

УДК 519.683

DOI: [10.26102/2310-6018/2021.34.3.008](https://doi.org/10.26102/2310-6018/2021.34.3.008)

Программное приложение для предсказания здоровья ребенка, рожденного при помощи вспомогательных репродуктивных технологий, по анамнезу матери

С.Л. Синотова¹, О.В. Лимановская², А.Н. Плаксина³, В.А. Макутина⁴

^{1,2}ФГАОУ ВО «УрФУ имени первого Президента России Б.Н. Ельцина»,
Екатеринбург, Российская Федерация

³ФГБОУ ВО «Уральский государственный медицинский университет Министерства здравоохранения Российской Федерации», Екатеринбург, Российская Федерация

⁴АО «Центр семейной медицины», Екатеринбург, Российская Федерация

Резюме. Вспомогательные репродуктивные технологии (ВРТ) много лет помогают обрести ребенка при невозможности зачатия естественным путем. Успешным протокол ВРТ можно считать не только при наступлении беременности, но и при успешном ее завершении – рождении здорового ребенка. В статье описано создание программного приложения для сотрудников центров ВРТ, помогающее сделать прогноз результата протокола, включающий в себя вероятность наступления беременности, прогноз возможных осложнений при ее течении, прогноз срока и способа родоразрешения, а также группы здоровья (1-5) рожденного ребенка. Для создания приложения использовались данные о 854 протоколах, осуществленных в 2016-2018 годах, в результате которых родилось 464 ребенка. Анализ их здоровья содержит информацию в возрасте от рождения до трех лет. Приложение использует шестнадцать бинарных классификаторов, девять из которых реализуют многоклассовые классификации срока родов, способа родоразрешения, группы здоровья ребенка. Для реализации мультиклассового вывода использовалась стратегия «один против всех». Для проверки качества использовалась кросс-валидация. Остальные 7 классификаторов предсказывают вероятность наступления беременности и возникновение ее осложнений: истмико-цервикальная недостаточность, гипертонические расстройства, предлежание плаценты, гестационный сахарный диабет, нарушения количества околоплодных вод и преждевременный разрыв плодных оболочек. Все модели построены на языке python на основе случайного леса. Интерфейс создан при помощи библиотек PyQt5 и QtDesigner.

Ключевые слова: машинное обучение, вспомогательные репродуктивные технологии, экспертная система, программное приложение, предсказание здоровья ребенка

Для цитирования: С.Л. Синотова, О.В. Лимановская, А.Н. Плаксина, В.А. Макутина. Программное приложение для предсказания здоровья ребенка, рожденного при помощи вспомогательных репродуктивных технологий, по анамнезу матери. *Моделирование, оптимизация и информационные технологии.* 2021;9(3). Доступно по: <https://moitvivr.ru/ru/journal/pdf?id=1021> DOI: 10.26102/2310-6018/2021.34.3.008

Software application for predicting the health status of a child born with the use of assisted reproductive technologies, according to the mother anamnesis

S.L. Sinotova¹, O.V. Limanovskaya², A.N. Plaksina³, V.A. Makutina⁴

^{1,2} FSAEI HE «UrFU named after the first President of Russia B.N. Yeltsin», Ekaterinburg,
Russian Federation

³ FSBEI HE «USMU of the Ministry of Health of the Russian Federation», Ekaterinburg,
Russian Federation

⁴The Family Medicine Centre, Ekaterinburg, Russian Federation

Abstract: For many years, assisted reproductive technologies (ART) have been helping to conceive a child when this is not possible naturally. We can consider the ART protocol to be successful not only upon pregnancy but also in the case of its successful completion: the birth of a healthy child. The article describes the creation of a software application for employees of ART centers, which helps to predict the outcome of the protocol, including the probability of pregnancy, the forecast of possible complications during its course, the forecast of the time and method of delivery, and the health group (1-5) of the born child. To create the application, we used data on 854 protocols implemented in 2016-2018, because of which 464 children were born. The analysis of their health contains information from birth to three years of age. The application uses sixteen binary classifiers, nine of which implement multiclass classifications of the term of delivery, the delivery method, and children's health groups. The one versus all strategy was used to implement multiclass withdrawal. Cross-validation was used to check the quality. The remaining seven classifiers predict the likelihood of pregnancy and the occurrence of its complications: cervical incompetence, hypertensive disorders, placenta previa, gestational diabetes mellitus, violations of the amount of amniotic fluid, and premature rupture of the membranes. We have built all the models based on the random forest algorithm using the Python programming language. The interface was created using the PyQt5 and QtDesigner libraries.

Keywords: machine learning, assisted reproductive technologies, expert system, software application, child health status prediction

For citation: S.L. Sinotova, O.V. Limanovskaya, A.N. Plaksina, V.A. Makutina. Software application for predicting the health status of a child born with the use of assisted reproductive technologies, according to the mother's anamnesis. *Modeling, Optimization and Information Technology*. 2021;9(3). Available from: <https://moitvvt.ru/ru/journal/pdf?id=1021> DOI: 10.26102/2310-6018/2021.34.3.008 (In Russ).

Введение

В течение многих лет вспомогательные репродуктивные технологии (ВРТ) помогают зачать ребенка, когда это невозможно естественным путем. Рождение ребенка и его здоровье определяют успешность протокола ВРТ для пациентов. В настоящий момент ведется интенсивная работа по созданию моделей машинного обучения, позволяющих прогнозировать живорождения в результате ВРТ [1, 2, 3]. Многие статьи посвящены анализу состояния здоровья детей, рожденных после применения ВРТ [4, 5, 6].

Например, прогнозирование живорождений с использованием модели многомерной логистической регрессии представлено в [4]. Модель построена на 8 предикторах, метрика ROC-AUC составляет 0,62. Авторы [7] построили модель XGBoost для прогнозирования рождения живого ребенка в результате ЭКО. Модель также использует 8 предикторов и имеет средний балл точности $0,70 \pm 0,003$. Однако авторы данных работ не прогнозируют возможные осложнения течения беременности и состояние здоровья ребенка.

Модели машинного обучения используются для прогнозирования различных проблем со здоровьем ребенка, но они не прослеживают связь с историей матери и течением беременности [8, 9, 10]. Нет также и работ, которые представляли бы модели, сочетающие задачу прогнозирования как рождения ребенка по протоколу ВРТ, так и состояния его здоровья на основе анализа данных анамнеза матери.

В данной статье представлена модель предсказания вероятности рождения ребенка в результате протокола ВРТ и состояния его здоровья, построенная на основе анализа данных анамнеза женщин, течения и завершения их беременностей и информации о здоровье рожденных детей до трех лет. На основе данной системы создано программное приложение для практического использования.

Материалы и методы

Для работы были использованы данные из Автоматизированной системы «Региональный акушерский мониторинг (АС «РАМ») (свидетельство о государственной программе для ЭВМ № 2017619189 от 16.08.2017), включающие в себя данные о 854 протоколах, осуществленных в 2016-2018 годах, в результате которых родилось 464 ребенка. Здоровье детей описано в возрасте от рождения до трех лет в виде имеющихся диагнозов, описанных по МКБ-10.

Приложение представляет собой совокупность бинарных классификаторов, построенных с помощью RandomForestClassifier из библиотеки Scikit-Learn [11, 12]. Эта модель из-за небольшого размера выборок и лучших результатов по сравнению с логистической регрессией.

Для выделения из текстового описания течения беременности и анамнеза женщины признаков, пригодных для построения моделей, была использована его предобработка, включающая в себя применение морфологического анализатора rymorphy2 [13] и классов CountVectorizer, TfidfVectorizer библиотеки Scikit-Learn [14, 15], кластеризацию данных методом k-средних (KMeans) [16, 17]. Значимые факторы были выбраны из анамнеза женщин при помощи алгоритма Boruta [18, 19] и представлены в виде 48 бинарных переменных.

Балансировка данных выполнялась при помощи алгоритма ADASYN [20, 21]. Данный способ балансировки основан на подходе over-sampling [21], который заключается в произвольном дублировании и добавлении в выборку данных миноритарного класса. Алгоритм дополняет выборку синтетическими объектами, созданными на основе имеющихся. Отличие ADASYN от других подобных алгоритмов заключается в генерации большего набора данных для тех объектов, которые сложнее прогнозировать, чем для тех, которые прогнозируются легче. ADASYN фокусируется на генерации объектов рядом с исходными, которые ошибочно классифицированы с использованием kNN классификатора. В результате такого подхода ADASYN улучшает данные за счет сокращения смещения, вносимого дисбалансом классов и дает классификатору больше наблюдений, помогающих настроиться на сложные случаи [21].

Качество работы классификаторов измерено на тестовых данных и оценено с помощью метрик долей правильных ответов (точность) и площадь под ROC-кривой – AUC (для бинарных классификаторов).

Доля правильных ответов или правильность – метрика, дающая значение процента или доли верно предсказанных классов. В целом, если классы сбалансированы, то значение метрики можно применять для оценки бинарного классификатора, однако, лучше не ориентироваться только на нее. При несбалансированных данных значение правильности может вводить в заблуждение: например, если в данных содержится 5 % примеров отрицательного класса, то классификатор, всегда предсказывающий положительный класс будет иметь точность 95 %. Однако это не говорит о его хорошем качестве. Данную метрику удобно использовать для оценки качества многоклассового классификатора, предварительно убедившись в адекватности каждой модели в его составе.

AUC-ROC (area under ROC curve) – площадь под кривой ошибок. ROC-кривая показывает долю ложно положительных примеров (false positive rate) – объектов, которые ошибочно относятся к первому классу в сравнении с долей истинно положительных примеров от общего количества положительных примеров (true positive rate). Метрика AUC позволяет одним числом охарактеризовать ROC-кривую и является более целесообразной, чем точность при решении задач несбалансированной классификации. Значение для бинарного классификатора, делающего предсказания

наугад, стремится к 0.5, в то время как AUC идеально предсказывающей модели равна единице.

Оценка значения метрик получена на тестовых данных с помощью использования перекрестной проверки [22, 23]. Это более надежный способ оценки обобщающей способности модели (качества прогнозирования на данных, не участвовавших в обучении), чем произвольное разбиение на тестовую и обучающую выборки. При произвольном разбиении существует вероятность попадания всех трудно классифицируемых (или легко классифицируемых) объектов в одну часть данных, что сделает оценку качества необоснованно высокой или низкой. При использовании перекрестной проверки все объекты данных попадут и в тестовую часть, и в обучающую. Следовательно, для получения высокого качества метрики, обобщающая способность моделей должна быть также высокой.

Способ предполагает разбиение данных некоторое число раз и построение такого же числа моделей. В работе была использована кросс-валидация на 5 блоков (фолдов) с помощью стратегии StratifiedKFold, реализованной в библиотеке sklearn [24]. Стратегия заключается в том, что сначала данные разбиваются на пять примерно одинаковых по размеру частей, сохраняя соотношения классов в каждом блоке примерно таким, как в исходной выборке. Далее строится 5 моделей: первая модель использует блоки 2-5 для обучения, а блок 1 – для тестирования, вторая – блоки 1, 3-5 для обучения и 2 – для тестирования и так далее. В конечном итоге получают 5 оценок качества для обучающей и тестовой выборок. Итоговая оценка – среднее значение по выборке с учетом стандартного отклонения.

Интерфейс программы был создан при помощи PyQt5 (версия 5.9.2) – набора расширений для создания графического интерфейса на базе Qt5 для языка программирования Python [25] и Qt Designer – инструмент для быстрого создания графических пользовательских интерфейсов с виджетами из среды Qt GUI [26]. Предварительное сохранением моделей выполнено с помощью библиотеки pickle [27].

Результаты

В составе приложения имеется 16 бинарных классификаторов, девять из которых реализуют многоклассовые классификации срока родов, способа родоразрешения, группы здоровья ребенка. Все классификаторы имеют параметры, представленные в Таблице 1.

Первый бинарный классификатор (модель 1) – предсказание наступления беременности в процентах. Для его построения используются данные всех протоколов без их дублирования.

Следующие 4 классификатора (модели 2-5) предсказывают наступление возможных осложнений в течении беременности: истмико-цервикальную недостаточность, гипертензивные расстройства, предлежание плаценты и гестационный сахарный диабет. Как и модель предсказания беременности, они строятся на значимых факторах, выбранных из анамнеза женщин.

Далее реализованы классификаторы, предсказывающие нарушения количества околоплодных вод и преждевременный разрыв плодных оболочек (модели 6-7). Они построены не только с учетом анамнеза, но и результатов работы моделей 2-5.

За предсказание способа родоразрешения отвечают модели 8-10, представляющие собой реализацию «One vs. all» [28] многоклассового классификатора, содержащего классы «плановое кесарево сечение», «экстренное кесарево сечение», «самостоятельные роды». Для предсказания срока родов используются модели 11-13: «прерывание беременности», «преждевременные роды», «срочные роды», которые реализуют второй

многоклассовый классификатор системы. Построение моделей 8-13 основано на значимых факторах анамнеза и результатах работы моделей 2-7.

Таблица 1 – параметры всех бинарных классификаторов системы
 Table 1 – parameters of all binary classifiers of the system

№	Параметр	Значение
1	n_estimators	100
2	criterion	gini
3	max_depth	None
4	min_samples_split	2
5	min_samples_leaf	1
6	min_weight_fraction_leaf	0.0
7	max_features	auto
8	max_leaf_nodes	None
9	min_impurity_decrease	0.0
10	min_impurity_split	None
11	bootstrap	True
12	oob_score	False
13	n_jobs	None
14	random_state	None
15	verbose	0
16	warm_start	False
17	class_weight	None
18	ccp_alpha	0.0
19	max_samples	None

Три последние главные классификатора (модели 14-16) реализуют многоклассовую классификацию прогноза группы здоровья ребенка от рождения до трех лет. Это классификаторы групп здоровья 1-2, 3-4, 5. Для их построения используются значимые факторы женского анамнеза и результаты моделей 2-13.

В 1-ю группу входят здоровые дети с нормальным физическим и психическим развитием без анатомических дефектов, функциональных и морфофункциональных отклонений, во 2-ю группу входят дети, не страдающие хроническими заболеваниями, но имеющие некоторые функциональные и морфофункциональные нарушения. В 3-ю и 4-ю группы входят дети, страдающие хроническими заболеваниями в стадии клинической ремиссии, активной стадии и стадии нестабильной клинической ремиссии с частыми обострениями. В 5-ю группу здоровья входят дети, страдающие тяжелыми хроническими заболеваниями, с редкими клиническими ремиссиями, дети-инвалиды [29].

Одним из важных факторов для большинства моделей является количество плодов во время беременности. Система выводит прогнозы модели отдельно для одноплодной и многоплодной беременности.

Количество предикторов и размер выборки, использованных для построения классификаторов, качество их работы на тестовых данных показаны в Таблице 2. Доля правильных ответов мультиклассового классификатора срока родов – 93 %, способа родов – 75 %, групп здоровья – 92 %.

Пример работы приложения представлен на рисунке 1. На нем можно видеть блоки признаков, используемых для прогнозирования. Все признаки заполняются с помощью элемента интерфейса переключатель (radioButton), позволяющего выбрать одно из допустимых значений признака («да» / «нет» для большинства).

Таблица 2 – описание классификаторов
Table 2 – description of classifiers

№	Название классификатора	Количество признаков	Общее количество объектов данных до балансировки	Количество объектов данных после балансировки	AUC	Доля правильных ответов
1	наступление беременности	21	854	(0, 588), (1, 572)	0,969±0,007	0,917±0,029
2	истмико-цервикальная недостаточность	14	572	(0, 492), (1, 501)	0,891±0,013	0,812±0,046
3	гипертензивные расстройства	26	536	(0, 487), (1, 495)	0,964±0,022	0,914±0,013
4	предлежание плаценты	31	536	(0, 483), (1, 503)	0,976±0,015	0,923±0,012
5	гестационный сахарный диабет	14	572	(0, 422), (1, 438)	0,827±0,056	0,765±0,032
6	нарушения количества околоплодных вод	28	536	(0, 495), (1, 494)	0,977±0,008	0,942±0,021
7	преждевременный разрыв плодных оболочек	34	572	(0, 541), (1, 542)	0,989±0,034	0,968±0,017
8	плановое кесарево сечение	15	536	(0, 368), (1, 369)	0,765±0,067	0,716±0,065
9	экстренное кесарево сечение	13	536	(0, 383), (1, 390)	0,732±0,054	0,687±0,068
10	самостоятельные роды	9	536	(0, 349), (1, 335)	0,724±0,070	0,684±0,072
11	прерывание беременности	35	572	(0, 550), (1, 543)	0,985±0,015	0,949±0,050
12	преждевременные роды	31	572	(0, 461), (1, 494)	0,933±0,042	0,835±0,077
13	срочные роды	26	572	(0, 452), (1, 425)	0,851±0,065	0,766±0,070
14	группы здоровья ребенка 1-2	28	464	(0, 404), (1, 400)	0,965±0,018	0,926±0,023
15	группы здоровья ребенка 3-4	17	464	(0, 355), (1, 369)	0,879±0,026	0,818±0,062
16	группа здоровья ребенка 5	41	464	(0, 429), (1, 435)	0,991±0,008	0,972±0,010

После заполнения признаков и нажатия на кнопку «Рассчитать», в выделенном поле появятся результаты прогнозирования: вероятность наступления беременности в процентах и, отдельно для многоплодной и одноплодной беременности, прогнозируемые осложнения беременности, срок и способ родов и группа здоровья ребенка.

Программа <input checked="" type="radio"/> ЭКО <input type="radio"/> Криопереносы	Перенесенные или имеющиеся заболевания	ИМТ <input type="radio"/> недостаток <input type="radio"/> избыток <input checked="" type="radio"/> норма <input type="radio"/> ожирение
Номер попытки ВРТ <input checked="" type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 <input type="radio"/> 4 и более	Аденомиоз <input type="radio"/> да <input checked="" type="radio"/> нет	Возраст <input type="radio"/> до 26 лет <input checked="" type="radio"/> 31 - 35 лет <input type="radio"/> 26 - 28 лет <input type="radio"/> 36 - 40 лет <input type="radio"/> 29 - 30 лет <input type="radio"/> более 40
Акушерский анамнез	Гастрит <input type="radio"/> да <input checked="" type="radio"/> нет	Хирургические вмешательства
Внематочная беременность <input type="radio"/> да <input checked="" type="radio"/> нет	Аутоиммунный тиреоидит (АИТ) <input type="radio"/> да <input checked="" type="radio"/> нет	Выскабливание <input type="radio"/> да <input checked="" type="radio"/> нет
Неразвивающаяся беременность <input type="radio"/> да <input checked="" type="radio"/> нет	Аллергия <input type="radio"/> да <input checked="" type="radio"/> нет	Гистерорезектоскопия <input type="radio"/> да <input checked="" type="radio"/> нет
Самостоятельные роды <input type="radio"/> да <input checked="" type="radio"/> нет	ДФКМ <input type="radio"/> да <input checked="" type="radio"/> нет	Гистероскопия <input checked="" type="radio"/> да <input type="radio"/> нет
Кесарево сечение <input type="radio"/> да <input checked="" type="radio"/> нет	Гипотиреоз <input type="radio"/> да <input checked="" type="radio"/> нет	Лапароскопия <input checked="" type="radio"/> да <input type="radio"/> нет
Выкидыш на раннем сроке <input type="radio"/> да <input checked="" type="radio"/> нет	ВИЧ / гепатит В <input type="radio"/> да <input checked="" type="radio"/> нет	Лапаротомия <input type="radio"/> да <input checked="" type="radio"/> нет
Поздний выкидыш <input type="radio"/> да <input checked="" type="radio"/> нет	Другие хронические заболевания <input type="radio"/> да <input checked="" type="radio"/> нет	Резекция в мочеполовой системе <input type="radio"/> да <input checked="" type="radio"/> нет
Аборт на раннем сроке <input checked="" type="radio"/> да <input type="radio"/> нет	Наследственный анамнез (наличие у ближайших родственников)	Прогноз по анамнезу
Преждевременные роды <input type="radio"/> да <input checked="" type="radio"/> нет	Онкология <input type="radio"/> да <input checked="" type="radio"/> нет	
Срочные роды <input type="radio"/> да <input checked="" type="radio"/> нет	Сахарный диабет <input type="radio"/> да <input checked="" type="radio"/> нет	
Вредные условия труда <input type="radio"/> да <input checked="" type="radio"/> нет	Гипертония \ ишемия \ инсульт \ тромбозы <input type="radio"/> да <input checked="" type="radio"/> нет	
Регулярные менструации без препаратов <input checked="" type="radio"/> да <input type="radio"/> нет		
Нормальная продолжительность цикла <input checked="" type="radio"/> да <input type="radio"/> нет		
Применялась гормональная контрацепция <input checked="" type="radio"/> да <input type="radio"/> нет		
Есть нарушения цикла <input type="radio"/> да <input checked="" type="radio"/> нет		
Гормональное лечение в наст. время <input type="radio"/> да <input checked="" type="radio"/> нет		

вероятность наступления беременности: 66.0%
при одноплодной беременности:
истмико-цервикальная недостаточность
срочные роды
плановое кс
группа здоровья ребенка: 3, 4
при многоплодной беременности:
истмико-цервикальная недостаточность
преждевременные роды
плановое кс
группа здоровья ребенка: 3, 4

Рисунок 1 – Пример работы приложения
Figure 1 – An example of how the application works

Обсуждение

Полученное качество классификаторов приемлемо. Наиболее низкое качество (до 0,8 по ROC-AUC) имеют классификаторы предсказания способа родов. Классификаторы предсказания групп здоровья ребенка имеют высокое качество: более $0,879 \pm 0,026$ по метрике ROC-AUC и более $0,818 \pm 0,062$ по точности.

Заключение

Было разработано уникальное приложение, ориентированное не только на прогнозирование исхода протокола ВРТ в виде беременности или живорождения, но также способное прогнозировать состояние здоровья ребенка после рождения.

Представленное приложение можно использовать перед протоколом как вспомогательный инструмент для прогнозирования его успеха, выражающегося в наступлении беременности, сроке и способе родов, а также состоянии здоровья рожденного ребенка. В настоящее время приложение проходит добавление прогнозов после этапов стимуляции, оплодотворения и переноса эмбрионов, а также тестирование, результаты которого можно использовать для уточнения классификаторов и интерфейса при необходимости.

Благодарности

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 19-37-90121.

Acknowledgments

The reported study was funded by RFBR, project number 19-37-90121.

ЛИТЕРАТУРА

1. David J McLernon, Ewout W Steyerberg, Egbert R te Velde, Amanda J Lee, Siladitya Bhattacharya. Predicting the chances of a live birth after one or more complete cycles of in vitro fertilisation: population based study of linked cycle data from 113 873 women. *BMJ*. 2016;355(8082) DOI: 10.1136/bmj.i5735
2. Porcu, G., Lehert, P., Colella, C, Giorgetti C. Predicting live birth chances for women with multiple consecutive failing IVF cycles: a simple and accurate prediction for routine medical practice. *Reproductive Biology and Endocrinology*. 2013;11(1) DOI: 10.1186/1477-7827-11-1
3. Katarina Kebbon Vaegter, Tatevik Ghukasyan Lakic, Matts Olovsson, Lars Berglund, Thomas Brodin, Jan Holte. Which factors are most predictive for live birth after in vitro fertilization and intracytoplasmic sperm injection (IVF/ICSI) treatments? Analysis of 100 prospectively recorded variables in 8,400 IVF/ICSI single-embryo transfers. *Fertility and Sterility*. 2017;107(3):641-648.e2 DOI: 10.1016/j.fertnstert.2016.12.005.
4. Dhillon R.K., McLernon D.J., Smith P.P., Fishel S., Dowell K., Deeks J.J., Bhattacharya S., Coomarasamy A. Predicting the chance of live birth for women undergoing IVF: a novel pretreatment counselling tool. *Hum Reprod*. 2016;31(1):84-92 DOI: 10.1093/humrep/dev268
5. Bergh C., Wennerholm U.B. Long-term health of children conceived after assisted reproductive technology. *Ups J Med Sci*. 2020;125(2):152-157 DOI: 10.1080/03009734.2020.1729904
6. Heijligers M., Peeters A., van Montfoort A., Nijsten J., Janssen E., Gunnewiek F.K., de Rooy R., van Golde R., Coonen E., Meijer-Hoogeveen M., Broekmans F., van der

- Hoeven M., Arens Y., de Die-Smulders C. Growth, health, and motor development of 5-year-old children born after preimplantation genetic diagnosis. *Fertil Steril*. 2019;111(6):1151-1158 DOI: 10.1016/j.fertnstert.2019.01.035
7. Qiu J., Li P., Dong M., Xin X., Tan J. Personalized prediction of live birth prior to the first in vitro fertilization treatment: a machine learning method. *J Transl Med*. 2019;17:317 DOI: 10.1186/s12967-019-2062-5
 8. Zare S., Thomsen M.R., Nayga R.M., Goudie A. Use of Machine Learning to Determine the Information Value of a BMI Screening Program. *American Journal of Preventive Medicine*. 2021;60(3):425-433. DOI: 10.1016/j.amepre.2020.10.016
 9. Bose S., Kenyon C.C., Masino A.J. Personalized prediction of early childhood asthma persistence: A machine learning approach. *PLOS ONE*. 2021;16(3):1-17 DOI: 10.1371/journal.pone.0247784
 10. Pedregosa et al. Scikit-learn: Machine Learning in Python. *JMLR*. 2011;12:2825-2830.
 11. Breiman L. Random Forests. *Machine Learning*. 2001;45:5-32
 12. Классификатор RandomForestClassifier из библиотеки Scikit-Learn. Доступно по: <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html> (дата обращения 15.07.2021)
 13. Морфологический анализатор pymorphy2. Доступно по: <https://pymorphy2.readthedocs.io/en/latest/> (дата обращения 15.07.2021)
 14. CountVectorizer из библиотеки Scikit-Learn. Доступно по: https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.CountVectorizer.html#sklearn.feature_extraction.text.CountVectorizer (дата обращения 15.07.2021)
 15. TfidfVectorizer из библиотеки Scikit-Learn. Доступно по: https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.TfidfVectorizer.html#sklearn.feature_extraction.text.TfidfVectorizer (дата обращения 15.07.2021)
 16. Arthur D., Vassilvitskii S. K-means++: the advantages of careful seeding. *Proc. of the Eighteenth Annu. ACM-SIAM Symp. on Discrete Algorithms, New Orleans, Louisiana, USA*. 2007:1027-1035 DOI:10.1145/1283383.1283494
 17. Кластеризация K-Means из библиотеки Scikit-Learn. Доступно по: <https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html?highlight=k%20mean#sklearn.cluster.KMeans> (дата обращения 15.07.2021)
 18. Kurba M.B., Rudnicki W.R. Feature Selection with the Boruta Package. *Journal of Statistical Software*. 2010;36(11):1-13 DOI: 10.18637/jss.v036.i11
 19. The Boruta algorithm implementation in Python. Доступно по: https://github.com/scikit-learn-contrib/boruta_py (дата обращения 15.07.2021)
 20. He, H.; Bai, Y.; Garcia, E.A.; Li, S. ADASYN: Adaptive synthetic sampling approach for imbalanced learning. *Proceedings of the 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), Hong Kong, China*. 2008; 1322–1328. DOI: 10.1109/IJCNN.2008.4633969
 21. Lemaitre G., Nogueira F., Aridas C.K. Imbalanced-learn: Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning. *JMLR*. 2017;18(17):1-5
 22. Kohavi R. A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence, San Mateo, CA*. 1995;2(12):1137-1143
 23. Stone M. Cross-Validatory Choice and Assessment of Statistical Predictions. *Journal of the Royal Statistical Society, Series B (Methodological)*. 1974;36(2):111-147

24. Реализация кросс-валидации StratifiedKFold в библиотеке Scikit-Learn. Доступно по: https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.StratifiedKFold.html (дата обращения 15.07.2021)
25. Набор библиотек Python для создания графического интерфейса на базе Qt5. Доступно по: <https://pypi.org/project/PyQt5/> (дата обращения 15.07.2021)
26. Qt Designer. Доступно по: <https://build-system.fman.io/qt-designer-download> (дата обращения 15.07.2021)
27. Библиотека Pickle. Доступно по: <https://docs.python.org/3.6/library/pickle.html> (дата обращения 15.07.2021)
28. Rifkin R., Klautau A. In Defense of One-Vs-All Classification. *JMLR*. 2004;5:101-104
29. Приказ Минздрава РФ № 621 «О комплексной оценке состояния здоровья детей», 2003

REFERENCES

1. David J McLernon, Ewout W Steyerberg, Egbert R te Velde, Amanda J Lee, Siladitya Bhattacharya. Predicting the chances of a live birth after one or more complete cycles of in vitro fertilisation: population based study of linked cycle data from 113 873 women. *BMJ*. 2016;355(8082) DOI: 10.1136/bmj.i5735
2. Porcu, G., Lehert, P., Colella, C, Giorgetti C. Predicting live birth chances for women with multiple consecutive failing IVF cycles: a simple and accurate prediction for routine medical practice. *Reproductive Biology and Endocrinology*. 2013;11(1) DOI: 10.1186/1477-7827-11-1
3. Katarina Kebbon Vaegter, Tatevik Ghukasyan Lakic, Matts Olovsson, Lars Berglund, Thomas Brodin, Jan Holte. Which factors are most predictive for live birth after in vitro fertilization and intracytoplasmic sperm injection (IVF/ICSI) treatments? Analysis of 100 prospectively recorded variables in 8,400 IVF/ICSI single-embryo transfers. *Fertility and Sterility*. 2017;107(3):641-648.e2 DOI: 10.1016/j.fertnstert.2016.12.005.
4. Dhillon R.K., McLernon D.J., Smith P.P., Fishel S., Dowell K., Deeks J.J., Bhattacharya S., Coomarasamy A. Predicting the chance of live birth for women undergoing IVF: a novel pretreatment counselling tool. *Hum Reprod*. 2016;31(1):84-92 DOI: 10.1093/humrep/dev268
5. Bergh C., Wennerholm U.B. Long-term health of children conceived after assisted reproductive technology. *Ups J Med Sci*. 2020;125(2):152-157 DOI: 10.1080/03009734.2020.1729904
6. Heijligers M., Peeters A., van Montfoort A., Nijsten J., Janssen E., Gunnewiek F.K., de Rooy R., van Golde R., Coonen E., Meijer-Hoogeveen M., Broekmans F., van der Hoeven M., Arens Y., de Die-Smulders C. Growth, health, and motor development of 5-year-old children born after preimplantation genetic diagnosis. *Fertil Steril*. 2019;111(6):1151-1158 DOI: 10.1016/j.fertnstert.2019.01.035
7. Qiu J., Li P., Dong M., Xin X., Tan J. Personalized prediction of live birth prior to the first in vitro fertilization treatment: a machine learning method. *J Transl Med*. 2019;17:317 DOI: 10.1186/s12967-019-2062-5
8. Zare S., Thomsen M.R., Nayga R.M., Goudie A. Use of Machine Learning to Determine the Information Value of a BMI Screening Program. *American Journal of Preventive Medicine*. 2021;60(3):425-433. DOI: 10.1016/j.amepre.2020.10.016
9. Bose S., Kenyon C.C., Masino A.J. Personalized prediction of early childhood asthma persistence: A machine learning approach. *PLOS ONE*. 2021;16(3):1-17 DOI: 10.1371/journal.pone.0247784
10. Pedregosa et al. Scikit-learn: Machine Learning in Python. *JMLR*. 2011;12:2825-2830.

11. Breiman L. Random Forests. *Machine Learning*. 2001;45:5-32
12. The RandomForestClassifier from the Scikit-Learn library. Available at: <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html> (accessed 15.07.2021)
13. Morphological analyzer pymorphy2. Available at: <https://pymorphy2.readthedocs.io/en/latest/> (accessed 15.07.2021)
14. The CountVectorizer from the Scikit-Learn library. Available at: https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.CountVectorizer.html#sklearn.feature_extraction.text.CountVectorizer (accessed 15.07.2021)
15. The TfidfVectorizer from the Scikit-Learn library. Available at: https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.TfidfVectorizer.html#sklearn.feature_extraction.text.TfidfVectorizer (accessed 15.07.2021)
16. Arthur D., Vassilvitskii S. K-means++: the advantages of careful seeding. *Proc. of the Eighteenth Annu. ACM-SIAM Symp. on Discrete Algorithms, New Orleans, Louisiana, USA*. 2007:1027-1035 DOI:10.1145/1283383.1283494
17. The K-Means clustering from the Scikit-Learn library. Available at: <https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html?highlight=k%20mean#sklearn.cluster.KMeans> (accessed 15.07.2021)
18. Kurba M.B., Rudnicki W.R. Feature Selection with the Boruta Package. *Journal of Statistical Software*. 2010;36(11):1-13 DOI: 10.18637/jss.v036.i11
19. The Boruta algorithm implementation in Python. Available at: https://github.com/scikit-learn-contrib/boruta_py (accessed 15.07.2021)
20. He, H.; Bai, Y.; Garcia, E.A.; Li, S. ADASYN: Adaptive synthetic sampling approach for imbalanced learning. *Proceedings of the 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), Hong Kong, China*. 2008; 1322–1328. DOI: 10.1109/IJCNN.2008.4633969
21. Lemaitre G., Nogueira F., Aridas C.K. Imbalanced-learn: Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning. *JMLR*. 2017;18(17):1-5
22. Kohavi R. A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence, San Mateo, CA*. 1995;2(12):1137-1143
23. Stone M. Cross-Validatory Choice and Assessment of Statistical Predictions. *Journal of the Royal Statistical Society, Series B (Methodological)*. 1974;36(2):111-147
24. Implementation of the StratifiedKFold cross-validation strategy in the Scikit-Learn library. Available at: https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.StratifiedKFold.html (accessed 15.07.2021)
25. A set of Python libraries for creating a graphical interface based on the Qt5 platform. Available at: <https://pypi.org/project/PyQt5/> (accessed 15.07.2021)
26. The Qt Designer. Available at: <https://build-system.fman.io/qt-designer-download> (accessed 15.07.2021)
27. The Pickle library. Available at: <https://docs.python.org/3.6/library/pickle.html> (accessed 15.07.2021)
28. Rifkin R., Klautau A. In Defense of One-Vs-All Classification. *JMLR*. 2004;5:101-104
29. Order of the Ministry of Health of the Russian Federation # 621 “O kompleksnoj ocenke sostoyaniya zdorov’ya detej” [About comprehensive assessment of the stage of health children] (in Russ). 2003

ИНФОРМАЦИЯ ОБ АВТОРАХ / INFORMATION ABOUT THE AUTHORS

Синотова Светлана Леонидовна, аспирант, младший научный сотрудник кафедры интеллектуальных информационных технологий института фундаментального образования ФГАОУ ВО «УрФУ имени первого Президента России Б.Н. Ельцина», Екатеринбург, Российская Федерация.

e-mail: sveta.volkova92@mail.ru

ORCID: [0000-0002-4318-5223](https://orcid.org/0000-0002-4318-5223)

Лимановская Оксана Викторовна, кандидат химических наук, доцент кафедры интеллектуальных информационных технологий института фундаментального образования ФГАОУ ВО «УрФУ имени первого Президента России Б.Н. Ельцина», Екатеринбург, Российская Федерация.

e-mail: o.v.limanovskaia@urfu.ru

ORCID: [0000-0002-2084-3916](https://orcid.org/0000-0002-2084-3916)

Плаксина Анна Николаевна, кандидат медицинских наук, ассистент кафедры физической и реабилитационной медицины ФГБОУ ВО «Уральский государственный медицинский университет Минздрава РФ», Екатеринбург, Российская Федерация.

e-mail: burberry20@yandex.ru

ORCID: [0000-0002-3119-478X](https://orcid.org/0000-0002-3119-478X)

Макутина Валерия Андреевна, кандидат биологических наук, заведующая лабораторией клинической эмбриологии АО «Центр семейной медицины», Екатеринбург, Российская Федерация.

e-mail: makutina_v@rambler.ru

ORCID: [0000-0003-1127-2792](https://orcid.org/0000-0003-1127-2792)

Svetlana Leonidovna Sinotova, P.G., Junior Researcher of The Department of Intellectual Information Technologies, Institute of Fundamental Education, FSAEI HE «UrFU named after the first President of Russia B.N.Yeltsin», Ekaterinburg, Russian Federation.

Oksana Viktorovna Limanovskaya, Candidate of Chemical Sciences, Associate Professor of The Department of Intellectual Information Technologies, Institute of Fundamental Education, FSAEI HE «UrFU named after the first President of Russia B.N.Yeltsin», Ekaterinburg, Russian Federation.

Anna Nikolaevna Plaksina, Candidate of Medical Sciences, Assistant of The Department of Physical and Rehabilitation Medicine, FSBEI HE «USMU of the Ministry of Health of the Russian Federation», Ekaterinburg, Russian Federation.

Valerija Andreevna Makutina, Candidate of Biological Sciences, Head of the Laboratory of Clinical Embryology of The Family Medicine Centre, Ekaterinburg, Russian Federation

Статья поступила в редакцию 07.06.2021; одобрена после рецензирования 16.09.2021; принята к публикации 27.09.2021.

The article was submitted 07.06.2021; approved after reviewing 16.09.2021; accepted for publication 27.09.2021.