

АНОТАЦІЯ

Яворський А. Б. Методи та алгоритми аналізу біосигналів людини на основі машинного навчання. – Кваліфікаційна наукова праця на правах рукопису.

Дисертація на здобуття наукового ступеня доктора філософії за спеціальністю 121 «Інженерія програмного забезпечення» (12 – Інформаційні технології). – Київський національний університет імені Тараса Шевченка. – Київський національний університет імені Тараса Шевченка, Київ, 2022.

Метою роботи є підвищення якості аналізу електрокардіограм (ЕКГ) шляхом розробки методів та алгоритмів більш точного визначення відхилень у попередньо оброблених біосигналах людини з метою мінімізації загроз здоров'ю та вчасного впливу для покращення якості життя. Основою розроблених методів є підходи на базі машинного навчання, які було апробовано на різноманітних доступних великих вибірках даних, що довело високу ефективність запропонованих методів.

Життя і здоров'я людини є пріоритетом конституцій багатьох держав, і загалом задачею багатьох галузей. Разом з тим, за статистикою Всесвітньої організації охорони здоров'я більше чверті смертей спричинено серцево-судинними захворюваннями. При цьому, вчасна діагностика та втручання фахівців могли б значно покращити ситуацію. Отже, більш точний аналіз та неперервний моніторинг у групах ризику допоміг би покращити якість життя людей, особливо з групи ризику щодо серцево-судинних захворювань. Таким чином, більше мобільних пристроїв у користуванні, які дозволяють у реальному часі проводити моніторинг ЕКГ, та точніші методи діагностики, з іншого боку, дозволили б досягти поставленої мети і зменшити кількість смертей.

Для досягнення мети дослідження в роботі було: проведено порівняльний аналіз існуючих методів (State of the Art) виявлення відхилень ЕКГ та їх особливостей; визначено основні показники та критерії ефективності, за якими розроблено та оптимізовано побудовані моделі та алгоритми; розроблено нові

ефективні за визначеними критеріями моделі та алгоритми аналізу ЕКГ; досліджено їх показники. При виконанні поставлених завдань використовувались: методи та засади інформаційних технологій, зокрема інженерії програмного забезпечення та комп'ютерних наук; методи обробки інформації, зокрема біосигналів людини; методи та підходи до розробки програмного забезпечення; методи машинного навчання, зокрема штучні нейронні мережі та випадковий ліс дерев рішень. Важливо зазначити, що отримані результати наукового дослідження набули практичного застосування. Розроблені в процесі дослідження моделі, методи та алгоритми показали високу ефективність, про що свідчать відповідні наукові статті за темою дисертації.

Розроблені моделі дозволяють досить точно виявляти серцеву аритмію, а саме фібриляцію передсердь, за рахунок підбору репрезентативних параметрів та характеристик ЕКГ та ефективного навчання випадкового лісу дерев прийняття рішень.

Наукова новизна одержаних результатів полягає в розробці математичних моделей та алгоритмів точного та швидкого виявлення відхилень ЕКГ з метою підвищення ефективності аналізу стану серцево-судинної системи людини та класифікації виявлених відхилень.

Наукова новизна одержаних результатів:

Вперше:

- Розроблено модель визначення відхилень ЕКГ з кращими показниками міри точності F_1 .
- Розроблено інформаційну технологію на основі методів машинного навчання штучних нейронних мереж та випадкового лісу дерев рішень для точного й швидкого виявлення відхилень на ЕКГ.

Удосконалено: розроблені в процесі проведення дисертаційного дослідження та випробувані раніше моделі для розв'язання поставленої задачі.

Основні результати дисертації пройшли апробацію, доповідались на міжнародних конференціях та семінарах. Основні положення, викладені в

дисертації, доведено до рівня програмної реалізації та опубліковано у міжнародних виданнях, отже вони підлягають широкому застосуванню як у проведенні подальших досліджень, так і у практичній діяльності. Прикладне значення отриманих результатів полягає в отриманій апробованій моделі, втіленій у програмній системі, яка може бути застосована для виявлення відхилень у ЕКГ у режимі моніторингу в реальному часі. Теоретична частина та запропоновані моделі та методи впроваджено на факультеті комп'ютерних наук та кібернетики Київського національного університету імені Тараса Шевченка в курсах з аналізу даних «Основи Data Mining» бакалаврської програми «Інформатика» та курсі «Актуальні проблеми Data Mining» магістерської академічної програми «Інформатика». Результати дисертаційної роботи можуть бути використані для подальших теоретичних та практичних досліджень у сфері інформаційних технологій, у навчальному процесі при підготовці курсів із обробки даних, спеціальних курсів з обробки сигналів.

Ключові слова: біомедичні сигнали, моніторинг здоров'я, обробка медичних сигналів, серцева діяльність, запис електричної активності серця, електрокардіограма, ЕКГ, характеристики ЕКГ, аналіз ЕКГ, хвороби серця, вади серця, частота серцевих скорочень, фібриляція передсердь, серцева аритмія, дерево прийняття рішень, випадковий ліс, реабілітація, обробка сигналів, діагностування захворювань, виявлення відхилень, машинне навчання, нейронні мережі, конволюційні нейронні мережі, змагальні нейронні мережі, ризики, моделювання, ранжування, приватність, паралельні обчислення, ефективність обчислень, метадані.

ANNOTATION

Yavorskyi A. B. Methods and Algorithms for Human Biosignals Analysis Using Machine Learning. – Qualifying scientific work on the rights of a manuscript.

Thesis for a scientific degree of Doctor of Philosophy in specialty 121 "Software engineering" (12 – Information technologies). – Taras Shevchenko National University Kyiv. – Taras Shevchenko National University of Kyiv, Kyiv, 2022.

The aim of the work is to improve the quality of electrocardiogram (ECG) analysis by developing methods and algorithms for more accurate determination of deviations in pre-processed human biosignals in order to minimize health threats and to impact the quality-of-life improvement in a timely manner. The basis of the developed methods are approaches based on machine learning, which were tested on various available large datasets, which proved the high efficiency of the proposed methods.

Human life and health are the advantage of constitutions of many states, and in general, the task for many industries and research. However, according to the statistics of the World Health Organization, more than a quarter of deaths are caused by cardiovascular diseases. At the same time, early diagnosis and intervention of professionals could significantly improve the situation. Therefore, more accurate analysis and continuous monitoring in groups at risk would help to improve the human's quality of life, especially those at risk for cardiovascular diseases. Thus, more mobile devices in use, which allowed for real-time ECG monitoring, and development of more accurate diagnostic methods, on the other hand, would allow to achieve the goal and reduce the number of deaths.

In order to achieve the goal of the study, the following was performed: a comparative analysis of existing State-of-the-Art methods for detecting ECG abnormalities and their features; the main performance indicators and criteria are determined, according to which the constructed models and algorithms are developed and optimized; new models and algorithms for ECG analysis that are effective for the defined criteria were developed, and their indicators were studied. When performing

the tasks for the goal achievement, the following were used: methods and principles of information technologies, in particular software engineering and computer science; information processing methods, in particular for human biosignals processing; methods and approaches for software development; machine learning methods, including artificial neural networks and random forest solutions. It is important to note that the obtained results of scientific research have gained practical application. The models, methods and algorithms developed in the research process showed high efficiency, as evidenced by the author's relevant scientific papers on the topic of the thesis.

The scientific novelty of the obtained results is in the development of mathematical models and algorithms for the accurate and rapid detection of ECG abnormalities aimed to increase the efficiency of the analysis of the human cardiovascular system and the classification of detected abnormalities.

The developed models allow fairly accurate detection of cardiac arrhythmia, namely atrial fibrillation, due to the selection of representative parameters and characteristics of the ECG (ECG features), and effective learning of a random forest of decision trees.

Scientific novelty of the obtained results:

For the first time:

- a model for the determination of ECG deviations with the best indicators of the F_1 accuracy measure has been developed,
- information technology based on machine learning methods of artificial neural networks and a random forest of decision trees for accurate and fast deviation detection on the ECG has been developed.

Improved: developed during the dissertation research and tested preliminary models for solving the given problem.

The main results of the dissertations were reported and presented at international conferences and seminars. The main provisions outlined in the thesis have been developed to the level of program implementation and published in international

publications, so they are subject to extensive development both in further research and in practical activities. The applied value of the obtained results lies in the developed and tested model embodied in a software system, that can be applied to analyze the ECG in the real-time monitoring mode. The theoretical part and the proposed models and methods were implemented at the Faculty of Computer Sciences and Cybernetics of Taras Shevchenko National University of Kyiv in the courses on data analysis "Fundamentals of Data Mining" of the bachelor's program "Informatics" and the course "Actual problems of Data Mining" of the master's academic program "Informatics ". The results of the dissertation work can be used for further theoretical and practical research in the field of information technologies, in the educational process when preparing courses on data processing, special courses on signal processing.

Keywords: biomedical signals, health monitoring, medical signal processing, heart activity, heart electrical activity record, electrocardiogram, ECG, ECG features, ECG analysis, heart disease, heart disabilities, heart rate, atrial fibrillation, cardiac arrhythmia, decision tree, random forest, rehabilitation, signal processing, disease diagnostics, deviation detection, machine learning, neural networks, convolution neural networks, adversarial neural networks, risks, modelling, ranking, privacy, parallel computing, calculation efficiency, metadata.