



Нейронные сети в задачах аналитического обеспечения систем социально-гигиенического и экологического мониторинга

Б.И. Марченко^{1,2}, Н.К. Плуготаренко¹, О.А. Семина¹

¹ФГАОУ ВО «Южный федеральный университет» Минобрнауки России, ул. Б. Садовая, д. 105/42, г. Ростов-на-Дону, 344006, Российская Федерация

²Филиал ФБУЗ «Центр гигиены и эпидемиологии в Ростовской области» в г. Таганроге, ул. Б. Проспект, д. 16-а, г. Таганрог, 347930, Российская Федерация

Резюме

Введение. Обеспечение дальнейшего роста эффективности системы социально-гигиенического мониторинга требует интеграции современной методологии анализа риска здоровью населения с комплексом адаптированных унифицированных традиционных и инновационных аналитических методов, а также тесного информационного взаимодействия с системой экологического мониторинга.

Цель: апробация и оценка точности прогнозирования заболеваемости злокачественными новообразованиями посредством искусственной нейронной сети.

Материалы и методы. Анализируемые временные ряды представлены сведениями из статистических отчетных форм о злокачественных новообразованиях в городе Таганроге Ростовской области. Использованы две модели прогнозирования – регрессионная и посредством искусственной нейронной сети прямого распространения сигнала типа многослойного персептрона. Искусственная нейронная сеть со 117 нейронами в скрытом слое создана в среде пакета прикладных программ Matlab R2021a с набором инструментов для синтеза и анализа нейронных сетей Neural Network Toolbox, при ее обучении применен алгоритм Левенберга – Марквардта.

Результаты. Апробация двух моделей прогнозирования выполнена на обучающих выборках различной продолжительности – 34 года и 15 лет. При сравнительной оценке точности прогнозов на 2018 и 2019 гг. рассчитывались абсолютные и относительные погрешности. Точность нейросетевой модели прогнозирования выше, чем регрессионной модели, как по сумме злокачественных новообразований, так и по большинству отдельных локализаций. Так, абсолютные погрешности прогнозов на 2018 год при применении регрессионной и нейросетевой моделей составляли 17,05 и 1,49 на 100 тысяч населения, на 2019 год – соответственно 39,07 и 4,42 на 100 тысяч населения. Точность прогноза снижалась при уменьшении продолжительности временного ряда и удалении от границ обучающей выборки.

Заключение. Искусственная нейронная сеть типа многослойного персептрона прямого распространения сигнала обеспечивает более точные прогнозы при использовании минимума входной информации по сравнению с результатами применения регрессионной модели, что является ее несомненным преимуществом.

Ключевые слова: нейронная сеть прямого распространения сигнала, обучение нейронных сетей, социально-гигиенический мониторинг, экологический мониторинг, аналитическое обеспечение, злокачественные новообразования.

Для цитирования: Марченко Б.И., Плуготаренко Н.К., Семина О.А. Нейронные сети в задачах аналитического обеспечения систем социально-гигиенического и экологического мониторинга // Здоровье населения и среда обитания. 2021. Т. 29. № 11. С. 23–30. doi: <https://doi.org/10.35627/2219-5238/2021-29-11-23-30>

Сведения об авторах:

✉ **Марченко** Борис Игоревич – д.м.н., доцент; профессор Института нанотехнологий, электроники и приборостроения ФГАОУ ВО «Южный федеральный университет» Минобрнауки России, кафедра техносферной безопасности и химии; врач-эпидемиолог отделения социально-гигиенического мониторинга филиала ФБУЗ «Центр гигиены и эпидемиологии в Ростовской области» в г. Таганроге; e-mail: borismarch@gmail.com; ORCID: <http://orcid.org/0000-0001-6173-329X>.

Плуготаренко Нина Константиновна – к.т.н., доцент; заведующая кафедрой техносферной безопасности и химии Института нанотехнологий, электроники и приборостроения ФГАОУ ВО «Южный федеральный университет» Минобрнауки России; e-mail: plugotarenko@mail.ru; ORCID: <http://orcid.org/0000-0003-0853-8510>.

Семина Олеся Александровна – магистрант Института нанотехнологий, электроники и приборостроения ФГАОУ ВО «Южный федеральный университет» Минобрнауки России; кафедра техносферной безопасности и химии; e-mail: semina@sfedu.ru; ORCID: <https://orcid.org/0000-0003-4263-5103>.

Информация о вкладе авторов: *Марченко Б.И.* – концепция и разработка дизайна исследования; управление проектом; сбор материала и формирование баз данных; статистическая и аналитическая обработка; написание текста статьи; утверждение окончательного варианта статьи; *Плуготаренко Н.К.* – концепция исследования; ресурсы и программное обеспечение; методология; обучение нейронной сети; проверка и редактирование материала; *Семина О.А.* – обзор публикаций по теме статьи; обучение нейронной сети; аналитическая обработка материала; ответственность за целостность всех частей статьи.

Финансирование: исследование не имело спонсорской поддержки.

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Статья получена: 27.05.21 / Принята к публикации: 03.11.21 / Опубликовано: 30.11.21

Neural Networks for the Tasks of Analytical Support of Public Health and Environment Monitoring Systems

Boris I. Marchenko^{1,2}, Nina K. Plugotarenko¹, Olesja A. Semina¹

¹Southern Federal University, 105/42 Bolshaya Sadovaya Street, Rostov-on-Don, 344006, Russian Federation

²Branch of the Center for Hygiene and Epidemiology in the Rostov Region in the City of Taganrog, 16-a Bolshoi Prospekt Street, Taganrog, 347930, Russian Federation

Summary

Introduction: Ensuring a further improvement of efficiency of the public health monitoring system requires integration of the modern health risk analysis methodology with a complex of adapted unified traditional and innovative analytical methods and data exchange with the environmental monitoring system.

Objectives: The study aimed to test and assess the accuracy of predicting the incidence of malignant neoplasms using an artificial neural network.

Materials and methods: The analyzed time series are presented by information from statistical reporting forms on malignant neoplasms in the city of Taganrog, Rostov Region. We applied a regression model and a forecasting modeling technique based on a feedforward artificial neural network of a multilayer perceptron type. An artificial neural network with 117 neurons in a hidden layer was created in the environment of the Matlab R2021a application package with a set of tools for

the synthesis and analysis of neural networks Neural Network Toolbox using the Levenberg-Marquardt algorithm for its learning.

Results: Approbation of two forecasting models was carried out on learning samples of different duration including 15 and 34 years. In a comparative assessment of the accuracy of forecasts for 2018 and 2019, absolute and relative errors were estimated. The accuracy of the neural network forecasting model was higher than that of the regression model both for the total of malignant neoplasms and for most cancer sites. The absolute errors of forecasts for 2018 when using regression and neural network models were 17.05 and 1.49 per 100,000 population, for 2019 – 39.07 and 4.42, respectively. The prediction accuracy dropped with a decrease in the time series duration and an increase in the distance from the boundaries of the learning sample.

Conclusions: The feedforward artificial neural network of the multilayer perceptron type provides more accurate predictions using minimal input information compared to the regression model, which is its undoubted advantage.

Keywords: feedforward neural network, training neural networks, public health monitoring, environmental monitoring, analytical support, malignant neoplasms.

For citation: Marchenko BI, Plugotarenko NK, Semina OA. Neural networks for the tasks of analytical support of public health and environment monitoring systems. *Zdorov'e Naseleniya i Sreda Obitaniya*. 2021; 29(10):23–30. (In Russ.) doi: <https://doi.org/10.35627/2219-5238/2021-29-10-23-30>

Author information:

✉ Boris I. Marchenko, Dr. Sci. (Med.), Associate Professor; Professor, Department of Technosphere Safety and Chemistry, Institute of Nanotechnologies, Electronics and Equipment Engineering, Southern Federal University, the Department of Technosphere Safety and Chemistry; epidemiologist, Branch of the Center for Hygiene and Epidemiology in the Rostov Region in the city of Taganrog; e-mail: borismarch@gmail.com; ORCID: <https://orcid.org/0000-0001-6173-329X>.

Nina K. Plugotarenko, Cand. Sci. (Eng.), Associate Professor; Head of the Department of Technosphere Safety and Chemistry, Institute of Nanotechnologies, Electronics and Equipment Engineering, Southern Federal University; e-mail: plugotarenko@mail.ru; ORCID: <https://orcid.org/0000-0003-0853-8510>.

Olesja A. Semina, Master's Degree Student, Department of Technosphere Safety and Chemistry, Institute of Nanotechnologies, Electronics and Equipment Engineering, Southern Federal University; e-mail: semina@sfsu.ru; ORCID: <https://orcid.org/0000-0003-4263-5103>.

Author contributions: Marchenko B.I. developed the research conception and design, managed the project, did data collection and developed databases, did statistical and analytical data processing, and wrote the manuscript; Plugotarenko N.K. developed the research concept, was responsible for resources, software, and methodology, and did neural network learning; Semina O.A. did a literature review on the topic, neural network learning, and analytical data processing; all authors reviewed the results and approved the final version of the manuscript.

Funding: The authors received no financial support for the research, authorship, and/or publication of this article.

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

Received: May 27, 2021 / Accepted: November 3, 2021 / Published: November 30, 2021

Введение. В обеспечении санитарно-эпидемиологического благополучия и гигиенической безопасности населения России ключевое значение приобретает совершенствование национальной системы социально-гигиенического мониторинга (СГМ), что призвано обеспечить минимизацию воздействия факторов неопределенности при разработке и принятии адекватных ситуации управленческих решений в условиях относительного дефицита исходной информации. Успешное решение задач в различных отраслях гигиены, эпидемиологии и экологии определяет приоритетность инновационной методологической базы как инструментария, расширяющего информационно-аналитический потенциал ученых и практиков. Это нашло свое отражение в проекте «Концепции развития системы государственного социально-гигиенического мониторинга в Российской Федерации на период до 2030 года», в котором сформулированы актуальные задачи, решение которых призвано обеспечить существенное повышение эффективности информационно-аналитического обеспечения СГМ и результативности мероприятий риск-менеджмента; постулируется межведомственный и междисциплинарный характер системы СГМ, включая его тесные взаимосвязи с экологическим мониторингом (ЭМ) [1].

Анализ риска здоровью населения, обусловленного влиянием вредных и опасных факторов среды обитания, объединяющий в себе процедуры оценки риска, риск-менеджмента и информирования о риске, относится к наиболее актуальным междисциплинарным направлениям научной и практической деятельности. Так, к числу стратегических приоритетов в решении проблем в области экологии, что призвано обеспечить сохранение и укрепление здоровья нации, отнесена задача существенного увеличения научной поддержки деятельности по обеспечению экологической безопасности населения России, в том числе на основе широкого внедрения современной мето-

дологии оценки рисков здоровью [2]. В качестве научной основы организации и ведения СГМ рассматривается количественная характеристика причинно-следственных связей в системе «среда обитания – здоровье населения» с выходом на прогноз состояния здоровья населения, формируемого в условиях воздействия комплекса экологических, медико-биологических и социально-экономических факторов. Успешное решение поставленных задач диктует необходимость дальнейшего совершенствования методического, кадрового, информационно-аналитического и аппаратно-программного обеспечения; существенного расширения региональных и федерального информационных фондов СГМ [3–10]. Это нашло свое отражение в концепции научного обеспечения деятельности Роспотребнадзора на 2021–2025 гг., предусматривающей внедрение и развитие наукоемких технологий аналитических исследований, методов моделирования и прогнозирования ситуаций, основанных на цифровой информационной поддержке оптимизационных управленческих решений в целях оценки, динамического наблюдения с выходом на прогноз и минимизации рисков здоровью населения [11].

Современная методология анализа риска здоровью населения должна гармонично дополняться комплексом других адаптированных для ведения СГМ аналитических методов – как традиционных, так и инновационных. Подобная интеграция призвана обеспечить рост результативности системы СГМ на федеральном, региональном и муниципальном уровнях. Оптимизация изучения причинно-следственных связей в системе «окружающая среда – здоровье населения» с оперированием значительными по объему и сложными по структуре информационными массивами требует унификации технологий аналитических исследований. Находит свое применение модифицированный метод анализа эпидемиологического риска с выходом на частную

и комплексную оценку состояния здоровья населения в системе СГМ [12]. Перспективным является применение многомерных статистических методов, которые позволяют трансформировать многомерные массивы анализируемых данных в пригодный для предметной интерпретации вид и выявлять закономерности в обилии исходной информации, например, факторного анализа при процедуре структурирования данных, кластерного анализа в целях классификации и множественного регрессионного анализа для экстраполяции [13, 14]. Разрабатываются принципы применения при ведении СГМ и в процедуре оценки риска здоровью населения современные географические информационные систем (ГИС) и доступных сведений космического мониторинга совместно с результатами наземного динамического наблюдения – инструментального и лабораторного контроля показателей качества и безопасности среды обитания [15, 16].

Одним из направлений совершенствования аналитического обеспечения гигиенических и эпидемиологических исследований может стать применение искусственных нейронных сетей (ИНС), представляющих собой программно или аппаратно реализованные математические модели, функционирующие подобно биологическим нейронным сетям. ИНС организуются как массивы взаимосвязанных искусственных нейронов – простых по устройству процессоров, которые за счет их объединения в достаточно большую сеть приобретают способность решать сложные задачи. Основным преимуществом ИНС перед традиционными алгоритмами является то, что они не программируются, а обучаются. К результатам успешного обучения относится способность ИНС к выявлению сложных линейных и нелинейных зависимостей, к процедурам обобщения, к параллельной обработке информации, а также универсальность механизма обработки данных, адаптивность и устойчивость к ошибкам. Разработано значительное число типов ИНС, находящихся применение для решения задач по аппроксимации функций, классификации и распознавания образов, прогнозирования, идентификации, оценивания и ассоциативного управления; они применяются в различных сферах: от распознавания лиц и голоса до диагностики патологии человека [17–19]. Одним из преимуществ ИНС при прогнозировании является то, что отсутствует необходимость построения тренда

для исследуемого временного ряда, выделения в нем циклической и случайной компонент. ИНС в процессе обучения, используя минимальный объем входных данных, самостоятельно определяет соответствующие параметры и учитывает их при расчете прогноза. Предварительная настройка ИНС в основном сводится к определению количества нейронов в скрытых слоях и выбору алгоритма обучения [20].

Цель исследования: апробация и оценка точности прогнозирования заболеваемости злокачественными новообразованиями посредством искусственной нейронной сети.

Материалы и методы. Анализируемые временные ряды представлены сведениями из статистических отчетных форм № 35 (1984–2015 гг.) и № 7 (2016–2019 гг.) о злокачественных новообразованиях (ЗН) в г. Таганроге Ростовской области. При аналитической обработке исходных данных использован комплекс методов вариационной статистики. Реализованы две модели прогнозирования – регрессионная и нейросетевая. Регрессионная модель предусматривала подбор аппроксимирующих функций с построением теоретических линий тенденций методом наименьших квадратов и их экстраполяцию на среднесрочную перспективу [21, 22]. Нейросетевая модель выполнена как ИНС прямого распространения (без обратных связей) типа многослойного персептрона (Feed-Forward MLP) с архитектурой из последовательных слоев нейронов – входного, скрытого и выходного, осуществляющих преобразование входного сигнала в выходной (рис. 1).

ИНС была сформирована в среде пакета прикладных программ Matlab R2021a с набором инструментов для синтеза и анализа нейронных сетей Neural Network Toolbox. Для обучения ИНС применен алгоритм Левенберга – Марквардта (Levenberg–Marquardt method), предназначенный для оптимизации параметров нелинейных регрессионных моделей при решении задач о наименьших квадратах [23, 24]. Критериями оптимизации являлись значения среднеквадратичных ошибок на обучающих выборках (динамические ряды из годовых показателей частоты ЗН). По результатам оценки обучения ИНС определено, что наиболее точный прогноз обеспечивается при 117 нейронах в скрытом слое. При аналитической обработке применены профессиональный пакет статистических программ IBM SPSS Statistics (Statistical Package for Social Science) version 19.0,

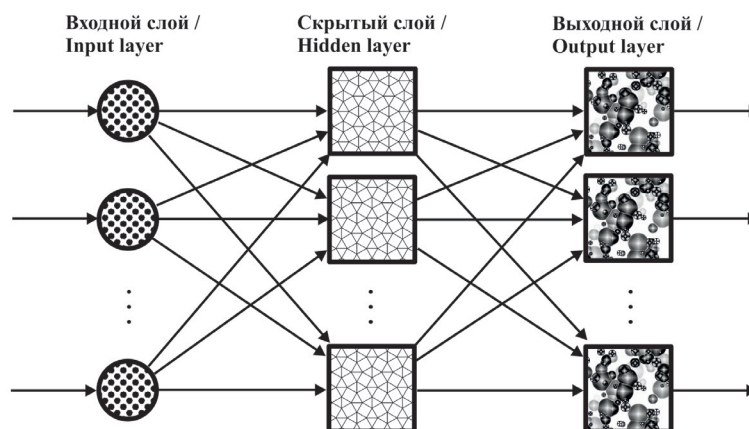


Рис. 1. Архитектура искусственной нейронной сети прямого распространения сигнала типа многослойного персептрона
Fig. 1. Architecture of a feedforward artificial neural network of a multilayer perceptron type

а также программное обеспечение собственной разработки — специализированный программный комплекс Turbo oncologist 2.01 и программа Turbo Dynamics 1.02 (среднесрочное экстраполяционное прогнозирование).

Результаты. Результаты анализа многолетней динамики показателей с выходом на прогноз в рамках гигиенических, эпидемиологических и экологических исследований служат важным критерием выбора приоритетов при разработке управленческих решений и базисом ретроспективной, по принципу обратной связи, оценки эффективности реализованных мероприятий. Сравнительная оценка точности прогнозов, полученных с применением регрессионной и нейросетевой моделей, выполнена на основе временных рядов из годовых показателей частоты ЗН (сумма локализаций, кожа без меланомы; ободочная железа; трахея, бронхи и легкое; желудок; ободочная кишка; тело матки). Выбор ЗН для сравнительного анализа точности регрессионной и нейросетевой моделей прогнозирования обусловлен тем, что данная патология характеризуется высокой социально-экономической значимостью, тяжестью клинического течения и высокой долей в структуре причин смертности, доказанной для отдельных локализаций экологической зависимостью и наличием достоверных сведений о регистрируемой онкологической заболеваемости. Актуальность проблемы ЗН для населения Таганрога определяется неблагоприятной ситуацией по онкологической заболеваемости при среднемноголетнем уровне за 2006–2020 гг. (480,11 ‰), превышающем показатель для городского населения Ростовской области (373,98 ‰) в 1,28 раза (первое ранговое место), а также сформировавшейся с 1985 года устойчивой тенденцией к росту частоты ЗН при среднегодовом темпе прироста +2,03 % (рис. 2).

В структуре ЗН за 2006–2020 гг. приоритетной локализацией является кожа (без учета меланомы) при удельном весе 15,08 %, на втором ранговом месте находится молочная железа (12,50 %), на третьем — трахея, бронхи и легкое (9,14 %); затем

следуют ободочная кишка (6,84 %), предстательная железа (6,14 %) и желудок (6,04 %).

Апробация обеих моделей прогнозирования выполнена на обучающих выборках различной продолжительности — 34 года (1984–2017 гг.) и 15 лет (2003–2017 гг.). В обоих случаях прогнозировались показатели частоты ЗН на 2018 и 2019 гг. При реализации регрессионной модели проводился выбор оптимальных вариантов аппроксимации из набора, включающего шесть функций (прямолинейную, логарифмическую, степенную, экспоненциальную, параболы 2-го и 3-го порядка) с последующим экстраполяционным прогнозированием по теоретическим линиям тенденций, рассчитанных методом наименьших квадратов. Разработанная программа ИНС, моделирующая многослойный перцептрон прямого распространения сигнала, обеспечивала построение прогнозов при обучении по алгоритму Левенберга — Марквардта и фиксированном числе нейронов в скрытом слое, равном 117. Результатами работы программы ИНС являются данные спрогнозированных временных рядов. На заключительном этапе проводилась сравнительная оценка точности полученных обоими методами среднесрочных прогнозов частоты ЗН по величинам их абсолютных (ΔX , ‰) и относительных (ΔY , %) погрешностей.

Точность нейросетевой модели прогнозирования по данным обучающих выборок продолжительностью 34 года оказалась выше, чем регрессионной как по сумме ЗН, так и по большинству отдельных локализаций. Так, абсолютная погрешность прогноза суммарной заболеваемости ЗН на 2018 год с применением нейронной сети (1,49 ‰) меньше, чем при регрессионной модели (17,05 ‰) в 11,44 раза при относительных погрешностях соответственно 0,29 и 3,35 %; абсолютная погрешность прогноза на 2019 год для нейросетевой модели (4,42 ‰) меньше, чем для регрессионной (39,07 ‰) в 8,84 раза при относительных погрешностях 0,89 и 7,89 %. Абсолютные погрешности прогнозов частоты ЗН отдельных локализаций на 2018 год, полученных с применением регрессионной

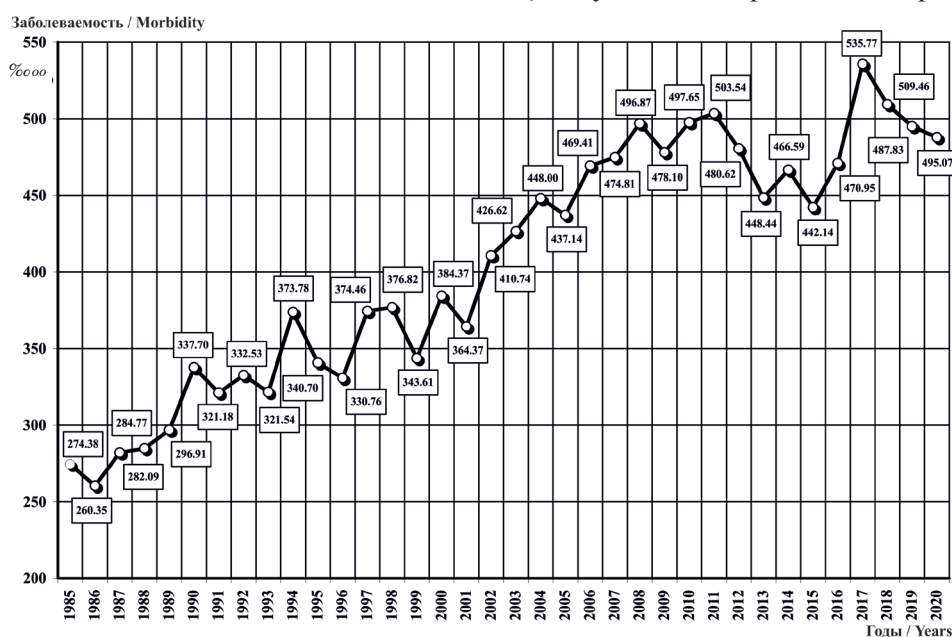


Рис. 2. Динамика заболеваемости злокачественными новообразованиями в городе Таганроге (на 100 000 населения)

Fig. 2. Cancer incidence rates in the city of Taganrog (per 100,000 population)

модели, также превышали соответствующие показатели альтернативного метода в 1,37–26,61 раза, относительные погрешности – на 5,56–42,19 %. Среди прогнозов на 2019 год с применением нейронной сети менее точный результат получен только в отношении ЗН ободочной кишки, а для ЗН молочной железы отмечено полное совпадение результатов обеих моделей (табл. 1).

Преимущество нейросетевой модели в точности прогнозов на 2018 год по сравнению с

регрессионной моделью в основном сохраняется при меньшей, пятнадцатилетней продолжительности обучающей выборки. Так, абсолютные погрешности соответствующих прогнозов для суммы ЗН различаются в 30,29 раза (0,63 и 19,08 ‰), а относительные погрешности – на 3,62 %. В то же время несколько более точный прогноз общей онкологической заболеваемости на 2019 год получен с применением регрессионной модели. Применение нейросетевой модели по

Таблица 1. Сравнительная оценка точности прогнозов частоты злокачественных новообразований методами экстраполяции по теоретическим линиям тенденций и применения нейронной сети прямого распространения по данным за 1984–2019 гг.

Table 1. Comparative assessment of the accuracy of cancer incidence forecasts by extrapolation along theoretical lines of trends and using a feedforward neural network according to data for 1984–2019

Годы / Years	Наименования показателей / Indicators	Локализации злокачественных новообразований / Tumor sites						
		Всего / Total	Желудок / Stomach	Кишка ободоч- ная / Colon	Легкие, брон- хи, трахея / Lung, bronchi, trachea	Кожа (без меланомы) / Skin (except melanoma)	Железы молочные / Breast	Тело матки / Uterus
Фактические зарегистрированные относительные показатели частоты злокачественных новообразований на 100 тысяч населения / Registered cancer incidence rates 100 thousand population (‰)								
2018	Заболеваемость злокачественными новообразованиями / Cancer incidence (‰)	509,46	29,59	25,99	37,99	85,98	79,98	19,19
2019	Заболеваемость злокачественными новообразованиями / Cancer incidence (‰)	495,07	23,67	26,88	46,54	97,09	58,17	20,06
Прогнозы частоты злокачественных новообразований методом экстраполяции по теоретическим линиям тенденций за 1984–2019 гг. / Cancer incidence rates predicted by extrapolation along theoretical lines of trends based on data for 1984–2019								
2018	Прогнозы заболеваемости / Predicted cancer incidence rates (‰)	526,51	23,53	37,87	43,40	71,35	68,71	25,65
	Абсолютные погрешности прогнозов заболеваемости / Absolute errors of predicted rates (ΔХА, ‰)	17,05	6,06	11,88	5,41	14,63	11,27	6,46
	Относительные погрешности прогнозов заболеваемости / Relative errors of predicted rates (ΔYA, %)	3,35	20,50	45,70	14,23	17,02	14,10	33,66
2019	Прогнозы заболеваемости / Predicted cancer incidence rates (‰)	534,14	22,30	38,67	43,33	72,07	70,09	26,18
	Абсолютные погрешности прогнозов заболеваемости / Absolute errors of predicted rates (ΔХА, ‰)	39,07	1,37	11,79	3,21	25,02	11,92	6,12
	Относительные погрешности прогнозов заболеваемости / Relative errors of predicted rates (ΔYA, %)	7,89	5,78	43,87	6,90	25,77	20,48	30,53
Прогнозы частоты злокачественных новообразований методом нейронной сети прямого распространения типа многослойного персептрона по данным за период 1984–2019 годы / Cancer incidence rates predicted using the technique of a feedforward neural network of the multilayer perceptron type (Feed-Forward MLP) based on data for 1984–2019								
2018	Прогнозы заболеваемости / Predicted cancer incidence rates (‰)	507,97	25,17	25,08	35,84	85,43	75,71	22,50
	Абсолютные погрешности прогнозов заболеваемости / Absolute errors of predicted rates (ΔХА, ‰)	1,49	4,42	0,91	2,15	0,55	4,27	3,31
	Относительные погрешности прогнозов заболеваемости / Relative errors of predicted rates (ΔYA, %)	0,29	14,94	3,50	5,66	0,64	5,34	17,25
2019	Прогнозы заболеваемости / Predicted cancer incidence rates (‰)	499,49	23,50	39,50	46,12	88,14	70,09	19,90
	Абсолютные погрешности прогнозов заболеваемости / Absolute errors of predicted rates (ΔХА, ‰)	4,42	0,17	12,62	0,42	8,95	11,92	0,16
	Относительные погрешности прогнозов заболеваемости / Relative errors of predicted rates (ΔYA, %)	0,89	0,72	46,95	0,90	9,22	20,49	0,80
Соотношения абсолютных погрешностей прогнозов частоты злокачественных новообразований методами экстраполяции по теорети- ческим линиям тенденций (ΔХА) и применения нейронной сети прямого распространения (ΔХВ) по данным за период 1984–2019 гг. / Ratios of absolute errors of predicted cancer incidence rates by extrapolation along theoretical lines of trends (ΔХА) and using a feedforward neural network (ΔХВ) based on data for 1984–2019								
2018	Соотношение / ratio ΔХА/ΔХВ	11,44	1,37	13,05	2,51	26,61	2,64	1,95
2019	Соотношение / ratio ΔХА/ΔХВ	8,84	8,05	0,93	7,65	2,80	1,00	38,27
Разница относительных погрешностей прогнозов частоты злокачественных новообразований методами экстраполяции по теоретичес- ким линиям тенденций (ΔYA) и применения нейронной сети прямого распространения (ΔYB) по данным за период 1984–2019 гг. / Difference of relative errors of predicted cancer incidence rates by extrapolation along theoretical lines of trends (ΔYA) and using a feedforward neural network (ΔYB) based on data for 1984–2019								
2018	Разница / Difference ΔYA–ΔYB (%)	3,05	5,56	42,19	8,57	16,38	8,76	16,42
2019	Разница / Difference ΔYA–ΔYB (%)	7,00	5,06	–3,08	6,00	16,55	–0,01	29,73

результатам прогнозирования на 2018 год оказалось предпочтительнее в отношении частоты ЗН для всех учитываемых отдельных локализаций, за исключением трахеи, бронхов и легкого, на 2019 год — для ЗН желудка, кожи без учета меланомы и тела матки (табл. 2).

Заключение. Апробация и результаты сравнительного анализа среднесрочных прогнозов заболеваемости злокачественными новообразованиями свидетельствуют о более высокой их точности при использовании искусственной

нейронной сети типа многослойного персептрона прямого распространения сигнала по сравнению с результатами применения регрессионной модели с экстраполяционным прогнозированием по теоретическим линиям тенденций, рассчитанных методом наименьших квадратов. Наиболее точные прогнозы были получены на выходе искусственной нейронной сети со 117 нейронами в скрытом слое, обученной по алгоритму Левенберга — Марквардта. Нейросетевая модель позволила получить достаточно точные прогнозы при использовании

Таблица 2. Сравнительная оценка точности прогнозов частоты злокачественных новообразований методами экстраполяции по теоретическим линиям тенденций и применения нейронной сети прямого распространения по данным за 2003–2019 гг.

Table 2. Comparative assessment of the accuracy of predicted cancer incidence rates by extrapolation along theoretical lines of trends and using a feedforward neural network based on data for 2003–2019

Годы / Years	Наименования показателей / Indicators	Локализации злокачественных новообразований / Cancer sites						
		Всего / Total	Желудок / Stomach	Кишка ободоч- ная / Colon	Легкие, брон- хи, трахея / Lung, bronchi, trachea	Кожа (без меланомы) / Skin(except melanoma)	Железы молочные / Breast	Тело матки / Uterus
Фактические зарегистрированные относительные показатели частоты злокачественных новообразований на 100 тысяч населения / Registered relative rates of malignant neoplasms per 100 thousand population (‰000)								
2018	Заболеваемость злокачественными новообразованиями / Cancer incidence rates (‰000)	509,46	29,59	25,99	37,99	85,98	79,98	19,19
2019	Заболеваемость злокачественными новообразованиями / Cancer incidence rates (‰000)	495,07	23,67	26,88	46,54	97,09	58,17	20,06
Прогнозы частоты злокачественных новообразований методом экстраполяции по теоретическим линиям тенденций за 2003–2019 гг. / Predicted cancer incidence rates by extrapolation along theoretical lines of trends based on data for 2003–2019								
2018	Прогнозы заболеваемости / Predicted incidence rates (‰000)	490,38	20,36	37,15	38,10	69,45	70,24	23,33
	Абсолютные погрешности прогнозов заболеваемости / Absolute errors of predicted rates (ΔХА, ‰000)	19,08	9,23	11,16	0,11	16,53	9,74	4,14
	Относительные погрешности прогнозов заболеваемости / Relative errors of predicted rates (ΔYA, %)	3,75	31,19	42,94	0,29	19,23	12,18	21,57
2019	Прогнозы заболеваемости / Predicted incidence rates (‰000)	491,62	18,92	37,48	37,23	69,57	71,83	23,42
	Абсолютные погрешности прогнозов заболеваемости / Absolute errors of predicted rates (ΔХА, ‰000)	3,45	4,75	10,60	9,31	27,52	13,66	3,36
	Относительные погрешности прогнозов заболеваемости / Relative errors of predicted rates (ΔYA, %)	0,70	20,07	39,43	20,00	28,34	23,48	16,75
Прогнозы частоты злокачественных новообразований методом нейронной сети прямого распространения типа многослойного пер- септрона по данным за период 2003–2019 годы / Predicted cancer incidence rates using the technique of a feedforward neural network of the multilayer perceptron type (Feed-Forward MLP) based on data for 2003–2019								
2018	Прогнозы заболеваемости / Predicted incidence rates (‰000)	508,83	21,56	27,09	42,86	79,20	70,46	18,50
	Абсолютные погрешности прогнозов заболеваемости / Absolute errors of predicted rates (ΔХА, ‰000)	0,63	8,03	1,10	4,87	6,78	9,52	0,69
	Относительные погрешности прогнозов заболеваемости / Relative errors of predicted rates (ΔYA, %)	0,12	27,14	4,23	12,82	7,89	11,90	3,60
2019	Прогнозы заболеваемости / Predicted incidence rates (‰000)	512,92	21,36	39,60	32,90	85,79	75,70	23,08
	Абсолютные погрешности прогнозов заболеваемости / Absolute errors of predicted rates (ΔХА, ‰000)	7,85	2,31	12,72	13,64	11,30	17,53	3,02
	Относительные погрешности прогнозов заболеваемости / Relative errors of predicted rates (ΔYA, %)	3,61	9,76	47,32	29,31	11,64	30,14	15,05
Соотношения абсолютных погрешностей прогнозов частоты злокачественных новообразований методами экстраполяции по теорети- ческим линиям тенденций (ΔХА) и применения нейронной сети прямого распространения (ΔХВ) по данным за период 2003–2019 гг. / Ratios of absolute errors of predicted cancer incidence rates by extrapolation along theoretical lines of trends (ΔХА) and using a feedforward neural network (ΔХВ) based on data for 2003–2019								
2018	Соотношение / ratio ΔХА/ΔХВ	30,29	1,15	10,15	0,02	2,44	1,02	6,00
2019	Соотношение / ratio ΔХА/ΔХВ	0,19	2,06	0,83	0,68	2,44	0,78	1,11
Разница относительных погрешностей прогнозов частоты злокачественных новообразований методами экстраполяции по теоретиче- ским линиям тенденций (ΔYA) и применения нейронной сети прямого распространения (ΔYB) по данным за период 2003–2019 гг. / Difference of relative errors of predicted cancer incidence rates by extrapolation along theoretical lines of trends (ΔYA) and using a feedforward neural network (ΔYB) based on data for 2003–2019								
2018	Разница / Difference ΔYA–ΔYB (%)	3,62	4,06	38,71	–12,53	11,34	0,28	17,98
2019	Разница / Difference ΔYA–ΔYB (%)	–2,91	10,31	–7,89	–9,30	16,71	–6,65	1,69

минимума входной информации — только временных рядов из годовых показателей частоты злокачественных новообразований без учета их особенностей (тренда, циклической и случайной компонент) и влияния внешних факторов. Таким образом, точность полученных прогнозов позволяет предположить, что правильно обученная нейронная сеть обладает способностью восстанавливать функцию без какой-либо дополнительной информации, что является несомненным преимуществом нейросетевых моделей по сравнению с другими способами среднесрочного прогнозирования. По результатам апробации нейронной сети можно сделать вывод о прямой зависимости точности прогноза от продолжительности исходного временного ряда и обратной зависимости — от расстояния между прогнозируемым элементом и границей обучающей выборки.

Список литературы

1. Попова А.Ю., Кузьмин С.В., Гурвич В.Б. и др. Информационно-аналитическая поддержка управления риском для здоровья населения на основе реализации концепции развития системы социально-гигиенического мониторинга в Российской Федерации на период до 2030 года // *Здоровье населения и среда обитания*. 2019. № 9 (318). С. 4–12. doi: <https://doi.org/10.35627/2219-5238/2019-318-9-4-12>
2. Попова А.Ю. Стратегические приоритеты Российской Федерации в области экологии с позиций сохранения здоровья нации // *Здоровье населения и среда обитания*. 2014. № 2 (25). С. 4–7.
3. Рахманин Ю.А., Леванчук А.В., Копытенкова О.И. Совершенствование системы социально-гигиенического мониторинга территорий крупных городов // *Гигиена и санитария*. 2017. Т. 96. № 4. С. 298–301. doi: <https://doi.org/10.18821/0016-9900-2017-96-4-298-301>
4. Зайцева Н.В., Жданова-Заплесвичко И.Г., Землянова М.А., Пережогин А.Н., Савиных Д.Ф. Опыт организации и проведения санитарно-эпидемиологических исследований по выявлению и документированию нарушений здоровья населения с качеством атмосферного воздуха в зонах влияния хозяйствующих субъектов // *Здоровье населения и среда обитания*. 2021. № 1 (334). С. 4–15. doi: <https://doi.org/10.35627/2219-5238/2021-334-1-4-15>
5. Зайцева Н.В., Май И.В., Клейн С.В., Кирьянов Д.А. Методические аспекты и результаты оценки демографических потерь, ассоциированных с вредным воздействием факторов среды обитания и предотвращаемых действиями Роспотребнадзора, в регионах Российской Федерации // *Здоровье населения и среда обитания*. 2018. № 4 (301). С. 15–20. doi: <https://doi.org/10.35627/2219-5238/2018-301-4-15-20>
6. Попова А.Ю., Гурвич В.Б., Кузьмин С.В., Мишина А.Л., Ярушин С.В. Современные вопросы оценки и управления риском для здоровья // *Гигиена и санитария*. 2017. Т. 96, № 12. С. 1125–1129. doi: <https://doi.org/10.18821/0016-9900-2017-96-12-1125-1129>
7. Попова А.Ю., Зайцева Н.В., Май И.В. К вопросу об имплементации оценки качества жизни населения в систему социально-гигиенического мониторинга // *Анализ риска здоровью*. 2018. № 3. С. 4–12. doi: <https://doi.org/10.21668/health.risk/2018.3.01>
8. Кузьмин С.В., Гурвич В.Б., Диконская О.В. и др. Социально-гигиенический мониторинг и информационно-аналитические системы обеспечения оценки и управления риском для здоровья населения и риск-ориентированной модели надзорной деятельности // *Гигиена и санитария*. 2017. Т. 96. № 12. С. 1130–1136. doi: <https://doi.org/10.18821/0016-9900-2017-96-12-1130-1136>
9. Ракитский В.Н., Авалиани С.Л., Новиков С.М., Шашина Т.А., Додина Н.С., Кислицин В.А. Анализ риска здоровью при воздействии атмосферных загрязнений как составная часть стратегии уменьшения глобальной эпидемии неинфекционных заболеваний // *Анализ риска здоровью*. 2019. № 4. С. 30–36. doi: <https://doi.org/10.21668/health.risk/2019.4.03>
10. Зайцева Н.В., Май И.В., Кирьянов Д.А., Горяев Д.В., Клейн С.В. Социально-гигиенический мониторинг на современном этапе: состояние и перспективы развития в сопряжении с риск-ориентированным надзором // *Анализ риска здоровью*. 2016. № 4. С. 4–16. doi: <https://doi.org/10.21668/health.risk/2016.4.01>
11. Попова А.Ю., Зайцева Н.В., Кузьмин С.В., Май И.В. Приоритеты научной поддержки деятельности санитарно-эпидемиологической службы в области гигиены: поиск ответов на известные угрозы и новые вызовы // *Анализ риска здоровью*. 2021. № 1. С. 4–14. doi: <https://doi.org/10.21668/health.risk/2021.1.01>
12. Айдинов Г.Т., Марченко Б.И., Синельникова Ю.А. Применение комплексной оценки состояния здоровья населения в задачах совершенствования системы социально-гигиенического мониторинга // *Гигиена и санитария*. 2016. Т. 95. № 10. С. 980–985. doi: <https://doi.org/10.18821/0016-9900-2016-95-10-980-985>
13. Айдинов Г.Т., Марченко Б.И., Софьяникова Л.В., Синельникова Ю.А. Применение многомерных статистических методов при выполнении задач совершенствования информационно-аналитического обеспечения системы социально-гигиенического мониторинга // *Здоровье населения и среда обитания*. 2015. № 7 (268). С. 4–8.
14. Айдинов Г.Т., Марченко Б.И., Синельникова Ю.А. Многомерный анализ структуры и долевого вклада потенциальных факторов риска при злокачественных новообразованиях трахеи, бронхов и легкого // *Анализ риска здоровью*. 2017. № 1. С. 47–55. doi: <https://doi.org/10.21668/health.risk/2017.1.06>
15. Студеникина Е.М., Стёпкин Ю.И., Клепиков О.В., Колнет И.В., Попова Л.В. Проблемные вопросы использования географических информационных систем в социально-гигиеническом мониторинге и риск-ориентированном надзоре // *Здоровье населения и среда обитания*. 2019. № 6 (315). С. 31–36. doi: <https://doi.org/10.35627/2219-5238/2019-315-6-31-36>
16. Епринцев С.А., Клепиков О.В., Шекоян С.В. Дистанционное зондирование Земли как способ оценки качества окружающей среды урбанизированных территорий // *Здоровье населения и среда обитания*. 2020. № 4 (325). С. 5–12. doi: <https://doi.org/10.35627/2219-5238/2020-325-4-5-12>
17. Гафаров Ф.М., Галимянов А.Ф. Искусственные нейронные сети и приложения. Казань: Издательство Казанского университета, 2018. 121 с.
18. Круглов В.В., Борисов В.В. Искусственные нейронные сети. Теория и практика. М.: Горячая линия — Телеком, 2001. 382 с.
19. Шляхов Г.Г. Искусственные нейронные сети и их приложения // *Вестник Пермского университета*. Серия: Информационные системы и технологии. 2007. № 10 (15). С. 99–102.
20. Червяков Н.И., Тихонов Э.Е. Применение нейронных сетей для задач прогнозирования и проблемы идентификации моделей прогнозирования на нейронных сетях // *Нейрокомпьютеры: разработка, применение*. 2003. № 10-11. С. 25–31.
21. Сепетлиев Д. Статистические методы в научных медицинских исследованиях. М.: Медицина, 1968. 420 с.

22. Марченко Б.И. Здоровье на популяционном уровне: статистические методы исследования. Таганрог: Сфинкс, 1997. 432 с.
23. Пархоменко С.С., Леденева Т.М. Обучение нейронных сетей методом Левенберга – Марквардта в условиях большого количества данных // Вестник Воронежского государственного университета. Серия: Системный анализ и информационные технологии. 2014. № 2. С. 98–106.
24. Пархоменко С.С. О сокращении времени обработки большого количества данных нейронными сетями методом Левенберга – Марквардта // Международный научно-исследовательский журнал. 2014. № 1-1 (20). С. 80–83.

References

1. Popova AYU, Kuzmin SV, Gurvich VB, et al. Data-driven risk management for public health as supported by the experience of implementation for development concept of the social and hygienic monitoring framework in the Russian Federation up to 2030. *Zdorov'e Naseleniya i Sreda Obitaniya*. 2019;(9(318)):4–12. (In Russ.) doi: 10.35627/2219-5238/2019-318-9-4-12
2. Popova AYU. Strategic priorities of the Russian Federation in the field of ecology from the position of preservation of health of the nation. *Zdorov'e Naseleniya i Sreda Obitaniya*. 2014;(2(251)):4–7. (In Russ.)
3. Rakhmanin YuA, Levanchuk AV, Kopytenkova OI. Improvement of the system of social and hygienic monitoring of territories of large cities. *Gigiena i Sanitariya*. 2017;96(4):298–301. (In Russ.) doi: 10.18821/0016-9900-2017-96-4-298-301
4. Zaitseva NV, Zhdanova-Zaplesvichko IG, Zemlyanova MA, Perezhogin AN, Savinykh DF. Experience in organizing and conducting epidemiological studies to detect and prove the causal relationship between ambient air quality and health disorders in the population of industrially contaminated sites. *Zdorov'e Naseleniya i Sreda Obitaniya*. 2021;(1(334)):4–15. (In Russ.) doi: 10.35627/2219-5238/2021-334-1-4-15
5. Zaitseva NV, May IV, Klein SV, Kiryanov DA. Methodological aspects and results of estimation of demographic loss associated with harmful influence of environmental factors and preventive activities of Rospotrebnadzor in regions of the Russian Federation. *Zdorov'e Naseleniya i Sreda Obitaniya*. 2018;(4(301)):15–20. (In Russ.) doi: 10.35627/2219-5238/2018-301-4-15-20
6. Popova AYU, Gurvich VB, Kuzmin SV, Mishina AL, Yarushin SV. Modern issues of the health risk assessment and management. *Gigiena i Sanitariya*. 2017;96(12):1125–1129. (In Russ.) doi: 10.18821/0016-9900-2017-96-12-1125-1129
7. Popova AYU, Zaitseva NV, May IV. On implementation of population life quality assessment into social-hygienic monitoring system. *Health Risk Analysis*. 2018;(3):4–12. (In Russ.) doi: 10.21668/health.risk/2018.3.01
8. Kuzmin SV, Gurvich VB, Dikonskaya OV, et al. Socio-hygienic monitoring and information analysis systems supporting the health risk assessment and management and a risk-focused model of supervisory activities in the sphere of securing sanitary and epidemiologic public welfare. *Gigiena i Sanitariya*. 2017;96(12):1130–1136. (In Russ.) doi: 10.18821/0016-9900-2017-96-12-1130-1136
9. Rakitskii VN, Avaliani SL, Novikov SM, Shashina TA, Dodina NS, Kislitsin VA. Health risk analysis related to exposure to ambient air contamination as a component in the strategy aimed at reducing global non-infectious epidemics. *Health Risk Analysis*. 2019;(4):30–36. (In Russ.) doi: 10.21668/health.risk/2019.4.03
10. Zaitseva NV, May IV, Kiryanov DA, Goryaev DV, Kleyn SV. Social and hygienic monitoring today: state and prospects in conjunction with the risk-based supervision. *Health Risk Analysis*. 2016;(4):4–16. (In Russ.) doi: 10.21668/health.risk/2016.4.01
11. Popova AYU, Kuzmin SV, Zaitseva NV, May IV. Priorities in scientific support provided for hygienic activities accomplished by a sanitary and epidemiologic service: how to face known threats and new challenges. *Health Risk Analysis*. 2021;(1):4–14. (In Russ.) doi: 10.21668/health.risk/2021.1.01
12. Aydinov GT, Marchenko BI, Sinelnikova YuA. The application of complex assessment of the health status of the population in tasks of the improvement of the socio-hygienic monitoring system. *Gigiena i Sanitariya*. 2016;95(10):980–985. (In Russ.) doi: 10.18821/0016-9900-2016-95-10-980-985
13. Aydinov GT, Marchenko BI, Sofyanikova LV, Sinelnikova YuA. The application of multidimensional statistical methods in the tasks of improving of information and analytical providing of the system socio-hygienic monitoring. *Zdorov'e Naseleniya i Sreda Obitaniya*. 2015;(7(268)):4–8. (In Russ.)
14. Aydinov GT, Marchenko BI, Sinelnikova YuA. Multivariate analysis of structure and contribution per shares made by potential risk factors at malignant neoplasms in trachea, bronchial tubes and lung. *Health Risk Analysis*. 2017;(1):47–55. (In Russ.) doi: 10.21668/health.risk/2017.1.06
15. Studenikina EM, Stepkin YuI, Klepikov OV, Kolnet IV, Popova LV. Problematic issues of the geographic information systems use in socio-hygienic monitoring and risk-based supervision. *Zdorov'e Naseleniya i Sreda Obitaniya*. 2019;(6(315)):31–36. (In Russ.) doi: 10.35627/2219-5238/2019-315-6-31-36
16. Yepintsev SA, Klepikov OV, Shekoyan SV. Remote sensing of the Earth as a method of assessing environmental quality of urban areas. *Zdorov'e Naseleniya i Sreda Obitaniya*. 2021;(4(325)):5–12. (In Russ.) doi: 10.35627/2219-5238/2020-325-4-5-12
17. Gafarov FM, Galimyanov AF. [Artificial Neural Networks and Applications]. Kazan: Kazan University Publ.; 2018. (In Russ.)
18. Kruglov VV, Borisov VV. [Artificial Neural Networks. Theory and Practice]. Moscow: Goryachaya Liniya – Telekom Publ.; 2001. (In Russ.)
19. Shlyakhov GG. Artificial neural networks and its applications. *Vestnik Permskogo Universiteta. Series: Information Systems and Technologies*. 2007;(10(15)):99–102. (In Russ.)
20. Chervyakov NI, Tikhonov EE. Neural networks application for problems of forecasting and a problem of identification of forecasting's models. *Neyrokomputery: Razrabotka, Primenenie*. 2003;(10-11):25–31. (In Russ.)
21. Sepetliev D. [Statistical Methods in Scientific Medical Research]. Moscow: Meditsina Publ.; 1968. (In Russ.)
22. Marchenko BI. [Public Health: Statistical Research Methods]. Taganrog: Sfinks Publ.; 1997. (In Russ.)
23. Parkhomenko SS, Ledeneva TM. Training neural networks of the method Levenberg-Marquardt in larger the amount of data. *Vestnik Voronezhskogo Gosudarstvennogo Universiteta. Series: Systems Analysis and Information Technologies*. 2014;(2):98–106. (In Russ.)
24. Parkhomenko SS. A Levenberg-Marquardt algorithm execution time reducing in case of large amount of the data. *Mezhdunarodnyy Nauchno-Issledovatel'skiy Zhurnal*. 2014;(1-1(20)):80–83. (In Russ.)

