

**СВЕДЕНИЯ О РЕЗУЛЬТАТАХ
ПУБЛИЧНОЙ ЗАЩИТЫ ДИССЕРТАЦИИ**

Гришиной Любови Сергеевны

**«Методы и алгоритмы интеллектуальной поддержки принятия решений в
медицинской практике на основе обработки естественных языков»
на соискание ученой степени кандидата технических наук
по научной специальности 2.3.1. Системный анализ, управление и
обработка информации, статистика**

ВЫПИСКА ИЗ ПРОТОКОЛА № 5

заседания диссертационного совета Д 24.2.352.03
от 27 сентября 2024 г.

Заседание проводил председатель диссертационного совета – доктор технических наук, профессор Боровский А.С.

Из 13 членов диссертационного совета присутствовали 13 человек, из них 12 докторов наук представляют научную специальность 2.3.1. Системный анализ, управление и обработка информации, статистика (технические науки)

1. Боровский Александр Сергеевич – председатель диссертационного совета, доктор технических наук, профессор, 2.3.1. Системный анализ, управление и обработка информации, статистика (технические науки);

2. Пищухин Александр Михайлович – заместитель председателя диссертационного совета, доктор технических наук, профессор, 2.3.1. Системный анализ, управление и обработка информации, статистика (технические науки);

3. Парфёнов Денис Игоревич – ученый секретарь диссертационного совета, кандидат технических наук, 2.3.1. Системный анализ, управление и обработка информации, статистика (технические науки);

4. Аралбаев Ташбулат Захарович – доктор технических наук, профессор, 2.3.1. Системный анализ, управление и обработка информации, статистика (технические науки);

5. Бахарева Надежда Федоровна (*в удаленной форме*) – доктор технических наук, профессор, 2.3.1. Системный анализ, управление и обработка информации, статистика (технические науки);

6. Болодурина Ирина Павловна (*в удаленной форме*) – доктор технических наук, профессор, 2.3.1. Системный анализ, управление и обработка информации, статистика (технические науки);

7. Зубкова Татьяна Михайловна – доктор технических наук, профессор, 2.3.1. Системный анализ, управление и обработка информации, статистика (технические науки);

8. Костин Владимир Николаевич – доктор технических наук, доцент, 2.3.1. Системный анализ, управление и обработка информации, статистика (технические науки);

9. Логунова Оксана Сергеевна (в удаленной форме) – доктор технических наук, профессор, 2.3.1. Системный анализ, управление и обработка информации, статистика (технические науки);

10. Припадчев Алексей Дмитриевич – доктор технических наук, профессор, 2.3.1. Системный анализ, управление и обработка информации, статистика (технические науки);

11. Соловьев Николай Алексеевич – доктор технических наук, профессор, 2.3.1. Системный анализ, управление и обработка информации, статистика (технические науки);

12. Султанов Наиль Закиевич (в удаленной форме) – доктор технических наук, профессор, 2.3.1. Системный анализ, управление и обработка информации, статистика (технические науки);

13. Тугов Виталий Валерьевич – доктор технических наук, доцент, 2.3.1. Системный анализ, управление и обработка информации, статистика, технические науки.

ПОВЕСТКА ДНЯ:

Публичная защита диссертации *Гришиной Любови Сергеевны* на тему «Методы и алгоритмы интеллектуальной поддержки принятия решений в медицинской практике на основе обработки естественных языков» на соискание ученой степени кандидата технических наук по научной специальности 2.3.1. Системный анализ, управление и обработка информации, статистика.

РЕШИЛИ:

По результатам публичной защиты на заседании 27 сентября 2024 г. диссертационный совет принял решение присудить Гришиной Любови Сергеевне ученую степень кандидата технических наук.

При проведении тайного голосования с использованием информационно-коммуникационных технологий диссертационный совет в количестве 13 человек, из них 12 докторов наук по научной специальности 2.3.1 Системный анализ, управление и обработка информации, статистика (технические науки), участвовавших в заседании, из 13 человек, входящих в состав совета, дополнительно введены на разовую защиту 0 человек, проголосовали: за – 13, против – 0.

Председатель
диссертационного совета
24.2.352.03
д-р техн. наук, профессор

Ученый секретарь
диссертационного совета
24.2.352.03
канд. техн. наук

27.09.2024



А.С. Боровский

Д.И. Парфёнов

ЗАКЛЮЧЕНИЕ ДИССЕРТАЦИОННОГО СОВЕТА 24.2.352.03,
СОЗДАННОГО НА БАЗЕ ФЕДЕРАЛЬНОГО ГОСУДАРСТВЕННОГО
БЮДЖЕТНОГО ОБРАЗОВАТЕЛЬНОГО УЧРЕЖДЕНИЯ ВЫСШЕГО
ОБРАЗОВАНИЯ «ОРЕНБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»
МИНИСТЕРСТВА НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ
ФЕДЕРАЦИИ, ПО ДИССЕРТАЦИИ
НА СОИСКАНИЕ УЧЕНОЙ СТЕПЕНИ КАНДИДАТА НАУК

аттестационное дело № _____

решение диссертационного совета от 27.09.2024 г. № 5

О присуждении Гришиной Любови Сергеевны, гражданину РФ, ученой степени кандидата технических наук.

Диссертация «Методы и алгоритмы интеллектуальной поддержки принятия решений в медицинской практике на основе обработки естественных языков» по научной специальности 2.3.1 Системный анализ, управление и обработка информации, статистика принята к защите 28.06.2024 г. (протокол заседания № 3) диссертационным советом 24.2.352.03, созданным на базе Федерального государственного бюджетного образовательного учреждения высшего образования «Оренбургский государственный университет» Министерства науки и высшего образования Российской Федерации, 460018, Оренбургская область, г. Оренбург, просп. Победы, д. 13, созданного приказом Минобрнауки России от 14 февраля 2023 г. № 235/нк, с изменениями в соответствии с приказом Минобрнауки России от 12 декабря 2023 г. № 2298/нк.

Соискатель Гришина Любовь Сергеевна 19 мая 1997 года рождения, в 2020 г. окончила магистратуру ФГБОУ ВО «Оренбургский государственный университет» по направлению подготовки 01.04.02 Прикладная математика и информатика, профиль «Оптимизация и оптимальное управление». В 2023 году освоила программу подготовки научно-педагогических кадров в аспирантуре ФГБОУ ВО «Оренбургский государственный университет» по направлению 02.06.01 Компьютерные и информационные науки (профиль Системный анализ, управление и обработка информации (в информатике, вычислительной технике и

автоматизации). Диплом об окончании аспирантуры выдан ФГБОУ ВО «Оренбургский государственный университет» в 2023 году.

Работает старшим преподавателем Федерального государственного бюджетного образовательного учреждения высшего образования «Оренбургский государственный университет» Министерства науки и высшего образования Российской Федерации.

Диссертация выполнена на кафедре прикладной математики ФГБОУ ВО «Оренбургский государственный университет» Министерства науки и высшего образования Российской Федерации.

Научный руководитель – доктор технических наук, профессор, заведующий кафедрой прикладной математики Ирина Павловна Болодурин, ФГБОУ ВО «Оренбургский государственный университет».

Официальные оппоненты:

1. Куприянов Александр Викторович, доктор технических наук, доцент, директор института информатики и кибернетики Федерального государственного автономного образовательного учреждения высшего образования «Самарский национальный исследовательский университет имени академика С.П. Королева»

2. Каширина Ирина Леонидовна, доктор технических наук, доцент, профессор кафедры математических методов исследования операций Федерального государственного бюджетного образовательного учреждения высшего образования «Воронежский государственный университет»

дали положительные отзывы на диссертацию.

Ведущая организация Федеральное государственное бюджетное образовательное учреждение высшего образования «Уфимский университет науки и технологий», г. Уфа, в своем положительном отзыве, подписанном Шахматовой Гюзель Радиковной, д.т.н., доцентом, заведующим кафедрой «Вычислительной математики и кибернетики», утвержденным проректором по инновационной деятельности ФГБОУ ВО «Уфимский университет науки и технологий», к.т.н., доцентом, Г.К. Агеевым, указала, что область исследования диссертационной работы Гришиной Л.С. соответствует научной специальности 2.3.1 Системный

анализ, управление и обработка информации, статистика. Диссертационная работа обладает научной новизной и практической значимостью, написана автором самостоятельно, структура работы соответствует требованиям ГОСТ Р 7.0.11-2011 и в полной мере отражает основные результаты и положения, выносимые на защиту. Диссертация Гришиной Любови Сергеевны на соискание ученой степени кандидата технических наук является законченной научно-квалификационной работой, в которой содержится решение научной задачи, имеющей значение для развития теории и практики разработки программных комплексов интеллектуальной поддержки принятия решений в медицинской практике на основе методов обработки естественных языков.

Диссертация отвечает требованиям ВАК РФ и критериям п.9 Положения «О присуждении ученых степеней», утвержденного постановлением Правительства Российской Федерации от 24 сентября 2013 г. № 842 (в ред. Постановления Правительства РФ от 25.01.2024 № 62), а ее автор, Гришина Любовь Сергеевна, заслуживает присуждения ученой степени кандидата технических наук по научной специальности 2.3.1. Системный анализ, управление и обработка информации, статистика.

Соискатель имеет более 100 опубликованных работ, в том числе по теме диссертации опубликовано 8 работ, из них 3 – в изданиях, рекомендованных ВАК РФ и отечественных изданиях, которые входят в международные базы данных и системы цитирования, а также 2 работы в изданиях, индексируемых Scopus и Web of Science; получено 1 свидетельство о государственной регистрации программы для ЭВМ.

Наиболее значимые работы по теме диссертации:

В рецензируемых журналах из списка ВАК и отечественных изданиях, которые входят в международные базы данных и системы цитирования

1. Разработка модели генерации клинических рекомендаций для пациентов на основе неструктурированных текстовых данных / Л.С. Гришина, И.П. Болодурина, // Научно-технический вестник Поволжья, 2023. - № 8. - С. 53-56.

2. Разработка модели управления потоком пациентов с

сердечно-сосудистыми заболеваниями методами интеллектуального анализа данных / И.П. Болодурина, А.М. Назаров, Д.И. Кича, Л.С. Забродина (Гришина), А.Ю. Жигалов // Вестник ЮУрГУ. Серия: Компьютерные технологии, управление, радиоэлектроника, 2020. - Т. 20, № 2. - С. 105-115.

3. L. S. Grishina, A. Yu. Zhigalov, I. P. Bolodurina, E. L. Borshhuk, D. N. Begun, Yu. V. Varennikova, "Investigation of the efficiency of graph data representation for a cardiovascular disease predictive model by deep learning methods", *Dal'nevost. Mat. Zh.*, 22:2 (2022), 179–184.

В изданиях, индексируемых в Scopus и Web of Science

4. Bolodurina, I.; Shukhman, A.; Legashev, L.; Grishina, L.; Zhigalov, A. Extracting and Processing of Russian Unstructured Clinical Texts for a Medical Decision Support System. *Eng. Proc.* 2023, 33, 41.

5. Development of a Model for Predicting Treatment of Cardiovascular Diseases Based on Machine Learning Methods / I. P. Bolodurina, D. I. Parfenov, A. Yu. Zhigalov, L. S. Zabrodina (Grishina) // Proceedings of the 2nd International Scientific and Practical Conference "Modern Management Trends and the Digital Economy: from Regional Development to Global Economic Growth", 16-17 April, 2020, Yekaterinburg, Russia - P. 984-989. - 6 с.

В диссертации отсутствуют недостоверные сведения об опубликованных соискателем ученой степени работах, в которых изложены основные научные результаты диссертации; соискатель ссылается на авторов и источники заимствования.

На диссертацию и автореферат поступили **положительные** отзывы:

- **ведущей организации** Федеральное государственное бюджетное образовательное учреждение высшего образования «Уфимский университет науки и технологий», г. Уфа. *Замечания:* 1) Для оценки точности моделей машинного обучения при прогнозировании укрупненной группы заболевания использована метрика сбалансированной точности (*balanced accuracy*), которая позволяет равнозначно оценивать редкие и наиболее распространенные классы заболеваний. Однако, эффективность прогнозных моделей в медицине также необходимо

рассматривать с точки зрения клинической значимости ошибок, чтобы оценить ложно-положительные и ложно-отрицательные результаты прогноза отдельных заболеваний (метрики *Precision*, *Recall* и *F1-Score*). 2) При построении интеллектуальной модели генерации индивидуальных листов назначений и рекомендаций к лечению автором сделано допущение, что протоколы МИС и уровень квалификации врачей соответствует стандартам. Но данное условие целесообразно подтвердить на практике, например, по результатам экспертной оценки на соответствие протоколов клиническим рекомендациям Министерства здравоохранения РФ. 3) В работе производится оценка эффективности принятия решений в медицинской практике на основе анализа данных средней длительности и структуры приёма больных врачом-терапевтом с учетом основных трудовых операций, необходимых к проведению на первичном осмотре. При этом не рассматривается вопрос об адекватности результатов прогнозирования диагноза и генерации рекомендаций на повторном приеме или при обращении за коррекцией лечения незаконченного клинического случая. 4) В качестве рекомендации по расширению области применения разработанного метода, на наш взгляд, целесообразно рассмотреть возможность применения предложенного подхода в отношении более широкого класса заболеваний (болезней нервной системы, эндокринной системы, расстройства питания, нарушения обмена веществ и др.), а также расширения обучающего набора данных за счет включения медико-географических особенностей территорий отдельных регионов, не ограничиваясь пациентами Оренбургской области.

- **официального оппонента** доктора технических наук, доцента, Куприянова Александра Викторовича, директора института информатики и кибернетики Федерального государственного автономного образовательного учреждения высшего образования «Самарский национальный исследовательский университет имени академика С.П. Королева». *Замечания:* 1) В рамках разработки модели данных амбулаторных карт пациентов для обработки документов МИС автор выбрал за основу иерархическую модель данных (раздел 2.2), которая среди прочих особенностей характеризуется ограниченной гибкостью и сложностью управления.

При этом в диссертационной работе отсутствует исследование альтернативных моделей, которые могут быть более эффективными в задачах интеллектуального анализа данных. Выбор модели данных должен быть обусловлен конкретными требованиями к доступным данным и приложениям. 2) В главе 3 указывается, что для оценки качества моделей машинного обучения прогнозирования укрупненных групп заболеваний проводится перекрестная проверка (k-fold кросс-валидация). Важно отметить, что выбор гиперпараметра k данного алгоритма может привести к некорректному представлению о качестве модели, например, к оценке с высокой дисперсией. При этом в рамках исследования значение для k зафиксировано ($k = 5$) и не предоставлено подробное исследование данной конфигурации. 3) В описании результата из таблицы 3.4 указано, что «Модель [логистической регрессии] имеет наименьшее стандартное отклонение». При этом, из таблицы очевидно, что у большинства остальных моделей (за исключением LinearSVC) отклонение меньше. 4) В рамках реализации алгоритма автоматической генерации индивидуальных листов назначений и рекомендаций автор утверждает, что для согласования с клиническими стандартами обучение языковой модели проводится на основе архивных данных (раздел 4.1). Однако, для используемого набора не проведено исследование влияния истории наблюдения за состоянием здоровья пациента в течении всего периода лечения и его длительности, которое напрямую связано с оценкой актуальности данных и может влиять на качество моделей машинного обучения. 5) Отсутствует обоснование вывода о том, что полученные значения метрики BLEU по униграммам, биграммам и триграммам «довольно большие» (раздел 4.4), для этих показателей отсутствуют референсные значения. 6) Отсутствует обоснование корректности выбора метрики для конкретной прикладной области. Например, изменение дозировки препарата (таблица 4.1) в нескольких цифрах повлечёт за собой минимальное снижение метрики BLEU, но с точки зрения врачебного назначения может быть жизненно важным отличием. 7) В главе 5 указывается, что для обеспечения безопасности и конфиденциальности медицинских данных использованы методы SSL-шифрования и OAuth аутентификации. Однако в диссертационной работе отсутствует анализ

современных методов шифрования и аутентификации, которые зависят от конкретных требований проекта и уровня необходимой безопасности.

- **официального оппонента** доктора технических наук, доцента Кашириной Ирины Леонидовны, профессора кафедры математических методов исследования операций Федерального государственного бюджетного образовательного учреждения высшего образования «Воронежский государственный университет».

Замечания: 1) В главе 1 указывается, что обработка больших данных медицинских информационных систем и построение языковых моделей генерации текста требует использования большого объема вычислительных ресурсов GPU для проведения обучения моделей. Однако, в диссертационной работе не представлено подробное описание использованного оборудования и его характеристик, достаточных для реализации разрабатываемых моделей ИИ. 2) Автор утверждает, что система ЕГИСЗ имеет сложную техническую инфраструктуру и среди прочих проблем анализа данных этой системы выделяет наличие в документах опечаток, некорректных сокращений и многовариантные аббревиатуры (раздел 1.4). Однако, количественная оценка исходных данных на наличие выделенных проблем не приводится и не указываются методы и алгоритмы их обработки. 3) В рамках реализации алгоритма выгрузки файлов протоколов из медицинской информационной системы автор указывает, что реализует подход с выделением нескольким скриптам непересекающегося пула протоколов случаев. Однако не до конца остается ясным, на каком этапе реализуется распараллеливание в алгоритме (табл. 2.6). 4) В главе 3 указывается, что для построения модели прогнозирования заболеваний по МКБ-10 на основе слабоструктурированных данных за основу выбирается укрупненная группа заболеваний. Однако, для эффективной генерации рекомендации к лечению не менее важно знать уточненный шифр конкретного заболевания, который автором не учитывается. 5) В работе имеется ряд опечаток и стилистических неточностей (стр. 9, 13, 35, 48, 51, 74 и т.д.).

Получено **пять положительных** отзывов на автореферат:

1. **Федеральное государственное бюджетное образовательное учреждение высшего образования «Поволжский государственный университет**

телекоммуникаций и информатики» (г. Самара), доктор технических наук, профессор, Заслуженный работник Высшей школы РФ, заведующий кафедрой управления в технических системах **Тарасов Вениамин Николаевич**. Замечания: Рис. 5 «Матрица ошибок для модели Logistic Regression»: не пояснено, какой объем данных относительно исходного набора данных выбран для обучения, какой для валидации и построения матрицы ошибок; Стр. 12-13. Для токенизации текста использован метод ВРЕ (Byte Pair Encoding), однако, после применения алгоритма ВРЕ, не указаны характеристики сформированного словаря V для корпуса текстов D .

2. **Федеральное государственное бюджетное образовательное учреждение высшего образования «Московский авиационный институт (национальный исследовательский университет)»** (г. Москва), кандидат технических наук, доцент, доцент кафедры «Космические телекоммуникации» МАИ, руководитель центра визуализации и спутниковых информационных технологий ФНЦ НИИСИ РАН **Мамросенко Кирилл Анатольевич**. Замечания: в тексте автореферата не представлена полная структура заявленного алгоритма рекурсивного считывания информации из медицинских протоколов в формате XML. Указано, что «реализована стратегия вложенных словарей для сохранения данных и построены соответствующие иерархические деревья записей». При этом отсутствует информация о технических характеристиках – времени обработки и объема памяти для обработки и хранения данных.

3. **Федеральное государственное бюджетное образовательное учреждение высшего образования «Комсомольский-на-Амуре государственный университет»** (г. Комсомольск-на-Амуре), кандидат технических наук, доцент, доцент кафедры «Прикладная математика» **Гордин Сергей Александрович**. Замечания: в автореферате приведена формула не дифференцируемого критерия BLEU (стр. 14), который является вполне надёжным, но приведенные значения по n -граммам не сопровождаются анализом полученных результатов; хотя автором и предложены вполне надёжные алгоритмы прогнозирования групп заболеваний и генерации индивидуальных листов назначений, но какова причина снижения работы с документацией в среднем всего

на 6,9%?

4. **Федеральное государственное бюджетное образовательное учреждение высшего образования «Воронежский государственный технический университет»** (г. Воронеж), доктор технических наук, профессор, Декан факультета экономики, менеджмента и информационных технологий, заведующий кафедрой управления **Баркалов Сергей Алексеевич**. Замечания: из автореферата не понятно, чем обоснован выбор автором показателя затрат рабочего времени при работе с документацией на приеме врачей-терапевтов в качестве обобщающего показателя эффективности принятия решений в медицинской практике; не ясно, использовались ли в исследованиях при обработке данных МИС протоколы незавершенных клинических случаев заболеваний, а также учитывались ли метаданные (например, время визита, причина обращения). В случае неоднородности исходных данных, применялись ли методы обработки и нормализации для обеспечения их совместимости и возможности анализа?

5. **Федеральное государственное бюджетное учреждение науки «Институт автоматизации и процессов управления Дальневосточного отделения Российской академии наук»** (г. Владивосток), доктор технических наук, член-корреспондент РАН, заместитель директора по научной работе **Грибова Валерия Викторовна**. Замечания: как диссертант планирует обеспечивать актуальность анализируемой информации, ведь исторические данные быстро устаревают (известно, что методы лечения часто меняются, - в среднем, по оценкам специалистов, каждые полгода происходит корректировка метода лечения, по некоторым заболеваниям чаще)? Для формирования персонифицированного лечения требуется учитывать множество различных признаков, которые часто отсутствуют в медицинских записях (известно, что качество и полнота их очень низкая), как автор планирует «обходить» эту проблему, учитывая то, что лечение должно быть персонифицированным? Считаю, что в работе не сформулированы ограничения применяемого метода: необходимо четко обозначить перечень заболеваний и/или методов лечения, для которых он применим. Вместе с тем очевидно, что диссертантом проделана большая работа, использованы современные

методы и подходы.

Выбор официальных оппонентов и ведущей организации обосновывается их достижениями в данной отрасли наук, наличием публикаций в соответствующей сфере исследования и способностью определить научную и практическую ценность диссертации.

Диссертационный совет отмечает, что на основании выполненных соискателем исследований:

- **разработаны:** концептуальная модель анализа клинических данных и поддержки принятия решений для автоматизации процессов заполнения ЭМК с учетом индивидуализации этапов оказания медицинских услуг; иерархическая модель данных амбулаторных карт пациентов для обеспечения семантической интероперабельности при обработке ЭМК; метод и алгоритм прогнозирования укрупненных групп заболеваний на основе слабоструктурированных текстовых данных ЭМК пациентов; метод и алгоритм автоматической генерации индивидуальных листов назначений и рекомендаций к лечению для автоматизации процессов заполнения документов; прототип автоматизированного программного комплекса интеллектуальной поддержки принятия решений в медицинской практике;

- **предложен:** рекурсивный подход к обработке разношаблонных файлов протоколов медицинских случаев лечения с расширением XML, включающий алгоритмы автоматической выгрузки обезличенных данных ЭМК и извлечения информации на основе последовательного анализа наличия содержимого и создания дерева записи оказанной услуги в медицинской организации; иерархическая модель данных амбулаторных карт пациентов, описывающая отношения между сущностями различных слабоструктурированных документов медицинских информационных систем, необходимых для структурирования данных, обеспечения семантической интероперабельности, а также построения обучающего набора данных и разработки систем искусственного интеллекта; метод прогнозирования группы заболеваний пациентов на основе алгоритмов обработки естественных языков (NLP) и модели логистической регрессии, обученной на базе уникального

узкоспециализированного корпуса медицинских текстов; метод автоматической генерации индивидуальных шаблонов листа назначений и рекомендаций к лечению на основе применения современной предобученной нейросетевой модели трансформеров GPT-3;

- экспериментально **доказана** эффективность использования разработанных методов и алгоритмов прогнозирования укрупненной группы заболеваний по МКБ-10 и генерации индивидуального листа назначений и рекомендаций к лечению на множестве реальных деперсонализированных данных электронных медицинских карт, полученных из базы данных медицинских организаций Оренбургской области.

Теоретическая значимость исследования обоснована тем, что:

- **разработана** иерархическая модель данных, представленная кортежем множеств текстовых документов и описанием уровней их иерархии, которая отражает отношения между основными сущностями слабоструктурированных данных информационных систем и обеспечивает их совместимость в процессе обмена и обработки;

- **применительно к проблематике диссертации** решены задачи разработки и реализации концептуальной модели анализа клинических данных и поддержки принятия решений в медицинской практике, обеспечивающей интеграцию различных классов статистических и генеративных моделей искусственного интеллекта применительно к обработке больших слабоструктурированных данных;

- **изложено** описание методов и моделей обработки естественных языков, прогнозирования классов заболеваний и построения генеративных моделей языкового моделирования, объединенных предложенной в работе концептуальной моделью анализа клинических данных, учитывающей границы компетенций экспертов в системе поддержки принятия врачебных решений;

- **раскрыты** проблемы неоднозначности узкоспециализированных терминов и многовариативности выхода моделей генерации текста при обработке текстовых документов информационных систем для решения задач поддержки принятия решений;

- **изучена** возможность реализации «однонаправленных» языковых моделей

генерации текста и модификации современных предобученных нейросетевых моделей трансформеров для решения задачи генерации индивидуальных листов назначений и рекомендаций к лечению;

- **проведена модернизация** существующих алгоритмов машинного обучения и языковых моделей на базе архитектуры трансформер для решения задач классификации и языкового моделирования: исследована архитектура, конфигурации и параметры моделей для обеспечения наиболее высоких показателей сбалансированной точности и высокой степени близости сгенерированного текста к эталонному.

Значение полученных соискателем результатов исследования для практики подтверждается тем, что:

- **разработаны** методы прогнозирования укрупненных групп заболеваний на основе слабоструктурированных данных ЭМК и генерации индивидуальных листов назначений и рекомендаций к лечению. Результаты **внедрены** в практику медицинских исследований организационно-методического отдела ГАУЗ «Оренбургской областной клинической больницы имени В.И. Войнова» и ГАУЗ «Бузулукской больницы скорой медицинской помощи им. академика Н.А. Семашко», а также в учебный процесс ФГБОУ ВО «Оренбургского государственного медицинского университета» на кафедре «Общественного здоровья и здравоохранения №1»;

- **определены** пределы практического использования, отраженные в доступном для прогнозирования множестве укрупненных групп заболеваний пациентов и множестве входных жалоб пациентов, отражающего формат заполнения протоколов ЭМК, а также обозначены перспективы будущих исследований, затрагивающих использование языковых архитектур большей размерности и предобученные модели ИИ на специализированных медицинских данных;

- **создан** прототип автоматизированного программного комплекса интеллект-туальной поддержки принятия решений в медицинской практике. Применение программной реализации разработанных методов позволило:

производить автоматизированный анализ состояния пациента и прогнозировать укрупненную группу заболевания с последующей генерацией индивидуального листа назначений и рекомендаций к лечению, что позволило снизить затраты рабочего времени на написание документации;

- **представлены** рекомендации по использованию веб-фреймворка для доступа к построенным алгоритмам и моделям ИИ и предложения по дальнейшему совершенствованию СППР, включающие расширение прогнозируемого множества заболеваний.

Оценка достоверности результатов исследования выявила:

- **для экспериментальных работ** в процессе обучения предложенных моделей машинного обучения и моделей глубокого обучения, а также работы прототипа системы поддержки принятия решений в медицинской практике использовалось серверное оборудование научно-исследовательского института цифровых интеллектуальных технологий Федерального государственного бюджетного образовательного учреждения высшего образования «Оренбургский государственный университет»;

- **теория (методы и модели)** основаны на использовании известных положений, подходов и принципов обработки больших слабоструктурированных текстовых данных, методов обработки естественного языка, математической статистики и нейросетевого моделирования, а также результатах вычислительных экспериментов;

- **идея базируется** на результатах анализа слабоструктурированных текстовых данных электронных медицинских карт из информационных систем, информация в которых сохраняется в различных шаблонах представления в различное время посещения пациентами медицинских учреждений и с различным временным интервалом;

- **использовано** сравнение полученных с помощью предложенных методов и моделей результатов с результатами рассмотренных в диссертации методов и моделей прогнозирования укрупненной группы заболеваний по МКБ-10 и генерации индивидуального листа назначений и рекомендаций к лечению;

- **установлено**, что метод прогнозирования группы заболеваний с использованием методов NLP имеет сбалансированную точность 85,20% (стандартное отклонение при перекрестной проверке $\pm 1.07\%$), метод генерации индивидуальных листов назначений и рекомендаций к лечению в рамках диагностированных заболеваний в среднем имеет метрику BLEU1 = 0,668 и BLEU2 = 0,357. При работе с документацией на основе результатов СППР затраты рабочего времени снизились в среднем на 6,9%;

- **использованы** современные методы *обработки* больших разнородных медицинских данных, методы системного и структурного *анализа*, методы обработки естественного языка, машинного обучения и нейросетевые технологии, а также *оценивания эффективности* разработанного прототипа системы поддержки принятия управленческих решений.

Личный вклад соискателя состоит в: анализе текущего уровня исследований по проблеме извлечения и структурирования знаний из ЭМК для интеллектуальной поддержки принятия решений при диагностике и лечении заболеваний; разработке концептуальной модели анализа клинических данных и поддержки принятия решений; разработке иерархической модели данных амбулаторных карт пациентов для сбора информации из разношаблонных XML-документов; разработке метода прогнозирования укрупненной группы заболеваний пациентов на основе алгоритмов обработки естественных языков и модели логистической регрессии; разработке метода автоматической генерации индивидуальных шаблонов листа назначений и рекомендаций к лечению на основе модели глубокого обучения GPT-3; разработке прототипа системы поддержки принятия решений путем объединения в единый процесс подпроцессов, начиная от первичной обработки данных МИС, формирования исходного датасета на основе рекурсивного подхода к анализу разношаблонных документов, и заканчивая обучением моделей машинного обучения для прогнозирования заболеваний и языковых моделей генерации назначений и рекомендаций к лечению; проведением экспериментов, *выполненных автором лично*, а также в обработке и интерпретации экспериментальных данных, подготовке научных статей и материалов докладов на российских и международных

научных конференциях.

Диссертационный совет пришел к выводу о том, что в диссертации:

- соблюдены установленные Положением о присуждении ученых степеней критерии, которым должна отвечать диссертация на соискание ученой степени;
- отсутствуют недостоверные сведения об опубликованных соискателем ученой степени работах, в которых изложены основные научные результаты диссертации;
- соискатель ссылается на авторов и источники заимствования;
- оригинальность диссертационной работы составляет 72,23%.

Диссертационная работа Гришиной Любови Сергеевны «Методы и алгоритмы интеллектуальной поддержки принятия решений в медицинской практике на основе обработки естественных языков» соответствует п. 9 Положения о присуждении ученых степеней (утвержденного Постановлением Правительства Российской Федерации от 24 сентября 2013 года № 842, в редакции с изменениями), предъявляемых к кандидатским диссертациям.

Тема и содержание исследований соответствует паспорту научной специальности ВАК 2.3.1. Системный анализ, управление и обработка информации, статистика.

Таким образом, представленная диссертация на соискание ученой степени кандидата технических наук является научно-квалификационной работой, в которой изложено решение научной задачи, имеющей значение для развития теории и практики разработки программных комплексов интеллектуальной поддержки принятия решений в медицинской практике на основе методов обработки естественных языков.

В ходе защиты диссертации были высказаны замечания. Соискатель Гришина Л.С. согласилась с замечаниями и привела собственную аргументацию.

На заседании 27.09.2024 г. диссертационный совет принял решение: за решение научной задачи, имеющей значение для развития теории и практики разработки программных комплексов интеллектуальной поддержки принятия решений в медицинской практике на основе методов обработки естественных

языков, присудить Гришиной Л.С. ученую степень кандидата технических наук по научной специальности 2.3.1. Системный анализ, управление и обработка информации, статистика.

При проведении тайного голосования диссертационный совет в количестве 13 человек, из них 12 докторов наук (12 докторов наук по научной специальности 2.3.1 – Системный анализ, управление и обработка информации, статистика), участвовавших в заседании, из 13 человек, входящих в состав совета, дополнительно введены на разовую защиту - 0 человек, проголосовали: за - 13, против - 0.

Председатель
диссертационного совета



А.С. Боровский

Ученый секретарь
диссертационного совета

Д.И. Парфёнов

27 сентября 2024 года