

На правах рукописи



Серобабов Александр Сергеевич

**ПОДДЕРЖКА ПРИНЯТИЯ РЕШЕНИЙ В СИСТЕМЕ РАННЕЙ
ДИАГНОСТИКИ ЗАБОЛЕВАНИЙ**

Специальность 2.3.1. – Системный анализ,
управление и обработка информации, статистика

АВТОРЕФЕРАТ
диссертации на соискание ученой степени
кандидата технических наук

Омск – 2023

Работа выполнена в федеральном государственном автономном образовательном учреждении высшего образования «Омский государственный технический университет».

Научный руководитель: **Денисова Людмила Альбертовна,**
доктор технических наук, доцент

Официальные оппоненты: **Бодин Олег Николаевич,**
доктор технических наук, профессор кафедры
техническое управление качеством федерального
государственного бюджетного образовательного
учреждения высшего образования «Пензенский
государственный технологический университет», г.
Пенза

Авдеенко Татьяна Владимировна,
доктор технических наук, профессор кафедры
теоретической и прикладной информатики федерального
государственного бюджетного образовательного
учреждения высшего образования «Новосибирский
государственный технический университет», г.
Новосибирск

Ведущая организация: Бюджетное учреждение высшего образования Ханты-Мансийского автономного округа - Югры «Сургутский государственный университет», г. Сургут

Защита состоится «28» февраля 2024г. в 15-00 на заседании диссертационного совета 24.2.350.07, созданного на базе ОмГТУ, по адресу: 644050, Омск, пр. Мира, 11, Главный корпус, ауд. П-202.

С диссертацией можно ознакомиться в библиотеке ОмГТУ и на сайте www.omgtu.ru.

Отзыв на автореферат в двух экземплярах, заверенный печатью, просьба направлять по адресу: 644050, г. Омск, пр. Мира, 11, ученому секретарю диссертационного совета 24.2.350.07. Тел.: (3812) 65-24-79, e-mail: dissov_omgtu@omgtu.ru.

Автореферат разослан «__» _____ 20__ года.

Ученый секретарь диссертационного совета
24.2.350.07, канд. техн. наук, доцент



А.С. Грицай

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы исследования. В настоящее время одной из основных задач развития системы отечественного здравоохранения является повышение качества предоставляемых услуг за счет внедрения в работу медицинских учреждений средств автоматизированного анализа данных пациентов и систем поддержки принятия решений. При этом согласно Стратегии развития здравоохранения в Российской Федерации на период до 2025 года особое внимание уделяется использованию современных интеллектуальных технологий обработки и интерпретации собранной о пациенте информации, позволяющих в условиях неполноты и неопределенности данных своевременно диагностировать заболевания. Необходимость в автоматизации диагностики существует для многих распространенных заболеваний, одно из них – неалкогольная жировая болезнь печени (НАЖБП). Согласно данным эпидемиологических исследований, проведенных в России, распространенность НАЖБП среди взрослого населения в 2007 году составила 27%, а в 2015 году повысилась до 37%. Поэтому создание системы поддержки принятия решений, способной улучшить качество ранней диагностики заболевания (на примере НАЖБП) и снизить нагрузку на медицинских работников, является актуальной задачей.

Состояние вопроса. Повышение требований к качеству медицинской диагностики обуславливает необходимость разработки систем поддержки принятия решений (СППР). Развитием теории принятия решений в разное время занимались такие отечественные и зарубежные ученые как Кобринский Б.А., Тарасенко Ф.П., Глушков В.М., Saaty T.L., Simon H.A., Kahmen D. и др.

Значительный вклад в разработку систем врачебной диагностики внесли Осипов Г.С. (динамическое прогнозирование и диагностика заболеваний), Долганов А.Ю. (диагностика нарушений вегетативной системы), Колмогоров А.Н., Горбаня А.Н. (нейросетевая диагностика заболеваний), Немков А.Г. (диагностика нарушения в неврологии). Несмотря на широкий круг решенных задач диагностики заболеваний, практически отсутствуют решения, применяемые для выявления заболевания на ранней стадии развития. В связи с этим становится очевидной потребность в разработке технических решений, которые помогут улучшить качество медицинского сопровождения за счет своевременной постановки диагноза.

Целью диссертационной работы является создание и обоснование методик и алгоритмов анализа данных пациентов, характеризующих симптомы

заболеваний, решающих задачу диагностирования стадий заболевания и обеспечения поддержки принятия врачебных решений.

Для достижения цели в работе поставлены и решены следующие **задачи**:

1. Анализ проблем информатизации и автоматизации при принятии врачебных решений для постановки диагнозов заболеваний.

2. Разработка методик и алгоритмов анализа данных для формирования набора значимых параметров обследования пациентов на основе корреляционного анализа и экспертных оценок, обеспечивающего повышение точности ранней диагностики заболевания.

3. Разработка методики и алгоритмов на основе нечеткого логического вывода, обеспечивающих повышение точности диагностики НАЖБП в сравнении с традиционными методами в условиях неполноты имеющихся данных, характеризующих симптомы заболевания.

4. Создание методики анализа данных пациентов для формирования набора замещающих параметров при отсутствии некоторых результатов обследования пациентов.

5. Создание программного комплекса системы поддержки принятия врачебных решений и проведение экспериментальных исследований, подтверждающих эффективности разработанных алгоритмов.

Научная новизна. В процессе исследований получены следующие новые научные результаты.

1. Установлено, что для определения стадии заболевания НАЖБП предложенным гибридным алгоритмом формирование пространства значимых параметров (при экспертной оценке) следует осуществлять по четырем критериям (точность полученных значений, уровень достоверности доказательности связи параметра с заболеванием, информативность параметра, статистическая взаимосвязь). В свою очередь, экспертная оценка врача дополняет статистическую и помогает определить значимые параметры при ранней диагностике заболевания НАЖБП. В результате проведенной оценки на основании принятых критериев выявлено, что при классификации следует использовать такие параметры медицинского обследования как: L_{lep} (лептин), L_{obr} (рецепторы, воспринимающие лептин), D_{nash} (наличие неалкогольного стеатогепатита). Установлено, что каждый из параметров L_{lep} , L_{obr} , D_{nash} характеризует стадию заболевания НАЖБП независимо друг от друга, что подтверждает проведенный факторный анализ, который опровергает возможность сжатия пространства параметров без потери информации.

2. На основании выявленных значимых параметров и метода нечеткой кластеризации получены новые результаты, которые представляют собой сформированные функциональные зависимости между входными мультипликативными параметрами и лингвистическими оценками входных параметров врачом. Это позволило установить численные границы и определить степени принадлежности для значений значимых параметров к заданным лингвистическим термам.

3. Впервые предложена методика для классификации стадии заболевания НАЖБП на основе использования теории нечетких множеств и паттерн-анализа (для получения мультипликативных параметров), что на основе полученных значимых параметров позволило сформировать мультипликативные параметры L_{lept} и L_{obrm} , использование которых дало возможность разграничить пространство и исключить пересечение близлежащих стадий. Такой подход к формированию групп пациентов позволил повысить точность классификации на 8% в сравнении с классическим методом.

Практическая значимость работы заключается в разработке:

– программной реализации системы поддержки принятия решений для классификации стадий заболеваний при ранней диагностике НАЖБП. Предложенная система позволяет классифицировать легкую, среднюю и тяжелую стадию фиброза при НАЖБП, уменьшить временные затраты на диагностирование, повысить объясняемость формируемых диагностических заключений, а также обеспечить единообразие хранения данных о пациентах;

– программной реализации системы поддержки принятия решений при классификации пациентов по степени приверженности к медицинскому сопровождению. Предложенная система дополняет методику доктора медицинских наук Н.А. Николаева, улучшает сопровождение больного во время лечения, что приводит к уменьшению затрат, улучшению качества и увеличению продолжительности жизни больного.

Внедрение результатов исследований. Разработанный комплекс программ для автоматической классификации стадий НАЖБП внедрен в информационную инфраструктуру БУЗОО «Госпиталь для ветеранов войн» (г. Омск), что позволило повысить эффективность диагностирования заболевания на ранней стадии развития.

Объектом исследования является процесс принятия решений при диагностике заболеваний.

Предметом исследования являются методы, подходы и алгоритмы принятия решений при диагностике заболевания в условиях неполноты информации.

Методология исследования базируется на основах системного анализа, методах теории вероятностей и математической статистики; теории принятия решений; интеллектуальных технологиях, включая разделы нечеткой логики и методы кластеризации.

Основные результаты, полученные автором и выносимые на защиту:

1. Гибридная методика выявления значимых параметров для определения стадии НАЖБП. Особенностью методики является формирование набора параметров при совместном использовании корреляционного анализа (зависимости стадии заболевания от параметров обследования пациентов) и аналитической иерархии показателей, характеризующих НАЖБП (по экспертным оценкам врача), что позволяет повысить точность диагноза.

2. Методика и алгоритм поиска замещающих значений параметров (при отсутствии одного из значимых параметров), основанные на выявлении регрессионных зависимостей (между отсутствующим и замещающими параметрами). При этом улучшение диагностических свойств системы обеспечивается за счет расширения набора параметров, пригодных для диагностики НАЖБП, что позволяет повысить точность диагностики в условиях неполноты данных.

3. Методика формирования входных данных для определения стадии НАЖБП, основанная на совместном использовании паттерн-анализа данных (для получения разделимых групп значимых параметров) и нечеткой кластеризации параметров пациентов (для получения функций принадлежности к стадии НАЖБП).

4. Модель и алгоритм принятия решений при диагностике степени заболевания печени, основанные на нечетком логическом выводе. Особенностью модели является то, что определение стадии заболевания выполняется на основе нечеткой базы данных (полученной при кластеризации входных данных) и базы правил, построенной с использованием знаний врачей-экспертов. Кроме того, модель принятия решений на основе нечеткого логического вывода использована для оценки степени приверженности пациентов к назначенному лечению.

Соответствие паспорту специальности. Диссертация соответствует областям исследований: п. 2 «Формализация и постановка задач системного анализа, оптимизации, управления, принятия решений, обработки информации и искусственного интеллекта», п. 4 «Разработка методов и алгоритмов решения задач системного анализа, оптимизации, управления, принятия решений, обработки информации и искусственного интеллекта», п. 5 «Разработка специального математического и алгоритмического обеспечения систем анализа, оптимизации, управления, принятия решений, обработки информации и искусственного интеллекта».

Достоверность полученных результатов. Обоснованность и достоверность теоретических результатов, положений и выводов, полученных в диссертационной работе, базируются на использовании апробированных научных положений и методов исследования, корректном применении математического аппарата, согласованности новых результатов с известными теоретическими положениями. Обоснованность и достоверность прикладных результатов диссертации подтверждается результатами апробации и внедрения предложенных методик и алгоритмов при проектировании системы поддержки принятия решений для диагностики заболевания печени.

Апробация результатов исследования. Результаты работы отражались в научных докладах, которые представлялись на международных и всероссийских конференциях: Ural Symposium on Biomedical Engineering, Radioelectronics and Information Technology (USBREIT) (Екатеринбург, 2020); XV International Scientific and Technical Conference: Applied Mechanics and Systems Dynamics, AMSD 2021 (Омск, 2021); XIII Всероссийской научно-практической конференции «Информационные технологии и автоматизация управления» (Омск, 2022); IV Всероссийской с международным участием научно-практической конференции «Информационные технологии и математическое моделирование» (Омск, 2022); X Всероссийской научно-технической конференции «Россия молодая: передовые технологии – в промышленность» (Омск, 2023), VII Международная научно-практическая конференция «Мехатроника, автоматика и робототехника» (Санкт-Петербург, 2023).

Публикации по теме исследования. По результатам исследований опубликовано 20 научных работ, в том числе 5 статей в рецензируемых журналах, рекомендуемых ВАК РФ, 2 статьи в изданиях, индексируемых в базах Scopus и Web of Science, 4 свидетельства о государственной регистрации программ для ЭВМ.

Личный вклад автора. Решение задач диссертации, разработанные алгоритмы и их программная реализация, экспериментальные и теоретические результаты, представленные в диссертации и выносимые на защиту, принадлежат лично автору.

Структура и объем работы. Диссертация состоит из введения, четырех глав, заключения, списка литературы из 148 наименований и пяти приложений. Общий объем работы составляет 176 страниц, включая 65 рисунков и 33 таблицы.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во введении обоснована актуальность темы диссертации, проведен анализ степени разработанности исследуемой научной проблемы и предложены подходы к ее решению, поставлена цель работы, сформулированы задачи исследования и основные результаты, выносимые на защиту, указаны научная новизна и практическая значимость работы, а также степень достоверности результатов.

Первая глава посвящена анализу состояния проблемы проектирования систем поддержки принятия решений при диагностике заболеваний. Изучены современные алгоритмы обработки данных и изменяющиеся требования к процедуре диагностики. Определена необходимость создания системы поддержки принятия решений, учитывающей специфику деятельности медицинского учреждения и позволяющей работать в условиях неполноты и неопределенности входных данных, а также обеспечивать эффективность принимаемых решений (за счет повышения точности, математической доказуемости и адекватности решений). Исследованы проблемы организации постановки диагноза в медицинских учреждениях и выявлена перспективность применения интеллектуальных технологий для поддержки принятия врачебных решений в медицинских системах.

Во второй главе представлены результаты разработки гибридной методики выявления значимых параметров на основе автоматизированного статистического анализа и экспертных оценок врача. При первичной обработке данные клинических исследований проверяются на наличие пропусков значений, очищаются от аномальных записей (ошибок занесения, наличие шумовых значений). Также выполняется проверка на наличие дублирующих данных, которые удаляются из исходной выборки. Данные группируются по категориям, рассчитываются статистические показатели, выявляются выбросы. Формируются следующие группы данных медицинских обследований: лабораторные параметры (L_{name} , индекс $name$ соответствует наименованию параметра); сопутствующие заболевания пациента (D_{name}); физиологические параметры (P_{name}).

Из всего множества параметров, характеризующих пациента, выбираются показатели, указывающие на наличие заболевания, или предикторы, предвещающие развитие заболевания. Для этого используется разработанный алгоритм выявления значимых параметров, характеризующих стадию заболевания пациента, представленный на рис. 1. Алгоритм основан на использовании двух процессов для получения совокупности аналитических и экспертных оценок и содержит несколько шагов. *Первый процесс* выполняет *аналитическое оценивание* следующим образом. *На первом шаге* алгоритма из базы данных считываются параметры пациентов (включая стадии заболевания) и вводятся требования (пользователем-врачом) к критериям достоверности принятых решений и количеству значимых параметров, необходимых для постановки диагноза. *На втором шаге* для каждого параметра пациента оценивается величина корреляционной связи между параметрами и стадией заболевания (рис. 2). Далее проверяется условие превышения коэффициентом корреляции порогового значения (принималось $b_1=0,7$, характеризующее высокую корреляционную связь

между параметрами по шкале Чеддока), при выполнении условия параметр добавляется в список значимых.

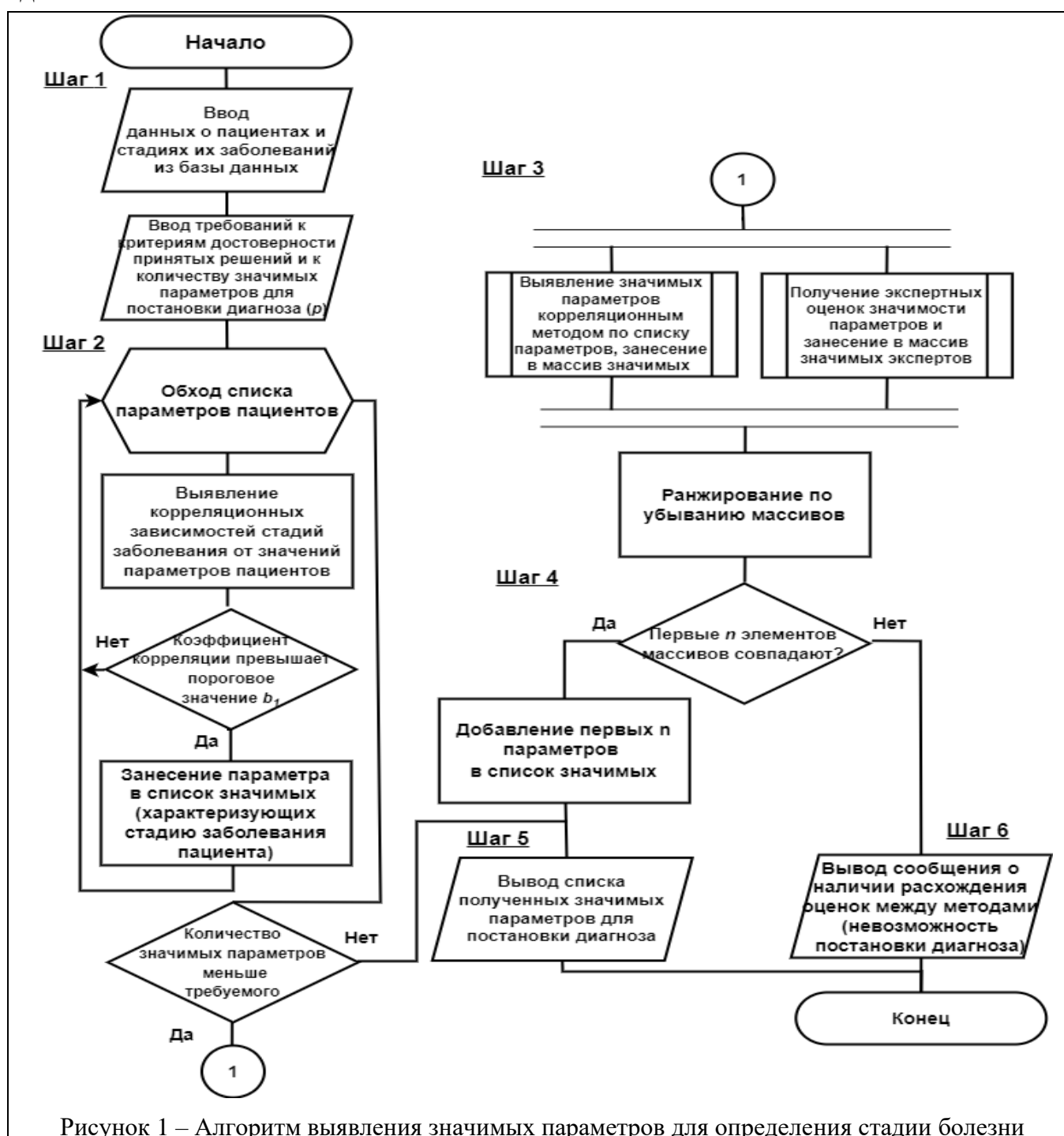
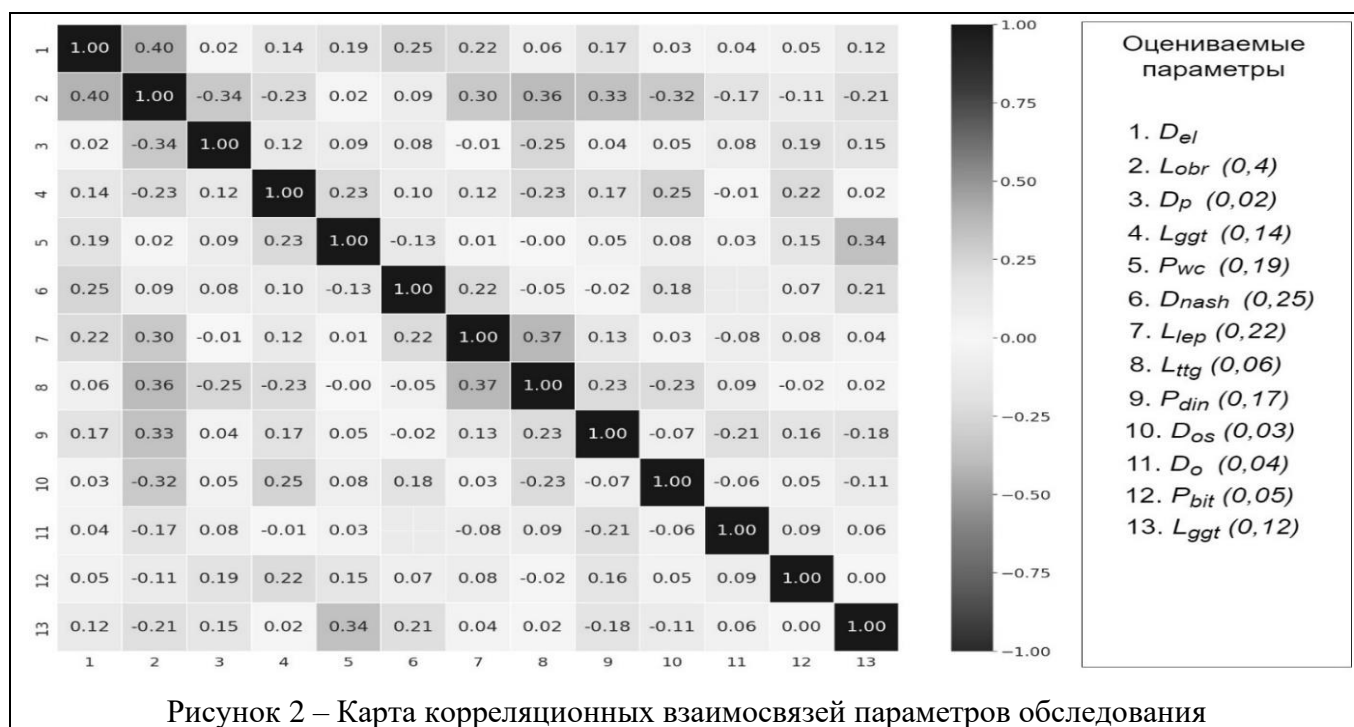
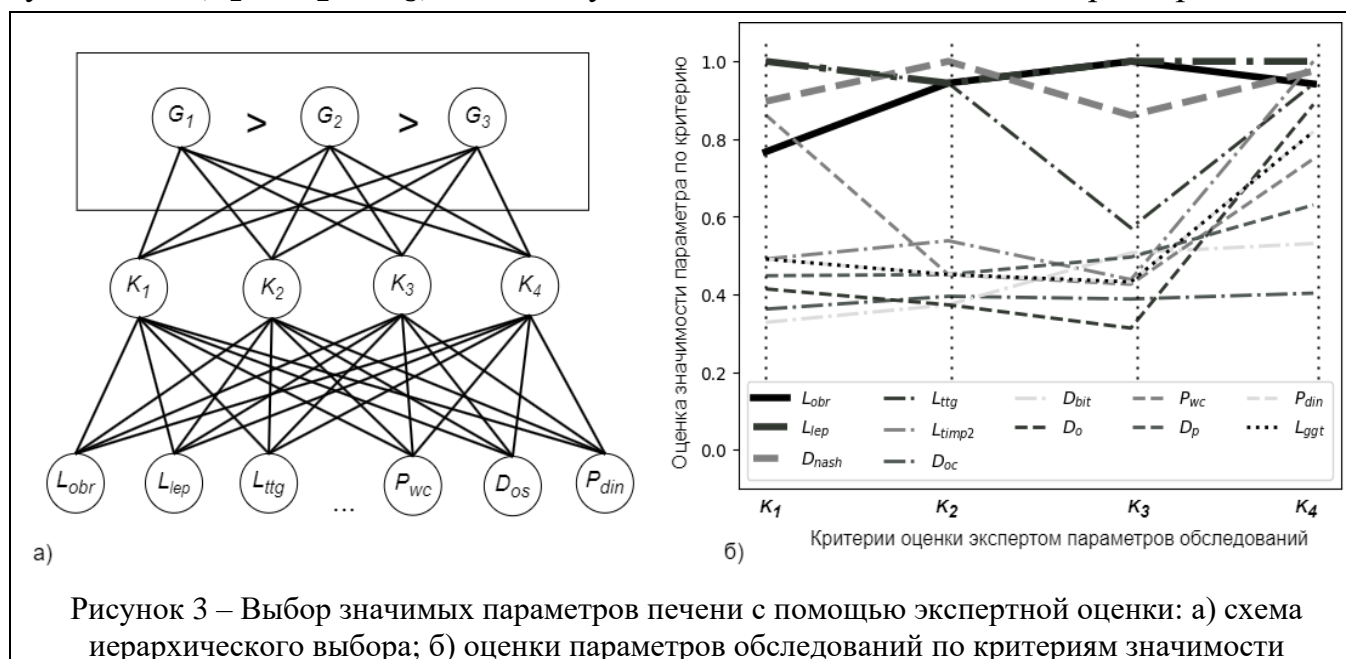


Рисунок 1 – Алгоритм выявления значимых параметров для определения стадии болезни

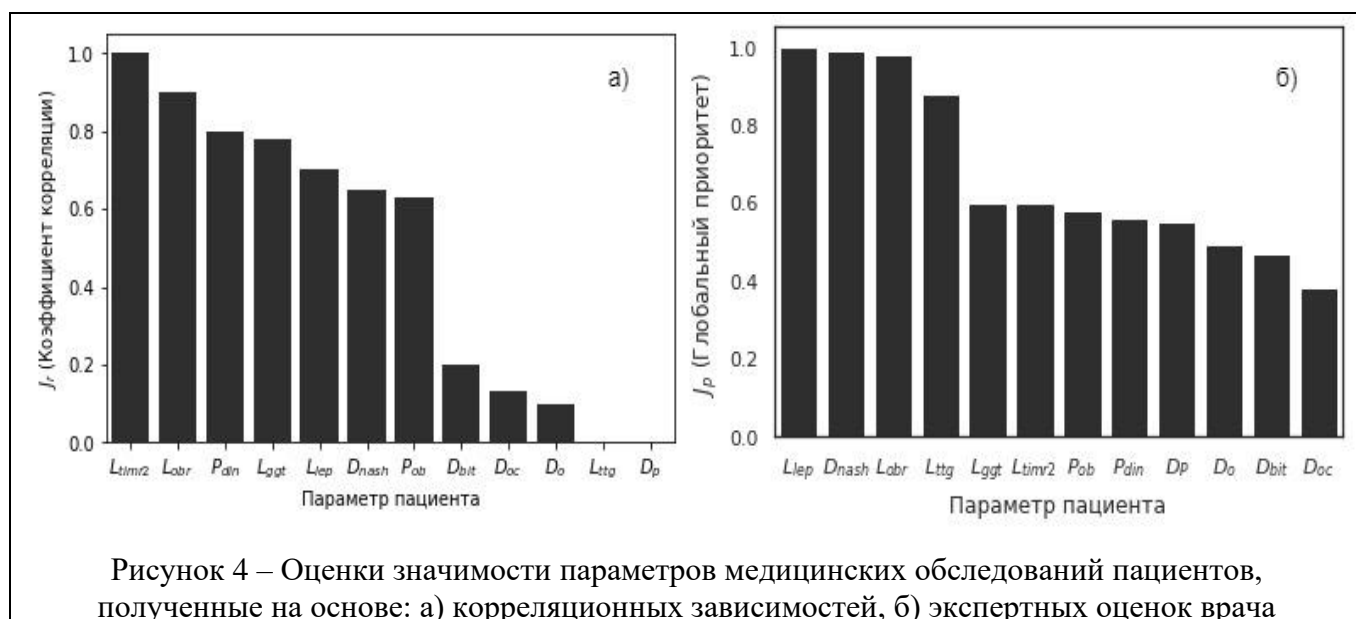
Если количество параметров в списке меньше установленного значения, то выполняется переход к *третьему шагу*, иначе – к *пятому шагу*. На *третьем шаге* реализуются два параллельных процесса. Первый процесс предусматривает понижение порогового значения (до $b_2=0,15$ – значения, характеризующего корреляционную связь между параметрами как слабую). Далее вычисляются корреляционные зависимости между параметрами для формирования (или дополнения) списка значимых параметров.



Второй процесс параллельно выполняет экспертное оценивание и реализован с целью повышения достоверности полученных аналитических результатов. Дополнительно к предложенным оценкам вычисляются значения значимых параметров на основе экспертной оценки (с помощью метода анализа иерархий) рис.3а. Врач-эксперт оценивает параметры обследования пациента по принятым критериям: K_1 – точность полученных значений, K_2 – уровень достоверности доказательности связи параметра с заболеванием, K_3 – информативность параметра, K_4 – статистическая взаимосвязь параметра с заболеванием. На основе этих критериев формируются глобальные оценки параметров, которые упорядочиваются по убыванию ($G_1 > G_2 > G_3$) и используются для оценки значимости параметров.

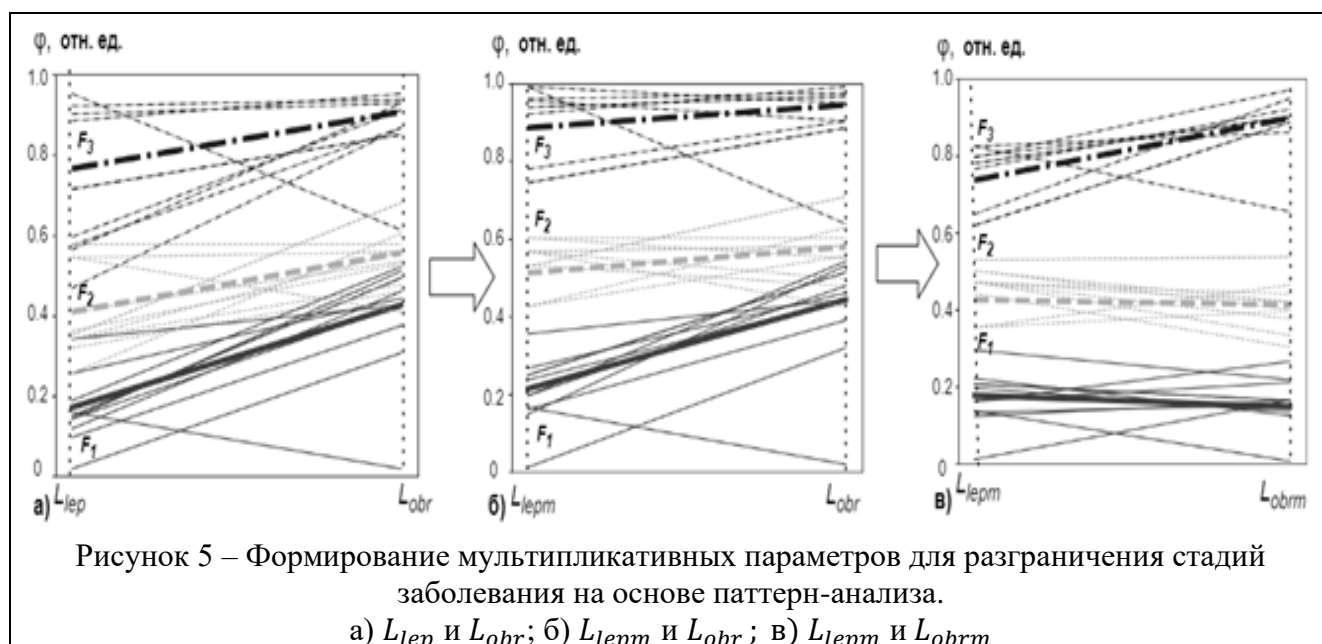


Полученные результаты оценивания для каждого параметра представлены на рис. 3б, где изображены нормированные оценки эксперта (от 0 до 1) по каждому из критериев K_i . Параметры L_{lep} , L_{obr} , D_{nash} выбранные в качестве значимых, имеют более высокие оценки, которые находятся в верхней части графика (утолщенные линии). Кроме того, видно, что параметр P_{wc} по критерию K_2 сравним с полученными значимыми параметрами (что может быть учтено при недостатке информации по основным параметрам). На основе полученной совокупности аналитических и экспертных оценок формируются массивы значимых параметров (упорядоченные по убыванию оценок). Согласованность полученных оценок проверяется на *четвертом шаге* алгоритма. Рис. 4 иллюстрирует результат проверки согласованности нормированных аналитических и экспертных оценок. На рис. 4а представлены гистограммы ранжированных оценок значимых параметров медицинских обследований пациентов, вычисленных на основе корреляционных зависимостей. На рис. 4б приведены гистограммы экспертных оценок практикующего врача, полученных с использованием аналитической иерархии. Далее производится проверка: если наименования первых p параметров совпадают (значение p вводится пользователем-врачом), то есть, параметры (полученные аналитически и экспертными процессами) согласованы и пригодны для постановки диагноза, то переходим к *пятому шагу*, иначе – к *шестому*. На *пятом шаге* алгоритма выводится список полученных значимых параметров (согласованных на четвертом шаге), которые заносятся в базу данных для дальнейшей классификации с их помощью стадии заболевания. На *шестом шаге* (в связи с отсутствием достаточного количества согласованных значимых параметров) выводится сообщение о расхождении оценок в результате выполнения алгоритма, что означает отсутствие возможности автоматически поставить диагноз.



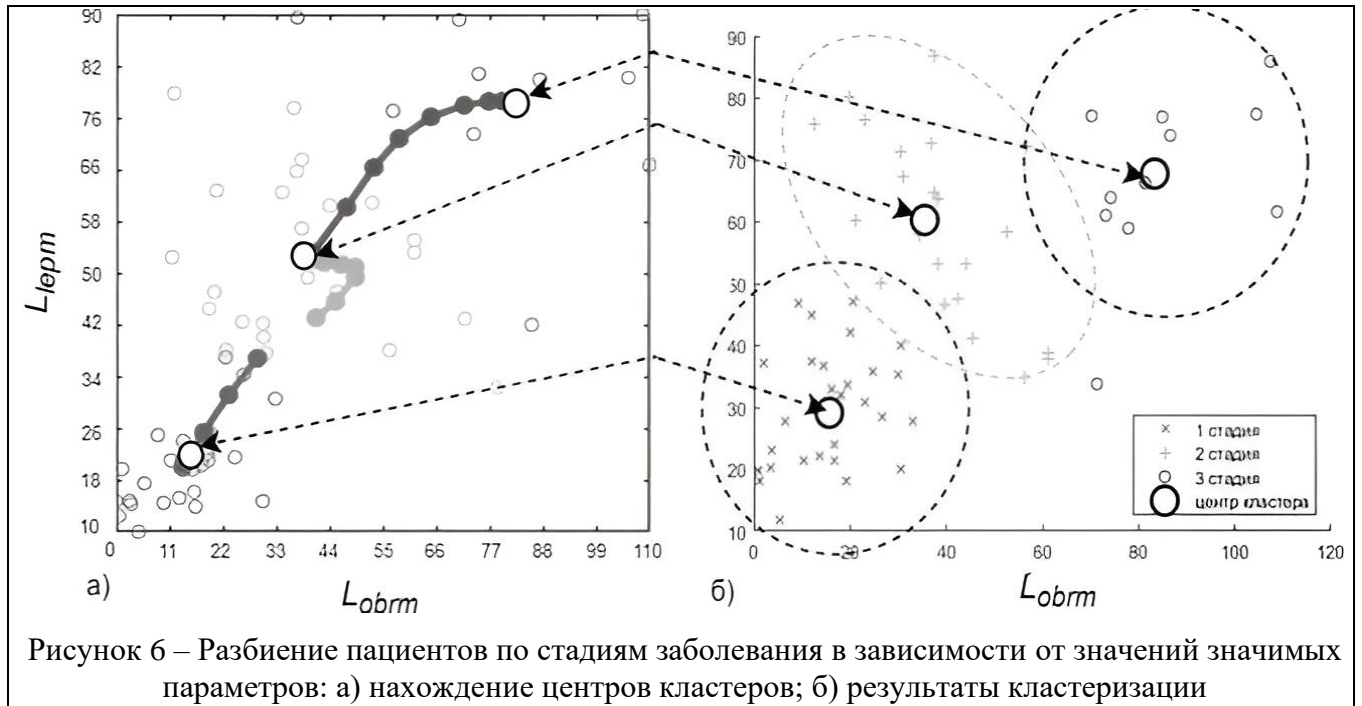
Далее пространство значимых параметров исследовалось с помощью факторного анализа на возможность его сжатия. Получено, что нельзя уменьшить количество значимых параметров без существенной потери информации для диагностики. Для обеспечения работоспособности системы при отсутствии или некорректности входных данных предложена методика формирования замещающих параметров на основе выявленных регрессионных зависимостей. В качестве характеристики, которая указывает на качество регрессионных моделей, был принят коэффициент детерминации $R^2 = 1 - \sum_{i=1}^n (y_i - \hat{y})^2 / \sum_{i=1}^n (y_i - \bar{y})^2$, где n – число пациентов, y_i – значение объясняемой переменной, \bar{y} – среднее арифметическое, \hat{y} – модельное значение. Построены регрессионные модели для замещения параметра L_{lep} (лептин) с помощью параметра L_e (количество эритроцитов в крови), а также для замещения параметра L_{obr} (содержание рецепторов, воспринимающих лептин) с помощью параметра L_{mmp9} (матриксная металлопротеиназа). Для этих параметров построены модели линейной и квадратичной регрессии. Получено, что модель квадратичной регрессии объясняет большее количество данных ($R^2 \approx 0,55$) по сравнению с линейной ($R^2 \approx 0,45$), поэтому квадратичная модель принята для определения замещающего параметра.

В третьей главе представлены результаты построения и обучения классификатора, основанного на нечетком логическом выводе. Входные данные для нечеткого классификатора предварительно обрабатывались для разграничения с помощью паттерн-анализа, основанного на выявлении сходства показателей, характеризующих стадию заболевания. В случае, если полученные паттерны для исследуемых параметров плохо разграничиваются, то привлекались дополнительные параметры (из базы данных медицинских обследований) с формированием мультипликативных параметров (рис. 5). Необходимо разграничить три стадии НАЖБП ($F_1 - F_3$). На рис. 5а представлены паттерны стадий заболевания печени по значимым параметрам L_{lep} и L_{obr} . Видно, что паттерны плохо разграничиваются: по каждому параметру имеются пересечения. Для устранения пересечений привлечены дополнительные данные и сформированы мультипликативные параметры. Так, путем умножения параметра L_{lep} на значение k_{gen} (коэффициент гендерной коррекции), сформирован параметр L_{lepm} , при использовании которого паттерны стали лучше разграничены в установленных пределах (рис. 5б).

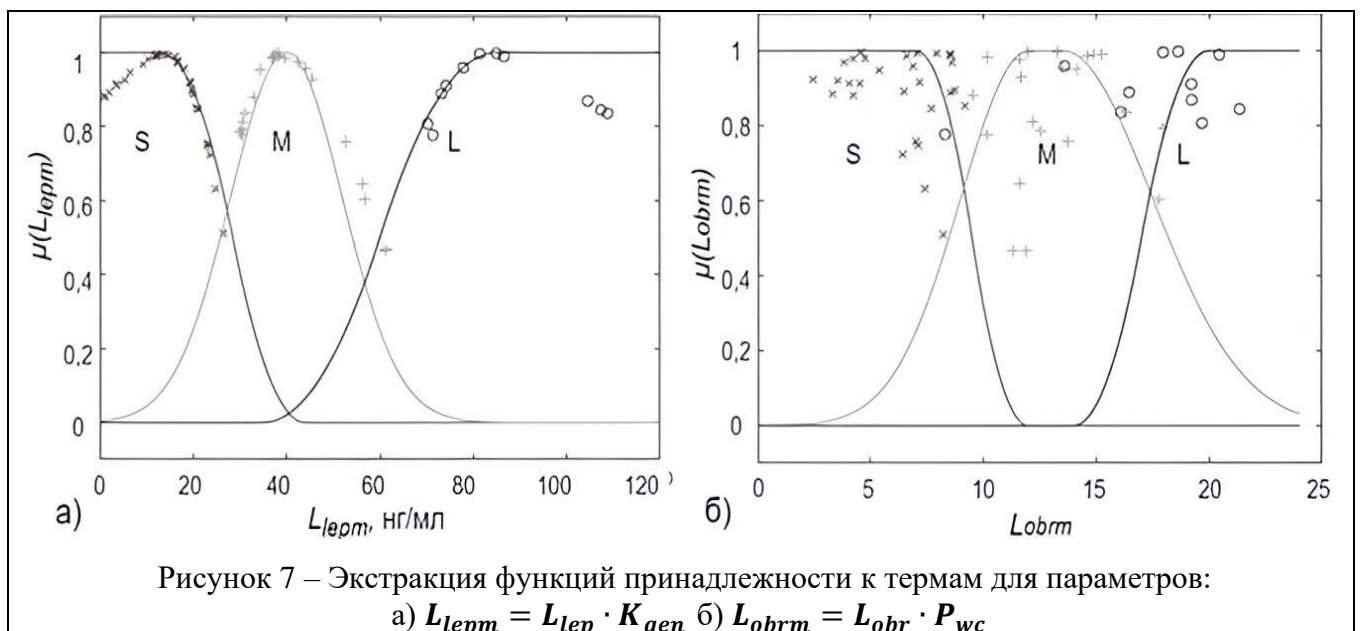


Мультипликативный параметр L_{obrm} , полученный путем умножения параметра L_{obr} на параметр P_{wc} (обхват талии), позволил увеличить межкластерные расстояния между паттернами стадий F_1 и F_2 , а также стадий F_2 и F_3 (рис. 5в). Таким образом, полученные мультипликативные параметры L_{obrm} и L_{lep} приняты для использования в качестве входных данных классификатора.

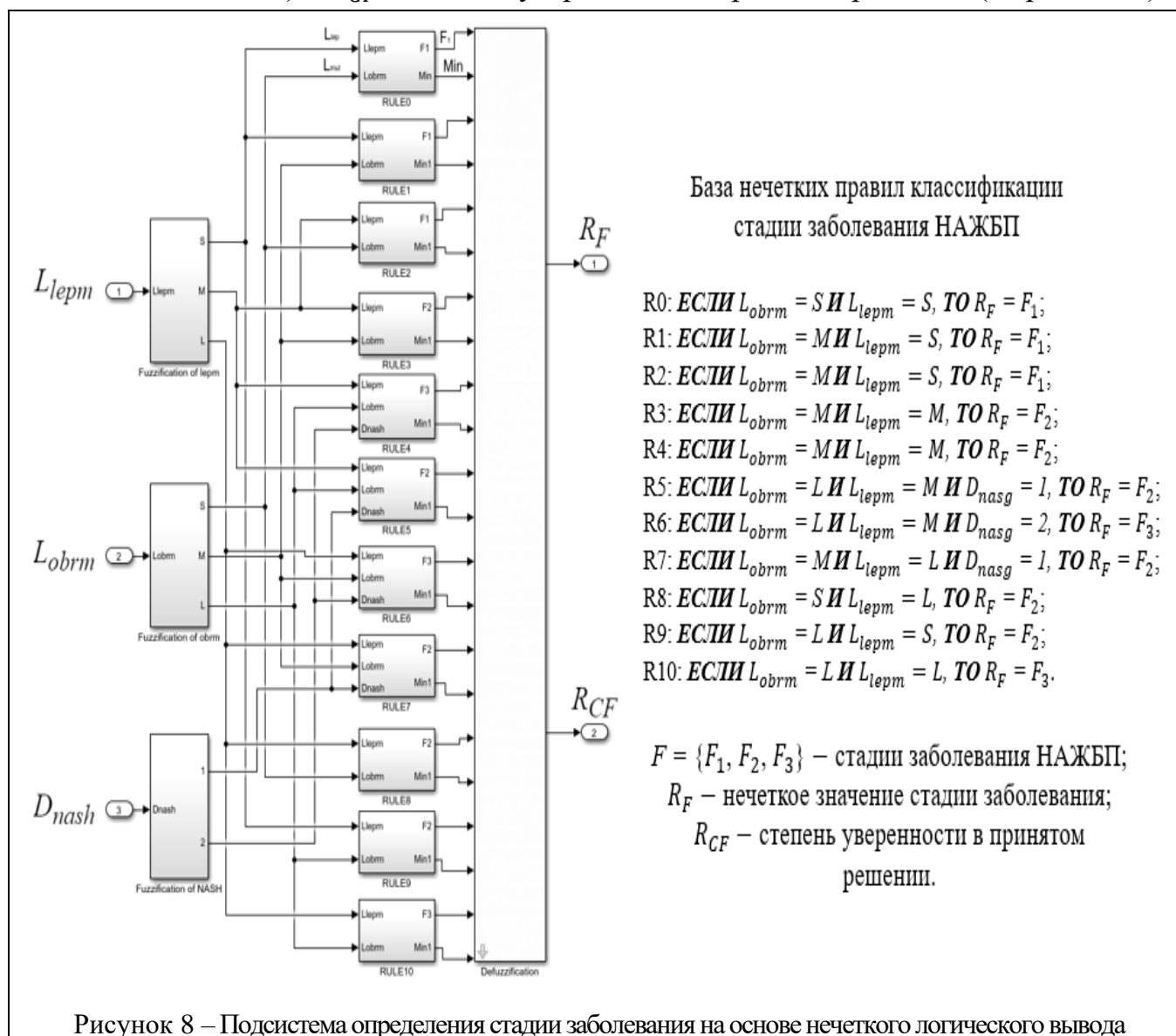
Для экстракции функций принадлежности классификатора использован алгоритм кластеризации нечетких c -средних, позволяющий определить степень принадлежности i -го пациента к j -му кластеру (стадии заболевания), то есть μ_{ij} , $i = \overline{1, n}$, $j = \overline{1, c}$, n – количество пациентов, c – количество кластеров (стадий заболевания). Разбиение на кластеры (рис. б) осуществлялось по значимым параметрам L_{lep} и L_{obrm} , сведенным в вектор X_i для каждого i -го пациента. При этом обеспечивается распределение объектов по кластерам в соответствии с выражением: $\sum_{j=1}^c \mu_{ij} = 1$, $i = \overline{1, n}$. Центры кластеров рассчитываются по выражению: $C_j = \sum_{i=1}^n ((\mu_{ij})^2 X_i) / \sum_{i=1}^n ((\mu_{ij})^2)$; $j = \overline{1, c}$. При определении расстояния между объектами используется Евклидова метрика. Процесс поиска центров кластеров является итерационным: выполняется перерасчет всех μ_{ij} до достижения требуемой точности (рис. ба). На рис. ба центры полученных кластеров обозначены кружками, линиями представлены пути их поиска. На рис. бб представлены результаты кластеризации с отмеченными областями принадлежности к кластерам. Видно, что площадь пересекающихся областей невелика, то есть кластеры представляют собой хорошо различимые группы. На основе полученных степеней принадлежности к каждому кластеру сформированы функции принадлежности для пациентов с разными стадиями заболевания.



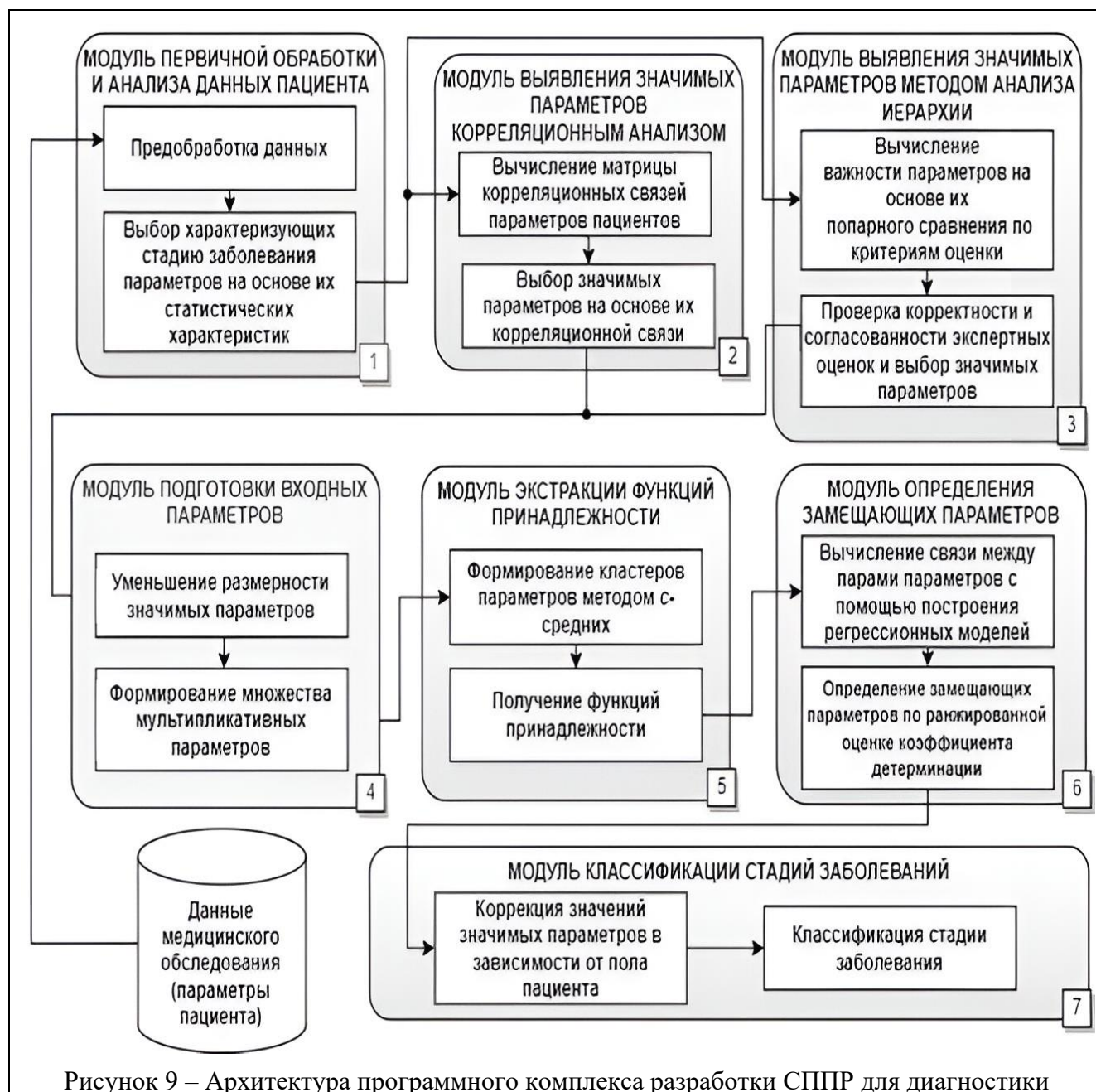
На рис. 7 приведены функции принадлежности каждого из параметров L_{lep} и L_{obrm} к трем принятым термам: S (*small*) – малое значение, M (*middle*) – среднее значение, L (*large*) – большое значение. Для крайних термов выбраны S и Z -образные функции, средний терм задается симметричной гауссовой функцией. Аппроксимация данных пациентов к принятым функциям принадлежности проводилась с помощью метода наименьших квадратов. Функции принадлежности используются для фаззификации входных параметров L_{lep} и L_{obrm} (преобразования в нечеткую форму с вычислением степеней принадлежности к термам) для дальнейшего применения в нечетком классификаторе.



На рис. 8 представлена подсистема классификатора, выполняющая определение стадии заболевания на основе нечеткого логического вывода. В построенной системе на входы классификатора поступает, кроме мультипликативных параметров (L_{obrm} , L_{lepm}), также уточняющий параметр D_{nash} (наличие заболевания неалкогольного стеатогепатита). Вычисленные степени принадлежности к термам для каждого входного параметра передаются в блоки нечетких правил ($RULE0 - RULE10$), которые формируют выходные переменные. Для получения «четкого» значения выходной переменной (номера стадии заболевания) выполняется дефаззификация (обратное преобразование нечетких переменных в четкие) по формуле: $R_F = \sum_{i=0}^{10} \mu_i(R_i)R_i / \sum_{i=0}^{10} \mu_i(R_i)$, где R_F – четкое значение выходной переменной (стадии заболевания); R_i – заключение i -го правила. В результате работы нечеткого классификатора формируются два выходных параметра: R_F – стадия заболевания (полученное действительное число округляется до ближайшего целого) и R_{CF} – степень уверенности в принятом решении (в процентах).



В четвертой главе представлены результаты создания программного комплекса, реализующего поддержку принятия врачебных решений (рис. 9).



С помощью модуля (1) данные обследования пациентов предобрабатываются: табличная система хранения заменяется на реляционную базу данных, а также исследуются статистические характеристики параметров (для выявления ошибок и пропусков в данных). Затем предобработанные данные поступают на входы модулей (2) и (3) для формирования множества значимых параметров на основе оценки их важности для диагностирования стадии заболевания. В модуле (2) оценка вычисляется методом корреляционного анализа, а в модуле (3) – определяется по экспертным оценкам на основе аналитической иерархии. Результаты работы модулей (2) и (3) проверяются на согласованность и

пригодность для постановки диагноза. Полученные значимые параметры подготавливаются для обучения классификатора с помощью модуля (4). Выполняется проверка на возможность уменьшения размерности множества значимых параметров, а также формируются мультипликативные параметры, которые содержат в себе больший объем информации, позволяя более четко разграничить стадии при диагностике заболевания. Для обеспечения работы системы с помощью модуля (6) отсутствующие параметры заменяются одним из замещающих параметров на основе построенных регрессионных моделей. В модуле (5) данные пациентов разбиваются на кластеры (на основе метода нечетких s -средних) и определяются степени принадлежности объектов соответствующему кластеру. Затем полученные множества степеней принадлежности аппроксимируют подходящими параметрами функций принадлежности. В модуле (7) полученные результаты используются для построения нечеткого классификатора, с помощью которого определяется стадия заболевания.

Проведено экспериментальное исследование, подтверждающее эффективность разработанной системы, при котором полученные системой результаты сравнивались с поставленными врачом диагнозами. Для исследования собраны и обработаны данные пациентов с предварительно диагностированной врачом стадией болезни (F_1 , F_2 или F_3), включающие параметры, из которых выбирались значимые параметры. Результаты тестирования для шести пациентов приведены в табл. 2. Степень уверенности в принятом системой решении рассчитана по формуле: $R_{CF} = 100 - \bar{R}_{CF}$, где $\bar{R}_{CF} = |[R_F] - R_F| \cdot 100\%$ – степень неуверенности в принятом решении, $[R_F]$ – округление до ближайшего целого. Строки таблицы, где установленная врачом стадия (по биопсии печени) совпадает с решением, принятым системой, выделены цветом.

Таблица 2–Результаты тестирования системы по значимым параметрам пациента в сравнении с диагнозом, поставленным по биопсии печени

№	L_{lep} (лептин)	L_{obr} (рецепторы лептина)	D_{nash}	P_{gen} (пол пациента)	P_{wc} (охват талии относительно нормы)	D_{el} (стадия заболевания)	R_F (F – стадия заболевания)	R_{CF} (степень уверенности, %)
1	86,6	4,5	1	2	1,05	1	1,052 (1)	94,8
2	11,904	12,252	2	1	1,1	2	1,689 (2)	68,9
3	20,631	9,189	2	1	1,21	3	1,634 (2)	63,4
...	
20	45,65	12,42	2	1	1,25	3	2,59 (3)	59
25	90,2	3,6	2	2	1,04	2	1,437 (1)	56,3
26	84,32	16,2	2	1	1,11	2	2,46 (2)	54

Для иллюстрации на рис. 10а (для тех же шести пациентов) приведены гистограммы, состоящие из двух частей. Каждый столбец гистограммы содержит

изображение стадии заболевания и степени уверенности ее определения. Черным цветом показан параметр R_F (выход классификатора). Серым цветом показана степень неуверенности в принятом решении. Горизонтальная черта на графике – окончательное принятое решение о стадии заболевания пациента (R_F , округленное до ближайшего целого). На рис. 10б показаны результаты сравнения принятых системой решений о стадии заболевания пациента с поставленными врачом диагнозами. Каждому пациенту соответствуют три столбца: светло-серый столбец – решение, принятое системой, о стадии заболевания печени (F); черным цветом представлен диагноз врача. Белый столбец соответствует степени уверенности в поставленном диагнозе (R_{CF} , %). Видно, что в четырех из шести рассматриваемых случаев принятые системой решения совпали с поставленными врачом диагнозами. При этом степень уверенности в постановке диагноза системой высока (от 65% до 95%). В двух случаях (когда поставленные системой и врачом-экспертом диагнозы не совпали) степень уверенности принятия системой решения ниже 65%.

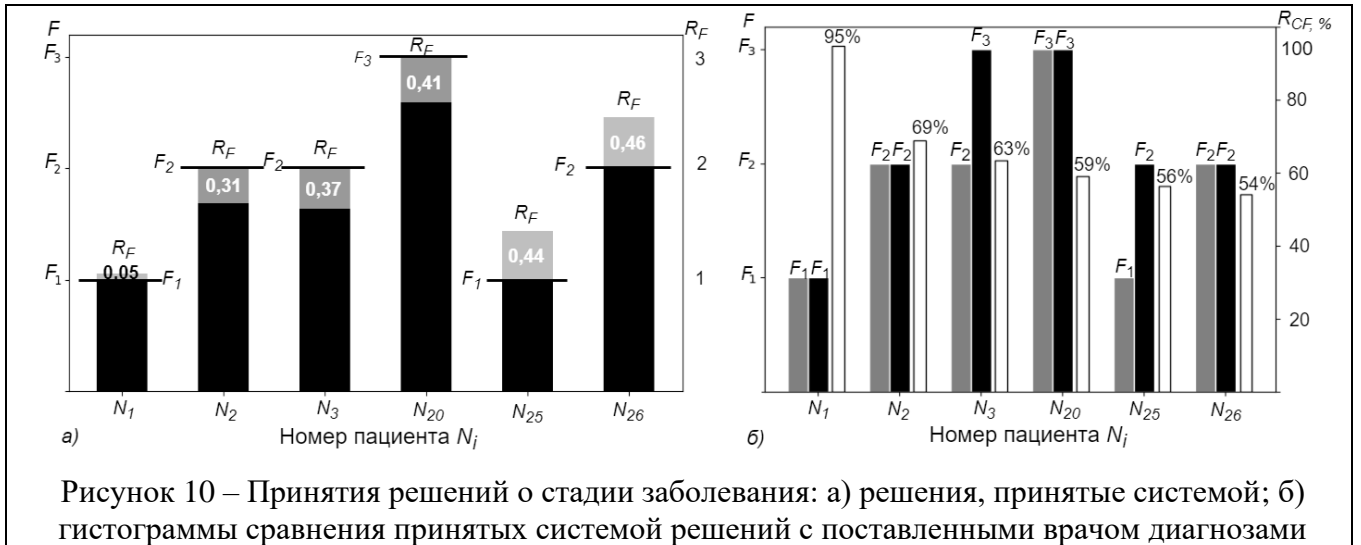


Рисунок 10 – Принятые решения о стадии заболевания: а) решения, принятые системой; б) гистограммы сравнения принятых системой решений с поставленными врачом диагнозами

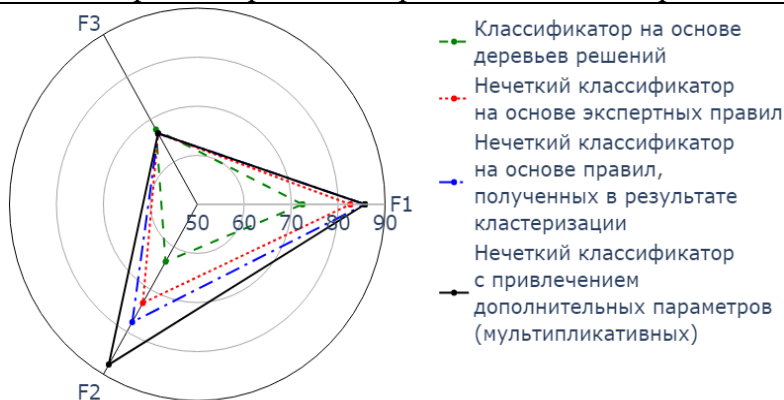


Рисунок 11 – Результаты классификации пациентов по стадиям заболевания различными классификаторам

На рис. 11 построена полярная диаграмма сравнения предлагаемого метода классификации с известными. На окружности по часовой стрелке указаны стадии заболевания, начиная с легкой (F_1) стадии и заканчивая стадией с осложнениями (F_3). Точки

соответствуют точности классификации (J_a) определенной стадии. Получено, что лучший результат (точность классификации $J_a = 84,6\%$) достигается с

использованием нечеткого классификатора с привлечением дополнительных параметров (с формированием мультипликативных) в сравнении с результатами, полученными нечетким классификатором (точность классификации $J_a = 70\%$) и классификатором на основе деревьев решений ($J_a = 68,3\%$).

Для подтверждения эффективности предлагаемого метода производилась оценка качества классификации на основе принятых метрик. Оценивалась полнота классификации (J_r), ошибка первого рода (доля случаев, когда отвергнута верная нулевая гипотеза), ошибка второго рода (выявляемая в случае, когда принята неверная нулевая гипотеза). Полученные оценки системы поддержки принятия решений на основе нечеткого классификатора представлены в табл. 3.

Таблица 3 – Значения оценок классификации стадий заболевания системой

Метрики оценки качества классификации	Стадии заболевания		
	F_1	F_2	F_3
Точность, J_a , %	85,7	87,5	66,7
Полнота, J_r , %	82,3	90	75
Ошибка первого рода, α , %	7,1	4,7	0
Ошибка второго рода, β , %	17,7	10	25

Получено, что наиболее точно классифицирована стадия F_2 . Это связано с тем, что на ранней стадии количественные показатели значительно ниже показателей других стадий. У всех пациентов со стадией F_3 (ошибка первого рода $\alpha = 0$) она определена верно. Таким образом, подтверждается эффективность предлагаемого метода.

В работе также исследована возможность применения модели принятия решений на основе нечеткого логического вывода для выявления групп пациентов на основе их степени приверженности к терапевтическим вмешательствам. Проверка точности классификации проводилась при сравнении значений классификатора с результатами врача. Получено, что точность классификации предложенной моделью составляет 85% (степень восприятия назначенного лечения для 170 из 200 пациентов определена верно). Таким образом, подтверждена точность и гибкость использования нечеткого классификатора для систем принятия врачебных решений.

ЗАКЛЮЧЕНИЕ

В результате проведенных исследований получены новые теоретические и практические результаты, направленные на повышение точности процесса диагностики заболеваний пациентов на ранней стадии путем поддержки принятия решений при назначении лечения.

1. Проведенный анализ проблем автоматической постановки диагноза *позволил выявить* необходимость в разработке систем поддержки принятия врачебных решений для обнаружения заболевания на ранней стадии в условиях неполноты информации о пациентах.

2. *Установлено*, что предложенная оценка параметров экспертами по четырем критериям (точность полученных значений, уровень достоверности доказательности связи параметра с заболеванием, информативность параметра, статистическая взаимосвязь) дополняет статистическую оценку и помогают определять значимые параметры при ранней диагностике заболевания НАЖБП.

3. *Выявлено*, что для определения стадии заболевания НАЖБП, предложенным гибридным алгоритмом формирования пространства значимых параметров, необходимо использовать следующие параметры медицинского обследования: L_{lep} (лептин), L_{obr} (рецепторы, воспринимающие лептин), D_{nash} (наличие неалкогольного стеатогепатита). Особенностью алгоритма является возможность поставить диагноз даже в случае слабой корреляционной зависимости между стадией заболевания и параметрами медицинского обследования пациента.

4. На основании выявленных значимых параметров и метода нечеткой кластеризации *получены новые результаты*, которые представляют собой сформированные функциональные зависимости между входными мультипликативными параметрами и лингвистическими оценками входных параметров врачом. Таким образом, установлены численные границы и определены степени принадлежности значимых параметров (и замещающих параметров) к заданным лингвистическим термам.

5. *Установлено*, что лучший результат классификации при диагностике НАЖБП (точность классификации) достигается с использованием нечеткого классификатора с привлечением дополнительных (мультипликативных) параметров. Точность классификации достигает 86%. Реализация этого классификатора помогает сократить временные затраты врачей на диагностику заболеваний у большого числа пациентов. Кроме того, реализованная система позволяет интерпретировать полученные результаты, что позволяет достичь объяснимости формируемых диагностических заключений.

ОСНОВНЫЕ РАБОТЫ, ОПУБЛИКОВАННЫЕ ПО ТЕМЕ ДИССЕРТАЦИИ

Статьи в научных журналах, рекомендованных ВАК при Минобрнауки России

1. **Серобабов, А.С.** Разработка экспертной системы ранней диагностики заболеваний: программные средства первичной обработки и выявление зависимостей / А.С. Серобабов, Е.В. Чебаненко, Л.А. Денисова, Т.С. Кролевец // Омский научный вестник. – 2018. – №4 (160). – С. 179-184.

2. **Серобабов, А.С.** Анализ входных параметров экспертной системы ранней диагностики заболевания / А.С. Серобабов // Вестник кибернетики. – 2020. – №4 (40). – С. 33-41.

3. **Серобабов, А.С.** Выбор ключевых параметров для диагностики заболевания печени на основе метода анализа иерархий / А.С. Серобабов // Вестник кибернетики. – 2022. №3 (47). – С. 57-65.

4. **Серобабов, А.С.** Разработка алгоритма выявления значимых параметров для определения стадии заболевания в системе поддержки принятия врачебных решений / А.С. Серобабов, Л.А. Денисова // Известия ТулГУ. – 2023. – №2. – С. 157-162.

5. **Серобабов, А.С.** Разработка системы поддержки принятия врачебных решений при назначении лечения пациенту / А.С. Серобабов, Л.А. Денисова, А.Л. Серобабова // Известия ТулГУ. – 2023. – №9. – С. 321-324.

Статьи в изданиях, индексируемых в базах Scopus и Web of Science

6. Chebanenko, E. Intelligent Processing of Medical Information for Application in the Expert system / E. Chebanenko, L. Denisova, **A. Serobabov** // USBEREIT 2020, – Yekaterinburg, – 2020. – P. 85-88.

7. **Serobabov, A.S.** Development of a medical expert system: Disease staging by a fuzzy classifier / A. S. Serobabov, L. A. Denisova // Journal of Physics: Conference Series: 15, Virtual, Online, 09–11 ноября 2021 года. – Virtual, Online, 2022. – P. 012030.

Статьи в других рецензируемых научных журналах

8. **Серобабов, А.С.** Разработка экспертной системы ранней диагностики заболеваний: регрессионный анализ входных параметров системы / А.С. Серобабов // Прикладная математика и фундаментальная информатика. – 2020. – Т. 7. – № 1. – С. 39-46.

9. **Серобабов, А.С.** Формирование диапазонов переменных экспертной системы с использованием дерева принятия решений / А.С. Серобабов // Journal of Advanced Research in Technical Science. – 2019. – № 17-2. – С. 161-166.

10. **Серобабов, А.С.** Анализ систем интеллектуального диагностирования заболевания у пациента / А.С. Серобабов // Прикладная математика и фундаментальная информатика. – 2019. – №4. – С. 58-69.

11. **Серобабов, А.С.** Анализ массива медицинских данных на наличие выбросов / А. С. Серобабов, А. Л. Серобабова // Мехатроника, автоматика и робототехника. – 2023. – № 11. – С. 198-201.

12. **Серобабов, А.С.** Проверка входных параметров экспертной системы на соответствие нормальному закону распределения / А.С. Серобабов // Проблемы и перспективы студенческой науки. – 2019. – № 2(6).

Статьи в материалах конференций

13. **Серобабов, А.С.** Определение интервалов термов входного параметра в медицинской экспертной системе диагностики на основе алгомеративной кластеризации / А.С. Серобабов // Информационные технологии и автоматизация управления / ОмГТУ – Омск, – 2022. – С. 248-252.

14. Мещеряков, В.А. Модель данных для визуализации организационной структуры вуза на сайте / В.А. Мещеряков, Д.С. Серобабов, **А.С. Серобабов** // Образование. Транспорт. Инновации. / СибАДИ – Омск, – 2021. – С. 776-781.

15. **Серобабов, А.С.** Разработка и верификация программного обеспечения системы принятия решения / А.С. Серобабов, Л. А. Денисова, А.Л. Серобабова // Россия молодая. – 2023. – № 1.

16. **Серобабов, А.С.** Построение регрессионных моделей для входных параметров экспертной системы и их замещающих значений / А.С. Серобабов // Системы управления, информационные технологии и математическое моделирование. / ОмГТУ – Омск, – 2022. – С. 78-86.

Государственная регистрация программ для ЭВМ

17. Свид. о гос. рег. прогр. для ЭВМ №2021667546 Российская Федерация. Автоматизированная система создания функций принадлежности на основе агломеративной иерархической кластеризации по методу Уорда с предположением о нормальном распределении кластеров №2021667546; заяв. 25.10.2021; опубл. 01.11.2021 / **А.С. Серобабов**.

18. Свид. о гос. рег. прогр. для ЭВМ №2020665568 Российская Федерация. Анализатор диапазонов параметров экспертной системы ранней диагностики заболевания печени № 2020665568; заяв. 18.11.2020; опубл. 27.11.2020 / **А.С. Серобабов**.

19. Свид. о гос. рег. прогр. для ЭВМ №2018616153 Российская Федерация. Автоматизированная система прогноза фиброза при неалкогольной жировой болезни печени № 2018616153; заяв. 03.04.2018; опубл. 24.05.2018 / Т.С. Кролевец, М.А. Ливзан, Н.А. Николаев, **А.С. Серобабов**, Е.В. Чебаненко; заявитель и патентообладатель ФГБОУ ВО ОмГМУ Минздрава России.

20. Свид. о гос. рег. прогр. для ЭВМ №2022681058 Российская Федерация. Автоматизированная система вычисления важности ключевых параметров диагностики заболевания неалкогольной жировой болезни печени методом анализа иерархии №2022681058; заяв. 01.11.2022; опубл. 09.11.2022 / **А.С. Серобабов**.

Печатается в авторской редакции
Подписано в печать 21.12.2023.
Формат 60×84/16.Тираж 100 экз.