

УДК 004.82

doi: 10.15622/rcai.2025.093

## ПРОГНОСТИЧЕСКАЯ СИСТЕМА МОНИТОРИНГА БОЛЬНЫХ ХРОНИЧЕСКИМ ЛИМФОЛЕЙКОЗОМ

Н.А. Игнатьев (*n\_ignatev@rambler.ru*)<sup>A</sup>

Е.Н. Згуральская (*iatu@inbox.ru*)<sup>B</sup>

М.В. Марковцева (*mmark7@yandex.ru*)<sup>C</sup>

<sup>A</sup> Национальный университет Узбекистана, Ташкент

<sup>B</sup> Ульяновский государственный технический университет, Ульяновск

<sup>C</sup> Ульяновский государственный университет, Ульяновск

Интеллектуальные системы мониторинга состояния здоровья в клинической практике оптимизируют оказание медицинской помощи пациентам. Такие системы позволяют с течением времени отслеживать изменения значений наблюдаемых параметров, вносить коррективы в прогноз заболевания. Результаты интеллектуального анализа, полученные в виде правил продукций, особенно полезны для представления экспертных знаний и автоматизации процессов принятия решений. Для формализации знаний разработана онтология, применяемая для мониторинга состояния пациентов в течение 5 лет с момента постановки диагноза хронический лимфолейкоз (ХЛЛ). Выделены основные связи понятий, достаточные для решения задач мониторинга состояния больных ХЛЛ в процессе лечения. Разработанная онтология входит в основу прогностической системы, где параметры мониторинга могут изменяться с учётом стадии лечения, клинических, лабораторных данных пациентов. Описан метод извлечения знаний, основанный на теории нечётких множеств. Описана система мониторинга, связанная с базой знаний.

**Ключевые слова:** онтология, правила продукций, база знаний, интеллектуальный анализ данных, хронический лимфолейкоз.

### Введение

Мониторинг состояния пациентов является неотъемлемой частью организации лечебно-диагностического процесса в современных условиях [Грибова и др., 2024]. На практике мониторинг позволит своевременно прояснить клиническую ситуацию пациента и при необходимости скор-

ректировать лечение больного. Полученные данные хранятся и используются для создания медицинских информационных, экспертных или рекомендательных систем, в которых накопленную информацию объединяют в базы знаний (БЗ), формируя на их основе методы решения задач [Богданова и др., 2024]. Система обобщения и классификации, которая реализована в интеллектуальных системах (ИС), позволяет сводить множество возможных ситуаций в многопризнаковом пространстве к небольшому числу типовых ситуаций, по которым выдвигаются гипотезы и производится выбор методов решения [Кобринский и др., 2020].

Навыки и опыт врача не представлены в строго формализованном виде, они являются результатом синтеза теоретических знаний и клинических наблюдений. С учётом этих особенностей в медицине связано построение ИС, т. е. систем, основанных на работе не с данными о больных, а со знаниями высоко профессиональных специалистов-экспертов [Кобринский, 2001]. Интеллектуальный анализ данных (ИАД), предлагая методы для выявления закономерностей, проверки гипотез и извлечения ценной информации, нивелирует имеющиеся экспертные противоречия. Гипотезы, выдвинутые врачами, проверяются методами ИАД и в случае их подтверждения включается в БЗ.

Желательным свойством при поиске закономерностей в базах и хранилищах данных является инвариантность результатов анализа к масштабам их измерений. При поиске закономерностей использовались интервальные методы, одним из универсальных ограничений которых является инвариантность к масштабам измерений данных [Згуральская, 2018]. Свойство инвариантности упрощает процедуру интерпретации результатов в терминах предметных областей. Это свойство даёт возможность выбора латентных признаков при моделировании процесса интуитивного принятия решений специалистами.

В [Игнатьев и др., 2020] рассматривается поиск логических закономерностей по описаниям объектов в спрямляющем пространстве. При синтезе латентных признаков этого пространства используются правила иерархической агломеративной группировки. Выбор пары признаков для объединения в группу реализуется по максимуму критерия разбиения значений признаков на непересекающиеся интервалы. Приводится аналитический вид арифметических выражений для расчета латентных признаков, используемых для обнаружения скрытых закономерностей на данных больных ХЛЛ.

В статье предлагается подход к построению системы мониторинга больных ХЛЛ с использованием БЗ, наполнение которой произведено методами ИАД.

## 1. Постановка задачи

ХЛЛ – одно из наиболее распространённых онкогематологических заболеваний. Высокая распространённость ХЛЛ у пациентов европейской популяции в пожилом (60–75 лет) и старческом (более 75 лет) возрасте представляет на сегодняшний день актуальную проблему. Мониторинг пациентов с ХЛЛ является важной задачей для оптимизации тактики ведения больных. Предложенные в последнее время способы прогнозирования и мониторинга ХЛЛ требуют сложных генетических исследований, которые не всегда являются доступными: мутационный статус генов варибельного региона иммуноглобулинов (IGHV), мутация TP53, делеция 13q14 и др. [Mayr et al., 2006]. Подобная ситуация диктует необходимость продолжения поиска общедоступного решения данного вопроса [Марковцева и др., 2023].

Анализ клинических данных методами ИАД позволяет оперировать более точной и персонализированной информацией о прогнозе заболевания, учитывая влияние на состояние пациента возможных коморбидных эффектов. На основании текущих индивидуальных определений параметров нельзя однозначно утверждать о возможном состоянии больных в последующем, т.е. сделать прогноз [Адмакин, 2016]. В задаче интеллектуального мониторинга требуется по информации о системе, состоянии которой вышло из гипотетической нормы, с помощью формализованных знаний определить показатели каких наблюдений и в какие моменты времени является индикатором, что состояние системы отличается от ожидаемого (прогнозируемого). Для этого требуется:

- собрать информацию (клинические данные) о пациентах с ХЛЛ, с момента постановки диагноза и в течении последующей пятилетней динамике заболевания;
- методами ИАД выявить наличие закономерностей в динамически изменяемых данных;
- обнаруженные закономерности представить в виде правил продукций;
- разработать блок объяснения процесса принятия решений.

## 2. Модель БЗ для мониторинга состояния больных

Для формализации знаний разработана онтология, применяемая для мониторинга состояния пациентов с момента постановки диагноза ХЛЛ и на протяжении всего заболевания.

Онтологию  $O_{HLL}$  можно представить в виде следующего упорядоченного набора элементов:

$$O_{HLL} = \langle G, I, SWRL \rangle,$$

где  $G$  – утверждения, соответствующие уровню схемы онтологии  $O_{HLL}$ ;

$I$  – множество утверждений, соответствующих индивидам онтологии;

$SWRL$  – множество продукционных правил на языке SWRL.

Множество  $G$  может быть определено кортежем, состоящим из следующих взаимосвязанных элементов:

$$G = \{G^C, G^O, G^D, G^A\},$$

где  $G^C$  – множество утверждений, характеризующих классы (концепты) онтологии;

$G^O$  – множество утверждений, описывающих объектные свойства, объявленные и используемые в онтологии для формирования связей между концептами;

$G^D$  – множество утверждений, описывающих свойства данных;

$G^A$  – множество аксиом, представляющих основные ограничения для элементов множеств  $G^C, G^O, G^D$ .

Онтологическая модель (ОМ) включает классы, отношения и ограничения, которые определяют предметную область, тем самым устанавливает единую терминологию для совместного использования информации о взаимодействии элементов, необходимых для мониторинга состояния пациентов в лечебно-диагностическом процессе.

В ОМ  $O_{HLL}$  представлены пять классов

, где:

- класс «Клинические факторы», содержащий сведения о возрасте, наличие сопутствующих заболеваний, наличие симптомов (лихорадка, ночная потливость, потеря веса, усталость) и т.п.

- класс «Лабораторные показатели» содержащий сведения о лабораторных исследованиях крови АЛТ, АСТ, билирубин-общий, непрямо-билирубин, креатинин, мочевины, глюкоза и т.п.;

- класс «Стадия заболевания», определение стадии по системе Binet, которые основаны на клинических и лабораторных факторах (количество лимфоцитов, вовлечение органов, уровень гемоглобина).

- класс «Риск», отнесение к данному классу производится на основе правил продукций, полученных методами ИАД и зависит от клинических, лабораторных показателей: креатинин, пол, возраст и стадии заболевания. Это позволяет механизму логического вывода автоматически классифицировать экземпляры этих классов, если они удовлетворяют заданным условиям.

- класс «Пациенты».

Правила продукций являются одним из методов представления и использования знаний в экспертных системах и в ИАД. Они особенно полезны для записи экспертных знаний и автоматизации процессов принятия решений.

Для формирования схемы БЗ и правил продукций на языке SWRL использован редактор онтологий Protégé [Horrocks et al., 2004]. В результате схема БЗ и множество продукций на языке SWRL представляются в виде онтологии на языке OWL. Фрагмент онтологии представлен на рис. 1.



Рис. 1. Фрагмент онтологии

Правила продукций на языке SWRL представлены на рис. 2.

	Query
Пациенты(?p) ^ имеет_СКФ(?p, ?skf) ^ имеет_срок_заболевания(?p, ?srok) ^ swrlb:lessThanOrEqual(?skf, 76) -> Риски_есть(?p)	
Пациенты(?p) ^ имеет_СКФ(?p, ?skf) ^ имеет_срок_заболевания(?p, autogen1:_года) ^ swrlb:greaterThan(?skf, 75) -> Рисков_нет(?p)	
Пациенты(?p) ^ имеет_СКФ(?p, ?skf) ^ имеет_срок_заболевания(?p, autogen0:_года) ^ swrlb:greaterThan(?skf, 70) -> Рисков_нет(?p)	
Пациенты(?p) ^ имеет_СКФ(?p, ?skf) ^ имеет_срок_заболевания(?p, autogen4:_года) ^ swrlb:greaterThan(?skf, 65) -> Рисков_нет(?p)	
Пациенты(?p) ^ имеет_СКФ(?p, ?skf) ^ имеет_срок_заболевания(?p, постановка_диагноза) ^ swrlb:greaterThan(?skf, 76) -> Рисков_нет(?p)	

Рис. 2. Правила продукций на языке SWRL

### 3. Применение методов ИАД для выявления закономерностей

Для поиска закономерностей применялись методы разбиения значений признаков на непересекающиеся интервалы. На базе этих методов стало возможным как вычисление обобщённых оценок объектов (латентных признаков) в разнотипном признаковом пространстве, так и меры их устойчивости [Згуральская, 2018]. Следует отметить, что применение альтернативных подходов для выявления закономерностей существенно ограничено особенностями процедуры сбора исходных данных.

В предлагаемой системе мониторинга используется такое фундаментальное свойство размеченных данных, как устойчивость. Вычисление устойчивости со множеством допустимых значений в  $(0,5;1]$  основано на теории нечётких множеств. Доказано, что свойство устойчивости присуще всем выборкам из генеральной совокупности. Свойство выполняется для количественных и номинальных (качественных) данных, является инвариантным к масштабам измерений. Инвариантность прослеживается при разбиении значений признаков на непересекающиеся интервалы с доминированием в каждом из них представителей одного из двух классов.

Количество интервалов и их состав определяются как решение задачи о минимальном покрытии по упорядоченной последовательности значений признака. Вычисление значений функции принадлежности по интервалам с учётом их оптимальных границ рассматривается как источник новых знаний о предметной области. Доказано [Игнатьев и др., 2023], что определяемый по многообразиям значений функции принадлежности показатель устойчивости признака сходится к фиксированной величине.

Было исследовано 207 пациентов (98 мужчин и 109 женщин) в возрасте от 45 до 80 лет с ХЛЛ А-С по Binet [Binet et al., 1981] и известной общей выживаемостью. Описание клинической значимости признаков, используемых в вычислительных экспериментах приводится в многочисленных публикациях в области медицины [Марковцева и др., 2023], [Pektaş et al., 2025].

Вычислительный эксперимент проводился в несколько этапов. На первом этапе исследовались показатели группы больных, прошедших начальное обследование. Последующие наблюдения проводились каждые двенадцать месяцев в течение пяти лет. За указанный промежуток времени численность исследуемой группы постепенно сокращалась вследствие смертности пациентов. Следует отметить, что проведение оценки состояния здоровья с меньшими временными интервалами было ограничено рядом факторов объективного характера.

Данные полученные на этапе постановки диагноза и в результате повторных обследований больных использовались для поиска закономерностей в выборках данных по значениям устойчивости.

Была получена закономерность, связанная со значениями латентного признака скорость клубочковой фильтрации (СКФ). Существует функциональная зависимость СКФ от значений возраста и креатинина в зависимости от гендерной принадлежности. Значение латентного признака СКФ рассчитывают по формуле [Марковцева, 2023]:

$$\frac{e}{\dots} \quad \frac{e}{\dots} \quad , \quad (1)$$

где: для женщин используются следующие значения: пол = 1.018, = - 0.329; = 0.7; для мужчин используются следующие значения: пол = 1, = - 0.411; = 0.9; для представителей негроидной расы: коэффициент «Раса» = 1.159.

Динамика заболевания по этапам анализировалась по результатам разбиения значений СКФ на непересекающиеся интервалы. Никаких предположений о плотности распределения данных не делалось. Количество интервалов, оптимальные значения их границ и состав входящих в них объектов не являлись изначально фиксированными. Использовался алгоритм метода [Згуральская, 2018] минимального покрытия значений количественного признака непересекающимися интервалами, в границах которых доминируют представители одного из двух классов.

Для мониторинга пациентов при повторных обследованиях будет достаточно задать значения измеримых показателей возраст и креатинин [Игнатьев и др., 2023]. При значении СКФ более 76 мл/ мин/1,73 м<sup>2</sup> в течение первого года заболевания, более 75 мл/мин/1,73 м<sup>2</sup> в течение второго года заболевания, более 70 мл/мин/1,73 м<sup>2</sup> в течение третьего года заболевания, более 68 мл/мин/1,73 м<sup>2</sup> в течение четвёртого и пятого годов заболевания пациент преодолест расчётные параметры медианы выживаемости для соответствующей стадии ХЛЛ по Binet. В противном случае – общая выживаемость пациента будет меньше расчётной медианы выживаемости по Binet.

В [Марковцева и др., 2023] приведены показательные клинические примеры. Выявлено, что значение СКФ у пациентов ХЛЛ планомерно снижается с течением заболевания (рис. 3) и его в динамике можно рассматривать в качестве значимого фактора мониторинга пациентов с ХЛЛ.

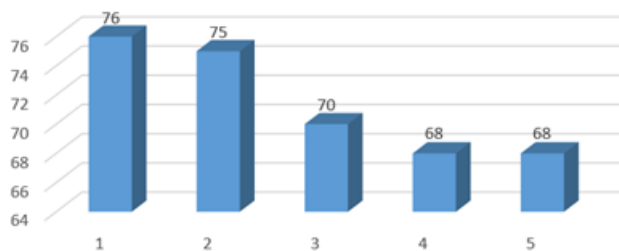


Рис. 3. Границы значения показателя СКФ при котором пациент преодолевает медиану выживаемости по классификации Binet в зависимости от этапа обследования

#### 4. Функциональная схема использования онтологии в системе мониторинга

Формализованные клинические знания и обнаруженные скрытые закономерности методами ИАД составляют основу системы мониторинга больных с ХЛЛ. Функциональная схема использования онтологии в системе мониторинга:

1. На этапе постановки диагноза врач вносит информацию о лабораторных и клинических показателях пациента в форму ввода прогностической системы.
2. На основе полученной информации вычисляется показатель СКФ по формуле (1).
3. Данные о пациенте трансформируются в формат онтологии, создается экземпляр класса пациенты (ФИО пациента, номер обследования).

4. Путем логического вывода на основе множества продукционных правил на языке SWRL в БЗ формируются новые знания.
5. Полученная информация поступает обратно к врачу, формируя отчет для специалиста с оценкой развития неблагоприятных событий (смерти) у пациентов с ХЛЛ (стадия заболевания, группа риска и т.п.).
6. Проводятся повторные обследования, информация о которых поступает в систему чем соответствующую форму.
7. Происходит поиск информации по онтологии о данном пациенте, если информация о предыдущих обследованиях не найдена, то переходим к п.1.
8. Переходим к п.2.

### Заключение

Разработанная прогностическая система, основанная на знаниях для стратификации рисков и мониторинга больных с ХЛЛ предназначена для автоматизированной оценки вероятности развития неблагоприятных событий (смерти) у пациентов и организации персонализированного мониторинга на основе этой оценки, позволит оптимизировать распределение ресурсов здравоохранения, улучшить результаты лечения.

### Список литературы

- [Грибова и др., 2024] Грибова В.В., Окунь Д.Б., Шалфеева Е.А. Применение графов знаний для клинического мониторинга процесса лечения // Онтология проектирования. – 2024. – Т. 14, № 4(54). – С. 504-517. – doi: 10.18287/2223-9537-2024-14-4-504-517.
- [Богданова и др., 2024] Богданова Д.Р., Шахматова Г.Р., Ниязгулов А.М. Структура информационного хранилища системы поддержки принятия клинических решений // Онтология проектирования. – 2024. – Т. 14, № 2(52). – С. 270-278. – doi: 10.18287/2223-9537-2024-14-2-270-278.
- [Кобринский и др., 2020] Кобринский Б.А., Хавкин А.И., Волюнец Г.В. Перспективы применения систем искусственного интеллекта в гастроэнтерологии // Экспериментальная и клиническая гастроэнтерология. – 2020. – № 179(7). – С. 109-117. – doi: 10.31146/1682-8658-ecg-179-7-109-117.
- [Кобринский, 2001] Кобринский Б.А. Искусственный интеллект и медицина: Возможности и перспективы систем, основанных на знаниях // Новости искусственного интеллекта. – 2001. – № 4. – С. 44-51.
- [Игнатьев и др., 2020] Игнатьев Н.А., Згуральская Е.Н., Марковцева М.В. Нелинейные преобразования признаков и поиск закономерностей на данных больных хроническим лимфолейкозом // VI Международная конференция. Информационные технологии и нанотехнологии (Самара, 26–29 мая 2020 г.): Труды конференции. В 4-х т. Т. 4. – Самара: Самарский национальный исследовательский университет имени академика С.П. Королева, 2020. – С. 123-128.



- [A new prognostic classification of chronic lymphocytic leukemia derived from a multivariate survival analysis, 1981] Binet J.L., Auquier A., Dighiero G., et al. A new prognostic classification of chronic lymphocytic leukemia derived from a multivariate survival analysis // *Cancer*. – 1981. – No. 48(1) – URL: [https://doi.org/10.1002/1097-0142\(19810701\)48:1<198::aid-cnecr2820480131>3.0.co;2-v](https://doi.org/10.1002/1097-0142(19810701)48:1<198::aid-cnecr2820480131>3.0.co;2-v) (дата обращения: 29.05.2025).
- [Марковцева и др., 2023] Марковцева М.В., Згуральская Е.Н. Прогнозирование общей выживаемости больных хроническим лимфолейкозом в динамике заболевания // *Качественная клиническая практика*. – 2023. – № 3. – С. 66-71.
- [Адмакин, 2016] Адмакин А.Л. Цепи Маркова – стохастическая модель анализа состояния тяжелообожженных // *Мед. биол. и соц. психол. пробл. безопасности в чрезв. ситуациях*. – 2016. – № 3. – С. 119-125. – doi: 10.25016/2541-7487-2016-0-3-119-125.
- [SWRL: A Semantic Web Rule Language Combining OWL and RuleML, 2024] Horrocks I., Patel-Schneider P.F., Boley H., Tabet S., Grosz B., Dean M. SWRL: A Semantic Web Rule Language Combining OWL and RuleML // *Member Submissions W3C*. 2004. URL: <https://www.w3.org/Submission/SWRL/> (дата обращения: 21.04.2025).
- [Згуральская, 2018] Згуральская Е.Н. Устойчивость разбиения данных на интервалы в задачах распознавания и поиск скрытых закономерностей // *Известия Самарского научного центра Российской академии наук*. – 2018. – Т. 20, № 4-3(84). – С. 451-455.
- [Игнатьев и др., 2023] Игнатьев Н.А., Акбаров Б.Х. Оценка близости структур отношений объектов обучающей выборки на многообразиях наборов латентных признаков // *Вестник Томского государственного университета. Управление, вычислительная техника и информатика*. – 2023. – № 65. – С. 69-78. – doi: 10.17223/19988605/65/7.
- [Mayr et. al., 2006] Mayr C., Speicher M.R., Kofler D.M. Chromosomal translocations are associated with poor prognosis in chronic lymphocytic leukemia // *Blood*. – 2006. – Vol. 107(2). – P. 742-51.
- [Pektaş et. al., 2025] Pektaş G., Gönül E., Öncü Ş. Chronic Lymphocytic Leukemia: Investigation of Survival and Prognostic Factors with Drug-Related Remission // *Diagnostics*. – 2025. – Vol. 15. – P. 728. – doi: 10.3390/diagnostics15060728.