

**УДК 004.891.3**

**В.М. Кіфер, І.В. Миколіук**

Тернопільський національний технічний університет імені Івана Пулюя, Україна

**АВТОМАТИЧНЕ ВИЯВЛЕННЯ АРИТМІЇ СЕРЦЯ НА  
ЕЛЕКТРОКАРДІОГРАМАХ З ДОПОМОГОЮ МАШИННОГО НАВЧАННЯ**

**V.M. Kifer, I.V. Mykoliuk**

**AUTOMATIC HEART FIBRILLATION DETECTION ON  
ELECTROCARDIOGRAMS USING MACHINE LEARNING**

Хвороби серця є однією з основних причин смертності в світі. Наприклад, в США 25% випадків смертності пов'язані з серцевими захворюваннями, з яких приблизно 47% відбуваються за межами лікарні. Що свідчить про те, що в багатьох випадках люди з серцевими недугами невчасно звертаються за допомогою. А в Україні серцеві недуги є на першому місці серед причин смертності.

Аритмії серця - це група порушень діяльності серця, пов'язані з розладом ритмічності, послідовності та сили скорочень серцевого м'яза. Аритмії серця зумовлені порушеннями його властивостей та можуть вказувати на присутність хвороб серця, зокрема, міокардит, міокардіосклероз тощо. Електрокардіограми є одним з основних способів дослідження ритму серця.

Запис ЕКГ є доволі складною процедурою що включає підключення 12 електродів до різних ділянок тіла. Для цього потрібно мати доволі велике обладнання. Проте, в останнє десятиліття було розроблено ряд кишенькових портативних пристроїв для запису ЕКГ, наприклад, пристрій AliveCor, принцип дії якого полягає у зчитуванні серцевої активності з кінчиків вказівного та середнього пальців обох рук.

Записані таким чином ЕКГ мають зазвичай більше шумів, зокрема, через недотримання умов запису а також через сторонні перешкоди, до яких відносять шуми від м'язової активності, шуми від електричних кіл тощо. Для того щоб очистити ЕКГ від шуму застосовується набір фільтрів. Першим є усунення базового рівня коли ми від сигналу віднімаємо середнє значення цього інтервалу. Оскільки ЕКГ серця містить набір частот від 0,5Гц до 40Гц, ми можемо застосувати фільтри низьких та високих частот. Таким чином можна прибрати шуми від скорочення м'язів. Після кожного кроку фільтрації здійснюватимемо нормалізацію сигналу до проміжку [-1;1].

ЕКГ складається з періодичних комплексів пов'язаних з діяльністю серцевих м'язів під час серцебиття. Кожен такий комплекс містить декілька основних частин: Р-хвиля, QRS-комплекс та Т-хвиля. QRS-комплекс складається з різкого перепаду між двома мінімумами в точках Q та S, та піком в точці R, загальною тривалістю 10-12мс та є найбільш стійким до шумів і тому його найлегше виявляти. Для виявлення QRS комплексів використано змінений алгоритм Пан-Томпкінсона. Сигнал, з якого прибрано низькі та високі частоти, піддається додатковій обробці що полягає у застосуванні фільтру похідних. Далі для усунення випадків, коли під час запису ЕКГ було інвертоване внаслідок неправильного розташування конекторів, застосуємо квадратичну функцію. Таким чином також вдається зменшити невеликі значення і збільшити великі (R-піки). Щоб прибрати випадки коли R та S піки значно поляризовані і утворюють дві дуже близькі високі хвилі, ми застосуємо фільтр бігучої середньої, що дозволить об'єднати такі два піки в один. В результаті, отримана функція матиме чіткі піки в позиціях R-піків, що будуть значно більші за середнє значення функції.

На основі виявлених комплексів будуються часові та частотні характеристики варіабельності серцевого ритму. До часових характеристик включено: середня та середнє квадратичне відхилення тривалості RR-інтервалів, квадрат середньої різниці тривалості RR-інтервалів, кількість та відсоток різниці тривалості RR-інтервалів, що відрізняються в довжині більше ніж на 20 та 50 мс, кількість ударів за хвилину. До частотних характеристик включено суми енергій різних частот: енергія дуже низьких частот (0-0.04), енергія низьких частот (0.04-0.15), та енергія високих частот (0.15-0.4), а також комбінації з цих енергій, зокрема, загальна сума енергій, співвідношення енергій низьких та високих частот, відсоток енергій низьких та високих частот відносно їхньої суми. Для поділу сигналу на спектри використано метод Велча, доступний в бібліотеці Scipy.Signal.

На основі виявлених R-піків, можна витягнути шаблони ЕКГ кожного з скорочень серця. З них, на основі середніх часових характеристик для різних етапів ЕКГ (0.25 с на Р-хвилю, 0.12 с на QRS-комплекс, 0.45с на Т-хвилю під час нормального серцебиття). Уточнення цих середніх характеристик для окремого запису ЕКГ можна здійснити за допомогою порівняння середньої довжини RR-інтервалу до довжини інтервалу під час нормального серцебиття.

Аналіз записів ЕКГ здійснимо з допомогою статистичних характеристик дискретних розподілів. Зокрема для кожного скорочення серця ми витягнемо проміжок з Р-хвилею та Т-хвилею та визначимо для них наступні характеристики: максимальне, середнє значення, співвідношення середнього до максимального значення та відсоток значень значень хвилі, більших за математичне сподівання, співвідношення математичних сподівань та максимальних значень Р та Т хвиль. Також обчислимо для обох хвиль середні квадратичні відхилення, коефіцієнт асиметрії та коефіцієнт ексцесу для розподілу.

Описаний вище набір характеристик використаємо для класифікації ЕКГ. Для створення моделі потрібно мати набір записів ЕКГ з вказаними типами аритмії серця. В даному проекті використано базу даних записів надану проектом PhysioNet в межах змагання Computing In Cardiology 2017. Даний набір містить 8528 записів ЕКГ з частотою 300Гц і довжиною від 9 до 60 с. Дані описані 4 мітками: миготіння передсердь (A-Fib), нормальний ритм, інший ритм та надто шумний сигнал. Дані є доволі незбалансованими. Нормальний ритм презентований 5154 записами, 2557 записів мають мітку іншого ритму, 771 містить миготіння передсердь та 46 з шумами. Оскільки таке співвідношення категорій змусить модель частіше класифікувати записи як нормальні, проведемо спочатку балансування даних. Зокрема, випадковим чином приберемо частину даних з нормальним ритмом та іншими ритмами. А також створимо більше менш представлених записів шляхом ротації сигналів.

Для побудови моделі застосовано алгоритм RandomForest, що складається з 60 дерев прийняття рішень. Імплементацию алгоритму взято з бібліотеки Scikit-learn. В якості оцінки моделі використано F1 score. Отримане значення метрики для побудованої моделі рівне 0.65.

#### Література

1. AF Classification from a short single lead ECG recording: the PhysioNet/Computing in Cardiology Challenge 2017 [електронний ресурс] - Режим доступу: <https://physionet.org/challenge/2017> (дата звернення: 10.02.2017)
2. Офіційна документація Scikit-learn [електронний ресурс] - Режим доступу: <http://scikit-learn.org/stable/documentation.html> (дата звернення: 15.03.2017)