

УДК 004.8 ГРНТИ 28.23

К.И. Шахгельдян, д.т.н., директор Института информационных технологий Федерального государственного бюджетного образовательного учреждения высшего образования «Владивостокский государственный университет экономики и сервиса» Министерства науки и высшего образования Российской Федерации, заведующая лабораторией анализа больших данных в биомедицине и здравоохранении Федерального государственного автономного образовательного учреждения высшего образования «Дальневосточный федеральный университет» Министерства науки и высшего образования Российской Федерации; тел.: +7-924-231-44-91; e-mail: carina.shahgeldyan@vvsu.ru

Б.И. Гельцер, д.м.н., профессор, член-корр. РАН, директор департамента клинической медицины Школы биомедицины Федерального государственного автономного образовательного учреждения высшего образования «Дальневосточный федеральный университет» Министерства науки и высшего образования Российской Федерации; тел.: (423) 245-17-83; e-mail: Boris.Geltser@vvsu.ru

И.А. Курпатов, аспирант департамента клинической медицины Школы биомедицины Федерального государственного автономного образовательного учреждения высшего образования «Дальневосточный федеральный университет» Министерства науки и высшего образования Российской Федерации; тел.: (423) 245-17-83; e-mail: kurpatov-i@mail.ru

Методы машинного обучения для дифференциальной диагностики болезней органов дыхания

10

Аннотация

Современные технологии обработки больших данных позволяют ученым проводить масштабные исследования в различных областях клинической медицины, используя методы машинного обучения. Наше исследование направлено на оценку возможностей их применения в области дифференциальной диагностики и прогнозирования заболеваний органов дыхания. Исследование включало сбор и обработку данных о пациентах с ХОБЛ, бронхиальной астмой и АСОС-синдромом (357 человек), а также здоровых лиц (256 человек). На основе регрессионных моделей и искусственных нейронных сетей (ИНС) построены модели оценки должных величин силы дыхательных мышц: индикаторы максимального экспираторного и инспираторного давления (MEP и MIP), а также индикатор интраназального давления (SNIP). Полученные с помощью ИНС модели существенно превосходят существующие аналоги. Для дифференциальной диагностики ХОБЛ/Астмы/АСОС-синдрома и оценки тяжести ХОБЛ были использованы ИНС и индикаторы силы дыхательных мышц, что позволило построить модели, имитирующие диагностику экспертов-пульмонологов, а также доказать информативность индикаторов силы дыхательных мышц (MEP, MIP и SNIP) при дифференциальной диагностике болезней органов дыхания.

Ключевые слова:

болезни органов дыхания, сила дыхательных мышц, ХОБЛ, искусственные нейронные сети, методы машинного обучения

В последние годы в медицинских организациях идет активный процесс накопления результатов клинических исследований в цифровой форме, содержащих разнородную информацию о пациентах, характеризующую динамику развития заболеваний и эффективность проводимой терапии. Современные технологии обработки больших данных позволяют ученым проводить масштабные исследования в различных областях клинической медицины, используя не только статистические методы, но и современные методы машинного обучения и искусственного интеллекта [1-5]. Клиническая медицина - одна из тех отраслей знаний, где потенциал применения информационных технологий раскрыт еще недостаточно глубоко. Несмотря на впечатляющие достижения в области машинного обучения и искусственного интеллекта, внедрение последних в практическую медицину ограничено. Развитие систем поддержки принятия врачебных решений требует использования новых знаний, методов и алгоритмов машинного обучения, которые необходимо интегрировать в современные интеллектуальные системы. Наше исследование направлено на оценку возможностей применения методов машинного обучения в области дифференциальной диагностики и прогнозирования заболеваний органов дыхания.

Хроническая обструктивная болезнь легких (ХОБЛ) и бронхиальная астма (БА) относятся к актуальным проблемам современной медицины, что обусловлено их высокой распространенностью и большими экономическими потерями для государства и общества. Среди заболеваний респираторной системы их доля составляет 15–25%. В 2014 году был впервые выделен АСОС-синдром (asthma-COPD overlap syndrome – синдром «перекреста» БА-ХОБЛ), который характеризуется сочетанием у пациентов клинико-функциональных признаков обеих нозологий. Дыхательные мышцы (ДМ) являются ключевым элементом респираторной системы, обеспечивающим в качестве «дыхательной помпы» альвеолярную вентиляцию. Силу ДМ относят к важнейшим индикаторам их функционального состояния, а также к независимым предикторам развития дыхательной недостаточности и сердечно-сосудистых осложнений. Определение силы ДМ чаще всего осуществляют путем измерения статических давлений при «закрытых» дыхательных путях во время максимального вдоха и выдоха: максимального инспираторного (MIP – Maximum Inspiratory Pressure), экспираторного (MEP – Maximum Expiratory Pressure) и интраназального (SNIP – Sniff Nasal Inspiratory Pressure) давлений. Ранняя диагностика дисфункции ДМ должна базироваться на точных представлениях о нормативных значениях их силовых

характеристик. Первая задача исследования состояла в разработке и сравнительной оценке моделей должных величин для показателей MIP, MEP, SNIP на основе методов машинного обучения. Вторая задача состояла в определении информативности индикаторов силы ДМ для верификации ХОБЛ/БА/АСОС-синдром. Третья задача состояла в анализе информативности показателей силы ДМ для оценки тяжести ХОБЛ на основе методов машинного обучения.

Дизайн исследования состоял в следующих шагах:

1. Сбор данных о здоровых лицах (256 человек), их антропометрических показателях и силе ДМ (MEP, MIP, SNIP).
2. Сбор данных о пациентах с ХОБЛ, БА и АСОС-синдромом (357 человек), а также их антропометрических данных, индикаторов силы ДМ, и других показателей: объема форсированного выдоха (ОФВ1), форсированной жизненной емкости легких (ФЖЕЛ), парциального давления углекислого газа (PaCO₂) и кислорода (PaO₂) в крови, фибриногена, шкалы выраженности одышки (mMRC) и САТ-тест.
3. Отбор и оценка ранее опубликованных другими авторами моделей силы ДМ на нашем датасете.
4. Разработка моделей силы ДМ, используя линейные регрессии и искусственные нейронные сети (ИНС), их сравнения с существующими аналогами с помощью оценки точности моделей.
5. Разработка моделей дифференциальной диагностики ХОБЛ/Астмы/АСОС-синдрома с - и без учета силы ДМ, сравнения точности моделей с целью оценки информативности индикаторов силы ДМ.
6. Построение моделей с помощью ИНС для оценки тяжести ХОБЛ, без использования индикаторов силы ДМ и с их учетом. Сравнение точности дифференциальной диагностики ХОБЛ с целью оценки информативности индикаторов силы ДМ.

Точность непрерывных моделей оценивалась с помощью средней относительной ошибки аппроксимации (СОАА) на обучающей и тестовой выборках, как 1-СОАА. Дополнительно использовали множественный коэффициент детерминации R². Точность дискретных моделей (классификация) определялась критерием Акаике, площадью под ROC-кривой и отношением корректно отнесенных объектов к общему числу объектов при критерии разделения=0,5. Все оценки точности выполняли на тестовой выборке.

Для персонифицированной оценки индикаторов

силы ДМ нами были отобраны несколько наиболее известных линейных регрессионных моделей, оценивающих индикаторы силы ДМ через возраст и пол [6-10]. Точность этих моделей на датасете здоровых лиц варьировала от 50 до 75%. Нами были построены линейные регрессионные модели, использующие помимо возраста и пола, некоторые антропометрические показатели. Так, для мужчин наиболее значимым предиктором была окружность предплечья, а для женщин - бедра. Точность этих моделей была несколько выше и варьировала от 67% до 84%. Ввиду проблемы мультиколлинеарности линейных регрессионных моделей, использовать одновременно несколько антропометрических показателей было некорректно. Мы построили и обучили ИНС для моделирования индикаторов силы ДМ, используя многослойную архитектуру (4-6 невидимых слоя) и большую часть антропометрических показателей, в том числе рост, вес, массу скелетной мускулатуры, окружности плеча, предплечья, груди, бедра, голени. Точность моделей на основе ИНС варьировала от 83% до 90%. Использование нескольких антропометрических показателей и нелинейность структуры сети позволило построить модели значительно превышающие в точности традиционные регрессионные модели индикаторов силы ДМ.

Для стратификации тяжести ХОБЛ нами были разработаны ИНС с использованием различных факторов: ОФВ1, ФЖЕЛ и их отношение, mMRC, САТ, возраст, антропометрические данные, а также газометрические параметры (SaO_2 , PaO_2 , $PaCO_2$), показатели системного воспаления (лейкоцитарные индексы и фибриноген) и индикаторы силы ДМ (MEP, MIP, SNIP). Наиболее значимыми являлись: ОФВ1, mMRC, ОМТ, $PaCO_2$, фибриноген и MIP. При этом в качестве базового предиктора во всех моделях использовался ОФВ1. Информативность MIP как предиктора стратификации ХОБЛ определяли путем сравнения точности моделей, включающих и не включающих данный индикатор. Архитектура разработанных ИНС была одноуровневой с числом нейронов от 2 до 19.

Полученные данные указывают на то, что при использовании для «машинной» стратификации тяжести ХОБЛ только показателя ОФВ1 точность результатов моделирования достигает 70%. Дополнение ИНС такими факторами, как ОМТ, mMRC, $PaCO_2$ и фибриноген улучшает качество верификации, а включение в разрабатываемые модели MIP во всех случаях повышало их точность на 10% -20%. При этом комбинация ОФВ1, относительной массы тела, $PaCO_2$ и MIP полностью исключала ошибку стратификации.

Для оценки информативности показателей

силы ДМ в качестве дополнительных диагностических маркеров, идентифицирующих группы здоровых и больных, а также различные клинические варианты бронхиальной обструкции, были построены логистические регрессионные модели. Предварительное тестирование прогностических свойств их потенциальных предикторов показало, что лучшими из них являются отношения ОФВ1/ФЖЕЛ и MEP/MIP. Сравнительный анализ точности стратификации показал, что качество «слепого» разделения обследованных на группы здоровых и больных существенно возрастает после включения в структуру модели с «классическим» предиктором ОФВ1/ФЖЕЛ отношения MEP/MIP. Критерий Акаике в этом случае снижается с 60,5 до 45,5 при возрастании площади под ROC-кривой с 0,98 до 0,99. Дифференциация БА и ХОБЛ была выполнена с помощью логистических регрессий, использующих отдельно ОФВ1/ФЖЕЛ и MEP/MIP. Точность моделей характеризуется критерием Акаике: 93,5 и 77,4 соответственно, а также площадью под ROC-кривой 0,78 и 0,84. Таким образом, более высокой точностью отличалась и модель, в которой использовался показатель MEP/MIP. Последний показатель обладает способностью разделять больных ХОБЛ и АСОС-синдромом, что невозможно сделать с помощью ОФВ1/ФЖЕЛ.

Внедрение в медицинскую практику разработанных моделей позволит расширить знания об особенностях клинического течения различных заболеваний, моделировать физиологические функции в условиях нормы и патологии, что было доказано участниками проекта на примере исследования силовых характеристик дыхательных мышц. Практическая значимость результатов исследования заключается также в совершенствовании алгоритмов дифференциальной диагностики различных клинических вариантов бронхиальной обструкции на основе технологий машинного обучения. Разработанные модели верификации с использованием в качестве новых предикторов силовых индикаторов ДМ существенно повышают точность диагностики этих состояний, а их внедрение в клиническую практику будет являться дополнительным инструментом, поддерживающим принятие обоснованных врачебных решений при оказании первичной и специализированной медицинской помощи. Необходимо отметить, что при расширении базы клинических данных модели могут совершенствоваться за счет корректировки весовых коэффициентов. Исследование выполнено при финансовой поддержке РФФИ в рамках научных проектов 18-29-03131.

СПИСОК ЛИТЕРАТУРЫ:

1. Zolfaghar K., Meadem N., Teredesai A., Roy S.B., Chin S.C. Big Data Solutions for Predicting Risk-of-Readmission for Congestive Heart Failure // Proc. IEEE International Conference on Big Data, USA. 2013.
2. Wang Y., Simon M.A, Bonde P., Harris B.U., Teuteberg J.J., Kormos R.L., and Antaki J.F. Decision tree for adjuvant right ventricular support in patients receiving a left ventricular assist device. // Journal of Heart and Lung Transplantation. 2012. №31.
3. Goto S., Kimura M., Katsumata Y., Goto S., Kamatani T., Susaki J., Ichihara G., Ko Fukuda K., Sano M. Artificial intelligence to predict needs for urgent revascularization from 12-leads electrocardiography in emergency patients // PLOS. 2019. 14(1).
4. Shakhgelyan K., Geltser B., Gmar D. Using machine learning to assess efficiency of tuberculosis service. Proceedings of 2017 Second Russian-Pacific Conference on Computer Technology and Applications (RPC'2017). DOI: 10.1109/RPC.2017.8168091
5. Гельцер Б.И., Шахгельдян К.И., Гмарь Д.В., Аршинов А.Ю., Медведев В.И., Транковская Л.В. Доступность фтизиатрической помощи в Приморском крае. Проблемы социальной гигиены, здравоохранения и истории медицины. 2017. Т. 25. № 4. С. 229-233.
6. Black L.F., Hyatt R.E. Maximal respiratory pressures: normal values and relationship to age and sex. American Review of Respiratory Disease. 99: 696-702. 1969.
7. Evans J.A., Whitelaw W.A. The assessment of maximal respiratory mouth pressures in adults. Respiratory care. 54 (10): 1348-1359. 2009
8. Gopalakrishna A., Vaishali K., Prem V., Aaron P. Normative values for maximal respiratory pressures in an Indian Mangalore population: A cross-sectional pilot study. Lung India. 28(4): 247-252. 2011.
9. Uldrey Ch., Fitting J.W. C in healthy subjects. Thorax. 50 (4): 371-375. 1995.
10. Wilson S.H., Cooke N.T., Edwards R.H., Spiro S.G. Predicted normal values for maximal respiratory pressures in caucasian adults and children. Thorax. 39 (7): 535-538. 1984.

UDC 004.8

13

B.I.Geltser¹, K.J.Shakhgelyan^{1,2}, I.G.Kurpatov¹¹Far Eastern Federal University, 690091, Vladivostok, Russian Federation;²Vladivostok State University Economics and Service, 690014, Vladivostok, Russian Federation**Authors information**

Boris I. Geltser – MD, professor, corresponding member of the Russian Academy of Sciences, Far Eastern Federal University; Head of the Clinical medicine department of Biomedicine School; tel.: (423) 245-17-83; e-mail: Boris.Geltser@vvsu.ru (<http://orcid.org/0000-0002-9250-557X>)

Karina I. Shakhgelyan – PhD, professor, Head of the Information technology Institute of the Vladivostok State University Economics and Service, Head of the Laboratory of Big Data analysis in biomedicine and healthcare Far Eastern Federal University; tel.: +7-924-231-44-91; e-mail: carina.shahgelyan@vvsu.ru (<http://orcid.org/0000-0002-4539-685X>)

Ilya G. Kurpatov – Far Eastern Federal University. Post-graduate student of the Clinical medicine department of Biomedicine School; tel.: (423) 245-17-83; e-mail: kurpatov-i@mail.ru (<http://orcid.org/0000-0002-4031-2979>)

Experience of using an improved model of the trauma registry for multiple trauma monitoring

Abstract. Modern technologies for data analysis allow scientists to conduct large-scale research in various fields of clinical medicine, using methods of machine learning. Our study is dedicated to assessing the possibilities of their application in the field of differential diagnosis and prediction of respiratory diseases. The study included the collection and processing of data on patients with COPD, asthma and ACOS syndrome (357 people), as well as healthy individuals (256 people). We built models for evaluating the reference values of respiratory muscle strength based on the regression and artificial neural networks (ANNs). These models included indicators of maximum expiratory and inspiratory pressure (MEP and MIP), as well as an indicator of intranasal pressure (SNIP). The models obtained using ANN significantly exceed existing analogues. We developed models for differential diagnosis of COPD / Asthma / ACOS syndrome and assessment of the severity of COPD. These models based on ANNs and respiratory muscle strength indicators allowed us to simulate the diagnosis of expert pulmonologists, as well as to prove significance of respiratory muscle strength indicators (MEP, MIP and SNIP).

Keywords: respiratory diseases, respiratory muscle strength, COPD, artificial neural networks, machine learning methods