

ОТЗЫВ

официального оппонента, доктора технических наук, доцента **Куприянова Александра Викторовича** на диссертационную работу **Гришиной Любови Сергеевны** на тему: «Методы и алгоритмы интеллектуальной поддержки принятия решений в медицинской практике на основе обработки естественных языков», представленную на соискание ученой степени кандидата технических наук по специальности 2.3.1 Системный анализ, управление и обработка информации, статистика

1. Актуальность исследования

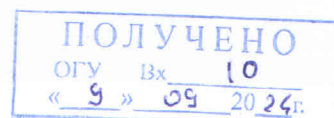
В числе важнейших задач, определенных в рамках стратегии научно-технологического развития Российской Федерации до 2030 г., описана необходимость внедрения систем искусственного интеллекта в здравоохранение с целью повышения доступности и качества медицинских услуг. Стратегия развития акцентирует внимание на необходимости перехода к персонализированным методам лечения, позволяющим оптимизировать процессы и управление ресурсами в медицинских учреждениях. Разработка интеллектуальных систем поддержки принятия врачебных решений позволит быстрее анализировать индивидуальные данные пациентов и разрабатывать более эффективные и персонализированные схемы лечения.

Как справедливо отмечает автор диссертации, в сфере здравоохранения, одной из самых значимых проблем, требующих поиска оптимизационных решений, является извлечение знаний из слабоструктурированных данных. Ввиду того, что большая часть медицинской информации представлена в виде текстов (например, врачебные заключения, истории болезни), применение методов обработки естественного языка (NLP) в этой области может помочь автоматизировать рутинные процессы, такие как кодирование диагнозов, составление отчетов и ведение документации, освобождая врачей для более важных задач.

Необходимо отметить, что автором диссертации в рамках 1 главы проведена серьезная аналитическая работа и, как следствие, подтверждена актуальность исследуемой проблематики, сформулированы цель и задачи исследования.

2. Обоснованность, достоверность и новизна научных положений, выводов и рекомендаций

Степень обоснованности научных положений, выводов и рекомендаций обеспечивается за счет корректного использования в работе методов первичной обработки данных, их анализа, правил определения зависимостей и математического моделирования, а также подтвержденной адекватностью



разработанных интеллектуальных моделей. Кроме того, степень обоснованности и достоверности обеспечивается согласованностью поставленных задач, проведенных исследований и сделанных выводов.

В работе использована качественная и достоверная информационно-эмпирическая база, содержащая большой объем данных, что также подтверждает обоснованность полученных научных положений, выводов и рекомендаций.

Для достижения поставленной цели в работе решены пять задач, решение каждой последующей задачи основывается на использовании предыдущих этапов, что подтверждает их взаимосвязь.

Доказательная база основных положений диссертационной работы усиливается достаточно обширным и убедительным списком использованных источников, а также работами самого соискателя, в которых отсутствуют противоречия с результатами ранее выполненных исследований.

Новизна исследования заключается в следующих полученных результатах:

- предложена концептуальная модель анализа клинических данных и поддержки принятия решений для автоматизации процессов заполнения ЭМК, позволяющая структурировать текстовые данные и внедрять интеллектуальные модели для формирования рекомендаций к лечению диагностированных заболеваний;

- разработана иерархическая модель данных амбулаторных карт пациентов, а также рекурсивные алгоритмы обработки разношаблонных XML-документов МИС для обеспечения семантической интероперабельности;

- разработаны метод и алгоритм прогнозирования группы заболеваний пациентов на основе узкоспециализированного неразмеченного корпуса текстов, методов обработки естественных языков и моделей машинного обучения (сбалансированная точность прогнозирования укрупненной группы заболеваний по МКБ-10 составляет $85,20\% \pm 1,07\%$);

- разработаны метод и алгоритм автоматической генерации индивидуальных листов назначений и рекомендаций к лечению для автоматизации процессов заполнения документов, который на основе предобученных языковых моделей трансформеров продемонстрировал высокую эффективность с метрикой BLEU1 = 0,668 и BLEU2 = 0,357;

- построен прототип автоматизированного программного комплекса интеллектуальной поддержки принятия решений в медицинской практике, который при работе с документацией снижает затраты рабочего времени в среднем на 6,9%.

В целом по содержанию результаты и выводы отражают решение основных задач исследования, являются обобщением теоретических и экспериментальных результатов и представляют несомненную ценность для науки и практики. Таким

образом, все научные положения, выводы и рекомендации обоснованы и достоверны.

3. Оценка содержания диссертации и ее завершенности

Диссертация состоит из введения, пяти разделов, заключения, списка литературы и приложений.

Работа содержит логически связанное подробное описание решения научных задач, поставленных в диссертации, применение результатов которых обеспечивает повышение эффективности принятия решений в медицинской практике за счет анализа слабоструктурированной текстовой информации электронных-медицинских карт методами обработки естественных языков.

Для постановки задач, решение которых отражено в диссертации, автором выполнен обзор научных работ, нормативной и методической документации в области здравоохранения. В рамках первой главы рассматривается опыт реализации современных инструментов для прогнозирования заболеваний и генерации индивидуальных листов назначений и рекомендаций. Кроме того, описана проблематика обработки слабоструктурированных данных медицинских информационных систем при решении задач поддержки принятия решений.

Во второй главе представлена концептуальная модель анализа клинических данных и поддержки принятия решений, а также иерархическая модель данных амбулаторных карт пациентов для обработки разношаблонных документов. Разработаны алгоритмы автоматической выгрузки данных и извлечения информации из разнородных XML-документов.

В рамках третьей главы проведена формализация задачи прогнозирования укрупненных групп заболеваний на основе методов машинного обучения, представлены алгоритмы формирования векторного представления текстовых данных, а также алгоритмы машинного обучения для прогнозирования укрупненных групп заболеваний. Эффективность моделей машинного обучения подтверждена экспериментально на основе метода перекрёстной проверки.

В четвертой главе решается задача языкового моделирования для автоматической генерации текста. Представлены алгоритмы токенизации листа назначений и рекомендаций, описаны языковые модели на базе архитектуры трансформер, проведено обучение моделей, эффективность которых подтверждена на основе оценки недифференцируемого критерия.

В пятой главе представлена внутренняя структура компонентов автоматизированного программного комплекса интеллектуальной поддержки принятия врачебных решений при диагностике и лечении заболеваний. Описаны подробно инструменты взаимодействия с внешними информационными системами. Эффективность разработанного программного комплекса исследована на основе

оценки затрат рабочего времени при внедрении результатов работы в конкретных медицинских организациях.

Корректность содержания и завершенность диссертационной работы Л.С. Гришиной подтверждается следующими признаками.

3.1. В диссертации присутствуют все необходимые атрибуты: анализ состояния вопроса с формулированием проблемы, на решение которой направлена работа; аналитические исследования; методика проведения исследований; результаты экспериментальных исследований; методические разработки; общие выводы и рекомендации; список использованной литературы; приложения.

3.2. Автореферат кратко, но полно и всеобъемлюще передает все основные идеи выполненной работы. Анализ автореферата вполне позволяет сделать все необходимые выводы о качестве выполненной работы.

3.3. Цель диссертационного исследования достигнута; все поставленные задачи успешно решены.

Таким образом, рассматриваемая диссертация Л.С. Гришиной является завершенной научно-квалификационной работой, обладающей внутренним единством, соответствующей научной и практической ценностью.

4. Полнота публикаций по теме исследования

Основные результаты, выносимые на защиту, достаточно полно отражены в 8 печатных работах автора, из них 3 опубликованы 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. Диссертация выполнена на современном научном уровне. Представленные материалы изложены в логической последовательности.

5. Соответствие автореферата содержанию диссертации

Автореферат диссертации в достаточной степени отражает основные положения диссертации. Стиль изложения способствует пониманию диссертации и позволяет объективно оценить личный вклад автора в полученные результаты исследования. Автореферат и диссертация оформлены и иллюстрированы в соответствии с принятыми правилами.

6. Замечания по диссертационной работе

6.1. 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.

6.2. 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.

6.3. 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.

6.4. 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.

6.6. 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.

6.7. 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.

6.8. 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.

Следует отметить, что приведенные выше замечания не снижают высокого качества исследований, выводов, результатов диссертации, её научной новизны и практической значимости.

7. Заключение

В диссертационной работе Гришиной Любови Сергеевны изложены новые научно обоснованные технологические решения и разработки, направленные на интеллектуальную поддержку принятия решений при диагностике и лечении заболеваний на основе применения методов обработки естественных языков для анализа электронных-медицинских карт пациентов, внедрение которых имеет существенное значение для развития цифрового здравоохранения.

Диссертация Гришиной Л.С. «Методы и алгоритмы интеллектуальной поддержки принятия решений в медицинской практике на основе обработки естественных языков» является законченной научно-квалификационной работой и полностью соответствует пунктам 9-11, 13, 14 «Положения о порядке присуждения ученых степеней», утвержденного Постановлением Правительства Российской Федерации №842 от 24 сентября 2013 года, а ее автор, Гришина Любовь Сергеевна, заслуживает присуждения ученой степени кандидата технических наук по специальности 2.3.1. «Системный анализ, управление и обработка информации, статистика».

Официальный оппонент:

Директор института информатики и кибернетики
федерального государственного автономного
образовательного учреждения высшего образования
«Самарский национальный исследовательский
университет имени академика С.П. Королёва»,
доктор технических наук, доцент



Куприянов Александр Викторович

15 августа 2024 года

ФИО: Куприянов Александр Викторович,

Адрес: 443086, Самарская область, г. Самара, Московское шоссе, д. 34.

Контактная информация: тел.: +7 (846) 335-18-26, e-mail: ssau@ssau.ru

Организация, должность: Федеральное государственное автономное
образовательное учреждение высшего образования «Самарский национальный
исследовательский университет имени академика С.П. Королёва», директор
института информатики и кибернетики, д.т.н.

Научная специальность докторской диссертации 05.13.18 – Математическое
моделирование, численные методы и комплексы программ

Дата составления отзыва: 15.08.2024 г.



Подпись Куприянов А.В. удостоверяю
Ученый секретарь Самарского университета



Васильева И.П.

« 15 » августа

2024 г.