

В.Ю. Головинова¹, С.Г. Киреев², П.К. Котенко³,
Ю.Л. Минаев⁴, И.Н. Штамбург¹, С.Г. Кузьмин¹

Нейросетевые модели прогнозирования заболеваемости в организованных коллективах

¹Военно-медицинская академия им. С.М. Кирова, Санкт-Петербург

²Городская больница № 40, Санкт-Петербург

³Всероссийский центр экстренной и радиационной медицины им. А.М. Никифорова МЧС России, Санкт-Петербург

⁴Самарский медицинский институт «РЕАВИЗ», Самара

Резюме. Представлены результаты применения искусственных нейронных сетей для прогнозирования заболеваемости в организованных коллективах на примере сотрудников Федеральной противопожарной службы Северо-Западного регионального центра Министерства Российской Федерации по делам гражданской обороны, чрезвычайных ситуаций и ликвидации последствий стихийных бедствий. Разработаны нейросетевые модели прогнозирования первичной заболеваемости и первичной заболеваемости болезнями органов дыхания. Установлено, что наилучший результат показали нейросетевые модели для прогнозирования заболеваемости болезнями органов дыхания, что связано с преимущественно инфекционным характером заболеваний данного класса. Искусственные нейронные сети показали высокую степень возможности создания статистически достоверных моделей для прогнозирования первичной заболеваемости, однако методика показывает свою эффективность в условиях стабильности существующей и сохраненной тенденций. Применение нейросетевой методики позволяет получить качественный прогноз, не уступающий классическим линейным регрессионным моделям. Методика искусственных нейронных сетей рекомендована для прогнозирования заболеваемости в организованных коллективах, в которых законодательством предусмотрена военная или приравненная к ней служба.

Ключевые слова: первичная заболеваемость, нейронные сети, нейросетевое прогнозирование, нейросетевые модели, организованные коллективы, Федеральная противопожарная служба, прогнозирование заболеваемости, искусственный интеллект.

Введение. В современных условиях руководители здравоохранения министерств и ведомств, где федеральным законодательством предусмотрена военная или приравненная к ней служба, располагая традиционной информационной базой и сложившимися принципами ее использования, все чаще сталкиваются с потребностью разработки эффективных методик анализа заболеваемости необходимых для принятия аргументированных стратегических решений, оперативного управления и планирования медицинской помощи. В здравоохранении в настоящее время наиболее часто и успешно используются такие методики анализа и прогнозирования заболеваемости, как экстраполяция, линейный регрессионный анализ, непрямолinéйные регрессии, ретроспективный эпидемиологический анализ [4, 5, 7, 8, 10, 12]. Однако в современной литературе отмечаются не только достоинства этих методик, но и их недостатки [4–6].

Наиболее перспективными, на наш взгляд, представляются разработка и внедрение в повседневную практику органов управления здравоохранением методик анализа и прогнозирования заболеваемости на основе искусственных нейронных сетей, представляющих собой математические модели, а также их программные или аппаратные реализации, построенные по принципу организации и функцио-

нирования биологических нейронных сетей – сетей нервных клеток живого организма.

Искусственные нейронные сети чаще используются в клинических исследованиях для прогнозирования развития конкретных заболеваний, осложнений, исходов, эффективности лечения [1, 3, 6, 9, 11]. Вместе с тем, исследования по использованию искусственного интеллекта для прогноза уровня заболеваемости представляют несомненный научный и практический интерес, но на сегодняшний день встречаются неоправданно редко [9].

Цель исследования. Оценить возможность применения методики искусственных нейронных сетей для прогнозирования заболеваемости в организованных коллективах.

Материалы и методы. Исследование проводилось на примере организованных коллективов сотрудников Федеральной противопожарной службы (ФПС) Северо-Западного регионального центра Министерства Российской Федерации по делам гражданской обороны, чрезвычайных ситуаций и ликвидации последствий стихийных бедствий (МЧС России). Материал исследования составили данные отчетов о зарегистрированных заболеваниях и их исходах и о причинах временной нетрудоспособности сотрудни-

ков за период с 2003 по 2008 г. Численность пожарных за каждый отчётный год составила в среднем 6 200 человек, а абсолютное число первичных заболеваний за исследуемый период, 16 245 случаев.

На подготовительном этапе выполнен анализ интенсивных показателей состояния здоровья сотрудников ФПС МЧС России путем расчета показателей частоты каждого вида заболеваемости на 1000 человек личного состава суммарно по всем классам заболеваний и по основным, наиболее значимым для данной когорты классам и группам болезней.

Математические эксперименты по разработке методики прогнозирования заболеваемости сотрудников на основе искусственных нейронных сетей проведены в пакете программного обеспечения Statistica 6.0 в три этапа. В ходе первого этапа была сформирована обучающая выборка, которую составили сведения о заболеваемости сотрудников ФПС МЧС России за 2003–2007 гг. по месяцам. Из состава обучающей выборки сеть автоматически исключала некоторое количество значений для создания контрольной и тестовой выборок, необходимых для контроля над процессом «обучения». Данные о заболеваемости в 2008 г. на сеть не подавались и были зарезервированы для проверки эффективности созданной модели прогнозирования и качества прогноза.

На втором этапе осуществлялось «обучение» искусственных нейронных сетей по подаваемому на сеть набору обучающих данных. Для выбора оптимального типа нейронной сети провели его обучение на моделях: линейной сети, многослойном персептроне и сети с радиальной базисной функцией. Программа автоматически выбирала из одной тысячи созданных моделей 5 наилучших [2]. Оптимальную архитектуру сети определяли в процессе эксперимента опытным путем. Определённые трудности в каждом математическом эксперименте составляли определение величины окна прогноза, количества нейронов во входном и скрытом слоях, функции активации на выходном слое. Выбор оптимальной модели осуществлялся на основе показателя отношение стандартных отклонений, которое представляет собой отношение стандартного отклонения ошибки прогноза к стандартному отношению исходных данных. Модель оценивали как удачную, если отношение стандартного отклонения приближалось к нулю [2].

Критерием успешного обучения являлось последовательное уменьшение ошибки на обучающем множестве. Критерием остановки процесса обучения служил рост ошибки на контрольном множестве при продолжающемся её уменьшении или остановке на обучающем множестве. Это свидетельствовало о «переобучении» сети, то есть сеть слишком близко аппроксимировала выборку, в результате чего снижалось качество прогноза при подаче на сеть новых данных [2, 12].

На третьем этапе осуществлялось тестирование модели, что достигалось путем сравнения прогнозируемых значений с набором известных данных, которые ранее на сеть не подавались. Верификация прогноза проводилась путем анализа качества прогноза с использованием абсолютных и сравнительных показателей точности прогноза (абсолютная $\Delta_{пр}$ и относительная $\bar{\epsilon}_{пр}$ ошибки, среднеквадратическая ошибка σ_t и коэффициента несоответствия на уровне среднего KN_t) [4].

В качестве обучающих данных приняты значения первичной заболеваемости сотрудников ФПС МЧС России за период с 2003 по 2007 г. по месяцам. На вход нейронной сети подавались 60 наблюдений, из них при постоянном перемешивании выделялись по 15% наблюдений для контроля и тестирования. С помощью мастера решений был запущен процесс «обучения». Наилучший результат из всех типов нейронной сети показал многослойный персептрон. Обучение его проводили методом сопряжённых градиентов и обратного распространения ошибки. При этом оптимальный результат был получен при совместном их применении. Опытным путем было найдено количество месяцев, на основе которых строился прогноз на один месяц.

Оценка эффективности применения разработанных моделей проведена путем сравнения качества прогнозов, построенных с помощью искусственных нейронных сетей и линейного регрессионного анализа [7, 12]. Математико-статистическая обработка полученных результатов проводилась в программе MS Excel-2007.

Результаты и их обсуждение. Показатели уровня первичной заболеваемости сотрудников ФПС МЧС России по ведущим классам болезней за период наблюдения представлены в таблице 1. Динамика уровня первичной заболеваемости

Таблица 1

Уровень первичной заболеваемости сотрудников ФПС МЧС России, %о

Показатель	Год						Средние данные
	2003	2004	2005	2006	2007	2008	
Первичная заболеваемость	624	475***	494	409***	409	345***	459±39,56
Болезни органов дыхания	352	288***	202***	216	249***	202***	251±20,47
Травматизм	140	87***	85	84	39***	34	78±15,71

Примечание: * – p<0,05; ** – p<0,01; *** – p<0,001 по сравнению с предыдущим годом.

характеризовалась достоверным снижением, а в структуре первичной заболеваемости преобладали болезни органов дыхания (57%), травмы, отравления и некоторые другие последствия воздействия внешних причин (16%).

Установлено, что для прогноза первичной заболеваемости значение на месяц эффективнее прогнозировать по данным за год. В течение одной попытки компьютер на основе заданных параметров создавал и анализировал 1000 моделей, из которых только 5 сохранялись как наиболее эффективные. Таким образом, проанализировано более 100 тысяч пробных моделей только для создания одной модели прогнозирования первичной заболеваемости. Кроме того, было определено, что та скорость обучения, которая заложена в программу по умолчанию, не позволяет эффективно «обучить сеть», для чего её уменьшили до 0,001. К улучшенному результату так же привело и изменение на выходе значения максимального стандартного отклонения и минимального среднего, а оптимальный результат был достигнут при значении, соответственно, 0,75 и 0,2.

В ходе математического эксперимента определили наилучшую модель обученной нейросети для прогноза первичной заболеваемости сотрудников, архитектура которой представлена на рисунке 1.

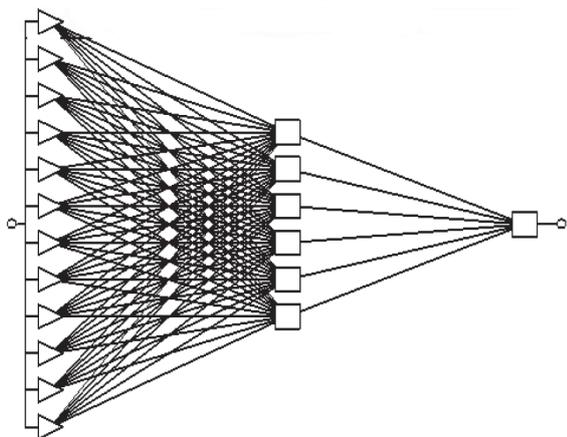


Рис. 1. Архитектура нейронной сети № 89

Структура наилучшей модели – это трехслойный персептрон, который содержит в выходном слое – 12 нейронов, в среднем – 6, во входном слое – 1. Контрольная и тестовая ошибки равны 0,08 и 0,05, отношение стандартных отклонений – 0,25.

Прогноз первичной заболеваемости первоначально провели на известные данные с 49-го наблюдения (после 4 лет с начала сбора данных) на 12 шагов, что соответствовало одному году (2007 г.). В последующем осуществили прогноз на область неизвестных сети данных, т.е. прогноз на двенадцать месяцев 2008 г. Для этого более целесообразным было начинать проекцию временного ряда с 61-го наблюдения на 12 шагов. На рисунке 2 представлены прогнозные и реальные значения уровня первичной заболеваемо-

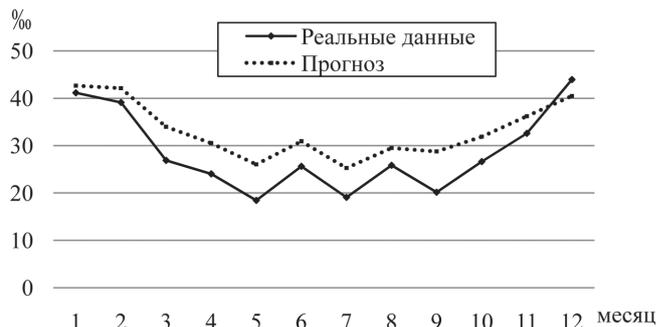


Рис. 2. Динамика первичной заболеваемости сотрудников ФПС МЧС России и нейросетевого прогноза на 2008 г.

сти. Описание прогноза первичной заболеваемости сотрудников ФПС МЧС России математически подтверждает распределение фактических значений по времени, однако прогноз предусматривал более высокие значения уровня первичной заболеваемости.

Прогнозная модель заболеваемости болезнями органов дыхания так же была создана с помощью трёхслойного персептрона. Низкие ошибки на обучающем (0,03), контрольном (0,05) и тестовом (0,07) множествах свидетельствовали о высоком качестве модели. Отношение стандартных отклонений равнялось 0,24 (сеть предсказывала ряд на 75%). «Обучение» сети проводилось методами обратного распространения и сопряженных градиентов.

Прогнозные значения первичной заболеваемости сотрудников болезнями органов дыхания в 2008 г., полученные с использованием нейросетевой модели № 22 (рис. 3), практически полностью совпадали с фактическими. На наш взгляд, это обусловлено тем, что в структуре болезней органов дыхания (X класс) 80% приходилось на долю острых респираторных инфекций и ангины, а 2% – на пневмонии. Преимущественно инфекционный характер заболеваний данного класса обеспечил положительный результат эксперимента. Это объясняет и наличие постоянной сезонной составляющей в данном временном ряду. В наиболее напряженные по эпидемической обстановке месяцы прогноз предусматривал более высокие значения (до 45%) первичной заболеваемости болезнями органов дыхания.

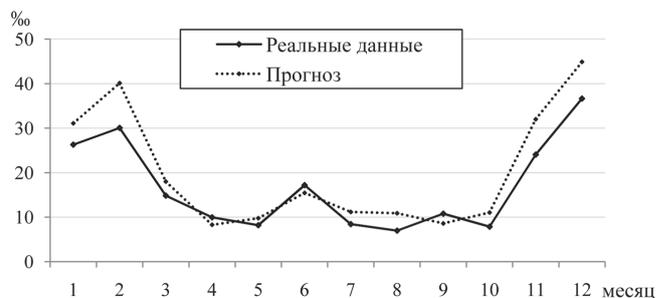


Рис. 3. Динамика первичной заболеваемости болезнями органов дыхания сотрудников ФПС МЧС России и нейросетевого прогноза на 2008 г.

Недостатком нейросетевой модели прогнозирования заболеваемости является отсутствие возможности наблюдения за скрытыми закономерностями в исследуемом временном ряду. Методика показывает свою эффективность в условиях стабильности существующей и сохранённой тенденций. Именно с этим связан факт недостаточно удачного эксперимента по созданию модели для прогнозирования травматизма сотрудников ФПС МЧС России. Тем не менее, вопрос о применении нейросетевой модели для прогнозирования травматизма целесообразно рассмотреть после дальнейшего исследования на больших массивах данных и временных промежутках.

Прогноз первичной заболеваемости сотрудников ФПС МЧС России на 2008 г. с использованием нейросетевой модели показал хорошее качество: средняя относительная ошибка 16,3%, $KN_1 < 1$. Прогноз первичной заболеваемости анализируемого контингента с помощью линейного регрессионного анализа имел удовлетворительное качество по средней относительной ошибке (27,9%). Однако, основываясь на значении KN_1 , который оказался выше 1, качество прогноза нельзя считать удовлетворительным. Прогноз первичной заболеваемости изучаемого контингента болезнями органов дыхания на 2008 г., выполненный с использованием нейросетевой модели, показал хорошее качество (средняя относительная ошибка 19,2%, $KN_1 < 1$), тогда как прогноз, полученный методом линейного регрессионного анализа, по средней относительной ошибке оказался неудовлетворительным (68,2%), $KN_1 > 1$ (табл. 2).

Таким образом, прогноз первичной заболеваемости пожарных, полученный с использованием искусственных нейронных сетей, показал хорошее качество, а в сравнении с качеством линейного регрессионного анализа установлено превосходство изучаемой методики. Проведенное исследование позволяет рекомендовать применение искусственных нейронных сетей для прогнозирования заболеваемости в организованных коллективах в качестве альтернативного подхода.

Выводы

1. Искусственные нейронные сети позволяют создать статистически достоверные модели для прогнозирования первичной заболеваемости.

2. Применение методики прогнозирования заболеваемости на основе искусственных нейронных сетей в организованных коллективах позволяет получить качественный прогноз, не уступающий классическим линейным регрессионным моделям.

3. Методику прогнозирования заболеваемости на основе искусственных нейронных сетей целесообразно рекомендовать для прогнозирования заболеваемости в организованных коллективах, в которых законодательством предусмотрена военная или приравненная к ней служба.

Литература

1. Андрияшкина, Е.В. Интеллектуализация диагностики критических состояний больных с бронхо-легочной недостаточностью на основе нейросетевого моделирования: дис. ... канд. тех. наук / Е.В. Андрияшкина – Воронеж, 2006. – 182 с.
2. Боровиков, В.П. Нейронные сети. Statistika Neural Networks: методология и технологии современного анализа данных / под ред. В.П. Боровикова. – М.: Горячая линия: Телеком, 2008. – 392 с.
3. Булка, А.П. Моделирование медицинских экспертных систем на основе нечёткой нейронной сети / А.П. Булка. – СПб.: ВМА, 2008. – 72 с.
4. Гамбаров, Г.М. Статистическое моделирование и прогнозирование: учеб. пособие / Г.М. Гамбаров [и др.]. – М.: Финансы и статистика, 1990. – 383 с.
5. Жидкова, О.И. Медицинская статистика / О.И. Жидкова. – М.: ЭКСМО, 2007. – 21 с.
6. Мазуров, В.И. Организация мониторинга состояния здоровья населения и военнослужащих в Соединенных Штатах Америки / В.И. Мазуров [и др.] // Вестн. Росс. воен.-мед. акад. – 2013. – № 3 (43). – С. 219–223.
7. Мельниченко, П.И. Ретроспективный эпидемиологический анализ и прогнозирование заболеваемости личного состава Вооруженных сил Российской Федерации: метод. указания / П.И. Мельниченко [и др.]. – М.: Воениздат, 2005. – 143 с.
8. Резванцев, М.В. Подходы к обоснованию задач Медицинского регистра военнослужащих, получивших боевые поражения / М.В. Резванцев [и др.] // Вестн. Росс. воен.-мед. акад. – 2013. – № 1 (41). – С. 165–170.
9. Фомин, А.М. Искусственные нейронные сети в решении задач анализа временного ряда / А.М. Фомин [и др.] // Нейроинформатика-2010: XII всерос. науч.-практ. конф. – М.: НИЯУ МИФИ, 2010. – С. 19–20.
10. Цыган, В.Н. Патофизиологическое обоснование путей модернизации качественного состава инфузий в остром периоде ожоговой болезни / В.Н. Цыган [и др.] // Вестн. Росс. воен.-мед. акад. – 2011. – № 4 (36). – С. 155–159.
11. Юсупов, А.Н. Применение искусственных нейронных сетей в оценке риска развития клинических проявлений атеросклероза у работников, занятых во вредных условиях труда:

Таблица 2

Показатели качества прогнозов, полученных с помощью нейросетевых моделей и линейного прогноза

Заболеваемость	Методика прогноза	Период прогноза	KN_1	$\bar{\Delta}пр, (\%)$	$\sigma_t, (\%)$	$\bar{\varepsilon}_{пр}, (\%)$
Первичная заболеваемость	НСМ	12	0,59	4,23	4,91	16,26
	ЛРА		1,07	7,47	8,98	27,91
Болезни органов дыхания	НСМ	12	0,42	3,35	4,31	19,28
	ЛРА		1,07	9,36	10,98	68,92

Примечание: НСМ – нейросетевые модели, ЛРА – линейный регрессионный анализ; $KN_1 < 1$ и $10 < \bar{\varepsilon}_{пр} < 20$ – хорошее качество прогноза; $20 < \bar{\varepsilon}_{пр} < 50$ – удовлетворительное, $\bar{\varepsilon}_{пр} > 50$ – неудовлетворительное [3].

автореф. дис. ... канд. биол. наук / А.Н. Юсупов – СПб.: ВМА, 2012. – 33 с.

12. Юнкеров, В.И. Математико-статистическая обработка данных медицинских исследований: 3-е изд., доп. / В.И. Юнкеров, С.Г. Григорьев, М.В. Резванцев. – СПб.: ВМА, 2011. – 318 с.

V.Yu. Golovinova, S.G. Kireyev, P.K. Kotenko, Yu.L. Minaev, I.N. Shtamburg, S.G. Kuzmin

Neural network models for prediction of morbidity in organized personnel groups

Abstract. *The results of artificial neural network model employment for prediction of morbidity in organized personnel groups through the example of Ministry of the Russian Federation for civil defense, emergencies and elimination of consequences of natural disasters Federal fire control service North-West regional center servicemen have been presented. There have been developed neural network models of primary morbidity rate prediction for all chapters of diseases. The best prediction result has been demonstrated for morbidity rate of respiratory diseases. That can be explained by predominantly infectious character of this chapter. The method shows its effectiveness in stable conditions and preserved tendencies. Artificial neural networks demonstrate their high effectiveness in developing statistically valid models for primary morbidity rate prediction. The use of the neural network model allows obtaining a quality prediction not worse than using the classic linear regression models. The artificial neural networks method is recommended for morbidity rate prediction in organized personnel groups in which the military service or equal with it is provided by the legislation.*

Key words: *primary morbidity, neural networks, neural network prediction, neural network models, Federal fire control service, morbidity prediction, artificial intelligence.*

Контактный телефон: 8-911-133-45-78; e-mail: nikagolova@yandex.ru