью глубокого обучения, используемой для анализа последовательных данных, таких как тексты или речевые сигналы. Она представляет собой модифицированную версию рекуррентной нейронной сети (RNN), в которой применяются механизмы ворот для более эффективной обработки и управления информацией, который имеет элемент активации 6(С) × 1(В), имеет атрибуты

InputWeights 18×75 , RecurrentWeights 18×6 , Bias 18×1 . Каждый блок GRU имеет элемент обновления и элемент сброса, и информация обрабатывается следующим образом:

$$r_t = \sigma(W_r \times [h_{t-1}, x_t]) \quad (9)$$

$$z_t = \sigma(W_z \times [h_{t-1}, x_t]) \quad (10)$$

$$\tilde{h}_t = \tanh(W_{\tilde{h}} \times [r_t \times h_{t-1}, x_t]) \quad (11)$$

$$h_t = (1 - z_t) \times h_{t-1} + z_t \times \tilde{h}_t \quad (12)$$

$$y_t = \sigma(W_\sigma \times h_t) \quad (13)$$

где σ – представляет функцию sigmoid,

z_t – элемент обновления модуля.

Функция sigmoid сводит значение элемента обновления к 0 или 1, т.е. независимо от того, запоминается значение предыдущего шага или отбрасывается. r_t – элемент сброса, чем меньше r_t , тем больше информации о предыдущем состоянии необходимо игнорировать,

W_σ – весовая матрица,

h_t, h_{t-1} – выходные и временные скрытые состояния в модуле.

Уровень GRU является одним из основных компонентов модели, способным кодировать последовательную информацию и фильтровать важную информацию, чтобы играть определенную роль в управлении потоком информации.

– Слой активации ReLU с $6(C) \times 1(B)$ элементами активации, который действует как нелинейный слой, помогает модели лучше соответствовать нелинейным данным, сокращает ненужные вычисления за счет увеличения разреженности модели и дополнительно предотвращает градиентных взрывов. Во время обратного распространения ошибки в сети значения градиента становятся очень большими, что приводит к слишком большому количеству обновлений весов и мешает сети сойтись. Проблему взрыва градиента можно эффективно снизить, добавив слой ReLU в нейронную сеть. При прохождении градиента через слой ReLU отрицательный градиент будет усечен до нуля, а положительные интервалы будут иметь постоянный градиент 1. Это помогает ограничить размер градиента и предотвратить его слишком большое значение.

– Слой Softmax с активирующими элементами размером $2(C) \times 1(B)$ отображает выходные данные полностью подключенного слоя на распределение вероятностей и выбирает категорию с наибольшей вероятностью в качестве выходного результата модели.

– Выходной слой с элементами активации $2(C) \times 1(B)$ и атрибутами весов 2×6 и смещения 2×1 может управлять выходным размером модели в полностью подключенном слое, и обычно размер соответствует количеству выходных категорий модели, которая используется для генерации результатов классификации в классификации.

– Уровень вывода классификации с элементами активации размером $2(C) \times 1(B)$ отображает выходные данные модели в фиксированное пространство выходных категорий, т.е. больные или здоровые, принадлежащие к предопределенному, соответствующему количеству категорий в задаче классификации.

– Пакетный градиентный спуск (BGD) используется для ускорения сходимости модели и выполнения обновления весов. Каждый пакет BGD содержит 64 образца для оптимизации параметров модели.

9. Направление полученных результатов диагностики в приложение по протоколу MQTT смартфона определенным пользователям или медицинским учреждениям для принятия окончательного решения о дальнейшем подтверждении БП.

Сбор данных осуществляется так же, как и в прототипе. Согласно сравнительному анализу с прототипом, было обнаружено, что при использовании 6-слойной модели GRU точность на тестовом наборе данных составила 86,67%, что на 2,41% превышает точность прототипа.

Таким образом, в предлагаемом способе обеспечивается минимизация временных затрат и повышение точности распознавания БП за счет удаленной диагностики при использовании нейронной сети в интернете вещей.

Источники информации:

1. Патент CN112820279, G10L15/063. Способ обнаружения болезни Паркинсона на основе динамических характеристик контекста голоса Цюань Чанцинъ, Луо Чживэй Жэнь Кан, Лин Юнь Чэнь. Код доступа: https://www-patentguru-com.translate.google.com/inventor/%E9%99%88%E4%BB%B2%E7%95%A5?_x_tr_sl=zh-CN&_x_tr_tl=ru&_x_tr_hl=ru&_x_tr_pto=sc Чжунлуэ / (Китай). – № CN202110270049.9А; заявл. 12.03.202.; опубл. 18.05.2021.

2. Sakar, C.O., Serbes, G., Gunduz, A., Tunc, H.C., Nizam, H., Sakar, B.E., Tutuncu, M., Aydin, T., Isenkul, M.E. and Apaydin, H., 2018. A comparative analysis of voice data processing algorithms for Parkinsonâ€™s disease classification and the use of the tunable Q-factor wavelet transform.

Формула

1. Способ диагностики болезни Паркинсона, основанный на анализе голосовых маркеров в сети интернет вещей, заключающийся в предварительной записи на смартфон голосов пациентов с установленной болезнью Паркинсона и обучении на их основе нейронной сети, удаленной записи голоса пациента с подозрением на наличие болезни Паркинсона отличающийся тем, что собранные аудиоданные пациента с подозрением на наличие болезни Паркинсона передают по протоколу MQTT на сервер интернета вещей с целью выделения, шумоподавления и сегментации этих данных, проводят классификацию и распознавание очищенных аудиоданных и выполняют постановку диагноза о наличии БП, далее направляют полученные результаты диагностики в приложения смартфона определенным пользователям или медицинским учреждениям.

2. Способ диагностики болезни Паркинсона, основанный на анализе голосовых маркеров в сети интернет вещей по пункту 1, отличающийся тем, что предварительную обработку полученных аудиоданных проводят при помощи выделения признаков голосовых данных, далее при помощи алгоритма PCA сокращают количество признаков и передают их на модель нейронной сети с шестью слоями GRU для диагностики болезни Паркинсона.

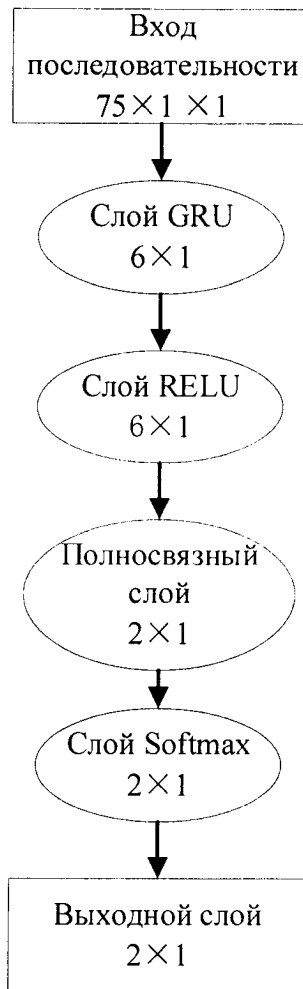


Рисунок 1 Структура нейронной сети

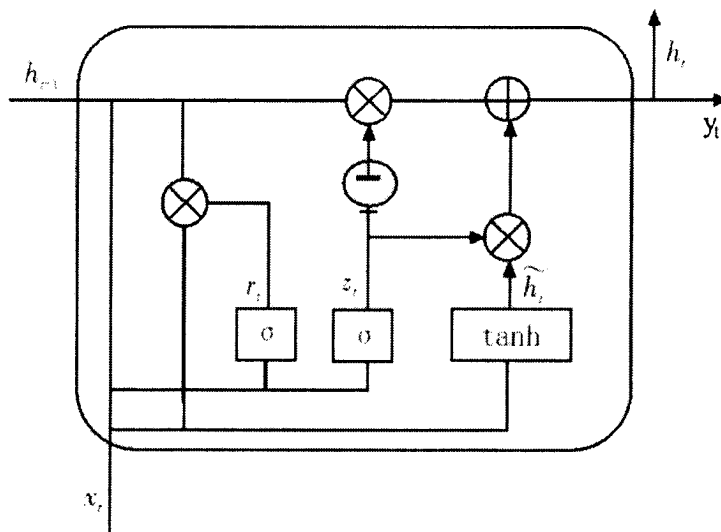


Рисунок 2 Структура одной модели GRU

ОТЧЕТ О ПАТЕНТНОМ ПОИСКЕ

(статья 15(3) ЕАПК и правило 42 Патентной инструкции к ЕАПК)

Номер евразийской заявки:

202392267**А. КЛАССИФИКАЦИЯ ПРЕДМЕТА ИЗОБРЕТЕНИЯ:**

МПК:

G10L 15/01 (2013.01)
G10L 25/00 (2006.01)
G16Y 40/20 (2020.01)

СПК:

G10L 15/063
G10L 25/30
G16Y 40/20

Согласно Международной патентной классификации (МПК)

Б. ОБЛАСТЬ ПОИСКА:

Просмотренная документация (система классификации и индексы МПК)

G10L 15/00-15/01, G10L 25/00, G06F 15/00-15/16, A61B 5/00-5/369, G06Q 50/00

Электронная база данных, использовавшаяся при поиске (название базы и, если возможно, используемые поисковые термины)
 Espacenet, EAPATIS, GOOGLE, PAJ, WIPO, «ПОИСКОВАЯ ПЛАТФОРМА» (РОСПАТЕНТ)

В. ДОКУМЕНТЫ, СЧИТАЮЩИЕСЯ РЕЛЕВАНТНЫМИ

Категория*	Ссылки на документы с указанием, где это возможно, релевантных частей	Относится к пункту №
X	RAZA M. et al, "Intelligent Framework for Indoor Healthcare Monitoring of Parkinson's Disease Patient", IEEE Journal on Selected Areas in Communications, vol. 39, issue 2, February 2021, 9 л., [онлайн] [найдено 17.11.2023]. Найдено в < https://ieeexplore.ieee.org/document/9186157 > реферат, разделы I-IV	1, 2
A	CARRON J. et al, "A mobile-assisted voice condition analysis system for Parkinson's disease: assessment of usability conditions", BIOMEDICAL ENGINEERING ONLINE, 2021, 24 л., [онлайн] [найдено 17.11.2023]. Найдено в < https://biomedical-engineering-online.biomedcentral.com/articles/10.1186/s12938-021-00951-y > ,	1, 2
A	US20220240846 A1 (HOWARD N.) 04.08.2022,	1, 2
A	US2020/375544 A1 (CELLOSCOPE LTD.) 03.12.2020	1, 2

 последующие документы указаны в продолжении

* Особые категории ссылочных документов:

«А» - документ, определяющий общий уровень техники

«D» - документ, приведенный в евразийской заявке

«E» - более ранний документ, но опубликованный на дату подачи евразийской заявки или после нее

«O» - документ, относящийся к устному раскрытию, экспонированию и т.д.

"P" - документ, опубликованный до даты подачи евразийской заявки, но после даты испрашиваемого приоритета"

«Т» - более поздний документ, опубликованный после даты приоритета и приведенный для понимания изобретения

«X» - документ, имеющий наиболее близкое отношение к предмету поиска, порочащий новизну или изобретательский уровень, взятый в отдельности

«Y» - документ, имеющий наиболее близкое отношение к предмету поиска, порочащий изобретательский уровень в сочетании с другими документами той же категории

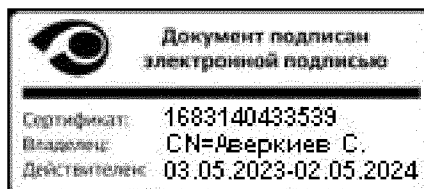
«&» - документ, являющийся патентом-аналогом

«L» - документ, приведенный в других целях

Дата проведения патентного поиска: 17 ноября 2023 (17.11.2023)

Уполномоченное лицо:

Начальник Управления экспертизы



С.Е. Аверкиев