

Прогнозирование хронического лимфолейкоза с использованием методов интеллектуального анализа данных

Марковцева М. В.¹, Згуральская Е. Н.²

¹ — ФГБОУ ВО «Ульяновский государственный университет», Ульяновск, Россия

² — ФГБОУ ВО «Ульяновский государственный технический университет», Ульяновск, Россия

Аннотация. *Актуальность.* Хронический лимфолейкоз (ХЛЛ) — одно из наиболее часто встречающихся лимфо-пролиферативных заболеваний европейской популяции с увеличением частоты в пожилом и старческом возрасте. Стандартные подходы к прогнозированию общей выживаемости у этой категории пациентов не учитывают наличие коморбидной патологии и имеют невысокую точность. Ввиду этого поиск параметров, влияющих на показатель общей выживаемости пациентов с ХЛЛ, представляет особую актуальность. *Цель исследования* — на этапе постановки диагноза ХЛЛ выделить факторы, влияющие на общую выживаемость больных ХЛЛ. *Материалы и методы.* Ретроспективно проанализированы данные 132 пациентов с ХЛЛ стадии А-С по Binet с известной общей выживаемостью. Задача решалась методами интеллектуального анализа данных, а именно с использованием логических алгоритмов классификации. *Результаты.* Скорость клубочковой фильтрации определена в качестве параметра, который объективно обосновывает отклонение реальных сроков общей выживаемости пациентов от расчётных по стандартной системе стадирования Binet. Для этого параметра сформировано if...then-правило, которое даёт возможность прогнозировать выживаемость пациента. При значении СКФ на момент постановки диагноза ХЛЛ более 76 мл/мин/1,73м² можно говорить о том, что пациент преодолит расчётные данные медианы выживаемости для соответствующей стадии ХЛЛ по Binet. В противном случае — общая выживаемость пациента ХЛЛ будет меньше расчётной медианы выживаемости по Binet. *Заключение.* Анализ проведённого исследования позволяет сделать вывод о целесообразности применения методов интеллектуального анализа данных в прогнозировании общей выживаемости пациентов с ХЛЛ. Приведённые в статье клинические примеры показывают их эффективность. По результатам проведённого исследования оформлена заявка на изобретение № 2022104419.

Ключевые слова: хронический лимфолейкоз; общая выживаемость; интеллектуальный анализ данных

Для цитирования:

Марковцева М. В., Згуральская Е. Н. Прогнозирование хронического лимфолейкоза с использованием методов интеллектуального анализа данных. *Качественная клиническая практика.* 2022;(3):31–34. <https://doi.org/10.37489/2588-0519-2022-3-31-34>

Поступила: 12 июля 2022 г. **Принята:** 17 июля 2022 г. **Опубликована:** 24 сентября 2022 г.

Chronic lymphocytic leukemia prediction using data mining methods

Markovtseva MV¹, Zguralskaya EN²

¹ — Ulyanovsk State University, Ulyanovsk, Russia

² — Ulyanovsk State Technical University, Ulyanovsk, Russia

Abstract. *Relevance.* Chronic lymphocytic leukemia (CLL) is one of the most common lymphoproliferative diseases of the European population with an increase in the elderly and senile age frequency. In this category of patients standard approaches to predicting overall survival do not take into account the presence of comorbid pathology and have low accuracy. In view of this, the search for parameters that affect the overall survival rate of patients with CLL is of particular relevance. The aim of the study is to identify factors affecting the CLL patients overall survival at the stage of CLL diagnosis. *Materials and methods.* The data of 132 CLL patients with stage A-C according to Binet with known overall survival were retrospectively analyzed. The problem was solved by data mining methods, namely using logical classification algorithms. *Results.* The glomerular filtration rate is defined as a parameter that objectively justifies the real terms deviation of the patients overall survival from the calculated ones according to the standard Binet staging system. For this parameter, an if...then rule is formed, which makes it possible to predict the patient's survival. If the GFR value at the time of diagnosis of CLL is more than 76 ml/min /1.73 m², we can say that the patient will overcome the calculated median survival data for the corresponding stage of CLL according to Binet. Otherwise, the overall survival of the CLL patient will be less than the estimated median survival according to Binet. *Conclusion.* The analysis of the study allows us to conclude that it is advisable to use data mining methods in predicting the patients overall survival with CLL. The clinical examples given in the article show their effectiveness. According to the study results, an application for invention No. 2022104419 was issued.

Keywords: chronic lymphocytic leukemia; survival prediction; data mining

For citations:

Markovtseva MV, Zgural'skaya EN. Chronic lymphocytic leukemia prediction using data mining methods. *Kachestvennaya klinicheskaya praktika = Good Clinical Practice*. 2022;(3): 31–34. (In Russ). <https://doi.org/10.37489/2588-0519-2022-3-31-34>

Received: July 12, 2022. **Accepted:** July 17, 2022. **Published:** September 24, 2022

Введение / Introduction

Хронический лимфолейкоз (ХЛЛ) — одно из наиболее распространённых лимфопролиферативных заболеваний европейской популяции с дебютом в среднем возрасте (44–60 лет) и увеличением частоты встречаемости в пожилом (60–75 лет) и старческом (более 75 лет) возрасте [1]. В настоящее время общепризнанными являются системы стадирования ХЛЛ *Rai KR и соавт.* (1975 г.) [2], *Binet JL и соавт.* (1981 г.) [3], которые позволяют рассчитать медиану выживаемости пациента в зависимости от величины опухолевой массы. Однако предлагаемые системы не берут во внимание коморбидную патологию, которая, как правило, наблюдается у пациентов пожилого возраста, в котором и диагностируется ХЛЛ. В результате в реальной клинической практике параметр общей выживаемости (ОВ) пациента может значимо отличаться от расчётной медианы выживаемости. По этой причине достоверность результатов прогнозирования ОВ по объёму опухолевой массы остаётся недостаточно высокой.

В последние годы разработаны современные способы прогнозирования ХЛЛ, однако они связаны с необходимостью проведения сложных исследований, которые не являются рутинными: мутационный статус генов варибельного региона иммуноглобулинов (IGHV), хромосомные aberrации, мутация TP53 и др. [4–9]. До последнего времени приходится констатировать факт высокой стоимости, трудоёмкости и недоступности этих методик для большинства пациентов в реальной клинической практике. Таким образом, достоверное прогнозирование ХЛЛ доступными способами сохраняет свою актуальность.

Цель исследования: на этапе постановки диагноза ХЛЛ выделить факторы, влияющие на ОВ пациентов.

Материалы и методы / Materials and methods

Ретроспективно исследованы данные 132 пациентов (60 мужчин и 72 женщины) в возрасте от 45 до 80 лет

с ХЛЛ А-С по Binet и известной ОВ, наблюдавшихся в гематологическом отделении ГУЗ «Ульяновская областная клиническая больница» в период с 01.2010 по 02.2020 года. Из исследования исключались пациенты с ВИЧ-инфекцией, туберкулёзом и иными паранеопластическими процессами. На момент постановки диагноза ХЛЛ учитывался возраст пациента, а также измерялись стандартные биохимические показатели: аланинаминотрансфераза (АЛТ), аспартатаминотрансфераза (АСТ), общий билирубин, непрямой билирубин, глюкоза, креатинин, мочевины, мочевиная кислота, лактатдегидрогеназа (ЛДГ). Дополнительно рассчитывали скорость клубочковой фильтрации (СКФ) пациента по формуле CKD-EPI [10].

Обработку полученных данных проводили с использованием методов интеллектуального анализа данных (ИАД), способных не только эффективно выявлять то или иное состояние, но и прогнозировать исследуемые эффекты [11]. Одним из пяти типов закономерностей, которые позволяют выявлять методы ИАД, является классификация. В основе логических алгоритмов классификации лежит принцип индуктивного вывода логических закономерностей или индукции правил. Создание и использование логических алгоритмов классификации основывается на выявлении в исходных данных закономерностей, из набора которых формируется решающая функция [12].

Объекты выборки были разделены на два непересекающихся класса: K_1 (фактическая выживаемость меньше прогнозируемой ОВ — 6 пациентов) и K_2 (фактическая выживаемость больше или равна прогнозируемой ОВ — 46 пациентов). Поиск логических закономерностей осуществлялся с использованием интервальных методов.

Упорядоченная последовательность по 132 значениям по каждому количественному признаку разбивалась на два непересекающихся интервала $[c_1; c_2]$, $(c_2; c_3]$, по значению критерия [13–15]

$$\left(\frac{\sum_{d=1}^2 \sum_{i=1}^2 u_i^d - (1)u_i^d}{\sum_{i=1}^2 |K_i|(|K_i| - 1)} \right) \left(\frac{\sum_{d=1}^2 \sum_{i=1}^2 u_i^d \left(m - |K_i| - \sum_{j=1}^2 u_j^d + u_i^d \right)}{\sum_{i=1}^2 |K_i|(m - |K_i|)} \right) \rightarrow \max_{c_1 < c_2 < c_3} \quad (1),$$

где $u_1^1, u_1^2 (u_2^1, u_2^2)$ — количество значений признака объектов из класса K_i , $i=1,2$ соответственно в интервалах $[c_1; c_2]$ и $(c_2; c_3]$. Выражение в левых скобках (1) представляет внутриклассовое сходство, в правых — межклассовое различие. Множество допустимых значений критерия (0; 1]. Если в каждом интервале содержатся все значения признака объектов одного класса, то критерий равен 1.

Результаты / Results

Значения критерия (1) и границы интервалов признаков представлены в табл. 1.

Таблица 1

Границы интервалов признаков при максимальном значении критерия 1

Table 1

The feature intervals boundaries at the maximum criterion 1 value

Название признака/ The parameter	Границы интервалов/ The intervals boundaries [c1; c2], (c2; c3]	Значение критерия (1)/ Criterion (1) value
Возраст	[45..54] (54..80]	0,3282
СКФ СКД-EPI	[42..74] (74..96]	0,8298
АЛТ	[2.6..21] (21..135.3]	0,2676
АСТ	[3.8..18] (18..217.9]	0,4393
Билирубин общий	[6..10] (10..45.3]	0,3537
Билирубин непрямой	[2.2..7.9] (7.9..75.1]	0,3389
Креатинин	[65..81] (81..114]	0,3718
Мочевина	[2.5..5.8] (5.8..11.3]	0,3550
Глюкоза	[4..4.69] (4.69..8.7]	0,3775

Наиболее значимая разделимость между классами была получена по показателю СКФ СКД-EPI при значении критерия (1), равном 0,8298. В качестве границы (порога) между классами использовалось значение $G=(c_2+b)/2$, где $b(b < c_2)$ — ближайшее к c_2 значение признака из $(c_2; c_3]$.

Для обоснования выбора порога между классами вычислялась устойчивость разбиения признака на непересекающиеся интервалы. При вычислении использовались значения функции принадлежности к интервалу $t(t=1,2)$ по классу $K_i, i=1,2$, определяемые как

$$f_i(t) = \frac{\eta_{it}}{\eta_{1t} + \eta_{2t}} \quad (2),$$

где $\eta_{it} = \frac{g_{it}}{g_{1t} + g_{2t}}$, g_{it} — число представителей класса K_i в t -м интервале.

С учётом того, что $f_i(t) = 1 - f_{3-i}(t), i=1,2$, устойчивость признака по множеству интервалов разбиения вычисляется как

$$U = \frac{1}{|K_1| + |K_2|} \sum_{t=1}^2 \left\{ \begin{aligned} &f_i(t)(v-u+1), f_i(t) > 0.5, \\ &(1-f_i(t))(v-u+1), f_i(t) < 0.5 \end{aligned} \right. \quad (3).$$

Множество допустимых значений устойчивости (3) принадлежат $[0,5; 1]$. Значения функции принадлежности (3) к интервалам по СКФ СКД-EPI приводятся в табл. 2.

Значения функции принадлежности (2) по признаку СКФ СКД-EPI

Таблица 2

Table 2

Values of the membership function (2) based on the GFR CKD-EPI

Интервал/ Interval	Класс/Class	
	K_1	K_2
[42,0; 74,0]	0,8888	0,1112
[78,0; 96,0]	0,0000	1,0000

Устойчивость разбиения по (3) для СКФ СКД-EPI составляла 0,911667.

Обсуждение / Discussion. Полученные результаты показывают, что при значении СКФ СКД-EPI на момент постановки диагноза более 76 мл/мин/1,73м² можно говорить о том, что пациент преодолеет расчётные параметры медианы выживаемости для соответствующей стадии ХЛЛ по Binet. В противном случае — общая выживаемость пациента будет меньше расчётной медианы выживаемости по Binet.

Для иллюстрации приводим клинические примеры.

Клинический пример 1. Больной З., 69 лет. Установлен диагноз: хронический лимфолейкоз, стадия В по Binet. Расчётная медиана выживаемости составляет 60 месяцев. При постановке диагноза определена СКФ по формуле СКД-EPI, которая составила 77 мл/мин/1,73м². Общая выживаемость пациента составила 71 месяц, что превышает расчётную.

Клинический пример 2. Больной Т., 55 лет. При текущем медицинском осмотре был поставлен диагноз хронического лимфолейкоза, стадия А по Binet. Расчётная медиана выживаемости составляет 120 месяцев. При постановке диагноза определена СКФ по формуле СКД-EPI, которая составила 94 мл/мин/1,73м². Пациент остаётся под наблюдением, в течение уже 227 месяцев, т.е. расчётная медиана выживаемости преодолена.

Клинический пример 3. Больной Н., 80 лет. Установлен диагноз: хронический лимфолейкоз, стадия С по Binet. Расчётная выживаемость составляет 24 месяца. При постановке диагноза СКФ по формуле СКД-EPI определена на уровне 47 мл/мин/1,73м². Общая выживаемость пациента составила 11 месяцев, что менее расчётной.

Основные выводы / Conclusion. Таким образом, значение параметра СКФ СКД-EPI на момент постановки диагноза ХЛЛ отражает сокращение либо увеличение срока ОВ пациента относительно расчётного показателя медианы выживаемости по системе стадирования Binet. Приведённые клинические примеры показывают эффективность анализа полученных результатов с использованием ИАД в прогнозировании общей выживаемости пациентов с ХЛЛ на практике. По результатам проведённого исследования оформлена заявка на изобретение № 2022104419.

ДОПОЛНИТЕЛЬНАЯ ИНФОРМАЦИЯ
ADDITIONAL INFORMATION

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Conflict of interest. The authors declare that there is no conflict of interest.

Финансирование. Исследование не имело спонсорской поддержки.

Funding. The study had no sponsorship.

СВЕДЕНИЯ ОБ АВТОРАХ
ABOUT THE AUTHORS

Марковцева Мария Владимировна
Автор, ответственный за переписку

e-mail: mmark7@yandex.ru

ORCID ID: <https://orcid.org/0000-0002-7299-5268>

SPIN-код: 3940-1833

к. м. н., доцент кафедры госпитальной терапии, ФГБОУ ВО «Ульяновский государственный университет», Ульяновск, Россия

Згуральская Екатерина Николаевна
e-mail: iatu@inbox.ru

ORCID ID: <https://orcid.org/0000-0003-1527-9692>

к. т. н., доцент кафедры информационных технологий и общенаучных дисциплин, ФГБОУ ВО «Ульяновский государственный технический университет», Ульяновск, Россия

Участие авторов. Марковцева М. В. — концепция исследования, литературный поиск, сбор и анализ материала, написание статьи, финальное утверждение рукописи; Згуральская Е. Н. — математическое моделирование, анализ результатов, написание текста.

Participation of authors. Markovtseva MV — research concept, literary search, collection and analysis of material, writing the article, final approval of a manuscript; Zguralskaya EN — mathematical modeling, analysis the results, writing the text.

Markovtseva Maria V.
Corresponding author

e-mail: mmark7@yandex.ru

ORCID ID: <https://orcid.org/0000-0002-7299-5268>

SPIN code: 3940-1833

PhD, Cand. Sci. Med., Associate Professor of the Department of Hospital Therapy, Ulyanovsk State University, Ulyanovsk, Russia

Zguralskaya Ekaterina N.
e-mail: iatu@inbox.ru

ORCID ID: <https://orcid.org/0000-0003-1527-9692>

PhD, Cand. Sci. Tech., Associate Professor of the Department of Information Technology and General Scientific, Ulyanovsk State Technical University, Ulyanovsk, Russia

Список литературы / References

1. Алгоритмы диагностики и протоколы лечения заболеваний системы крови / под ред. В. Г. Савченко. В 2 т. Т. 2. — М.: Практика, 2018. — 1264 с. [Algoritmy diagnostiki i protokoly lecheniya zaboolevaniy sistemy krovi / Ed. by VG Savchenko. In two volumes. Volume 2. Moscow: Practice, 2018. (In Russ).].
2. Rai KR, Sawitsky A, Cronkite EP et al. Clinical staging of chronic lymphocytic leukemia. *Blood*. 1975;46:219–34.
3. Binet JL, Auquier A, Dighiero G et al. A new prognostic classification of chronic lymphocytic leukemia derived from a multivariate survival analysis. *Cancer*. 1981;48(1):198–206. doi: 10.1002/1097-0142(19810701)48:1>198::aid-cnrc2820480131<3.0.co;2-v
4. Dohner H, Stilgenbauer S, Benner A, et al. Genomic aberrations and survival in chronic lymphocytic leukemia. *N Engl J Med*. 2000;343(26):1910–6. doi: 10.1056/nejm200012283432602
5. Mayr C, Speicher MR, Kofler DM, et al. Chromosomal translocations are associated with poor prognosis in chronic lymphocytic leukemia. *Blood*. 2006;107(2):742–51. doi: 10.1182/blood-2005-05-2093
6. Haferlach C, Dicker F, Schnittger S, et al. Comprehensive genetic characterization of CLL: A study on 506 cases analysed with chromosome banding analysis, interphase FISH, IgVH status and immunophenotyping. *Leukemia*. 2007;21(12):2442–51. doi: 10.1038/sj.leu.2404935
7. Никитин Е. А., Луговская С. А., Варламова Е. Ю., Обухова Т. Н., Судариков А. Б., Капланская И. Б., Бидерман Б. В., Наумова Е. В., Сидорова Ю. В., Почтарь М. Е., Кислицына Д. Г., Домрачева Е. В., Иванова В. Л., Ковалёва Л. Г., Птушкин В. В. Факторы, предсказывающие полную ремиссию и рефрактерность к терапии у первичных больных ХЛЛ, получающих режим FCR. *Гематология и трансфузиология*. 2012;57(3):16. Конгресс гематологов России. Москва, 2–4 июля 2012 года. [Nikitin EA, Lugovskaya SA, Varlamova EY, Obukhova TN, Sudarikov AB, Kaplanskaya IB, Biderman BV, Naumova EV, Sidorova YV, Pochtars ME, Xilichina DG, Domracheva EV, Ivanova VL, Kovaleva LG, Ptushkin VV. Faktory, predskazyvayushchie polnuyu remissiyu i refrakternost' k terapii u pervichnyh bol'nyh HLL, poluchayushchih rezhim FCR. *Gematologiya i transfuziologiya*. 2012;57(3):16. Kongress gematologov Rossii. Moskva, 2–4 iyulya 2012 goda. (In Russ).].
8. International CLL-IPI working group. An international prognostic index for patients with chronic lymphocytic leukaemia (CLL-IPI): a meta-analysis of individual patient data. *Lancet Oncol*. Lancet Publishing Group. 2016;17(6):779–90. doi: 10.1016/s1470-2045(16)30029-8
9. Обухова Т. Н., Кислова М. И., Никитин Е. А., Кислицына М. А., Бидерман Б. В., Тагирова М. К., Судариков А. Б., Птушкин В. В., Савченко В. Г. Структура и прогностическое значение делеции 13q14 при хроническом лимфолейкозе. *Гематология и трансфузиология*. 2022;67(1):75–89. [Obukhova TN, Kislova MI, Nikitin EA, Kislytsyna MA, Biderman BV, Tagirova MK, Sudarikov AB, Ptushkin VV, Savchenko VG. Structure and prognostic significance of 13q14 deletion in chronic lymphocytic leukemia. *Russian Journal of Hematology and Transfusiology (Gematologiya i transfuziologiya)*. 2022;67(1):75–89. (In Russ).]. doi: 10.35754/0234-5730-2022-67-1-75-89
10. Клинические рекомендации. Хроническая болезнь почек (ХБП). *Нефрология*. 2021;25(5):10–82. [Clinical recommendations. Chronic kidney disease (CKD). *Nephrology (Saint-Petersburg)*. 2021;25(5):10–82. (In Russ).].
11. Michie D, Spiegelhalter DJ, Taylor CC. Machine learning, Neural and Statistical Classification. 1994;13.
12. Антамошкин А. Н., Масич И. С. Обнаружение закономерностей в данных для распознавания объектов как задача условной псевдобулевой оптимизации. *Вестник СибГАУ*. 2015;16(1):16–21. [Antamoshkin AN, Masich IS. Detection of patterns in data for recognition of objects as a conditional pseudo-boolean optimization problem. *Vestnik SibGAU*. 2015;16(1):16–21. (In Russ).].
13. Згуральская Е. Н. Устойчивость разбиения данных на интервалы в задачах распознавания и поиск скрытых закономерностей. *Известия Самарского научного центра Российской академии наук*. 2018;204(3):451–5. [Zguralskaya EN. Sustainability of dividing data in intervals in the problems of recognition and searching for hidden laws. *SamSC RAS*. 2018;204(3):451–5. (In Russ).].
14. Игнатьев Н. А. Вычисление обобщенных показателей и интеллектуальный анализ данных. *Автоматика и телемеханика*. 2011;5:183–90. [Ignat'ev NA. Computing generalized parameters and data mining. *Autom. Remote Control*. 2011;72:5:1068–74. (In Russ).]. doi: 10.1134/S0005117911050146
15. Ignat'ev NA, Zguralskaya EN, Markovtseva MV. Searching for Hidden Patterns That Affect the Overall Patient Survival with Data Mining. *Scientific and Technical Information Processing*. 2021;48(6):461–6. doi: 10.3103/S014768822106006X