

# ИНСТРУМЕНТАРИЙ СИНТЕЗА БАЗЫ ЗНАНИЙ ДЛЯ СИСТЕМЫ ПОДДЕРЖКИ ПРИНЯТИЯ ДИАГНОСТИЧЕСКИХ РЕШЕНИЙ ЗАБОЛЕВАНИЙ ЛЕГКИХ В ПРОЦЕССЕ АУСКУЛЬТАЦИИ

## KNOWLEDGE BASE SYNTHESIS TOOLKIT FOR DIAGNOSTIC DECISION SUPPORT SYSTEM FOR LUNG DISEASES DURING AUSCULTATION

|                                                       |                                       |
|-------------------------------------------------------|---------------------------------------|
| ■ Артеменко Михаил Владимирович                       | ■ Artemenko Mikhail Vladimirovich     |
| ■ Федянин Виталий Иванович                            | ■ Fedyanin Vitaly Ivanovich           |
| ■ Шуткин Александр Николаевич                         | ■ Shutkin Alexander Nikolaevich       |
| ■ Квашнина Галина Анатольевна                         | ■ Kvashnina Galina Anatolyevna        |
| ■ Калугина Наталья Михайловна                         | ■ Kalugina Natalia Mikhailovna        |
| ■ Юго-Западный государственный университет            | ■ South-West State University         |
| ■ Воронежский государственный технический университет | ■ Voronezh State Technical University |

E-mail: [artem1962@mail.ru](mailto:artem1962@mail.ru)

### Резюме

Цель исследования - изучение возможностей конвергенции различных методов анализа данных и синтеза базы данных для системы поддержки принятия решений (на примере исследования дыхательного). Разработаны и проанализированы на примере диагностики различных форм астматического бронхита возможности использования следующих методов (и их конвергенции) синтеза решающих правил для базы знаний системы поддержки принятия диагностических решений: спектрально-диапазонный анализ; анализ функциональной организации спектра Фурье звукового сигнала аускультации легких; анализ биспектра регистрируемого сигнала; анализа специфичности графа связности информационных характеристик. В качестве примера рассмотрены решающие правила, построенные на выборках записей дыхательного шума взрослых людей «здоровые», легкая и средняя, тяжелая стадии астматического бронхита. На клиническом репрезентативном материале показано, что предлагаемые методы позволяют осуществлять структурно-параметрическую идентификацию решающих правил с приемлемой уверенностью (диагностическая эффективность полученных решающих правил составила  $0,91 \pm 0,05$ ). Это позволяет существенным образом уменьшить лечебно-профилактические затраты в дальнейшем. Результаты предлагается применять при разработке автоматизированных интеллектуальных систем поддержки принятия решений для превентивной диагностики деструктивных состояний (включая технологии индивидуальной телемедицины) и в учебном процессе повышения квалификации медицинских работников.

**Ключевые слова:** системы поддержки принятия диагностических решений, база знаний, методы синтеза решающих правил, конвергенция, болезни легких, астматический бронхит.

The aim of the study was to study the possibilities of convergence the various methods of data analysis and database synthesis for a decision support system (using the study of respiratory disease as an example). The possibilities of using the following methods (and their convergence) for synthesizing decision rules by the knowledge base of the diagnostic decision support system were developed and analyzed using the example of diagnosing various forms to asthmatic bronchitis: spectral-range analysis; analysis of the functional co-organization of the Fourier spectrum of the lung auscultation sound signal; analysis of the bispectrum of the recorded signal; analysis of the specificity of the connectivity graph of information characteristics. As an example, the decision rules built on samples of respiratory noise recordings by: healthy adults, mild and moderate, severe stages of asthmatic bronchitis are considered. Using clinical representative material, it is shown that the proposed methods allow for structural and parametric identification of decision rules with acceptable confidence (the diagnostic efficiency of the obtained decision rules was  $0.91 \pm 0.05$ ). This allows to significantly reduce treatment and preventive costs in the future. The results are proposed to be used in the development of automated intelligent decision support systems for preventive diagnostics of destructive conditions (including individual telemedicine technologies) and in the educational process of advanced training of medical workers.

**Key words:** diagnostic decision support systems, knowledge base, methods of synthesis of decision rules, convergence, lung diseases, asthmatic bronchitis.

**Библиографическая ссылка на статью**

Артеменко М.В., Федянин В.И., Шуткин А.Н., Квашнина Г.А., Калугина Н.М. Инструментарий синтеза базы знаний для системы поддержки принятия диагностических решений заболеваний легких в процессе аускультации // Innova. - 2024. - Т. 10. - № 4. - С.16-27.

**References to the article**

Artemenko M.V., Fedyanin V.I., Shutkin A.N., Kvashnina G.A., Kalugina N.M. Knowledge base synthesis toolkit for diagnostic decision support system for lung diseases during auscultation // Innova. - 2024. - T. 10. - № 4. - P.16-27.

Заболевания легких занимают одно из ведущих мест в статистике потерь трудоспособности во всех странах мира [14]. По данным института пульмонологии Минздравсоцразвития РФ доля заболеваний органов дыхания в стране составляет около 5-7%, скорость прироста заболевания в год по России составляет 21% [1, 8]. Донозологическая и превентивная диагностики в процессе скрининга существенно уменьшают затраты на лечебно-профилактические мероприятия. Для повышения качества и результативности медицинского обслуживания населения в стране Министерство здравоохранения определило интенсификацию научно-практических исследований в области применения современных информационных и компьютерных технологий в медицине. Между тем, несмотря на большое количество исследований в данном направлении (представленных, например портале [http://bono-esse.ru/blizzard/Socpom/Metod/metod\\_5\\_4.html](http://bono-esse.ru/blizzard/Socpom/Metod/metod_5_4.html)), актуальной остается проблема автоматизации принятия своевременных диагностических решений в клинической, превентивной медицине о возможностях развития деструктивных состояний и быстро прогрессирующих заболеваний легких.

Анализ существующих методов, технологий и алгоритмов, аппаратных и программных средств их информационной поддержки [4, 10] позволяет сделать вывод, что в них недостаточно внимания уделяется: ранжированию и формированию множества информативных признаков, извлекаемых из спектральных характеристик акустических и иных аналоговых сигналах от объектов биологической природы; результативной конвергенции решающих правил; применению показателей системной организации функции [2, 9] в качестве индикаторных переменных; совместному использованию результатов объективного и субъективного анализов; методов искусственного и гибридного интеллекта; применению современных средств и возможностей мобильной связи на этапе превентивной диагностики в процессе скрининга и-или индивидуальных телемедицинских

обследованиях. Это существенным образом сказывается на качестве и результативности применения существующих автоматизированных систем поддержки принятия решений в ходе лечебно-диагностического процесса.

Применяемые в настоящее время методы диагностики заболеваний легких (акустические, рентгенологические, томографические, эндоскопические, ультразвуковые (сонография), радиографические), имея свои преимущества и недостатки, слабо автоматизированы и субъективизированы (особенно акустические). Акустические методы диагностики (в основном – различные формы аускультации) во многом основываются на клиническом опыте медицинского работника их осуществляющего. Появившиеся в последнее десятилетие электронные стетоскопы и фонендоскопы [19] (и патенты) не реализуют диагностический процесс, выполняя функции регистрации, хранения и передачи информации о дыхательных шумах в исследовательские центры, и или индикации критических состояний (ФСЭ-1М, фирм STG Stetnoscope, Repository Medi-Wave InC, диагностический алгоритм BLUE и др.).

В клинической практике для диагностики легочных заболеваний на основе учета множества гетерогенных показателей применяются, например: программный комплекс ПФТ, АРМ для спирометрии СП-300, аппаратно-программные комплексы VentCheck, SomnoChek, комплекс АКДО, экспертная система "TRUTH". Ведущим в этом направлении в Российской федерации является Санкт-Петербургский НИИ Фтизиопульмонологии.

Для анализа непосредственно регистрируемой и латентной информации, полученной на основе спектра дыхательного шума в последнее время интенсивно увеличивается количество разработок и исследований в области использовать достижений искусственного интеллекта, позволяющие, в условиях слабоструктурированных и размытых данных, получать диагностические модели с приемлемым уровнем адекватности [12, 16, 18, 21, 22, 23].

Указанные проблемы обуславливают актуальность цели исследования – изучение

возможностей конвергенции различных методов анализа данных и синтеза базы данных для системы поддержки принятия решений (на примере исследования дыхательного шума для идентификации различных стадий астматического бронхита).

Построенная на полученных диагностических решающих правилах база знаний экспертного модуля системы поддержки принятия решений (СППР) позволит автоматически, в процессе аускультации, осуществлять превентивную диагностику заболеваний органов дыхания и выполнять роль тренажера в обучающем процессе повышения квалификационного уровня медицинских работников соответствующего профиля.

#### Материалы и методы.

Объектом приложения СППР являются дыхательные шумы, характерные для определенных видов деструктивных состояний легких. Дыхательный шум (ДШ), являясь нестационарным процессом с периодической составляющей, имеет сложный спектральный состав без четко выделяемой гармоник основного тона. При этом источник акустических сигналов обладает следующими особенностями [15]:

- источник звуков дыхания (ЗД) имеет определенную структуру, в которой выделяются отдельные анатомические органы,
- каждый входящий в структуру ЗД орган рассматривается как распределенный объект;
- источник ЗД находится в ограниченном пространстве, ткани которого (ткани «корпуса пациента») ослабляют акустические волны и влияют на их характеристики.

Диапазон уверенного прослушивания побочных ДШ, составляющий частоты от 180 до 1400 Гц, делится на три не перекрывающиеся области: низкочастотной (180 - 360 Гц), - среднечастотной (360 - 720 Гц) и высокочастотной (720 - 1440 Гц). Большинство заболеваний имеет характерные частотные диапазоны ДШ.

Поскольку любая запись временного ряда может теоретически содержать участки с артефактами, то рекомендуется регистрацию сигнала осуществлять в течении 30 секунд.

Пусть исходный сигнал после дискретизации и цифровой фильтрации представляет собой временной ряд  $X(t)$ . Для исследования предлагается сформировать дополнительно два временных ряда, характеризующие скорость и «насыщение» за определенное время:  $dX(t)$  и  $sX(t, T)$  - соответственно, оценки первой производной и

интеграл сигнала  $X(t)$  за время  $T$ .

Предлагаются следующие подходы к синтезу решающих правил с последующей конвергенцией результатов:

- спектрально-диапазонный анализ;
- анализ функциональной организации спектра;
- исследование трехкратной автокорреляционной функции (биспектра);
- анализа специфичности графа связности информационных характеристик.

Реализации указанных подходов предшествует формирование множества информативных признаков  $\{X_{inf}\}$ . Поскольку основным критерием является классификационная эффективность, то предлагается формировать указанное множество путем проведения *разведочного дискриминантного анализа* (РДА) [5, 11].

*Методология спектрально-диапазонного способа* заключается в использовании в качестве индикаторных переменных различных характеристик спектра мощности в информативных диапазонах спектра мощности сигналов  $x(t)$ ,  $dx(t)$  и  $Sx(t, T)$ . Информативные диапазоны выделяются: либо визуально в интерактивном режиме, либо, применяя селекционный критерий (в теории измерений рекомендуются максимум энтропии или логарифм среднеквадратичного отклонения [6]). В качестве кандидатов на индикаторные переменные предлагается рассматривать:  $\min(Z(t))$ ,  $\max(Z(t))$ ,  $M(Z(t))$ ,  $\sigma(Z(t))$ ,  $As(Z(t))$ ,  $Ex(Z(t))$  – соответственно, минимальное, максимальное, среднее, среднеквадратичное отклонение, асимметрия и эксцесс сигнала  $Z(t)$  и их различные соотношения ( $Z(t)$  – анализируемые сигналы  $x(t)$ ,  $dx(t)$  и  $Sx(t, T)$  в выделенных информативных диапазонах. Количество информативных диапазонов рекомендуется ограничить тремя - четырьмя.

После формирования значений исходного множества на обучающей выборке (с объектами-примерами различных альтернативных классов), на основе выбранных критериев селекции осуществляется включение признаков в искомый ансамбль информативных признаков. По значениям информативных признаков  $\{X_{inf}\}$  в альтернативных классах синтезируются решающие правила вида:

*ЕСЛИ  $A(F(\{X_{inf}\}))$  ТО Action, (1)*

где антецедент представляет собой, в общем случае, множество логических функционалов над элементами  $\{X_{inf}\}$ , а консеквент включает в себя множество действий Action - соотнесение к определенному классу

(заболеваний).

Во втором случае (анализ функциональной организации спектра) реализуется методология самоорганизационного моделирования с вычислением показателей системной организации функций показатели системной организации функций ISO. Первоначально реализуется спектрально-

$$ISOnd_j = \frac{\sum_{i=1}^N \frac{(p_{i,j} - \bar{p}_i)^2}{\sigma_i^2}}{N}, \quad ISOas_j = \frac{\sum_{i=1}^N \frac{(p_{i,j} - \bar{p}_i)^3}{\sigma_i^3}}{N}, \quad ISOex_j = \frac{\sum_{i=1}^N \frac{(p_{i,j} - \bar{p}_i)^4}{\sigma_i^4}}{N} \quad (2)$$

где:  $N$  – количество информативных признаков,  $\bar{p}_i$ ,  $\sigma_i^2$  – модальное значение и дисперсия признака  $p_i$ , соответственно;  $j$  – индекс объекта в обучающей выборке.

Формирование решающих правил путем исследования трехкратной автокорреляционной функции (ТАКФ) – биспектральной функции – основывается на гипотезе, что дыхательные шумы, характерные для определенных заболеваний легких, различаются кластерами биспектральных функций. Данный способ хорошо зарекомендовал себя при идентификации технических объектов [13] и не использовался ранее в медицинских ЭС.

Методология формирования решающих правил путем анализа специфичности графа связности признакового пространства заключается в следующем (модификация метода, изложенного в [20]). На обучающей выборке строятся графы связности между характеризующими объекты признаками в каждом из альтернативных классов. Фиксируются специфические для каждого класса цепочки вершин графа, ребра которых представляют собой значимые корреляции, которые принимаются за специфическими для класса. В процессе классификации для анализируемого объекта строится граф связности и анализируется его близость к специфическим графам. В качестве меры близости принимается величина, характеризующая количество совпадений со специфическими цепями по соответствующим матрицам смежности.

Описанная процедура хорошо реализуется на достаточно «длинных» аналоговых сигналах путем формирования вычисления характеристик, используя процедуру «скользящего окна».

диапазонная методология по формированию множество  $\{Xinf\}$ . Затем синтезируются решающие правила (1), предварительно сформировав множество значений  $\{ISO\}$ , включающее в себя элементы  $ISO = \{ISOnd, ISOas, ISOex\}$ ,

определяющиеся по формулам:

В качестве примера исследования возможностей описанных подходов в процессе исследования рассматривались классы: нормальное функционирование дыхательной системы: отсутствие легочных заболеваний -  $\omega_0$ ;  $\omega_1$  – легкая и средняя стадии астматического бронхита;  $\omega_2$  – тяжелая стадия.

В работах научной школы Завьялова А.В. отмечается, что физиологические и функциональные системы организма характеризуются определенной организацией (синергетикой) своих элементов как топологически, так и функционально. Поскольку любой наблюдаемый объект предполагает регистрацию характеристик (признаков), то предполагаем, что граф, отображающий значимые функциональные коррелянты между признаками, несет в своей топологии индикаторы кластеров состояний объекта и является для него специфичным. Указанный граф предлагается определять по следующему алгоритму.

1. Определяется множество информативных признаков  $\{Xinf\} = \{x_1, x_2, \dots, x_N\}$ . На экспериментальном материале формируется обучающая матрица, каждый элемент  $x_{ij}$  которой представляет собой значение признака  $j$  для  $i$ -го объекта. Последний столбец матрицы  $(N+1)$ -ый представляет собой идентификатор альтернативного класса.

2. Определяется матрица парных коэффициентов связности, в качестве которых предлагается использовать парные коэффициенты корреляции (если особенности структуры данных не обуславливают применение иных мер близости [7]) -  $R$ .

3. По полученной матрице  $R$  определяется матрица силы связи между признаками  $S$  по следующей формуле:

$$S_{ij} = \begin{cases} [(K * \tanh(\frac{R_{ij}}{R_p} - 1)) + 1], & \text{если } |R_{ij}| > R_p \\ 0, & \text{иначе} \end{cases}, \quad (3)$$

где  $R_p$  – пороговое значение определенного уровня статистической значимости.

При  $R=R_p$  ( $K=10$ ) значение силы связи равно 1, при превышении модуля корреляции порогового значения вдвое, сила связи составляет 6, - при росте корреляции асимптота сила связи равняется ( $K-1$ ). Тем самым, устраняются последствия эффекта ложной корреляции, возникающие при больших объемах выборок.

Элементы полученной матрицы  $S$  позволяют осуществить анализ изменения степени влияния системообразующих признаков для каждого из анализируемых альтернативных классов состояний, аналогично максимальному градиенту функциональных различий [3], но с меньшим количеством вычислений. Введем показатель «степень дифференциации парной связи» ( $SD2K$ ):

$$SD2K_{i,j} = \frac{S_{i,j}}{\sum_{j=1}^N S_{i,j}}, \quad (4)$$

где:  $N$  – количество признаков (размерность матрицы  $S$ ).

(Чем больше значение этого показателя, тем более значима в системе признакового пространства соответствующая парная связь в рассматриваемом классе.)

4. Осуществляется переход от матрицы  $S$  к матрице связности графа  $P$ , каждый элемент которого равен 0, если соответствующий

элемент матрицы  $S$  равен 0, и 1, в противном случае. Знак элемента матрицы  $P$  соответствует знаку элемента матрицы  $R$ .

5. Определяются специфические для каждого класса контуры и цепи в графе по матрице  $P$ . Специфический для каждого класса

граф задается матрицей смежности  $SP_{\omega_l}$ :

$$SP_{\omega_l} = P_{\omega_l} \oplus \prod_{l=0}^L P_{\omega_{l1}} \quad (5)$$

В формуле операция  $\prod_{l=0}^L P_{\omega_{l1}}$  означает поэлементное произведение матриц специфичности всех

классов  $P_{\omega_{l1}}$  ( $l=1, \dots, L$ ;  $L$  – количество классов).

Полученные на этапе обучения матрицы фиксируются в базе данных СППР.

Решающие правила для базы знаний определяются на экзаменационной выборке. Для каждого объекта  $i$ , входящего в альтернативные

классы  $\omega_l$  экзаменационной выборке формируется матрица смежности графов связи

между регистрируемыми признаками -  $SP_{\omega_l}^*$ . Определяется специфичность графа по формуле:

$$SE_{w_l,i} = \frac{\sum E(SP_{w_l} \& SP_{w_l,i}^*)}{\sum E(SP_{w_l})}, \quad (6)$$

где: знаком «&» обозначено произведение элементов матриц, символов операции  $\sum E(A)$  - подсчет количества единиц в матрице  $A$ .

В результате, каждый альтернативный класс характеризуется определенной специфичностью графов связности, которую

предлагается использовать для вычисления значений коэффициента относительной уверенности, в случае применения соответствующих решающих правил, по формуле:

$$\sigma U_{\omega_l,i} = \frac{\frac{SE_{w_l,i}}{\sum_{i=1}^L SE_{w_l,i}}}{\max(\frac{SE_{w_l,i}}{\sum_{i=1}^L SE_{w_l,i}})} \quad (7)$$



Значение относительного коэффициента уверенности лежит в диапазоне  $\sigma U_{\omega_i} \in [0,1]$ .

Если  $\sigma U_{\omega_i} = 1$  - то соотнесение объекта к классу  $\omega_i$  наиболее предпочтительно, чем к другим альтернативным классам.

ЕСЛИ  $\sigma U_{\omega_i}^* \geq \sigma U_{label_{\omega_i}}$  ТО исследуемый объект принадлежит классу  $\omega_i$  с уверенностью не меньше  $\sigma U_{label_{\omega_i}}$ ,

где  $\sigma U_{\omega_i}^*$  - относительная уверенность для анализируемого объекта.

Пороговые значения  $\sigma U_{label_{\omega_i}}$  задаются экспертами или определяется на экзаменационной выборке как модальные значения значений относительных уверенностей  $\sigma U_{\omega_i, i}$  в соответствующих классах.

Таким образом, в отличие от классического коэффициента уверенности, коэффициент относительной уверенности отражает дифференциальную, отличительную особенность результатов применения решающего правила. Тогда решающее правило в базе знаний фиксируется в виде:

$$y(t) = F(y(t-1), y(t-2), \dots, y(t-T_-), y(t+T_+)) \quad (8)$$

где:  $y(t)$  – дискреты аналогового сигнала в момент времени  $t$ ;  $T_-$  – максимальное время запаздывания,  $T_+$  – требуемое (анализируемое) упреждение во времени.

После идентификации функционала (11) для прогноза: либо выделяется из него элемент с необходимым временем прогноза  $y(t+T_+)$  (если существует обратная функция), либо с помощью имитационного моделирования

При классификации технических объектов различной природы по относительно коротким акустическим сигналам, достаточно хорошо зарекомендовал себя биспектральный анализ. В проведенных нами исследованиях, обнаружено, что более лучшими прогностическими возможностями являются функционалы, использующие в качестве аргументов как тренды «прошлого», так и «будущего», т.е. идентифицируются функционалы типа:

осуществляем оценку  $y(t+T_+)$ , обладающее свойством максимального правдоподобия.

Тройная автокорреляционная функция  $\hat{R}_x(k, l)$  (ТАКФ) позволяет на уровне анализа матрицы связности оценить как запаздывание так и упреждение, поскольку для ее вычисления используется формула:

$$\hat{R}_x(k, l) = \sum_{t=0}^{N-1} x(t) \cdot x(t+k) \cdot x(t+l) \quad (9)$$

где:  $N$  – количество дискретных отсчетов сигнала  $x(t)$  для которых строится ТАКФ;  $k, l$  – временные сдвиги.

Согласно рекомендациям [17] для построения биспектральной функции, сигналы  $x(t)$  реализуются  $m$  раз, а затем усредняются, применяя матрицу весовых коэффициентов для различных отсчетов  $k$  и  $l$ .

Модифицировав этот подход, предлагается следующий алгоритм на основе ТАКФ [24]:

1. Регистрируются сигналы  $x(t)_m$  /  $m=1, \dots, M_i$ , где  $M_i$  – количество объектов исследования в классе  $\omega_i$  ( $i=0$  соответствует классу здоровых пациентов);

2. Для сигналов класса  $\omega_0$  строятся

спектральные функции мощностей, по которым формируются множества информативных значений шагов  $k$  и  $l$ :  $kt=(k_1, k_2, \dots, k_{10})$ ,  $lt=(l_1, l_2, \dots, l_{10})$ , (если нет никакой дополнительной информации, то принимается  $l_i=k_i$ ). Индексы соответствуют первым 10 максимумам усредненных спектральных функций мощностей).

3. Из класса  $\omega_0$  случайным образом выбирается сигнал  $x(t)_k$ .

4. Для  $x(t)_k$  вычисляется матрица  $Rx_0$  по формуле (12), подставляя вместо  $k$  и  $l$ , значения векторов  $kk$  и  $ll$ , соответственно. Значения элементов матрицы  $Rx_0$  меньше по модулю 0,3 обнуляются. (Если в результате процедуры обнуления в матрице остается менее 9 (10%) не нулевых элементов лежащих вне главной

диагонали, то пороговое значение 0,3 соответствующим образом снижается).

Определяется обратная матрица  $R_0^{-1} = R x_0^{-1}$ , значения которой запоминаются в базе данных СППР.

5. Процедуры п.4 повторяются для всех сигналов кластера  $\omega_l$ , формируя матрицы  $Rx_{l,j}$  ( $l$

$$SR_{\omega_l, j} = \frac{\sum_{i=1}^N \sum_{m=1}^N IR_{\omega_l, j} - N}{N^2}. \quad (10)$$

Показатель  $SR_{\omega_l, j}$  принимается в качестве индикаторной переменной.

7. Формируются диагностические интервалы по индикаторной переменной, в качестве которых принимаются доверительные интервалы  $dSR_l$ .

8. Определяются коэффициенты уверенности решающих правил  $UR_j$ , качества

– индекс класса,  $j$  – индекс объекта в классе  $\omega_l$ ,  $l=0$ ).

6. Вычисляются индикаторные матрицы вида:  $IR_{\omega_l, j} = R x_{l, j} \cdot R x_0$ . Определяются вектора среднеквадратичных отклонений от единичной матрицы:

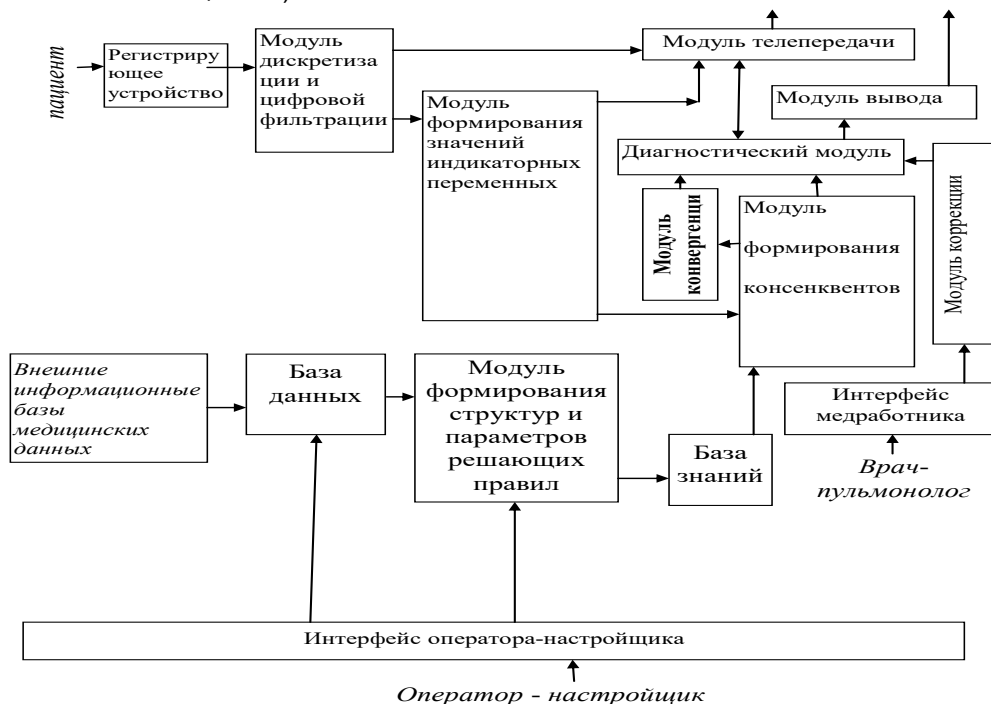
которых принимается диагностическая специфичность применения доверительного интервала в качестве условий активации соответствующих продуктов на экзаменационной выборке. Найденные параметры и структуры диагностических решающих правил производственного типа формируют базу знаний в виде конструкций:

(11) *ЕСЛИ  $SR^* \subset dsR$ , ТО состояние объекта*

соответствует классу  $\omega_i$  с уверенностью  $UR_i$

(где:  $SR^*$  - значение индикаторной переменной определенной в результате применения формул (12) и (13) для акустического сигнала анализируемого объекта (состояния респираторной системы пациента).

Архитектура СППР представлена на рисунке 1. Отличительной особенностью модели является возможность конвергенции консенквентов решающих правил, находящихся в базе данных.



**Рисунок 1** - Архитектура СППР врача-пульмонолога

### Результаты исследований.

Выборка исходных примеров для рассматриваемых классов составила (прослушивалось и регистрировалось бронхиальное дыхание):  $\omega_0$  - «здоровые» - 54 записи акустического сигнала дыхательного шума (прослушивалось бронхиальное дыхание),  $\omega_1$  - легкая и средняя стадии астматического бронхита – 82 записи,  $\omega_2$  – тяжелая стадия – 58 записей. Случайным образом были сформированы обучающая и экзаменационная выборки объемными соотношениями в соответствии с «золотым сечением» согласно рекомендациям: обучающая выборка для  $\omega_0$  – 33 записи, для  $\omega_1$  – 51 запись, для  $\omega_2$  – 36 записей. Каждая запись получила идентификатор  $x(t)_{l,k}$  ( $l$  – «номер» класса,  $k$  – номер записи в классе). Из записей была выделены участки временного промежутка в 20 секунд. Каждая запись в выборках задавалась идентификатором  $x(t)_{l,k}$  ( $l$  – «номер» класса,  $k$  – номер записи в классе). Из записей выделялись участки временного промежутка в 20 секунд. Все сигналы для соблюдения чистоты эксперимента отфильтрованы и оцифрованы с частотой дискретизации 2,8 кГц.

Обучающая и экзаменационные подвыборки формировались путем рандомизации. По энтропийному критерию выделены следующие информационные диапазоны частот:  $d_1$  – 5-120 Гц,  $d$  численные методы, определяют дискретные значения сигналов  $dx(t)_{l,k}$  и  $sx(t, T)_{l,k}$  – соответственно, первая производная и интеграл (за время  $T=10 \cdot t$ ,  $t=1/f_d$ ,  $f_d$  – частота дискретизации).

Первоначальный ансамбль признаков включал в себя 126 элементов - кандидатов на индикаторные переменные:  $\min(Z(t))$ ,  $\max(Z(t))$ ,  $M(Z(t))$ ,  $\sigma(Z(t))$ ,  $As(Z(t))$ ,  $Ex(Z(t))$  и их арифметические комбинации – соответственно, минимальное, максимальное, среднее, среднеквадратичное отклонение, асимметрия и эксцесс сигнала  $Z(t)$  и их различные соотношения ( $Z(t)$  – анализируемые сигналы  $x(t)$ ,  $dx(t)$  и  $Sx(t, T)$  в выделенных информативных диапазонах).

После применения метода РДА на основе линейных регрессий искомое множество информативных признаков уменьшилось вдвое и

составило 63 элемента -  $S_{i,d_j,m}$  ( $i=1, \dots, 7$ ;  $j=1, 2, 3$  – номера выделенных частотных диапазонов;  $m=1$  – соответствует сигналу  $x(t)_{l,k}$ ,  $m=2$  -  $dx(t)_{l,k}$  и  $m=3$  -  $sx(t, T)_{l,k}$ ),

$$\begin{aligned} \text{а именно: } S_{1,d_j,m} &= \frac{\sum_{f_{d_j}=1}^{F_{d_j}} sp_{f,d_j}(z(t))}{F}, \quad S_{2,d_j,m} = \sqrt{\frac{(\sum_{f_{d_j}=1}^{F_{d_j}} sp_{f,d_j}(z(t)) - s_{2,d_j,m})^2}{F}}, \quad S_{3,d_j,m} = \frac{S_{2,d_j,m}}{S_{1,d_j,m}}, \\ S_{4,d_j,m} &= \ln(s_{2,d_j,m}), \quad S_{5,d_j,m} = \frac{\max(sp_{f,d_j}(z(t))) - \min(sp_{f,d_j}(z(t)))}{S_{1,d_j,m}}, \\ S_{6,d_j,m} &= As(sp_{f,d_j}(z(t))), \quad S_{i,d_j,m} = Ex(sp_{f,d_j}(z(t))) \end{aligned}$$

Здесь:  $sp_{f,d_j}(z(t))$  - спектральные мощности сигнала  $z(t)$  в диапазоне  $d_j$  соответствующая дискрете спектра  $f_{d_j}$ ,  $f_{d_j}$  – количество частотных дискрет в диапазоне  $d_j$ ;  $As()$  и  $Ex()$  – соответственно, асимметрия и эксцесс.

Проведенный анализ корреляции частотных диапазонов по информативным признакам внутри них показал, что: между собой коррелируют  $d_1$  и  $d_2$ ,  $d_2$  и  $d_3$  диапазоны по

спектрам мощностей сигналов  $x(t)_{l,k}$ , первый диапазон по сигналу  $x(t)_{l,k}$  и сигналам  $dx(t)_{l,k}$  и  $sx(t)_{l,k}$ , третий диапазон и второй диапазоны по  $sx(t)_{l,k}$ .

Специфические для классов заболеваний цепочки графов связности представлены анализируемых характеристик спектральных

окон  $S_{i,d_j,m}$  (частотных диапазонов) представлены в таблице 1.



Таблица 1. Специфические цепочки графа связности

| Класс      | Специфические цепочки                     | m | d       |
|------------|-------------------------------------------|---|---------|
| $\omega_0$ | $S_{1,d_j,m} - S_{2,d_j,m} - S_{5,d_j,m}$ | 1 | 1, 2, 3 |
|            |                                           | 2 | 2, 3    |
|            |                                           | 3 | 1, 3    |
| $\omega_0$ | $S_{1,d_j,m} - S_{6,d_j,m}$               | 1 | 1, 2, 3 |
|            |                                           | 3 | 1, 3    |
| $\omega_1$ | $S_{2,d_j,m} - S_{5,d_j,m} - S_{6,d_j,m}$ | 2 | 1, 2, 3 |
|            |                                           | 3 | 1, 2, 3 |
| $\omega_1$ | $S_{4,d_j,m} - S_{5,d_j,m}$               | 2 | 1, 3    |
|            |                                           | 3 | 1, 2    |
| $\omega_2$ | $S_{3,d_j,m} - S_{5,d_j,m}$               | 1 | 1, 2    |
|            |                                           | 2 | 2, 3    |
|            |                                           | 3 | 1, 2    |
| $\omega_2$ | $S_{2,d_j,m} - S_{3,d_j,m}$               | 1 | 1,3     |
|            |                                           | 3 | 1, 3    |

Выявлено, что: значения характеристик для сигналов  $dx(t)_{l,k}$  и  $sx(t)_{l,k}$  у здоровых слабо коррелируют между собой. На легкой и средней тяжести астматического бронхита доминируют процессы, отражающие скорость и интенсивность изменения характеристик сигнала во всех рассматриваемых частотных диапазонах (организм находится в некоторых флуктуационных точках), в тяжелой стадии заболевания – коррелируют характеристики на более высоких частотах, что связано с физиологическими изменениями, связанными с прогрессирующей деструкцией, ослаблением стенок бронхов и увеличение объема мокроты в легких. Данный вывод не противоречит известным медицинским положениям развития заболевания.

Анализ корреляции признаков внутри диапазонов с одной стороны подтвердил ортогональность признакового пространства, с другой – выявил линейные значимые корреляции между дисперсией, максимальным значением и средним значением мощностей в окне спектра соответствующего частотного диапазона и между дисперсией и суммой мощностей сигнала в спектре. Это подтвердило гипотезу о том, что в

информативных диапазонах колебания мощностей спектра (а, следовательно, вариабельности амплитуд сигнала) прямо пропорционально связаны с амплитудой. Т.е., чем лучше слышится звук (больше амплитуда), тем больше он вариабелен и, следовательно, информативен. Проведенный анализ позволил уменьшить объем множества информативных признаков на 40%, удалив из него высоко коррелированные признаки.

Применяя предложенные методы (инструменты) получены следующие решающие правила (и результаты их верификации на экзаменационной выборке): по показателям системной организации функций – представлены в таблице 2, по специфическим структурам – таблица 3, по ТАКФ – таблица 4. Назовем данные группы решающих правил, соответственно: SOF, SS, TAF. В случае, если значения индикаторных переменных лежат за границами, указанных в таблицах, то требуется дополнительное обследование, поскольку высок риск более тяжелых стадий заболеваний астматического бронхита или других патологий легких.

Таблица 2. Решающие правила по показателям системной организации функций

| N | Анцедент                                     | консеквент       | Коэффициент уверенности $U_{sof}$ | ДЧ   | ДС   | ДЭ   |
|---|----------------------------------------------|------------------|-----------------------------------|------|------|------|
| 1 | $ISOnd \subset [0, 2]$                       | $O \in \omega_0$ | 0,91                              | 0,91 | 0,92 | 0,92 |
| 2 | $ISOas \subset [-2; 2]$                      | $O \in \omega_0$ | 0,85                              | 0,85 | 0,9  | 0,88 |
| 3 | $ISOex \subset [0; 4]$                       | $O \in \omega_0$ | 0,88                              | 0,88 | 0,91 | 0,9  |
| 4 | $ISOnd \subset ]2, 6]$                       | $O \in \omega_1$ | 0,52                              | 0,88 | 0,52 | 0,63 |
| 5 | $ISOas \subset ([-5, 7; -2[ \cup ]2; 5, 7])$ | $O \in \omega_1$ | 0,86                              | 0,86 | 0,87 | 0,87 |

|   |                                                    |                  |      |      |      |       |
|---|----------------------------------------------------|------------------|------|------|------|-------|
| 6 | $ISOex \subset ([-14; -4] \cup [4; 14])$           | $O \in \omega_1$ | 0,87 | 0,88 | 0,87 | 0,875 |
| 7 | $ISOnd \subset ]6, 12]$                            | $O \in \omega_2$ | 0,85 | 0,89 | 0,94 | 0,925 |
| 8 | $ISOas \subset ([-6, 3; -5, 7] \cup [5, 7; 6, 3])$ | $O \in \omega_2$ | 0,91 | 0,86 | 0,93 | 0,91  |
| 9 | $IISOex \subset ([-120; -14] \cup [14; 120])$      | $O \in \omega_2$ | 0,89 | 0,89 | 0,94 | 0,925 |

(номера 1-3 соответствуют сигналам  $x(t)$   $dx(t)$   $Sx(t, T)$ , соответственно)

**Таблица 3.** Результаты верификации решающих правил по показателям специфических структур графа связности

| N | Анцедент<br>$\sigma U_{\omega_1}^* \geq \sigma U_{\omega_1} label$ | консеквент       | Коэффициент<br>уверенности $U_{ss}$ | ДЧ   | ДС   | ДЭ   |
|---|--------------------------------------------------------------------|------------------|-------------------------------------|------|------|------|
| 1 | $\sigma U_{\omega_0} label = 0,9$                                  | $O \in \omega_0$ | 0,9                                 | 0,76 | 0,89 | 0,85 |
| 2 | $\sigma U_{\omega_0} label = 0,85$                                 | $O \in \omega_1$ | 0,85                                | 0,84 | 0,85 | 0,85 |
| 3 | $\sigma U_{\omega_0} label = 0,8$                                  | $O \in \omega_2$ | 0,8                                 | 0,78 | 0,87 | 0,84 |

$\sigma U_{\omega_1} label$

Примечание: пороговые значения коэффициента относительной уверенности определялись экспертами – медиками.

**Таблица 4.** Решающие правила по ТКАФ

| N | сигнал     | Анцедент: $SR^* \subset dsR_i$ | консеквент       | Коэффициент<br>уверенности<br>$U_{tkaf}$ | ДЧ   | ДС   | ДЭ    |
|---|------------|--------------------------------|------------------|------------------------------------------|------|------|-------|
| 1 | $x(t)$     | $dsR < 5$                      | $O \in \omega_0$ | 0,91                                     | 0,91 | 0,92 | 0,92  |
| 2 | $dx(t)$    | $dsR < 10$                     | $O \in \omega_0$ | 0,86                                     | 0,86 | 0,87 | 0,87  |
| 3 | $Sx(t, T)$ | $dsR < 15$                     | $O \in \omega_0$ | 0,86                                     | 0,86 | 0,93 | 0,91  |
| 4 | $x(t)$     | $dsR = 21 \pm 16$              | $O \in \omega_1$ | 0,82                                     | 0,82 | 0,89 | 0,87  |
| 5 | $dx(t)$    | $dsR = 700 \pm 680$            | $O \in \omega_1$ | 0,84                                     | 0,86 | 0,84 | 0,85  |
| 6 | $Sx(t, T)$ | $dsR = 24 \pm 8$               | $O \in \omega_1$ | 0,83                                     | 0,83 | 0,94 | 0,91  |
| 7 | $x(t)$     | $dsR = 92 \pm 34$              | $O \in \omega_2$ | 0,76                                     | 0,76 | 0,89 | 0,85  |
| 8 | $dx(t)$    | $dsR = 1020 \pm 930$           | $O \in \omega_2$ | 0,74                                     | 0,74 | 0,86 | 0,825 |
| 9 | $Sx(t, T)$ | $dsR = 114 \pm 81$             | $O \in \omega_2$ | 0,84                                     | 0,84 | 0,87 | 0,84  |

Где:  $x(t)$  – анализируемый сигнал, представленный в дискретной форме;  $dx(t)$  – первая производная сигнала  $x(t)$ ;  $Sx(t, T)$  – интеграл  $x(t)$  за интервал времени  $T$

Из таблицы 4 видно, что доверительные интервалы по некоторым правилам пересекаются, что не противоречит клиническим наблюдениям и соответствуют переходным состояниям между классами.

Для определения конвергенции решающих правил первоначально определяем значения порогов уверенности мажоритарного

голосования в группах SOF, SS, TAF одинаковыми и равными 0,75 (диагностическая эффективность неавтоматизированной диагностики стандартными способами при аускультации). Тогда, согласно алгоритму, предложенному в главе 2, получаем значения коэффициенты уверенности по группам – представлены в таблице 5.

Таблица 5. Групповые коэффициенты уверенности

|            | SOF  | SS   | TAF  |
|------------|------|------|------|
| $\omega_0$ | 0,88 | 0,9  | 0,86 |
| $\omega_1$ | 0,87 | 0,85 | 0,84 |
| $\omega_2$ | 0,91 | 0,8  | 0,84 |

На экзаменационной выборке получена матрица коэффициентов конкордации Кендалла:

|     | SOF  | SS   | TAF  |
|-----|------|------|------|
| SOF | 1    | 0.82 | 0.9  |
| SS  | 0.82 | 1    | 0.72 |
| TAF | 0.9  | 0.72 | 1    |

Получаем соответствующие весовые коэффициенты групповой уверенности: 0,86; 0,77; 0,8. Следовательно, значения коэффициентов уверенности в правильной классификации (диагностики) при конвергенции всех групп решающих правил составили: для класса  $\omega_0$  -0,88,  $\omega_1$  -0,85,  $\omega_2$  -0,85 (что на 12-15% лучше используемых в настоящее время традиционных методов).

#### Закключение.

Теоретическая и практическая значимость работы состоит в развитии технологических решений формирования базы знаний систем поддержки принятия решений, применяемых для улучшения качества медицинского обслуживания населения в ходе скрининга, массового обслуживания населения или индивидуальной медицины на основе конвергенции результатов автоматизированной рекомендательной диагностики. Предлагаемые методы позволяют идентифицировать и конвергировать решающие правила продукционного типа с клинически приемлемыми значениями показателей качества диагностики (классификации), в antecedенты которых входят: показатели системной организации функций, специфические структуры графа связности признакового пространства, различия трехкратных автокорреляционных функций, бальные оценки на вопросы анкеты. Полученные по предложенным методам решающие правила составляют основу системы поддержки принятия диагностических решений с приемлемыми качествами. Это позволит, проводя раннюю диагностику, существенным образом уменьшить затраты на лечебно-профилактические ресурсы в дальнейшем. Результаты исследования могут быть использованы при разработке автоматизированных интеллектуальных систем поддержки принятия решений для превентивной диагностики деструктивных состояний легких в процессе аускультации (включая технологии индивидуальной телемедицины) и в учебном процессе повышения квалификационного уровня медицинских работников.

#### Литература.

1. Антонов Н. С. и др. Динамика заболеваемости болезнями органов дыхания среди населения Российской Федерации в 2010–2022 гг // Медицина. – 2023. – Т. 11. – №. 3. – С. 1-17.
2. Артеменко М.В., Корневский Н.А., Дронова Т.А. Применение показателей системной организации в диагностическом процессе // Системный анализ и управление в биомедицинских системах. 2003. Т. 2. № 1. С. 16-20.
3. Артеменко М. В., Дронова Т.А., Калугина Н.М., Бойцов А.А.. Применение показателей системной организации в автоматизации диагностико-терапевтического процесса // Юго-Западного государственного университета. // Известия Юго-Западного государственного университета. Серия: Управление, вычислительная техника, информатика, Медицинское приборостроение – 2017. Т.7 №2(3). С.122-134
4. Борисов И. В. Совершенствование управления здравоохранением на основе цифровой трансформации : дис. – Автореферат диссертации на соискание ученой степени кандидата экономических наук. Владимир–2023, 2023
5. Ватьян А. С. и др. Интеллектуальная поддержка клинических решений при небольших выборках числа пациентов // Научно-технический вестник информационных технологий, механики и оптики. – 2023. – Т. 23. – №. 3. – С. 595-607.
6. Грибанов Д. Д. Общая теория измерений: монография Москва : ИНФРА-М, 2015. - 115 с.
7. Гайдышев И. П. Моделирование стохастических и детерминированных систем: Руководство пользователя программы AtteStat // БИ, 2015.–484 с. – 2015.
8. Единая государственная информационная система в сфере здравоохранения (ЕГИСЗ) // URL

<https://egis.rosminzdrav.ru/#firstPage>

9. Завьялов А.В. Системная организация физиологических функций в норме и патологии: Актовая речь на заседании Ученого совета Курского государственного медицинского университета 8 февраля 2005 г. - Курск, КГМУ, 2005. -20с.

10. Информационные технологии в здравоохранении  
<https://md.school/blog/informacionnye-tehnologii-v-medicine>

11. Корневский Н.А., Зарубина Т.В., Щелыкалина С.П., Артеменко М.В. Математические методы обработки медико-биологической информации. Статистика. Старый Оскол ООО "Тонкие наукоемкие технологии", 2024. - 416 с.

12. Мосин С. Г. Нейросетевая диагностика заболеваний сердечно-сосудистой системы на основе data-driven-метода //Программные продукты и системы. - 2024. - Т. 37. - №. 1. - С. 122-130.

13. Манохин А. Е. Идентификация цифровых последовательностей с интегральным биспектральным преобразованием //Ural Radio Engineering Journal. 2023. Vol. 7.№ 1. - 2023. - Т. 7. - №. 1. - С. 56-71.

14. Распространенность хронических обструктивных болезней легких //URL [https://gateway.euro.who.int/ru/indicators/hfa\\_403-2510-prevalence-of-chronic-obstructive-pulmonary-disease/#id=19388](https://gateway.euro.who.int/ru/indicators/hfa_403-2510-prevalence-of-chronic-obstructive-pulmonary-disease/#id=19388)

15. Тихоненко Д. О. и др. Новые подходы к акустическому анализу состояния организма человека //Вестник современных цифровых технологий. - 2021. - №. 6. - С. 29-36.

16. 1 Хусанов У. А. У. и др. Искусственный интеллект в медицине //Science and Education. - 2023. - Т. 4. - №. 5. - С. 772-782

17. Цифровая обработка сигналов и изображений в радиофизических приложениях. / Под ред. В.Ф. Кравченко. -М.: ФИЗМАТЛИТ, 2007. -544 с.

18. Шайханова А. и др. Интеграция искусственного интеллекта для обнаружения респираторных заболеваний в программно-аппаратный комплекс «ДИАГНОСТИКА НА ДОМУ» //Вестник КазАТК. - 2024. - Т. 135. - №. 6. - С. 272-282.

19. 6 лучших стетоскопов //URL <https://expertology.ru/6-luchshikh-stetoskopov/>

20. Artemenko M.V., Kalugina N.M., Dobrovolsky I.I The formation of a set of informative features based on the functional relationships between the data structure field observations //European Journal of Natural History. 2016. № 6. P. 43-48.

21. Chen J. et al. Artificial intelligence for heart sound classification: A review //Expert Systems. - 2024. - Vol. 41. - №. 4. - P. e13535.

22. Garanin A. A. et al. Digital stethoscope: a new era of auscultation //Digital Diagnostics. - 2024. - Vol. 5. - №. 4. - P. 808-818.

23. Kapetanidis P. et al. Respiratory Diseases Diagnosis Using Audio Analysis and Artificial Intelligence: A Systematic Review //Sensors. - 2024. -Vol. 24. - №. 4. - P. 1173.

24. Артеменко М.В., Калугина Н.М. Способ диагностики астматического бронхита в процесса аускультации легких взрослых людей // Патент на изобретение, Номер патента: RU 2766751 С1 Патентное ведомство: Россия. Год публикации: 2022.