

УДК 004.89

ПОДГОТОВКА ДАННЫХ ДЛЯ РЕШЕНИЯ ЗАДАЧИ КЛАССИФИКАЦИИ СЕРДЕЧНЫХ ЗАБОЛЕВАНИЙ МЕТОДАМИ МАШИННОГО ОБУЧЕНИЯ

Сергиенко И.Р. (Университет ИТМО)

Научный руководитель – преподаватель, кандидат технических наук, Русак А.В.

(Университет ИТМО)

Введение. Автоматизация диагностики заболеваний человека на данный момент, как никогда, актуальна и активно развивается [1, 2]. Популярными инструментами автоматизации являются методы машинного обучения для решений задач классификации. Особенность таких методов заключается в автоматизации нахождения вероятностных зависимостей в предоставляемых данных. Для того чтобы искомая вероятностная модель была адекватна и достаточно приближена к реальности, необходимо использовать значительный объем заранее обработанных или сырых данных. На данный момент существует множество баз данных с обработанными и аннотированными записями ЭКГ сердечных заболеваний в открытом доступе. Исследования показывают, что популярные методы машинного обучения, включая модели искусственных нейронных сетей, повышают качество прогнозирования при предварительной дополнительной обработке баз данных [3]. При этом в сфере здравоохранения такой обработке перед использованием методов машинного требуется быть не только точной, но и доступной и быстрой.

Основная часть. В работе решается задача обработки данных ЭКГ для использования методов машинного обучения классификации сердечных заболеваний. Предоставляются результаты обзорного анализа следующих баз данных:

- AF Termination Challenge Database, [4]
- Combined measurement of ECG, Breathing and Seismocardiograms, [5]
- CU Ventricular Tachyarrhythmia Database, [6]
- BIDMC Congestive Heart Failure Database, [7]
- PTB Diagnostic ECG Database, [8]
- St Petersburg INCART 12-lead Arrhythmia Database. [9]

Результаты анализа позволяют решать задачу разработки алгоритма для обработки баз как размеченных, так и сырых данных ЭКГ. Предлагается использовать популярные методы выделения признаков по QRS-комплексам ЭКГ. Туда входят статистические и частотные характеристики R-пиков и RR-интервалов. Использование R-пиков в QRS-комплексах обосновано возможностью автоматизации их выявления на не размеченных ЭКГ данных. Для этого существует множество различных алгоритмов [10]. Использование частотных и статистических характеристик RR-интервалов также встречается в работах по прогнозированию и иных, помимо сердечных, заболеваний человека [11].

Выводы. Проанализированы открытые базы данных ЭКГ. Разработан алгоритм обработки как размеченных, так и сырых данных ЭКГ. На основе открытых баз данных ЭКГ подготовлена и проанализирована единая база признаков для использования в решении задач классификации сердечных заболеваний.

Список использованных источников:

1. Aggarwal, M. IBM'S Watson Analytics for Health Care: A Miracle made true / M. Aggarwal, M. Madhukar. - IGI Global, 2017 – Режим доступа: <https://www.igi-global.com/gateway/chapter/164580>
2. Seshadri, D. Accuracy of Apple Watch for Detection of Atrial Fibrillation / D. Seshadri, B. Bittel, D. Browsky, P. Houghtaling, C. K. Drummond, M. Y. Desai, A. M. Gillinov. –

Circulation, 2020 – Режим доступа:

<https://www.ahajournals.org/doi/10.1161/CIRCULATIONAHA.119.044126>

3. Hong S. et al. Opportunities and Challenges in Deep Learning Methods on Electrocardiogram Data: A Systematic Review / S. Hong – статья - arXiv preprint arXiv:2001.01550. – 2019.
4. Moody, G.B. Spontaneous Termination of Atrial Fibrillation: A Challenge from PhysioNet and Computers in Cardiology 2004. / G.B. Moody. – статья - IEEE, 2004 –с. 101 – 104.
5. García-González, M. A. A Comparison of Heartbeat Detectors for the Seismocardiogram. / M. A. García-González, A. Argelagos-Palua, M. Fernandez-Chimeno, J. Ramos-Castro. – статья - Computing in Cardiology, 2013 – с. 461-464 – ISSN: 2325-8861.
6. Nolle, F.M. CREI-GARD, a new concept in computerized arrhythmia monitoring systems. / F. M. Nolle, F. K. Badura, J. M. Catlett, R. W. Bowser, M. H. Sketch. - статья - Computers in Cardiology, 1986 – с. 515-518.
7. Baim, D. S. Survival of patients with severe congestive heart failure treated with oral milrinone. / D. S. Baim, W. S. Colucci, E. S. Monrad, H. S. Smith, R. F. Wright, A. Lanoue, D. F. Gauthier, B. J. Ransil, W. Grossman, E. Braunwald. – статья - American College of Cardiology Journal, 1986 – с. 661-670.
8. Boussejot, R. Nutzung der EKG-Signaldatenbank CARDIODAT der PTB über das Internet. / R. Boussejot, D. Kreiseler, A. Schnabel. – статья - Biomedizinische Technik, 1995 – с. 317.
9. Yakushenko, E. St Petersburg INCART 12-lead Arrhythmia Database. / E. Yakushenko. - PhysioNet, 2008 - [Электронный ресурс]. – Режим доступа: <https://doi.org/10.13026/C2V88N>.
10. Рослякова А. В. Сравнительный анализ алгоритмов обнаружения R-зубца электрокардиосигнала / А. В. Росляков, П. Г. Чупраков - Вятский медицинский вестник, 2012. №2. URL: <https://cyberleninka.ru/article/n/sravnitelnyy-analiz-algoritmov-obnaruzheniya-r-zubtsa-elektrokardiosignala>.
11. Успенский, В. М. Информационная функция сердца : теория и практика диагностики заболеваний внутренних органов методом информационного анализа электрокардиосигналов / В. М. Успенский – Москва: Экономика и информатика, 2008. – 116 с. – ISBN 5893450755.