

Федеральное государственное автономное образовательное учреждение
высшего образования
«Уральский федеральный университет имени первого
Президента России Б.Н. Ельцина»

На правах рукописи



Синотова Светлана Леонидовна

РАЗРАБОТКА МАТЕМАТИЧЕСКОГО И АЛГОРИТМИЧЕСКОГО
ОБЕСПЕЧЕНИЯ СИСТЕМЫ ПОДДЕРЖКИ ПРИНЯТИЯ
ВРАЧЕБНЫХ РЕШЕНИЙ ДЛЯ ПРОГНОЗИРОВАНИЯ
ЗДОРОВЬЯ РЕБЕНКА, ЗАЧАТОГО ПРИ ПОМОЩИ
ВСПОМОГАТЕЛЬНЫХ РЕПРОДУКТИВНЫХ ТЕХНОЛОГИЙ

2.3.1. Системный анализ, управление и обработка информации,
статистика

АВТОРЕФЕРАТ
диссертации на соискание ученой степени
кандидата технических наук

Екатеринбург – 2023

Работа выполнена на кафедре интеллектуальных информационных технологий Института фундаментального образования ФГАОУ ВО «Уральский федеральный университет имени первого Президента России Б.Н. Ельцина».

Научный руководитель: кандидат физико-математических наук, доцент, **СОЛОДУШКИН Святослав Игоревич**.

Официальные оппоненты: **БЕРИКОВ Владимир Борисович**, доктор технических наук, доцент, ФГБУН Институт математики им. С. Л. Соболева Сибирского отделения Российской академии наук, г. Новосибирск, главный научный сотрудник;

ЦЫМБЛЕР Михаил Леонидович, доктор физико-математических наук, доцент, ФГАОУ ВО «Южно-Уральский государственный университет (национальный исследовательский университет)», г. Челябинск, заместитель директора научно-образовательного центра «Искусственный интеллект и квантовые технологии»;

ХАЛЪЯСМАА Александра Ильмаровна, кандидат технических наук, доцент, ФГАОУ ВО «Уральский федеральный университет имени первого Президента России Б.Н. Ельцина», г. Екатеринбург, заведующий лабораторией цифровых двойников в электроэнергетике Уральского энергетического института.

Защита диссертации состоится «21» февраля 2023 г. в 14:00 часов на заседании диссертационного совета УрФУ 2.3.12.13 по адресу: 620002, г. Екатеринбург, ул. Мира, 19, ауд. И-420 (зал Ученого совета).

С диссертацией можно ознакомиться в библиотеке и на сайте ФГАОУ ВО «Уральский федеральный университет имени первого Президента России Б.Н. Ельцина»: <https://dissovet2.urfu.ru/mod/data/view.php?d=12&rid=4265>

Автореферат разослан «_____» _____ 202_ года.

Ученый секретарь
диссертационного совета



Сафиуллин Николай Тахирович

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы исследования и степень ее разработанности. По данным Росстат в настоящее время в Российской Федерации проживает 35,1 млн. женщин репродуктивного возраста, при этом по сведениям ФГБУ «Национального медицинского исследовательского центра акушерства, гинекологии и перинатологии имени академика В.И. Кулакова» Министерства здравоохранения РФ около 8 млн. из них страдают бесплодием.

По данным Регистра ВРТ Российской ассоциации репродукции человека за 2018 г. наблюдается ежегодный прирост доли родов после применения вспомогательных репродуктивных технологий (ВРТ) от общего числа родов в РФ. Международный практический опыт также показывает, что ВРТ можно рассматривать как эффективную и востребованную меру повышения рождаемости (J. de Mouzon, S. Sunderam), но до сих пор имеют место противоречивые мнения по поводу здоровья рожденных детей (Г. М. Савельевой, Е. С. Кешишян, M. von Wolff, U. B. Wennerholm, S. McDonald, M. Hansen). Поэтому сотрудники центров ВРТ и врачи, занимающиеся детьми, зачатыми в результате применения ВРТ, ставят перед специалистами по анализу данных задачу создания системы поддержки принятия врачебных решений (СППВР), позволяющей прогнозировать осложнения течения беременности и родов, а также возникновение с течением времени заболеваний у ребенка, которые можно связать с наследственными факторами.

Несмотря на многочисленные исследования в области прогнозирования результата применения ВРТ (J. Qiu, G. Porcu, A. Goyal, M. B. Ratna, S. C. Esteves, S. M. Nelson, Y. Wu, L. Benoit и др.), имеющиеся на сегодняшний день способы прогнозирования исходов ВРТ не позволяют оценить именно здоровье детей и не проводят взаимосвязи с материнским здоровьем. Это обусловлено отсутствием накопленных данных, которые смогли бы предоставить информацию для таких исследований. С появлением в некоторых странах наборов данных, содержащих информацию о проведенных протоколах ВРТ, начинаются разработки СППВР для предсказания исходов ВРТ, например, в университете Монаша (Австралия) и G. Letterie G. и A. Mac Donald, что подчеркивает актуальность данной темы. Несмотря на то, что СППВР являются современным и востребованным инструментом персонифицированной медицины, в РФ системы прогнозирования здоровья ребенка, зачатого при помощи ВРТ, не существует в связи со сложностями в накоплении и обработке медицинской информации.

Цель исследования. На основе применения алгоритмов машинного обучения к медицинской документации осуществить разработку математического и алгоритмического обеспечения системы поддержки принятия врачебных решений для многоэтапного прогнозирования здоровья ребенка, зачатого при помощи вспомогательных репродуктивных технологий, а также создать алгоритмы

обработки, компоновки и анализа неструктурированных текстовых данных на примере данных медицинских организаций.

Для достижения поставленной цели были сформулированы и решены следующие **задачи**:

1) проанализировать состояние разработок в области создания математического и алгоритмического обеспечения для прогнозирования исходов ВРТ и формализовать постановку задачи принятия врачебных решений при проведении протокола ВРТ;

2) разработать алгоритмы обработки, компоновки и анализа разнородных текстовых данных на примере данных медицинских организаций для выделения признаков описания моделируемых объектов;

3) разработать структуру СППВР, которая позволяет осуществлять многоэтапное прогнозирование здоровья ребенка, зачатого при помощи ВРТ;

4) создать математическое и алгоритмическое обеспечение СППВР для многоэтапного прогнозирования здоровья ребенка, зачатого при помощи ВРТ, реализовать СППВР в виде программного приложения и апробировать СППВР в клинической практике.

Объект исследования – данные о проведенных протоколах ВРТ и здоровье рожденных в результате этих протоколов детей в возрасте до трех лет.

Предмет исследования – математическое и алгоритмическое обеспечение системы поддержки принятия врачебных решений для прогнозирования здоровья ребенка, рожденного при помощи ВРТ.

Научная новизна результатов диссертационного исследования заключается в следующем:

1) предложена структура СППВР, позволяющая моделировать многоэтапные процессы, которая применена к имеющимся данным о здоровье ребенка, зачатого при помощи ВРТ;

2) предложено математическое и алгоритмическое обеспечение для объединения и обработки необладающих функциональной совместимостью данных разных медицинских учреждений в единую базу на основе методов интеллектуального анализа данных;

3) разработано математическое и алгоритмическое обеспечение СППВР, состоящее из моделей машинного обучения, позволяющих на всех этапах проведения протокола ВРТ предсказать его исходы: от наступления беременности до здоровья ребенка в возрасте до трех лет.

Методология и методы исследований. В работе использованы методы обработки текстовой информации, прикладной статистики, алгоритмы отбора значимых признаков для моделей машинного обучения, алгоритмы машинного обучения с учителем, методы разработки программного приложения.

Теоретическая значимость заключается в разработке методов и алгоритмов интеллектуальной поддержки при принятии управленческих решений для прогнозирования здоровья ребенка, зачатого при помощи ВРТ. Методы и алгоритмы могут применяться для моделирования медицинских задач, которые формализуются в виде многоэтапных процессов.

Практическая значимость результатов работы заключается в разработке системы поддержки принятия врачебных решений и в создании программного приложения, готового к использованию в качестве вспомогательного инструмента принятия решений для сотрудников и пациентов центров ВРТ.

Основные положения, выносимые на защиту.

1. Предложенный метод обработки разнородных данных на основе алгоритмов токенизации коллекций текстовых медицинских документов, с использованием морфологического анализатора, исправления ошибок и кластеризации данных может использоваться для решения задач создания признакового описания объектов данных по текстовому описанию.
2. Разработанная структура системы поддержки принятия врачебных решений многоэтапного прогнозирования результатов использования ВРТ позволяет сделать прогноз группы здоровья ребенка с указанием диагнозов, по которым возможно возникновение осложнений. Структура предоставляет возможность уточнять прогноз, дополняя его новыми признаками, полученными с каждым следующим этапом протокола ВРТ.
3. Разработанное математическое и алгоритмическое обеспечение СППВР, состоящее из 63 моделей машинного обучения, позволяет осуществлять прогнозирование с уровнем точности, достаточным для клинического применения, а именно по результатам апробации средняя доля правильных ответов для прогнозируемых исходов составляет 84,5%, при этом 95% доверительный интервал – [78,7; 90,3].

Достоверность полученных в работе результатов обеспечивается корректным применением математических методов, подтверждается успешно проведенной апробацией разработанного математического и алгоритмического обеспечения системы поддержки принятия врачебных решений.

Внедрение результатов диссертационного исследования. Результаты диссертационной работы используются в АО «Центр семейной медицины», г. Екатеринбург, Россия (акт об использовании результатов от 11.04.2022).

Апробация работы. Основные положения и результаты работы докладывались и обсуждались на научных семинарах кафедры интеллектуальных информационных технологий ИнФО УрФУ, кафедры информационных технологий и систем управления ИРИТ-РТФ УрФУ, а также на конференциях: VII Международная молодежная научная конференция «Физика. Технологии. Инновации ФТИ-2020» Екатеринбург, Россия, 18–22 мая 2020; VIII Международная

молодежная научная конференция «Физика. Технологии. Инновации ФТИ-2021» Екатеринбург, Россия, 17–21 мая 2021.

Диссертационное исследование выполнено при финансовой поддержке гранта РФФИ № 19-37-90121 «Экспертная система прогнозирования здоровья детей, рожденных при помощи ВРТ».

Личный вклад. Все результаты исследований, составляющие основное содержание диссертации, получены автором самостоятельно. Вклад в совместных публикациях соавтора Лимановской О.В., являвшейся руководителем во время обучения в аспирантуре, заключается в постановке задач исследования. В работах, выполненных вместе с Плаксиной А.Н. и Макутиной В.А., соавторами выполнен сбор первичных данных, на основе которых было выполнено исследование, а также клиническая интерпретация полученных результатов. Прочими соавторами в работах диссертанта предоставлены исходные данные исследования и пояснения к ним. Во всех работах в соавторстве диссертантом внесен основной вклад по предобработке данных исследования, статистической обработке и анализу данных исследования, включая все результаты, связанные с применением машинного обучения к данным и разработкой программного приложения.

Публикации. По теме диссертации опубликовано 11 научных работ, в том числе 5 статей в рецензируемых научных изданиях, определенных ВАК РФ и Аттестационным советом УрФУ, из них 1 статья в издании, индексируемом в международных цитатно-аналитических базах WoS и Scopus, а также 1 свидетельство о государственной регистрации программы для ЭВМ.

Структура и объем работы. Диссертационная работа состоит из введения, четырех глав, заключения, списка сокращений, списка литературы, шести приложений. Работа изложена на 162 страницах, содержит 37 рисунков и 15 таблиц.

КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ

Во введении обоснована актуальность темы диссертационного исследования; сформулированы цель, задачи, объект и предмет исследования; показана научная новизна, теоретическая и практическая ценность результатов исследования.

В первой главе содержится анализ текущего состояния современных исследований в области прогнозирования различных результатов применения ВРТ, который показывает отсутствие моделей предсказания успешности протоколов ВРТ, выраженной в долгосрочном здоровье ребенка. Необходимость подобных моделей в практическом здравоохранении подтверждает актуальность исследования.

Далее описаны базы данных и информационные системы медицинских организаций, предоставивших обезличенные данные для проведения исследования. Показано, что данные неинтероперабельны, плохо структурированы, содержат ошибки и нуждаются в предобработке.

Описаны основные принципы работы проектируемой СППВР. Отмечено, что задача прогнозирования многоэтапна, каждый исход привязан к определенному временному промежутку (например, к наступлению беременности, ее течению или родам) и этапу протокола ВРТ. При этом решения, принимаемые на текущем этапе протокола ВРТ, влияют на последующие этапы. Поэтому после каждого этапа протокола ВРТ целесообразно уточнять прогноз для всех последующих исходов, формирующих в итоге группу здоровья ребенка.

Показано, что требуемое математическое и алгоритмическое обеспечение должно представлять собой набор иерархических моделей машинного обучения, позволяющих на основе прогнозов принимать персонифицированные управленческие решения по тактике ведения пациентки на каждом этапе протокола ВРТ.

Поставленная задача относится к классу задач машинного обучения с учителем, состоит в нахождении функций и разрешающих правил, которые по признаковому описанию объектов (женщин или детей) будут давать прогнозы исходов (результатов протоколов ВРТ).

Проведен анализ алгоритмов отбора признаков для машинного обучения и анализ методов, используемых для решения задачи машинного обучения с учителем. Описаны методы и алгоритмы, использованные в работе.

В заключении сформулированы цель и задачи исследования.

Во второй главе представлена созданная база данных, объединяющая информацию от всех медицинских организаций, и описана разработка алгоритмического обеспечения для предобработки исходных данных.

Целью подготовки данных является извлечение из них признакового описания объектов данных и целевых значений, которые в дальнейшем будут использованы для создания моделей машинного обучения СППВР.

Для того, чтобы использовать данные в поставленной задаче, требуется их значительная предобработка, поскольку данные

- 1) не обладают функциональной совместимостью;
- 2) недостаточно стандартизированы;
- 3) замусорены;
- 4) не формализованы;
- 5) плохо структурированы.

В первую очередь представлен алгоритм получения итогового набора данных. Цель работы этого алгоритма – сбор в единую базу полной информации о всех имеющихся объектах данных (детях). Информация должна включать в себя данные о матери, о течении беременности и родов, о здоровье ребенка с течением времени.

После устранения функциональной несовместимости была получена 821 оригинальная запись о женщинах и рожденных ими 836 детях на этапе родильного дома, а также о состоянии здоровья 464 детей в возрасте до трех лет. Из 821 женщин

полная информация была найдена только для 572; эта информация использовалась для построения моделей машинного обучения разрабатываемой СППВР. Для предсказания вероятности наступления беременности были использованы дополнительные данные, не включенные в 572 полученных образца. Для того, чтобы дополнить данные информацией о пациентках и протоколах, в результате которых беременность не наступила, были добавлены 282 протокола ВРТ, из которых 188 – протоколы экстракорпорального оплодотворения (ЭКО), 94 – протоколы криопереноса. Таким образом, общее число протоколов, включенных в анализ – 854.

Далее описан алгоритм обработки текстовых данных. Блок-схема алгоритма представлена на рисунке 1.

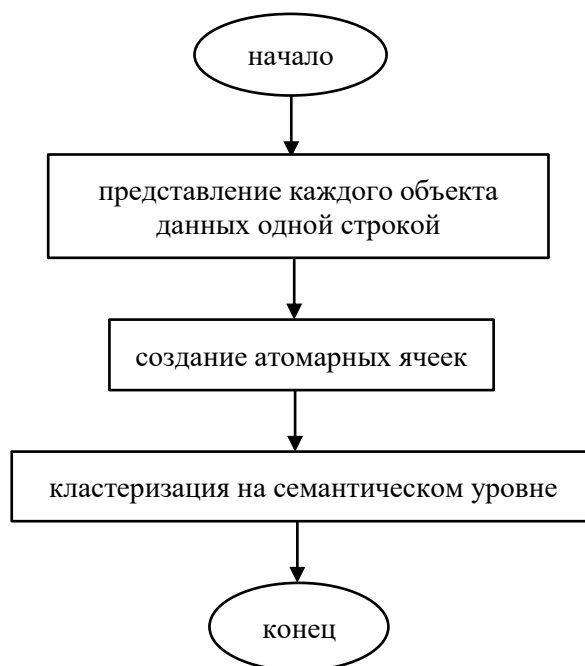


Рисунок 1. Блок-схема алгоритма обработки текстовых данных

Анализ структуры предоставленных данных позволил выделить этапы, прохождение которых последовательно упрощает задачу, делая данные более структурированными и чистыми.

Этап 1. Представление каждого объекта данных одной строкой

В исходных таблицах данных информация о некоторых пациентках могла быть представлена в нескольких строках таблицы, при этом каждая из строк таблицы соответствует отдельному клиническому случаю (например, у женщины могло быть прерывание беременности, выкидыш и кесарево сечение в течение жизни). Этап заключается в объединении в одну строку базы данных информации об одном объекте данных, разрозненной по разным строкам таблицы.

Этап 2. Создание атомарных ячеек

Суть действий этапа в создании гарантированного доступа к данным, то есть удовлетворении второго правила Кодда: каждое отдельное (атомарное) значение

данных должно быть логически доступно с помощью комбинации имени таблицы, значения первичного ключа (объекта данных) и имени столбца.

Полученная медицинская документация представляет собой произвольный текст, распределенный по ячейкам таблиц данных. Токенизация этого текста заключается в выделении из него отдельных слов, которые представляют собой признаки объектов данных, и формируют названия столбцов в полученных в результате токенизации таблицах. Поскольку слова в медицинской документации содержат опечатки и орфографические ошибки, то при токенизации число столбцов большое (до 4000) и требуется его уменьшить. Для этого выполняются следующие шаги:

- 1) приведение к нижнему регистру;
- 2) очистка от «мусора» – от знаков препинания и букв латинского алфавита;
- 3) лемматизация – приведение к нормальной форме;
- 4) токенизация;
- 5) удаление стоп-слов;
- 6) устранение ошибок написания.

Многие слова, полученные в результате токенизации медицинской документации – это варианты ошибочного написания слов и для сокращения числа столбцов нужно привести их к родительскому слову, устранить орфографические ошибки. Представлен код и блок-схема алгоритма устранения орфографических ошибок. Идея алгоритма в том, что слова близкие по расстоянию Левенштейна, заменяются на оригинальное слово. Блок-схема алгоритма показана на рисунке 2. Благодаря такой обработке удалось значительно сократить объем словаря, например, для акушерско-гинекологического анамнеза с 992 до 707 слов, а для хирургических вмешательств с порядка 4000 тысяч слов до порядка 2000 без потери информации.

Этап 3. Кластеризация на семантическом уровне

Кластеризация – объединение признаков в укрупненные группы с отсеиванием незначимых деталей. Если признаки будут не сгруппированы, то почти каждый пациент будет описан своим уникальным набором признаков, что затруднит обобщение и, следовательно, построение моделей машинного обучения. После группировки (кластеризации) пациенты будут объединены под укрупненными признаками (например, «внематочная беременность» включает в себя внематочную беременность слева, внематочную беременность справа, эктопическую беременность, трубную беременность и так далее, а признак «гинекологическая операция» объединяет все хирургические вмешательства в гинекологии). Кластеризация признаков проводилась на основе словарей, которые создавались совместно с врачами. Данные словари имеют самостоятельную ценность.

Применяя созданный алгоритм обработки и компоновки данных, удалось получить из исходных данных полное признаковое описание женщин и состояния

здоровья всех детей. Описаны особенности обработки каждой таблицы.

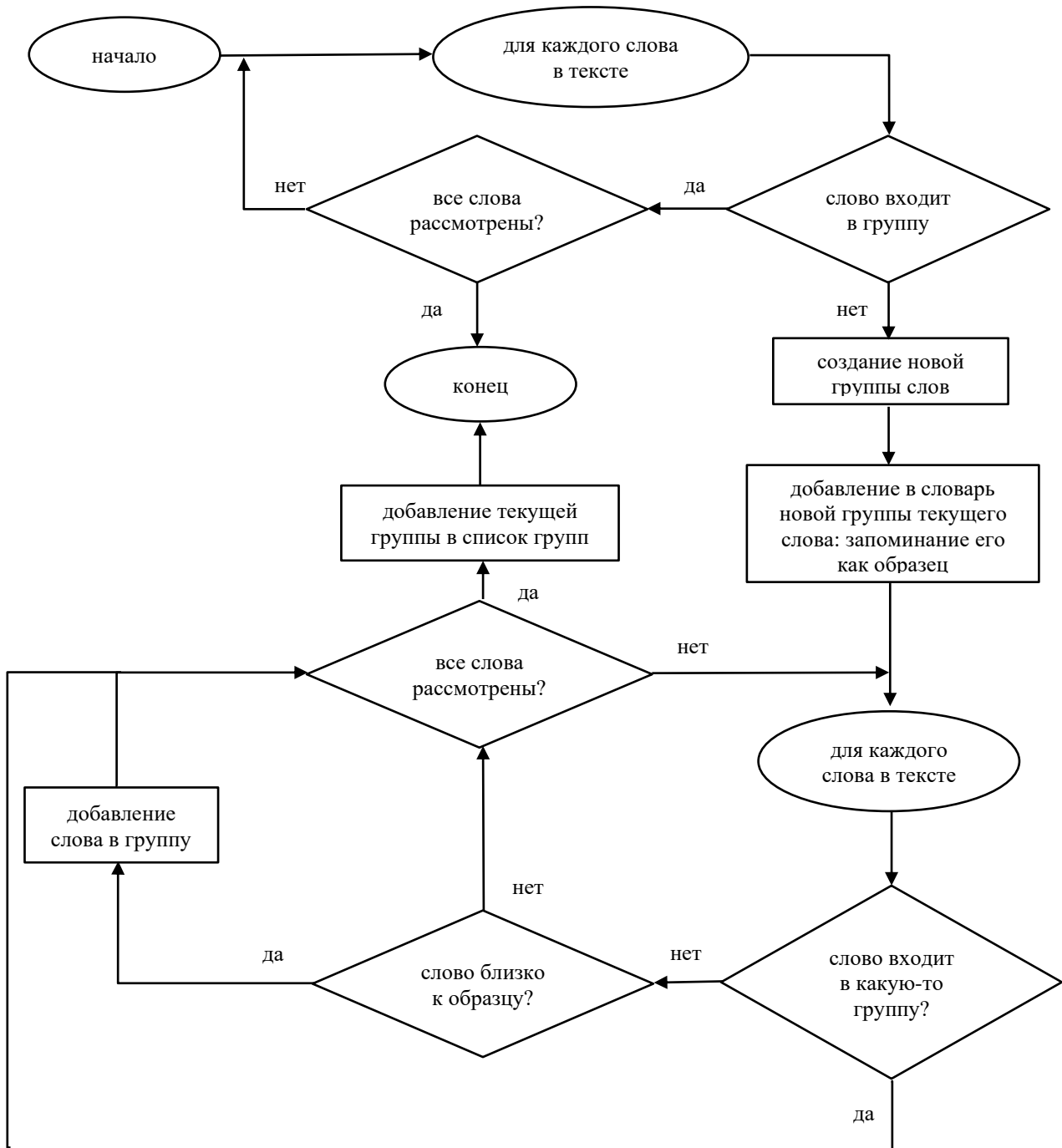


Рисунок 2. Блок-схема алгоритма устранения орфографических ошибок

Предложенный метод обработки разнородных данных на основе алгоритмов токенизации коллекций текстовых документов, с использованием морфологического анализатора, исправления ошибок и кластеризации данных может применяться для решения задач создания признакового описания объектов данных по текстовому описанию.

В третьей главе описана разработка математического и алгоритмического обеспечения СППВР для прогнозирования здоровья ребенка, зачатого при помощи ВРТ.

Согласно основным принципам работы СППВР, описанным в первой главе, разрабатываемые прогностические модели машинного обучения были сгруппированы по четырем блокам:

- 1) прогнозирование результата отбора пациентов;
- 2) прогнозирование результата овариаальной стимуляции;
- 3) прогнозирование результата оплодотворения;
- 4а) прогнозирование результата переноса (для программы ЭКО);
- 4б) прогнозирование результата переноса (для программы криопереноса).

В разделе 3.1 представлены входные и выходные данные для каждого блока СППВР.

Далее в работе описаны структура разрабатываемой СППВР (рисунок 3) и функциональная схема работы СППВР (рисунок 4).

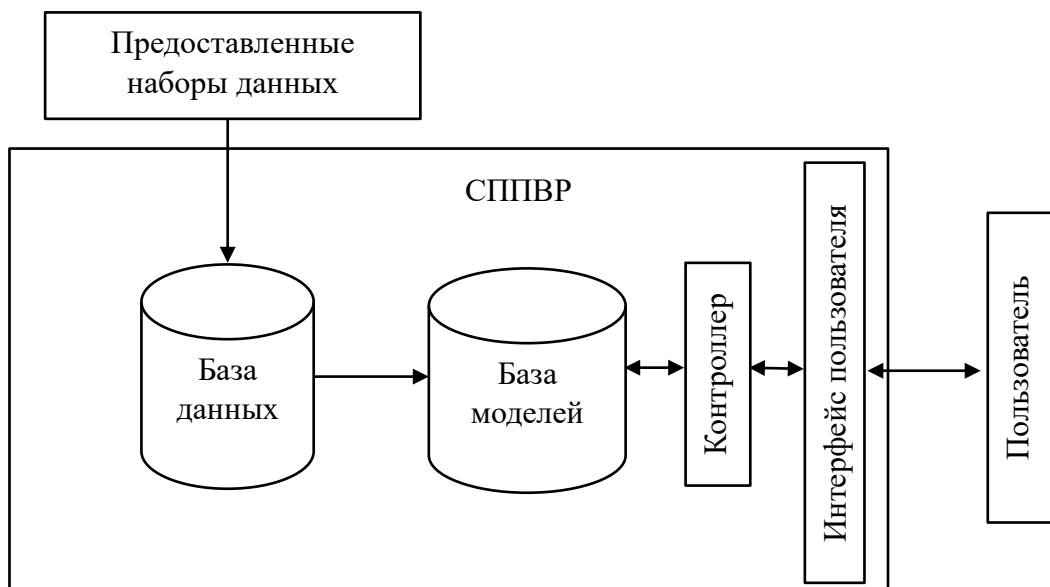


Рисунок 3. Структура разрабатываемой СППВР

Разработанная СППВР состоит из взаимодействующих между собой базы данных, базы моделей, контроллера и пользовательского интерфейса.

Признаковые описания объектов данных сохранены в базу данных. После применения алгоритмов машинного обучения с учителем к данным, получена база моделей, являющаяся математическим обеспечением разрабатываемой СППВР. Ключевая особенность разработанной СППВР в том, что это не просто набор несвязанных классификаторов и регрессионных моделей, а иерархия моделей машинного обучения, позволяющих выполнять последовательное уточнение прогнозов на всех этапах ВРТ. Логика взаимодействия ML-моделей между собой и порядок их работы составляют алгоритмическое обеспечение созданной СППВР и регулируются контроллером. Пользовательский интерфейс позволяет пользователю вводить данные, описывающие конкретный новый объект, для которого нужно дать прогноз. То есть с помощью пользовательского интерфейса врач взаимодействует с СППВР, описывая женщину, вступающую в протокол ВРТ.

Входные данные передаются через контроллер соответствующим ML-моделям. ML-модели проводят вычисления, то есть осуществляют прогнозирование, и результаты отображаются пользователю посредством пользовательского интерфейса. На основе отображаемых результатов пользователь СППВР может принимать персонализированные медицинские решения. Кроме того, пользователь может варьировать входные параметры и сравнивать соответствующие результаты.

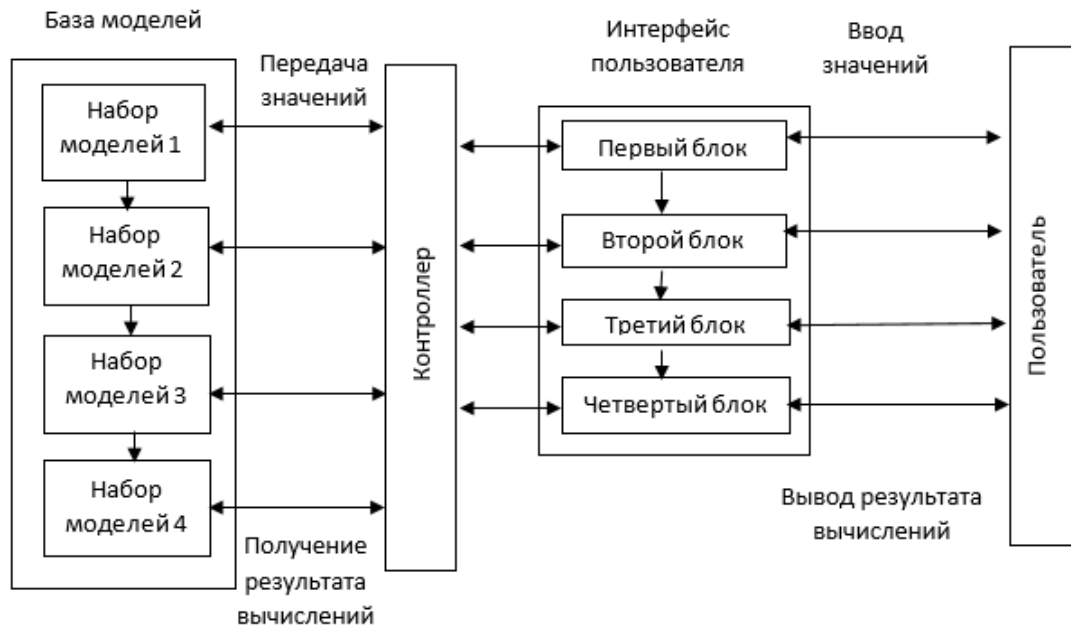


Рисунок 4. Функциональная схема работы системы

Согласно функциональной схеме работы системы, интерфейс пользователя предоставляет возможность вводить признаковое описание объектов для четырех различных блоков системы. Каждый блок отвечает за признаки, которые добавляются к описанию объекта на определенном этапе процедуры ВРТ.

Каждому блоку системы соответствует определенный набор моделей машинного обучения. Контроллер позволяет группировать входные данные в наборы признаков и отправлять в каждую из ML-моделей нужный набор. Также контроллер отвечает за последовательность прогнозирования в наборах моделей, то есть задает очередность построения прогнозов и передает спрогнозированные значения в следующие по логике модели, использующие эти значения в качестве предикторов. Пользователь получает вывод результатов вычислений для каждого блока.

Процесс создания базы моделей происходит вне СППВР. Он представлен далее в диссертации и на рисунке 5.

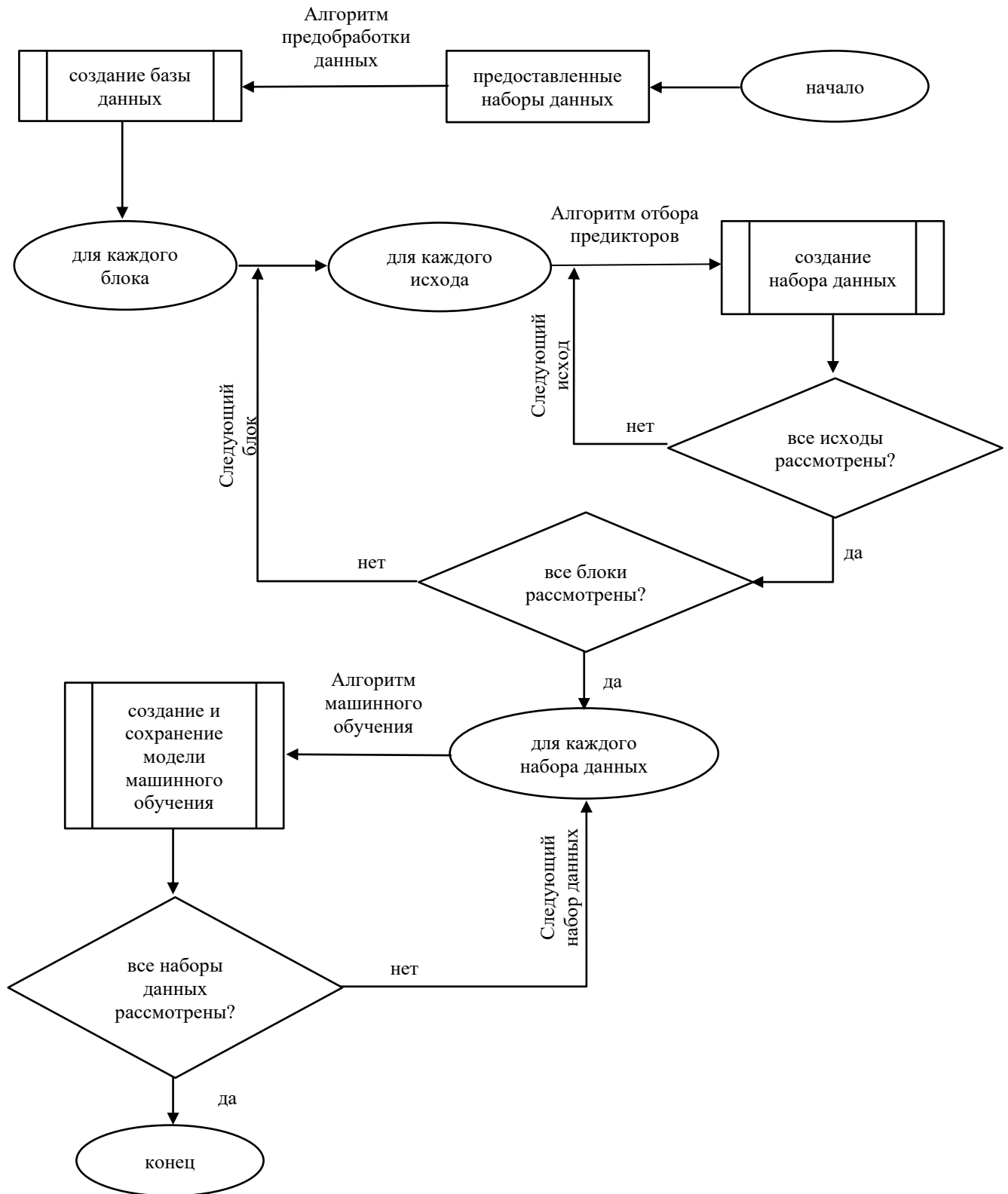


Рисунок 5. Алгоритм создания базы моделей

В разделе 3.3 перечислены все модели машинного обучения в составе СППВР и показано их распределение по блокам системы.

Первый блок («Прогнозирование по данным анамнеза») содержит 26 ML-моделей, позволяющих прогнозировать на первом этапе ВРТ весь набор исходов, формирующих здоровье ребенка.

Второй блок («Прогнозирование после стимуляции (для программы ЭКО)») соответствует второму вариативному этапу ВРТ и содержит 29 ML-моделей:

1) ML-модели из первого блока СППВР, необходимые для прогнозирования исходов, формирующих здоровье ребенка (15 моделей, прогнозируемые исходы которых не имеют новых значимых признаков на этом этапе ВРТ);

2) новые ML-модели, которые представляют собой модификации ML-моделей из первого блока и прогнозируют те же исходы, но имеют большее число значимых признаков, в связи с тем, что на данном этапе ВРТ добавляется информация о проведенной стимуляции яичников (11 моделей);

3) ML-модели для прогнозирования данных стимуляции (три модели).

Третий блок («Прогнозирование результатов оплодотворения (для программы ЭКО)») содержит четыре модели машинного обучения, прогнозирующие результаты третьего этапа ЭКО.

Четвертый-а блок («Прогнозирование после переноса для программы ЭКО») содержит 27 ML-моделей:

1) ML-модели из первого и второго блоков СППВР (16 моделей);

2) новые ML-модели – модификации аналогичных ML-моделей первого и второго блоков СППВР, но с учетом значимых признаков, полученных на третьем и четвертом этапах ЭКО (10 моделей);

3) бинарный классификатор для прогнозирования количества плодов.

Четвертый-б блок («Прогнозирование после переноса для программы Криопереносы») содержит 27 моделей:

1) ML-модели из первого и второго блоков СППВР (19 моделей);

2) новые ML-модели – модификации аналогичных ML-моделей первого и второго блоков СППВР, но с учетом значимых признаков, полученных на данном этапе протокола криопереноса (семь моделей);

3) бинарный классификатор для прогнозирования количества плодов.

Всего разработанная СППВР содержит 63 модели машинного обучения, из которых шесть являются регрессионными, пять – многоклассовыми классификаторами, остальные 52 – бинарными классификаторами.

В разделе 3.4 показан способ выбора алгоритма отбора признаков и алгоритма построения ML-модели на примере задачи прогнозирования наступления беременности.

В четвертой главе изложены результаты применения алгоритмов предобработки данных исследования, описано создание моделей машинного обучения на примере имеющихся данных и полученные результаты, описан процесс создания программного приложения, позволяющего работать с СППВР и представлены данные апробации разработанного приложения.

Благодаря использованию алгоритмов предобработки данных удалось описать все объекты данных 87 признаками и получить 33 исхода, подлежащих

прогнозированию. После проведения отбора признаков, удалось выделить 67 значимых признаков для всех ML-моделей системы.

Для каждой отдельной ML-модели приведены размеры выборки до и после балансировки, указаны наборы предикторов, приведена оценка качества прогнозирования на тестовых данных.

Описан интерфейс приложения и приведен скриншот примера его работы (рисунок 6). Приложение представляет собой единую форму, на которой расположены все четыре блока системы. Код основного файла интерфейса приложения содержит 505 строк. В отдельные модули вынесена загрузка моделей и функции, отвечающие за работу с пользовательским интерфейсом.

The screenshot displays the SPPVR application interface, divided into four main functional blocks:

- Program Selection (Программа):** Includes options for 'EKO' (selected) and 'cryopreservation'. Fields for 'Attempt number (BPT)', 'BMI', and 'Age'. A section for 'Obstetric history' with radio buttons for various conditions like 'Abortion in early term', 'Ectopic pregnancy', etc. A 'Prognosis by history' section shows predicted pregnancy rates (80% for single pregnancy, 8% for multiple) and risks (2% for cervical insufficiency, 4% for hypertension, etc.).
- Stimulation (program EKO) (СТИМУЛЯЦИЯ):** Includes 'Pre-treatment' options (Painkillers, Estrogen, ZGT, Gestagens, KOK protocol, etc.). 'Protocol' (long/short cycle, natural cycle) and 'Trigger' (diferelin, GnRH, ovidrel, etc.) selection. A 'Prognosis after stimulation' section shows expected follicle count (18), endometrial thickness (8-15 mm), and pregnancy probability (95% for single pregnancy, 8% for multiple).
- Fertilization (program EKO) (ОПЛОДОТВОРЕНИЕ):** Includes 'Number of mature oocytes' (8) and 'Fertilization type' (EKO, ICSI, EKO+ICSI). 'Medium' selection (Cook, CSSM, Irvine, Sage, Vitrolife). A 'Prognosis of fertilization result' section shows blastocyst counts at 3, 5, and 6 days.
- Transfer (ПЕРЕНОС):** Includes 'Number of embryos' (1 or 2) and 'Quality' (high/low). 'Transfer' timing (3, 5, or 6 days). 'Additional for cryopreservation' options (Irvine, Kitazato). 'Storage period' (up to 2 months or 5 years). A 'Prognosis after transfer' section shows overall pregnancy probability (95% for single pregnancy, 0% for multiple) and risks.

Рисунок 6. Скриншот СППВР с обрисовкой каждого блока системы

Опытная эксплуатация системы была проведена в клинике АО «Центр семейной медицины» г. Екатеринбург.

Валидация показала, что общая структура приложения соответствует ожидаемой, учтены все аспекты решаемой задачи, приложение понятно для пользователя и удобно в работе.

Для проверки на корректность работы системы (верификации) использовались данные о женщинах, прошедших 635 протоколов ВРТ с конца 2018 года по начало 2021 года. Информация об этих женщинах не использовалась при разработке системы. Группа здоровья ребенка оценена по диагнозам при рождении. Получены следующие результаты:

- 1) наступление беременности предсказано верно в 377 случаях (86,7 %);
- 2) способ родов предсказан верно в 218 случаях (65,1 %);
- 3) срок родов предсказан в неделях верно в 274 случаях (81,8 %) (ошибка предсказания ± 1 неделя);

4) осложнения текущей беременности предсказаны верно: истмико-цервикальная недостаточность – 391 случай (89,9 %), гипертензивные расстройства – 318 случаев (94,9 %), предлежание плаценты – 284 случаев (84,7 %), гестационный сахарный диабет – 346 случаев (79,5 %), нарушения количества околоплодных вод – 287 случаев (85,7 %), преждевременный разрыв плодных оболочек – 305 случаев (91,0 %);

5) группы здоровья рожденных детей предсказаны верно в 342 случаях (85,5 %).

Для оценки качества прогнозирования каждого из исходов построены соответствующие матрицы ошибок, рассчитаны чувствительность и специфичность, представлены ROC-кривые.

Клиника АО «Центр семейной медицины» г. Екатеринбург признала прохождение этапа опытной эксплуатации успешным, а программу рекомендованной для использования в лечебном процессе, о чем получен акт внедрения.

Таким образом, цель работы достигнута – разработано математическое и алгоритмическое обеспечение системы поддержки принятия врачебных решений для прогнозирования здоровья ребенка, зачатого при помощи ВРТ. Система позволяет осуществлять прогнозирование с уровнем точности, достаточным для клинического применения, а именно по результатам апробации средняя доля правильных ответов для основных исходов – 84,5% 95% ДИ [78,7;90,3].

ОСНОВНЫЕ РЕЗУЛЬТАТЫ РАБОТЫ И ВЫВОДЫ

В процессе диссертационного исследования решены следующие задачи:

1) проанализировано состояние разработок в области создания математического и алгоритмического обеспечения для прогнозирования исходов ВРТ. Результаты анализа говорят об актуальности поставленной задачи. Формализована постановка задачи принятия врачебных решений при проведении протокола ВРТ;

2) разработаны алгоритмы обработки, компоновки и анализа разнородных текстовых данных на примере данных медицинских организаций с целью исследования системных связей и закономерностей в данных и приведения их в удобный для применения машинного обучения вид;

3) разработаны методы и алгоритмы многоэтапного прогнозирования здоровья ребенка, зачатого при помощи ВРТ;

4) созданы прогностические модели машинного обучения и объединены в единую СППВР. Создано программное приложение, позволяющее работать с СППВР, и апробировать созданное математическое и алгоритмическое обеспечение в клинической практике;

5) тестирование СППВР на новых данных признано успешным.

Перспективы дальнейшей разработки темы исследования заключаются в применении созданного многоэтапного подхода и алгоритма подготовки данных к новым данным о проведенных протоколах ВРТ и о рожденных в результате детях; к прогнозированию здоровья детей, зачатых естественным путем. Планируется исследовать возможность использования предложенных подходов для моделирования медицинских задач, которые формализуются в виде многоэтапных процессов.

ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИОННОЙ РАБОТЫ

Статьи, опубликованные в рецензируемых научных журналах и изданиях, определенных ВАК РФ и Аттестационным советом УрФУ:

1. **Синотова С.Л.** Система поддержки принятия врачебных решений для прогнозирования исхода протокола вспомогательных репродуктивных технологий на различных этапах его проведения / **С.Л. Синотова**, С.И. Солодушкин, А.Н. Плаксина, В.А. Макутина // Моделирование, оптимизация и информационные технологии. – 2022. – № 10 (2). – 13 с. (1,01 п. л. / 0,8 п. л.).
2. **Синотова С.Л.** Программное приложение для предсказания здоровья ребенка, рожденного при помощи вспомогательных репродуктивных технологий / **С.Л. Синотова**, О.В. Лимановская, А.Н. Плаксина, В.А. Макутина // Моделирование, оптимизация и информационные технологии. – 2021. – № 9 (3). – 12 с. (0,82 п. л. / 0,31 п. л.).
3. **Sinotova S.L.** Selection from the obstetric history of women signs that affect the outcome of pregnancy obtained using IVF / **S.L. Sinotova**, O.V. Limanovskaya, A.N. Plaxina // AIP Conference Proceedings. – 2020. – Vol. 2313. – No. 070027. (0,32 п. л. / 0,21 п. л.). (Scopus, WoS).
4. **Синотова С.Л.** Сравнение эффективности различных методов отбора признаков для решения задачи бинарной классификации предсказания наступления беременности при проведении экстракорпорального оплодотворения / **С.Л. Синотова**, О.В. Лимановская, А.Н. Плаксина, В.А. Макутина // Моделирование, оптимизация и информационные технологии. – 2020. – № 8 (3). – 16 с. (1,27 п. л. / 0,32 п. л.).
5. **Синотова С.Л.** Применение корреляционного анализа для выявления факторов из анамнеза женщины, влияющих на исход беременности, полученной с помощью ЭКО / **С.Л. Синотова**, О.В. Лимановская, А.Н. Плаксина, В.А. Макутина // Моделирование, оптимизация и информационные технологии. – 2020. – № 8 (1). – 13 с. (0,85 п. л. / 0,42 п. л.).

Свидетельство о регистрации программы для ЭВМ:

6. Свидетельство о гос. регистрации программы для ЭВМ. Прогнозирование исходов и здоровья детей, зачатых при помощи вспомогательных репродуктивных технологий / Плаксина А.Н., **Синотова С.Л.**, Лимановская О.В., Макутина В.А. — № 2021618399; дата регистрации 26.05.2021 (Российская Федерация).

Другие публикации:

7. Плаксина А.Н. Факторы, определяющие исход беременности и здоровье детей, рожденных от женщин с заболеваниями мочеполовой системы при помощи вспомогательных репродуктивных технологий / А.Н. Плаксина, О.П. Ковтун, **С.Л. Синотова**, О.В. Лимановская, В.А. Макутина // Педиатрия им. Г.Н. Сперанского. – 2021. – Т. 100 (1). – С. 52–59. (0,83 п. л. / 0,17 п. л.).
8. Плаксина А.Н. Оценка здоровья детей, рожденных при помощи вспомогательных репродуктивных технологий, по данным региональных информационных медицинских систем / А.Н. Плаксина, О.П. Ковтун, О.Ю. Аверьянов, О.В. Кожевникова, **С.Л. Синотова**, О.В. Лимановская, В.А. Макутина // Неонатология: новости, мнения, обучение. – 2020. – № 8 (3). – С. 18–25. (0,90 п. л. / 0,13 п. л.).
9. Плаксина А.Н. Оценка физического развития и здоровья детей, зачатых при помощи ВРТ, имеющих ретинопатию недоношенных: популяционное одномоментное исследование / А.Н. Плаксина, О.П. Ковтун, Е.А. Степанова, Е.А. Дугина, В.А. Макутина, **С.Л. Синотова**, О.В. Лимановская // Российский педиатрический журнал. – 2020. – № 1 (2). – С. 18–24. (0,79 п. л. / 0,11 п. л.).
10. Ковтун О.П. Информационно-аналитические системы для оценки перинатальных исходов и состояния здоровья детей, рожденных при помощи вспомогательных репродуктивных технологий / О.П. Ковтун, А.Н. Плаксина, В.А. Макутина, Н.О. Анкудинов, Н.А. Зильбер, О.В. Лимановская, **С.Л. Синотова** // Российский вестник перинатологии и педиатрии. – 2020. – Т. 65(1). – С. 45–55. (0,72 п. л. / 0,10 п. л.).
11. **Синотова С.Л.** Отбор из акушерского анамнеза женщин признаков, влияющих на исход беременности, полученной с помощью ЭКО / **С.Л. Синотова**, О.В. Лимановская, А.Н. Плаксина // Сборник статей VII Международной молодежной научной конференции «Физика. Технологии. Инновации». – Екатеринбург: УрФУ, 2020. – С. 289–297. (0,36 п. л. / 0,12 п. л.).