

Selskyy P. R., Selskyy B. P. Алгоритм принятия решений для оптимизации прогнозирования развития заболеваний на первичном уровне с комплексным поэтапным применением информационных методик = The algorithm of decision-making to optimize the prediction of disease at the primary level with a complex phased application of information techniques. *Journal of Education, Health and Sport*. 2016;6(9):305-314. eISSN 2391-8306. DOI <http://dx.doi.org/10.5281/zenodo.61842>
<http://ojs.ukw.edu.pl/index.php/johs/article/view/3846>

The journal has had 7 points in Ministry of Science and Higher Education parametric evaluation. Part B item 755 (23.12.2015).
755 *Journal of Education, Health and Sport* eISSN 2391-8306 7

© The Author (s) 2016;

This article is published with open access at Licensee Open Journal Systems of Kazimierz Wielki University in Bydgoszcz, Poland
Open Access. This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited. This is an open access article licensed under the terms of the Creative Commons Attribution Non Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted, non commercial use, distribution and reproduction in any medium, provided the work is properly cited.
This is an open access article licensed under the terms of the Creative Commons Attribution Non Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted, non commercial use, distribution and reproduction in any medium, provided the work is properly cited.
The authors declare that there is no conflict of interests regarding the publication of this paper.
Received: 05.08.2016. Revised 25.08.2016. Accepted: 06.09.2016.

THE ALGORITHM OF DECISION-MAKING TO OPTIMIZE THE PREDICTION OF DISEASE AT THE PRIMARY LEVEL WITH A COMPLEX PHASED APPLICATION OF INFORMATION TECHNIQUES

P. R. Selskyy, B. P. Selskyy

State Institution of Higher Education “I. Horbachevsky Ternopil State Medical University of the Ministry of Health of Ukraine”, Ternopil, Ukraine

Summary

The paper describes the algorithm of decision-making to optimize the prediction of disease at the primary health care level with a complex phased application of information techniques. The approach is based on analysis of the average values of indicators, correlation coefficients, using multi-parameter neural network clustering, ROC-analysis and decision tree.

The data of 63 patients with arterial hypertension obtained at teaching and practical centers of primary health care were used for the analysis. The blood pressure at the first treatment of the patients was as follows: systolic – $(154,76 \pm 2,29)$ mm Hg, diastolic – $(92,94 \pm 1,04)$ mm Hg, pulse pressure – $(61,83 \pm 1,95)$ mm Hg. Indicators of the upper and lower blood pressure were significantly lower compared with those before the treatment: systolic – $(145,86 \pm 2,01)$ mm Hg ($P < 0,01$), diastolic – $(88,49 \pm 1,08)$ mm Hg ($p < 0,001$). In 13 (19,40 %) patients we observed worsening of the health condition and the development of complications. Correlation analysis revealed a direct correlation between the indices of the heart rate (+ 0,5), the systolic (+ 0,1), the diastolic (+ 0,4) and the pulse (+ 0,1) pressure at the first pressure and re-examination.

It has been established that neural network clusterization can effectively and objectively allocate patients into the appropriate categories according to the level of average indices of patient examination results. The cluster analysis data has revealed that the combination of high arterial tension measurement (during the first and repeated examination) provides a basis of the deteriorated patient

condition, while the combination of high age and heart rate measurement (polycardia) also plays a significant, but not the primary role for the disease prognosis.

Determination of the sensitivity and specificity of hemodynamic parameters, including blood pressure, and repeated during the initial survey conducted using ROC-analysis.

In order to obtain the rules in decision-making and, corresponding algorithms of clinical thinking, we applied the method of decision tree that can be used to predict the flow of other pathology. The recursive procedure is taken as a basis for research. The method is implemented in Netbeans development environment using Java programming language. Base training data are deployed on the MySQL server.

The algorithm of decision-making to optimize the prediction of disease at the primary level in order to adjust examination procedures and treatment based on the analysis of indicators of patient examination with a complex, gradual application of information procedures. The proposed algorithm can be used to determine the risk of developing health complications of various diseases that are most prevalent in a given region. The suggested methodology is easy to use and does not require large financial investments, which is important for the organization of primary health care in rural areas.

Keywords: the primary health care level, hypertension, information methods decision algorithm, neural network clustering, ROC-analysis, decision tree.

АЛГОРИТМ ПРИНЯТИЯ РЕШЕНИЙ ДЛЯ ОПТИМИЗАЦИИ ПРОГНОЗИРОВАНИЯ РАЗВИТИЯ ЗАБОЛЕВАНИЙ НА ПЕРВИЧНОМ УРОВНЕ С КОМПЛЕКСНЫМ ПОЭТАПНЫМ ПРИМЕНЕНИЕМ ИНФОРМАЦИОННЫХ МЕТОДИК

П. Р. Сельський, Б. П. Сельський

**Государственное высшее учебное заведение “Тернопольский государственный
медицинский университет имени И.Я. Горбачевского МОЗ Украины”, Тернополь,
Украина**

Резюме

В работе предложен алгоритм принятия решений для оптимизации прогнозирования развития заболеваний на первичном уровне предоставления медико-санитарной помощи с

комплексным поэтапным применением информационных методик. Подход основан на анализе средних значений показателей, коэффициентов корреляции, использовании многопараметрической нейросетевой кластеризации, ROC-анализа и дерева решений.

В качестве данных для анализа использованы результаты обследования 63 пациентов с артериальной гипертензией в учебно-практических центрах первичной медико-санитарной помощи. Показатели артериального давления при первом обращении пациентов были следующими: систолическое – $(154,76 \pm 2,29)$ мм. рт. ст., диастолическое – $(92,94 \pm 1,04)$ мм. рт. ст., пульсовое давление – $(61,83 \pm 1,95)$ мм. рт. ст. Показатели верхнего и нижнего артериального давления были значительно ниже по сравнению с аналогичными показателями до лечения: систолическое – $(145,86 \pm 2,01)$ мм. рт. ст. ($p < 0,01$), диастолическое – $(88,49 \pm 1,08)$ мм. рт. ст. ($p < 0,001$). В 13 (19,40 %) пациентов наблюдалось ухудшение состояния и развитие осложнений. При корреляционном анализе обнаружена прямая корреляционная связь между показателями пульса (+ 0,5), верхнего (+ 0,1), нижнего (+ 0,4) и пульсового (+ 0,1) давлений при первом и повторном обследовании.

Установлено, что нейросетевая кластеризация позволяет эффективно и объективно распределить пациентов в соответствующие категории по уровню средних показателей результатов обследования. В результате анализа кластерных портретов обнаружено, что именно сочетание высоких показателей артериального давления (при первом и повторном обследовании) дает основание прогнозировать ухудшение состояния пациентов, а сочетание высоких показателей возраста и пульса (тахикардия) имеют существенное, однако не первостепенное значение для прогноза.

Определение чувствительности и специфичности показателей гемодинамики, в частности артериального давления, при первичном и повторном обследовании проведено с помощью ROC-анализа.

С целью получения правил в процессе принятия решений, отвечающих алгоритмам клинического мышления, применена методика дерева решений, которая может быть использована и для прогнозирования течения другой патологии. За основу взята рекурсивную процедуру работы. Метод реализован в среде разработки Netbeans на языке программирования Java. Базу учебных данных развернуто на сервере MySQL.

Разработан алгоритм принятия решений для оптимизации прогнозирования развития заболеваний на первичном уровне с целью коррекции обследования и лечения на основе анализа показателей обследования больных с комплексным, поэтапным применением информационных методик.

Ключевые слова: первичный уровень медико-санитарной помощи, артериальная гипертензия, информационные методики, алгоритм принятия решений, нейросетевая кластеризация, ROC-анализ, дерево решений.

Вступление. Применение информационных технологий в медицине приобретает все большее значение [1, 2, 3]. На решение проблем внедрения инновационных медицинских информационных технологий направлен целый ряд исследований [3, 4, 5, 6]. Однако нерешенной остается проблема информатизации в сельской медицине. При этом для повышения эффективности первичной медицинской помощи важным является оптимизация прогнозирования течения заболеваний с использованием недорогих и доступных информационных методик.

Поэтому **целью** нашей работы было проанализировать результаты обследования пациентов с гипертензией на основе корреляционных показателей, многопараметрической нейросетевой кластеризации, ROC-анализа и дерева решений и разработать алгоритм принятия решений для оптимизации прогнозирования течения заболеваний на первичном уровне оказания медико-санитарной помощи.

Материалы и методы. В качестве данных для анализа использованы результаты обследования 63 больных с артериальной гипертензией в учебно-практических центрах первичной медико-санитарной помощи (УПЦПМСП) сел Гнилицы и Зарубинцы Тернопольской области Украины [7]. Группу контроля составляли 19 жителей данных населенных пунктов, у которых не зафиксировано никакой патологии. Статистическая обработка материала была проведена с использованием пакета программ "Microsoft Excel" (Microsoft Office 2003). Статистическая значимость различий между средними арифметическими и относительными величинами оценивалась по критерию Стьюдента-Фишера (t). При сравнении однотипных групп проводился корреляционный анализ с учетом коэффициента корреляции (r) с помощью метода квадратов Пирсона.

Для более глубокого анализа показателей обследования с целью прогнозирования течения заболевания использован нейросетевой подход с применением надстройки NeuroXL Classifier для программы Microsoft Excel. Программа NeuroXL Classifier (разработка компании AnalyzerXL) реализует самоорганизационные нейросети, выполняющие категорирование путем изучения трендов и взаимосвязей внутри данных. Несмотря на высокую эффективность,

нейросети часто не используются в силу своей сложности и длительного обучения, требуемого для их правильной реализации. NeuroXL Classifier устраняет такие барьеры, скрывая сложность методов на основе нейросетей и используя преимущества рабочих книг Microsoft Excel [8].

Для оптимизации прогнозирования заболеваний на первичном уровне использован анализ классификаций с применением ROC-кривых. ROC-кривая (англ. receiver operating characteristic, операционная характеристика приемника) – график, позволяющий оценить качество бинарной классификации, который отражает соотношение между долей верных положительных классификаций от общего числа положительных классификаций (англ. true positive rate – TPR) и долей ложных положительных классификаций от общего числа негативных классификаций (англ. false positive rate – FPR) при варьировании порога решающего правила. При этом TPR называют чувствительностью (sensitivity) алгоритма классификации, а FPR определяют как 1-специфичность (1-specificity). Соответственно, специфичностью (specificity) алгоритма классификации называют долю верных негативных классификаций (true negative rate – TNR) от общего числа негативных классификаций. Интерпретацию ROC дает показатель AUC (англ. area under ROC curve, площадь под ROC-кривой) – площадь, ограниченная ROC-кривой и осью доли ложных положительных классификаций. По сути, этот показатель является мерой качества классификатора [9].

Одним из подходов, отражающих естественный процесс мышления при дифференциальной диагностике, является метод индукции дерева решений. За основу взято рекурсивную процедуру работы [10]. Метод реализован в среде разработки Netbeans на языке программирования Java. Базу учебных данных развернуто на сервере MySQL.

Результаты исследования. Всего было обследовано 63 больных, среди которых – 15 мужчин и 48 женщин. Средний возраст больных составлял $(64,30 \pm 1,81)$ лет. Среднее значение положения электрической оси сердца составило $(37,83 \pm 1,92)^0$. Средний показатель пульса составлял $(78,24 \pm 1,15)$ ударов в минуту. При первом обращении пациентов показатели артериального давления были следующими: систолическое – $(154,76 \pm 2,29)$ мм. рт. ст., диастолическое – $(92,94 \pm 1,04)$ мм. рт. ст., пульсовое давление – $(61,83 \pm 1,95)$ мм. рт. ст. Средний показатель пульса ($(78,59 \pm 1,07)$ ударов в минуту) и пульсовое давление ($(57,06 \pm 1,57)$ мм. рт. ст.) при повторном обследовании статистически не отличались ($p > 0,05$). Показатели верхнего и нижнего артериального давления были значительно ниже по сравнению с аналогичными показателями до лечения: систолическое – $(145,86 \pm 2,01)$ мм. рт. ст. ($p < 0,01$), диастолическое – $(88,49 \pm 1,08)$ мм. рт. ст. ($p < 0,001$). В 13 (19,40 %) больных наблюдалось ухудшение состояния и развитие осложнений. При первом и повторном обследовании обнаружена прямая корреляционная связь между показателями пульса (+ 0,5), верхнего (+ 0,1), нижнего (+ 0,4) и пульсового (+ 0,1) давлений.

Анализ результатов обследования пациентов с гипертензией в УПЦПМСП на основе средних значений и корреляционных показателей выявил и статистически достоверное отличие частоты пульса в группах со стабильным течением болезни и в группе с ухудшением состояния. Установлено, что уровень тахикардии является существенным исходным показателем, который указывает на большую вероятность ухудшения состояния и, в частности, развитие осложнений. При этом выявлена прямая корреляционная связь между подавляющим большинством показателей гемодинамики при первом и повторном обследовании пациентов во всех группах с гипертензией свидетельствует о значении изменения частоты пульса и артериального давления как объективных маркеров течения заболевания, а, следовательно, и эффективности лечения. Выяснено, что анализ на основании средних значений и вычисления коэффициентов корреляции средних значений показателей возраста, положения электрической оси сердца, ряда гемодинамических показателей является первичным инструментом, который не дает установить значение сочетания изменения тех или иных параметров для прогнозирования течения заболевания в сторону ухудшения или улучшения. В то же время, эффективно и объективно распределить пациентов в соответствующие категории позволяет нейросетевая кластеризация. Анализ кластерных портретов обнаружил, что именно сочетание высоких показателей артериального давления (систолического, диастолического и пульсового) дает основание прогнозировать ухудшение состояния пациентов, тогда как сочетание высоких показателей возраста и пульса (тахикардия) имеют существенное, однако не первостепенное значение для прогноза.

Проведено определение чувствительности и специфичности показателей гемодинамики, в частности артериального давления, при первичном и повторном обследовании. С целью исследования данных показателей в качестве маркеров включения в группы риска ухудшения состояния и развития осложнений у больных с артериальной гипертензией проведен ROC-анализ. Установлено, что при прогнозировании течения гипертензии на основе объединенных изменений показателей артериального давления в качестве маркеров включения в группу риска ухудшения состояния и развития осложнений целесообразно использовать данные обследования до начала лечения. При этом ROC-анализ результатов исследования показателей систолического и пульсового артериального давления показал большую чувствительность при первичном обследовании больных по сравнению со вторым. Анализ пульсового артериального давления обнаружил также преобладание на большинстве уровней показателей специфичности при первичном обследовании. Соответственно, площадь, ограниченная ROC-кривой и осью доли ложных положительных классификаций, была больше при обследовании до назначенного лечения, что показывает высокое качество данного классификатора. В то же время на большинстве уровней систолического и диастолического артериального давления показатели

специфичности не отличались по результатам первого и второго обследований. Введение методики ROC-анализа для прогнозирования течения заболеваний при оказании первичной медицинской помощи с целью определения чувствительности и специфичности исследуемых показателей в разные периоды и по разным методикам является перспективным, учитывая доступность и простоту в использовании.

На заключительном этапе решения вопроса о включении пациентов в группу риска ухудшения состояния и развития осложнений было внедрено методике индуцирования дерева решений. Использован массив учебных данных, которые включали показатели гемодинамики, данные о возрасте и поле. Эта методика применена с целью получения правил в процессе принятия решений.

Тактика поэтапного анализа показателей обследования больных с артериальной гипертензией предложена в виде алгоритма принятия решений для оптимизации прогнозирования заболевания с целью коррекции обследования и лечения (рис. 1).

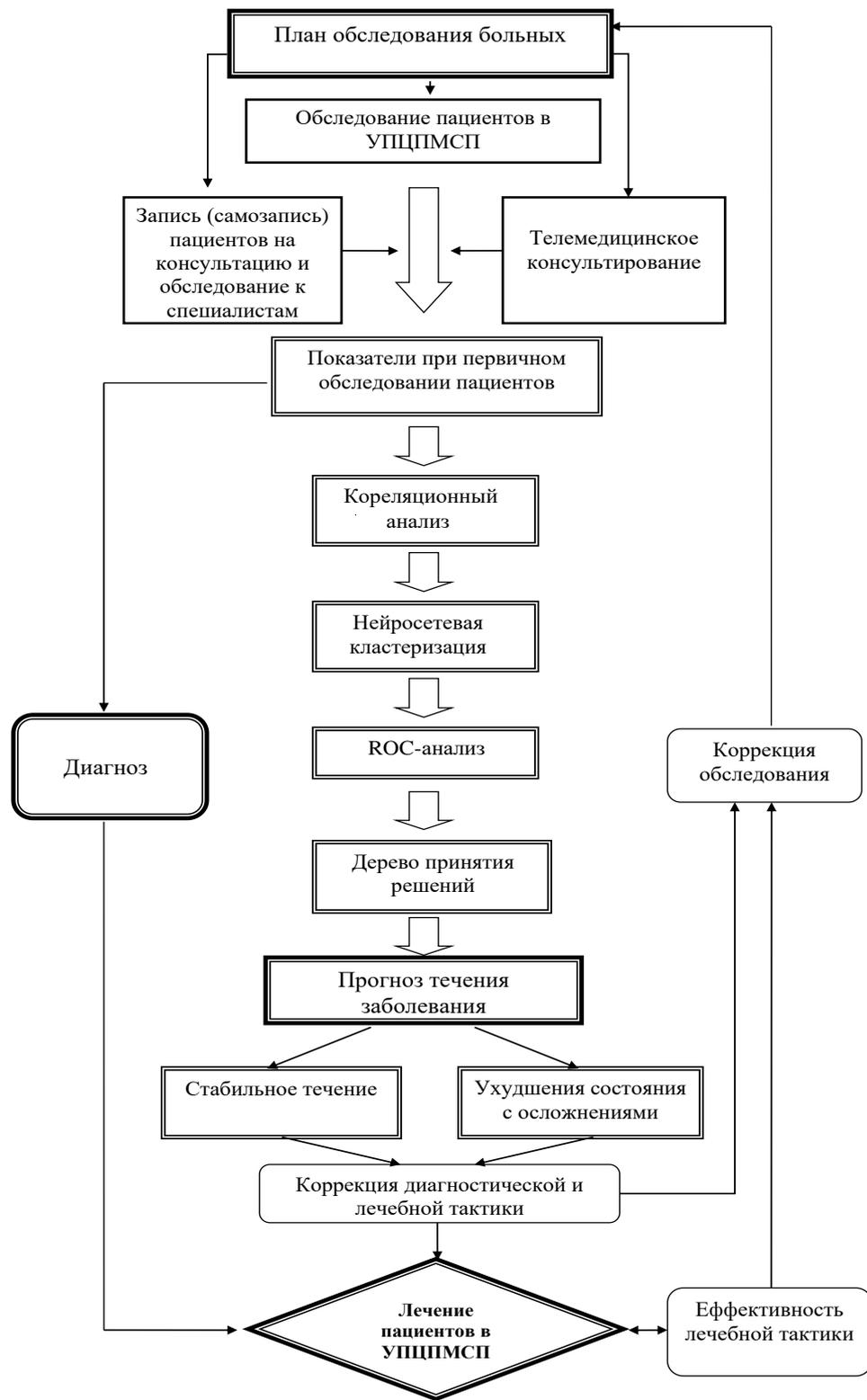


Рисунок 1 – Алгоритм принятия решений для оптимизации прогнозирования заболеваний при оказании первичной медицинской помощи с целью коррекции обследования и лечения

Выводы. 1. В работе предложена методика анализа результатов обследования пациентов с гипертензией в учебно-практических центрах первичной медико-санитарной помощи на основе средних значений, корреляционных показателей, алгоритмов нейросетевой кластеризации, ROC-анализа и дерева решений.

2. С целью эффективного и объективного распределения пациентов в соответствующие категории по уровню показателей результатов обследования применена нейросетевая кластеризация.

3. Определение чувствительности и специфичности показателей гемодинамики, в частности артериального давления, при первичном и повторном обследовании проведено с помощью ROC-анализа.

4. С целью получения правил в процессе принятия решений, отвечающих алгоритмам клинического мышления, применена методика дерева решений, которая может быть использована и для прогнозирования течения другой патологии.

5. Разработан алгоритм принятия решений для оптимизации прогнозирования развития заболеваний на первичном уровне с целью коррекции обследования и лечения на основе анализа показателей обследования больных с комплексным, поэтапным применением информационных методик.

6. Предложенный алгоритм может быть использован для определения групп риска развития осложнений при различных заболеваниях, которые являются самыми распространенными в том или ином регионе. При этом предложенные информационные методики просты в использовании и не требуют больших финансовых вложений, что важно при организации первичной медицинской помощи в сельской местности.

References

1. Health Informatization Concept in Ukraine / O. P. Mintzer, Y. V. Woronenko, L. Y. Babintseva [et al.] // Medical Informatics and Engineering. – 2012. – № 3. – P. 5-29.
2. Martsenyuk V. P. Information quality management system training in higher medical education / V. P. Martsenyuk, P. R. Selskyy. – Ternopil : TSMU, 2015. – 312 p.
3. Mintzer O. P. Information and technological problems of telemedicine consultations / O. P. Mintzer, V. V. Krasnov, H. Taher // Medical Informatics and Engineering. – 2011. – № 4. – P. 32-37.
4. Measuring patient-centered communication in patient-physician consultations: theoretical and practical issues / R. M. Epstein, P. Franks, K. Fiscella [et al.] // Soc. Sci. Med. – 2005. – № 61. – P. 1516-1528.

5. Martsenyuk V. P. On the software development environment of intelligent medical databases / V. P. Martsenyuk, N. O. Kravets // *Clinical Informatics and Telemedicine* – 2004. – № 1. – P. 47–53.
6. Vostrov G. N. Information model of remote medical services. The first message / G. N. Vostrov, O. P. Mintser, O. O. Pavlov [at all.] // *Medical Informatics and Engineering*. – 2010. - № 3. – P. 37-47.
7. Kovalchuk L. A. Results of the latest techniques of the educational process in the I. Horbachevsky Ternopil State Medical University and future plans / L. A. Kovalchuk // *Medical education*. – 2012. – № 2. – P. 11-17.
8. Martsenyuk V. P. Neural network forecasting assembly of medical students integrated licensing examination "Step 1" on the basis of current performance and integrated semester test examination / V. P. Martsenyuk, A. V. Semenets, O. O. Stahanska // *Medical Informatics and Engineering*. – 2010. – № 2. – P. 57-62.
9. Hanley J. A. Sampling variability of nonparametric estimates of the areas under receiver operating characteristic curves: an update / J. A. Hanley, K. O. Hajian-Tilaki // *Academic Radiology*. – 1997. – Vol. 4. – P. 49–58.
10. Han J. *Data Mining: Concepts and Techniques* / J. Han, M. Kamber. – Morgan Kaufmann, San Francisco, 2006. – 800 p.