

# ИНТЕЛЛЕКТУАЛЬНАЯ РЕКОМЕНДАТЕЛЬНАЯ СИСТЕМА ДЛЯ МАРШРУТИЗАЦИИ ПАЦИЕНТОВ ПО СИМПТОМАМ

## INTELLIGENT RECOMMENDATION SYSTEM FOR PATIENT ROUTING BASED ON SYMPTOMS

I. Ambrosov

*Summary.* An intelligent recommendation system has been developed for the automated routing of patients in healthcare institutions based on described symptoms. The use of natural language processing and machine learning techniques allows for recommending specialists for initial consultation. The proposed system improves the efficiency of medical care, reduces the workload of primary care physicians, and contributes to the optimization of healthcare resources.

*Keywords:* intelligent system, patient routing, machine learning, NLP, symptom processing, recommendation system, healthcare.

Амбросов Игорь Александрович  
ФГБОУ ВО «СПбГЭУ», г. Санкт-Петербург  
cvbzwq1@yandex.ru

*Аннотация.* Разработана интеллектуальная рекомендательная система для автоматизированной маршрутизации пациентов медицинских учреждений на основе описанных симптомов. Применение методов обработки естественного языка и машинного обучения позволяет рекомендовать профильных специалистов для первичного обращения. Предложенная система повышает эффективность оказания медицинской помощи, снижает нагрузку на врачей первичного звена и способствует оптимизации ресурсов здравоохранения.

*Ключевые слова:* интеллектуальная система, маршрутизация пациентов, машинное обучение, NLP, обработка симптомов, рекомендательная система, медицинская помощь.

### Введение

Состояние здравоохранения является важнейшим показателем социального и экономического развития страны. Российская медицинская система в последние годы сталкивается с серьёзными вызовами: кадровый дефицит, необходимость модернизации инфраструктуры и рост хронических заболеваний [7, с. 321]. Пандемия COVID-19 выявила структурные слабости системы и усилила нагрузку на первичное звено.

Одной из острых проблем стала маршрутизация пациентов: сложности в интерпретации симптомов и нехватка специалистов, которая приводит к задержкам диагностики и ухудшению качества помощи [3, с. 443]. Согласно данным Росстата [6] с 2019 по 2024 годы численность медицинских кадров, имеющих высшую и первую квалификационные категории сократилась на 13 %, а среднего медперсонала — на 21 %, что усугубляет доступность медицинской помощи (Рис. 1.).

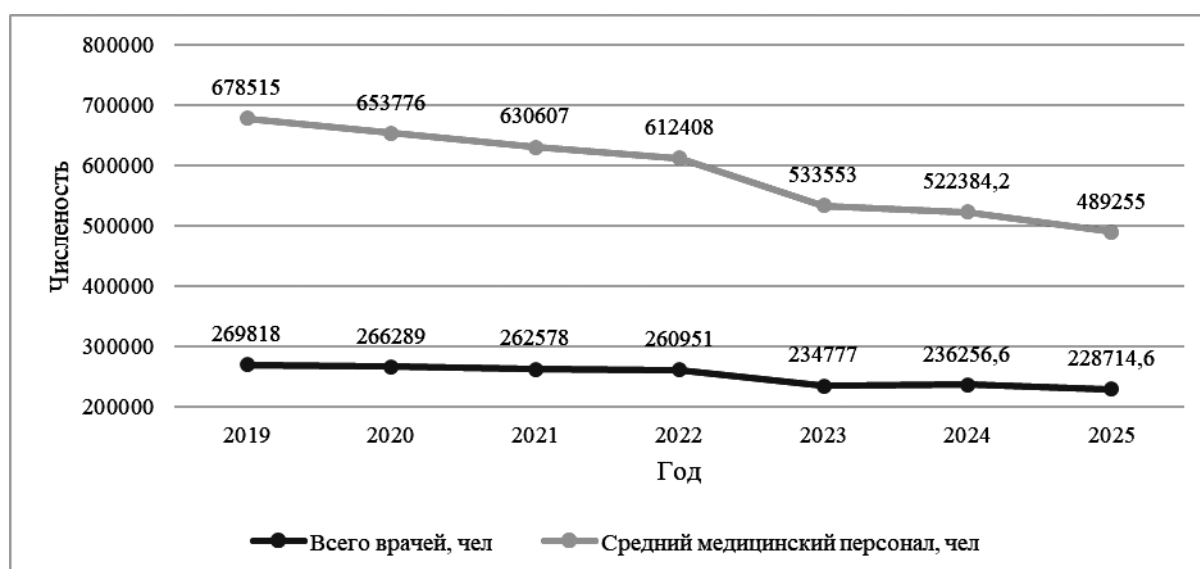


Рис. 1. Численность медицинских кадров, имеющих высшую и первую квалификационные категории с прогнозом на 2025 г., РФ, чел.

При сохранении текущей тенденции, в 2025 году прогнозируется дальнейшее снижение численности медицинских кадров: числа врачей на 3 %, а числа среднего медицинского персонала примерно на 6 % по сравнению с 2024 годом (Рис. 1).

Современные цифровые решения для маршрутизации пациентов часто неэффективны из-за отсутствия интеллектуального анализа текстовых описаний симптомов. В этих условиях всё большее значение приобретают интеллектуальные рекомендательные системы, основанные на технологиях обработки естественного языка (NLP: Natural Language Processing — Обработка естественного языка) и машинного обучения. Они позволяют сократить время поиска нужного специалиста, повысить эффективность первичной медицинской помощи и рационализировать использование ресурсов здравоохранения.

### Обзор литературы

Развитие интеллектуальных систем в медицине связано с интеграцией методов машинного обучения и обработки естественного языка для создания систем поддержки принятия врачебных решений (СППР). Эти технологии автоматизируют диагностику, подбор лечения и маршрутизацию пациентов, что особенно важно при дефиците узких специалистов.

Согласно Астаховой и Киселёвой [1, с. 665], интеллектуальные системы повышают эффективность клинических решений за счёт анализа симптомов, анамнеза и лабораторных данных на основе накопленных знаний.

Ахикян и Данилюк [2, с. 1394] демонстрируют устойчивость адаптивных моделей машинного обучения к разнородным медицинским данным, что особенно ценно в изменяющихся клинических условиях. Бритвина [3, с. 443] отмечает, что машинное обучение способствует переходу медицины от реактивной к превентивной модели за счёт анализа больших массивов данных.

Переломова [5, с. 540] обобщает различные подходы, подчёркивая высокую точность диагностики при правильной предобработке данных и выборе моделей, таких как деревья решений и нейронные сети.

Анализ литературы показывает, что маршрутизация пациентов на основе симптомов в основном реализуется через NLP-модели и мультимассовую классификацию. Однако большинство исследований сосредоточено на диагностике заболеваний, а не на первичном выборе специалиста. В данной работе сделан акцент на разработке рекомендательной системы для оптимизации первичных обращений пациентов (см. Приложение 1), с учётом вариативности описаний и особенностей медицинской практики.

Также следует отметить, что, согласно современным исследованиям, применение технологий искусственного интеллекта в медицинских информационных системах существенно расширяет возможности диагностики и маршрутизации пациентов, снижая вероятность врачебных ошибок и повышая качество обслуживания [7, с. 31].

Таким образом, обзор литературы подтверждает, что интеграция методов машинного обучения и обработки естественного языка является ключевым направлением в развитии систем поддержки принятия врачебных решений. Современные исследования демонстрируют высокую эффективность таких систем в автоматизации диагностики и маршрутизации пациентов, особенно в условиях дефицита медицинских кадров и роста требований к качеству медицинского обслуживания. Однако, несмотря на достигнутые успехи, проблема точного первичного направления пациента к профильному специалисту остаётся недостаточно проработанной. Настоящее исследование направлено на восполнение этого пробела путём создания рекомендательной системы, учитывающей лингвистическую вариативность симптомов и специфические особенности медицинской практики в России.

### Методология исследования

Разрабатываемая интеллектуальная система рекомендательного типа направлена на автоматическое определение наиболее подходящего медицинского специалиста на основе текста, вводимого пользователем. В основе системы лежит анализ естественного языка и применение методов машинного обучения. Целью является построение гибкого и расширяемого решения, способного интерпретировать описания симптомов и рекомендовать одного или нескольких релевантных врачей. Система включает этапы обработки текстовых данных, их преобразования в машинно-читаемый формат и последующей классификации (Рис. 2.).

Модуль векторизации отвечает за перевод неструктурированных описаний симптомов в числовые представления, что позволяет алгоритмам машинного обучения анализировать семантические связи между жалобами пациентов и профилями врачей. На этапе классификации модель определяет наиболее вероятные медицинские специальности, соответствующие введенным симптомам, с учетом возможной неоднозначности формулировок.

Такая архитектура (Рис. 2.) обеспечивает высокую адаптивность системы к различным типам входных данных, включая краткие описания и развернутые жалобы. Для обучения и оценки системы была создана собственная база (Рис. 3.), включающая следующую структуру: «Болезнь — Симптомы — Специальность».

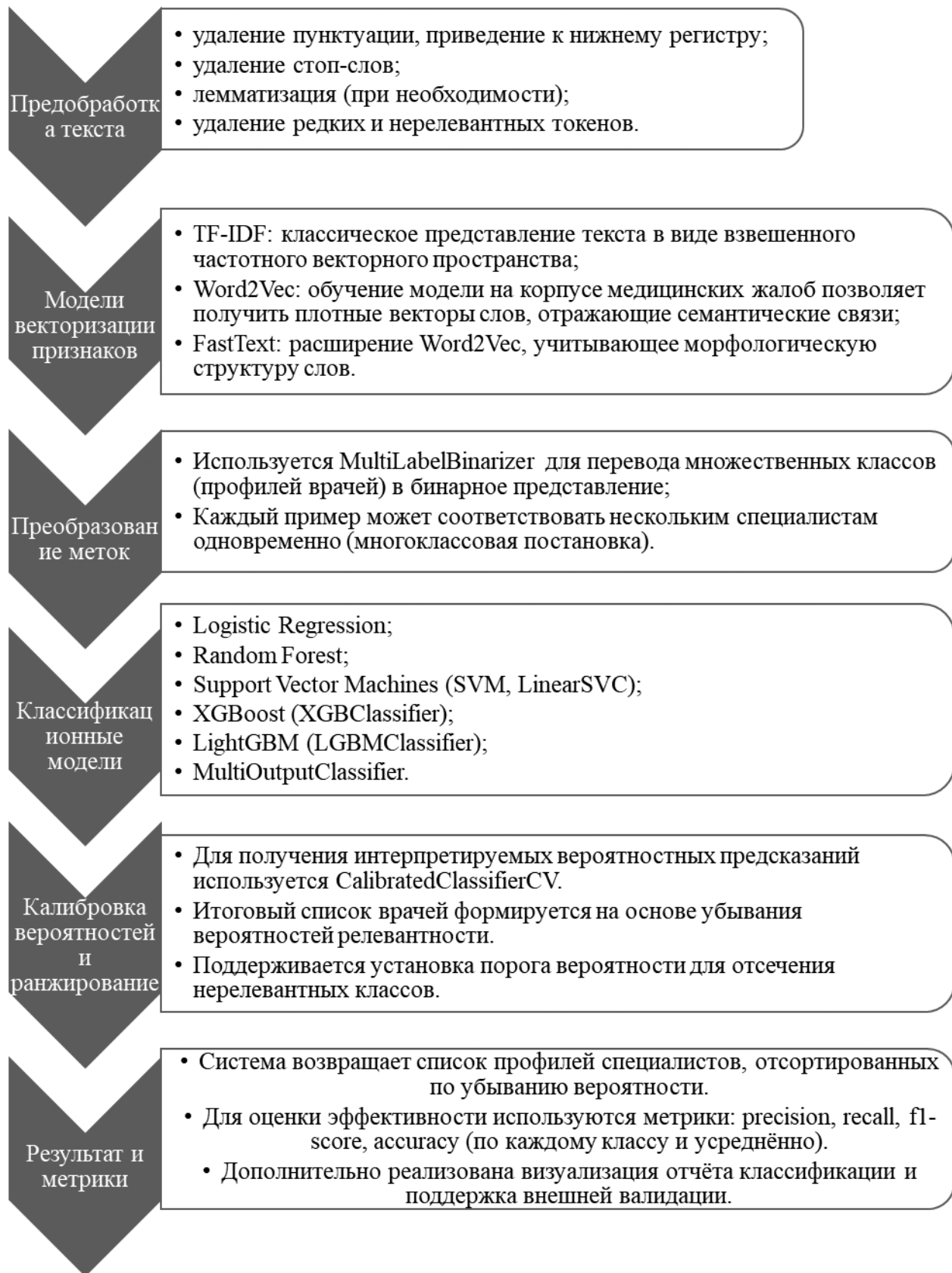


Рис. 2. Архитектура рекомендательной системы

№	Болезнь	Симптомы	Специальность
1	Полиомиелит	температура;головная боль;рвота;слабость в мышц...	Инфекционист
2	Птичий грипп	тяжелая дыхательная недостаточность;вирусная п...	Инфекционист
3	Ротавирусная инфекция	тошнота;рвота;диарея;схваткообразные боли в ве...	Инфекционист;педиатр
4	Сибирская язва (кишечная форма)	боли в области желудка;понос;рвота;общая инток...	Инфекционист
5	Сибирская язва (легочная форма)	озноб;температура;покраснение глаз;чихание;каш...	Инфекционист;пульмонолог
...	...	...	...

Рис. 3. Таблица данных

Данные собирались из открытых медицинских порталов таких как «Яндекс. Здоровье» [12], «Здоровье@Mail.ru» [11], «Минздрав РФ» [8], а также проходили верификацию при помощи консультирования с практикующими врачами. Таким образом, был сформирован сбалансированный и очищенный набор данных, пригодный для построения моделей машинного обучения.

#### Методы обработки и классификации текстовых данных

Для анализа текстовых описаний симптомов применялись современные методы обработки естественного языка и машинного обучения. Основное внимание было уделено двум подходам к векторизации текста: статистическому методу TF-IDF и нейросетевым моделям распределенных представлений слов. TF-IDF использовался в сочетании с различными алгоритмами классификации, включая логистическую регрессию, метод опорных векторов (SVM), случайный лес (Random Forest) и алгоритмы градиентного бустинга (XGBoost и LightGBM). Для учета семантических связей между симптомами применялись нейросетевые модели Word2Vec и FastText, которые особенно эффективны для работы с медицинской терминологией и учитывают морфологические особенности русского языка.

На этапе классификации использовалась многомерная модель, учитывающая возможность соответствия одного набора симптомов нескольким медицинским специальностям. В исследовании сравнивались различные подходы:

- Простые линейные модели (логистическая регрессия).
- Метод опорных векторов (SVM).
- Ансамблевые методы (случайный лес, XGBoost, LightGBM).
- Нейросетевые модели распределённых представлений (Word2Vec, FastText).

Особое внимание уделялось балансу между точностью классификации, устойчивостью к шуму в данных и интерпретируемостью результатов. Комбинация раз-

личных методов векторизации и классификации позволила создать гибкую систему, способную эффективно обрабатывать как структурированные, так и свободные текстовые описания симптомов, обеспечивая точные рекомендации по выбору медицинского специалиста.

#### Оценочные метрики

Для оценки качества модели применялись стандартные метрики классификации, рекомендованные в исследованиях по машинному обучению в медицинской диагностике [9].

В качестве ключевых метрик использовались Accuracy (Точность), Precision (Прогностическая ценность положительного результата), Recall (Чувствительность) и F1-score (Гармоническое среднее между Precision и Recall). Выбор именно этих показателей обусловлен их широким применением в задачах медицинской классификации и способностью комплексно оценивать качество модели.

Выбранный подход к оценке позволяет не только определить общую эффективность модели, но и проанализировать её способность правильно идентифицировать как положительные, так и отрицательные случаи. Подробное описание методологии расчёта и интерпретации метрик можно найти в соответствующих научных публикациях по машинному обучению в медицине [10].

#### Результаты классификации

В рамках исследования были протестированы различные модели рекомендаций медицинских специалистов на основе текстовых описаний симптомов. Сравнительный анализ выявил значительные различия в их эффективности, что отражено в таблице 1.

Наилучшие результаты среди протестированных подходов показала модель на основе TF-IDF и алгоритма Random Forest. Данная модель достигла максимального значения точности (Accuracy = 0,949), что отражает высокий уровень правильных предсказаний по всем классам. Кроме того, модель обеспечила уверенные показатели



по ключевым метрикам классификации: прогностическая ценность положительного результата (Precision) составила 0,586, чувствительность (Recall) — 0,486, а гармоническое среднее между ними (F1-score) достигло 0,531.

Таблица 1.

Сравнительные показатели качества моделей

№	Модель	Accuracy	Precision	Recall	F1-score
1	TF-IDF + Random Forest	0,949	0,586	0,486	0,531
2	FastText + Random Forest	0,164	0,260	0,626	0,367
3	TF-IDF + XGBoost	0,142	0,677	0,156	0,254
4	TF-IDF + Logistic Regression	0,111	0,117	0,599	0,119
5	Word2Vec + Random Forest	0,352	0,360	0,576	0,443
6	TF-IDF + SVM	0,279	0,298	0,521	0,324
7	TF-IDF + LightGBM	0,161	0,579	0,165	0,257

Такой уровень F1-score позволяет говорить о хорошем балансе между способностью модели выявлять релевантных специалистов и минимизацией ложных срабатываний. Высокий показатель Precision подтверждает, что большая часть выданных рекомендаций является обоснованной, что особенно важно в медицинских приложениях, где недопустимы ошибки при направлении пациентов.

Таким образом, модель TF-IDF + Random Forest продемонстрировала наилучшее соотношение точности и надёжности рекомендаций, что позволяет рекомендовать её в качестве основного решения для задачи автоматизированного подбора медицинских специалистов по текстовым описаниям симптомов.

С учётом чувствительности медицинских решений и возможной неоднозначности пользовательских описаний система реализует стратегию top-k рекомендаций: на выходе формируется список из N наиболее вероятных специальностей с указанием индивидуальной вероятности для каждой. Дополнительно введён порог уверенности. Если уверенность классификатора ниже заданного значения, система не даёт однозначной рекомендации или предлагает обратиться к терапевту для первичной диагностики.

### Заключение

Целью данного исследования являлась разработка интеллектуальной рекомендательной системы, способной по текстовому описанию симптомов пациента автоматически определять наиболее подходящего медицин-

ского специалиста. Результаты экспериментов показали, что предложенная архитектура системы, основанная на семантическом представлении симптомов через модель TF-IDF и использовании алгоритма Random Forest в многовыводной классификации, обеспечивает обоснованные и достоверные рекомендации. Система достигла показателей F1-мера 0.531, recall 0.486 и accuracy 0.949, что подтверждает достижение поставленной цели.

В рамках работы разработана интеллектуальная система маршрутизации пациентов, которая реализует комплексный анализ текстовых описаний симптомов с применением современных методов обработки естественного языка и машинного обучения. Основное отличие от существующих решений заключается в использовании оригинального алгоритма семантического анализа, который учитывает контекстные особенности и вариативность формулировок симптомов на русском языке. Предложенная система внедряет принципиально новый подход к формированию рекомендаций: вместо единичного направления пациенту предоставляется ранжированный список специалистов с указанием степени соответствия каждого из них описанной симптоматике. Это особенно важно при сложных или пересекающихся симптомах, где традиционные системы дают однозначные, но не всегда точные рекомендации.

Научная новизна исследования состоит в следующем:

- A. Впервые для задачи маршрутизации пациентов на русском языке реализована многометочная классификация с учётом семантики симптомов.
- B. Проведено комплексное сравнение векторизаций и классификаторов в рамках медицинской задачи, что позволило выявить оптимальные комбинации методов.

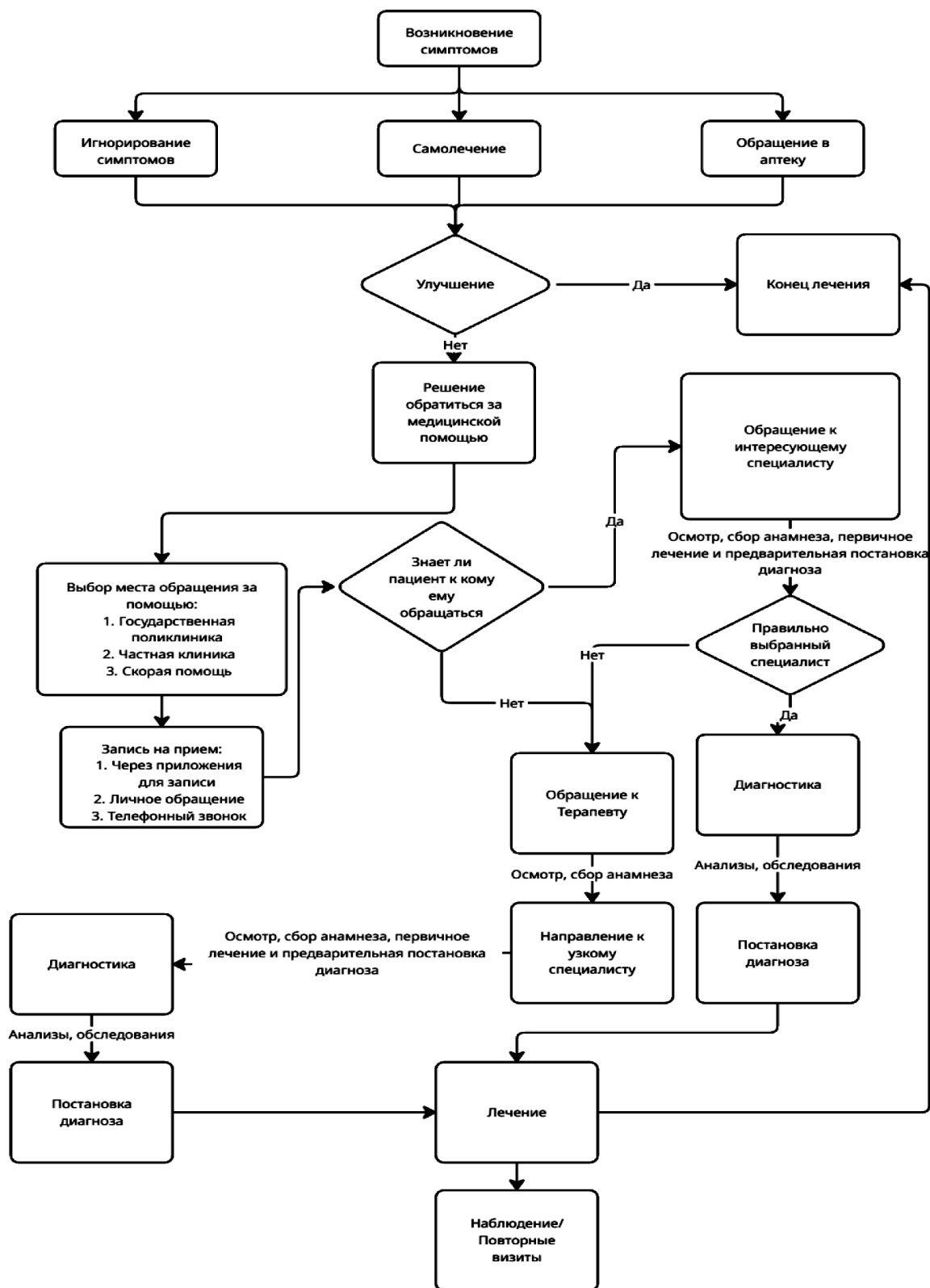
Таблица 2.

Прогресс исследования

Было	Сделано	Предстоит
Отсутствие систем на русских данных для рекомендаций врачей по тексту симптомов	Построена и протестирована система на Word2Vec + Random Forest с многовыводной классификацией	Расширение корпуса данных и новых медицинских профилей
Использование отдельных моделей без сравнения их эффективности	Сравнены TF-IDF, Word2Vec, FastText с классификаторами (SVM, Random Forest, XGBoost, LightGBM)	Внедрение языковых моделей глубокого обучения
Ограниченная обработка свободных текстов	Реализован алгоритм семантического анализа с учётом синонимов и морфологии	Интеграция с электронными медкартами и реализация активного обучения

# Приложение 1

## Маршрут пациента



Для наглядного представления итогов работы была сформирована таблица 2, которая отражает прогресс исследования:

Таким образом, цель исследования была достигнута: построена и протестирована интеллектуальная система, демонстрирующая работоспособность и перспективность применения в реальных условиях. Полученные результаты свидетельствуют о высокой потенциальной пользе предлагаемого подхода как инструмента оптимизации первичных обращений пациентов в системе

здравоохранения. Перспективы развития включают расширение базы симптомов и специалистов, внедрение механизмов активного обучения и интеграцию с данными электронных медицинских карт.

Практическая значимость подтверждается способностью системы снижать нагрузку на врачей первичного звена, сокращать число ошибочных обращений и повышать эффективность здравоохранения в условиях ограниченных ресурсов.

#### ЛИТЕРАТУРА

1. Астахова И.Ф., Киселева Е.И. Интеллектуальная поддержка принятия врачебных решений [Электронный ресурс] // Современные информационные технологии и ИТ-образование. — 2020. — Т. 16. — № 3. — С. 664–672. — Режим доступа: <https://cyberleninka.ru/article/n/intellektualnaya-podderzhka-prinyatiya-vrachebnyh-resheniy/viewer> (дата обращения: 20.04.2025).
2. Ахикян А.И., Данилюк С.С. Алгоритм машинного обучения адаптивный случайный лес и его применение [Электронный ресурс] // Вестник науки. — 2024. — Т. 1. — № 6 (75). — С. 1393–1402. — Режим доступа: <https://cyberleninka.ru/article/n/algorithm-mashinnogo-obucheniya-adaptivnyy-sluchaynyy-les-i-ego-primeneniye/viewer> (дата обращения: 20.04.2025).
3. Бритвина П.В. Машинное обучение в медицине: революция диагностики, лечения и персонализации [Электронный ресурс] // Вестник науки. — 2024. — Т. 4. — № 1 (70). — С. 442–444. — Режим доступа: <https://cyberleninka.ru/article/n/mashinnoe-obucheniye-v-meditsine-revolutsiya-diagnostiki-lecheniya-i-personalizatsii/viewer> (дата обращения: 20.04.2025).
4. Гришина Л.С. Методы и алгоритмы интеллектуальной поддержки принятия решений в медицинской практике на основе обработки естественных языков: автореферат дис. ... канд. техн. наук: 2.3.1 [Электронный ресурс]. — Оренбург, 2024. — 134 с. — Режим доступа: [https://www.osu.ru/ds/24\\_2\\_352\\_03/237/237\\_dis\\_01.pdf](https://www.osu.ru/ds/24_2_352_03/237/237_dis_01.pdf) (дата обращения: 20.04.2025).
5. Переломова П.А. Как искусственный интеллект помогает в диагностике и выборе стратегии лечения [Электронный ресурс] // Вестник науки. — 2025. — Т. 1. — № 3 (84). — С. 538–543. — Режим доступа: <https://cyberleninka.ru/article/n/kak-iskusstvennyy-intellekt-pomogaet-v-diagnostike-i-vybore-strategii-lecheniya/viewer> (дата обращения: 20.04.2025).
6. Федеральная служба государственной статистики. Здравоохранение в России [Электронный ресурс]. — 2023. — Режим доступа: <https://rosstat.gov.ru> (дата обращения: 20.04.2025).
7. Информационные технологии, вычислительные системы и искусственный интеллект в медицине: монография / под ред. М.С. Каганова [Электронный ресурс]. — Москва: ДПК Пресс, 2022. — 284 с. — Режим доступа: <https://dpk-press.ru/wp-content/uploads/2022/06/Информационные-технологии-вычислительные-системы-и-искусственный-интеллект-в-медицине.pdf> (дата обращения: 28.04.2025).
8. Клинические рекомендации Минздрава РФ [Электронный ресурс]. 2023. URL: <https://cr.minzdrav.gov.ru> (дата обращения: 02.04.2025).
9. Метрики качества работы моделей машинного обучения [Электронный ресурс] // Webiomed. URL: <https://webiomed.ru/blog/metriki-kachestva-raboty-modelei/> (дата обращения: 20.04.2025).
10. Оценка качества в задачах классификации и регрессии [Электронный ресурс] // Wiki НИУ ИТМО. — URL: [https://neerc.ifmo.ru/wiki/index.php?title=Оценка\\_качества\\_в\\_задачах\\_классификации\\_и\\_регрессии](https://neerc.ifmo.ru/wiki/index.php?title=Оценка_качества_в_задачах_классификации_и_регрессии) (дата обращения: 20.04.2025).
11. Справочник симптомов «Здоровье@Mail.ru» [Электронный ресурс]. URL: <https://health.mail.ru/disease> (дата обращения: 02.04.2025).
12. Яндекс. Здоровье: сервис самодиагностики [Электронный ресурс]. URL: <https://health.yandex.ru> (дата обращения: 02.04.2025).

© Амбросов Игорь Александрович (cvbzwq1@yandex.ru)

Журнал «Современная наука: актуальные проблемы теории и практики»