

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ТЕРНОПІЛЬСЬКИЙ НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ
ІМЕНІ ІВАНА ПУЛЮЯ
ФАКУЛЬТЕТ КОМП'ЮТЕРНО-ІНФОРМАЦІЙНИХ СИСТЕМ
І ПРОГРАМНОЇ ІНЖЕНЕРІЇ
КАФЕДРА ПРОГРАМНОЇ ІНЖЕНЕРІЇ

КІФЕР ВІКТОР МИХАЙЛОВИЧ

УДК 004.4

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

121 «Інженерія програмного забезпечення»

Автореферат

магістерської роботи на здобуття освітнього ступеня «магістр»

Тернопіль

2018

Проект виконано на кафедрі програмної інженерії Тернопільського національного технічного університету імені Івана Пулюя.

Керівник проекту: доктор фізико-математичних наук, професор
Петрик Михайло Романович,
Тернопільський національний технічний університет
імені Івана Пулюя

Захист відбудеться 20 лютого 2018 р. о 9⁰⁰ годині на засіданні екзаменаційної комісії №31 у Тернопільському національному технічному університеті імені Івана Пулюя за адресою: 46001, м. Тернопіль, вул. Руська, 56, навчальний корпус №1, ауд. 101

ЗАГАЛЬНА ХАРАКТЕРИСТИКА ПРОЕКТУ

Актуальність теми проекту. Серцеві хвороби є однією з головних причин смертності в світі. Вони становлять приблизно 25% випадків в США. Крім того, близько 47% з раптових смертей через хвороби серця відбуваються поза межами лікарень, тож можна припустити, що багато людей з хворобами серця не зважають на ранні прояви хвороби. Тому важливо надати можливість якісно та швидко проводити діагностику стану серця без потреби візиту до лікарні та проходження численних діагностик.

Наукова спільнота активно досліджує нові рішення для покращення виявлення хвороб серця на ранніх стадіях, коли їм легко запобігти за допомогою прийняття певних медикаментів або в складних випадках підготувати пацієнта до оперативного втручання. Зокрема, PhysioNet та науковий журнал Computing in Cardiology щорічно проводять змагання на тематику побудови рішень для класифікації біологічних сигналів (записи електрокардіограм, електроенцефалограм тощо) та виявлення певного набору хворіб. За основу проекту взято завдання зі змагання за 2017 рік, що полягає у класифікації електрокардіограм у чотири категорії: аритмія передсердь, нормальний синусоїдний ритм, інше відхилення ритму серця та шум.

Мета проекту. Створення моделі для автоматичного виявлення аритмії серця засобами машинного навчання.

Об'єкт, методи та джерела дослідження. Використання методів навчання для виявлення аритмії серця на записах електрокардіограм.

Практичне значення отриманих результатів. Розроблена система дозволяє проводити аналіз коротких записів ЕКГ з одним відведенням та класифікувати записи у чотири категорії: нормальний ритм, аритмія передсердь, інші відхилення нормального ритму та шум.

Структура проекту. Проект складається з розрахунково-пояснювальної записки та графічної частини. Розрахунково-пояснювальна записка складається з вступу, 6 частин, висновків, переліку посилань. Обсяг проекту: розрахунково-пояснювальна записка – 110 арк. формату А4, графічна частина – 12 слайдів.

ОСНОВНИЙ ЗМІСТ ПРОЕКТУ

У вступі проведено аналіз актуальності та мети проекту, поставлено задачі дослідження, наведена наукова новизна та практичне значення одержаних результатів.

В розділі “Аналіз предметної області” проводиться огляд літературних джерел з тематики машинного навчання та його застосування у різних сферах людської діяльності; зроблено загальний опис електричної активності серця, проведено аналіз існуючих методів обробки та класифікації ЕКГ з вказанням їх переваг та недоліків.

В розділі “Методи вирішення задачі класифікації ЕКГ” сформульовано вимоги до кінцевої системи; описано методологію проектування системи та здійснено підбір та опис методів реалізації системи, що включають методи попередньої обробки сигналів та витягнення фіч, технології побудови моделей машинного навчання.

В розділі “Реалізація системи” здійснено підбір технологій реалізації проекту, що включає вибір мови програмування, середовища розробки, бібліотек та фреймворків для обробки даних та побудови моделей машинного навчання; описано вимоги до електронно-обчислювальних машин, необхідних для реалізації проекту; представлено стандарти стилю коду; подано деталі реалізації основних компонент системи; описано процес навчання моделей та методологію їх тестування; розказано про вимоги до підготовки заявки на змагання PhysioNet; представлено результати оцінки моделей.

В розділі “Обґрунтування економічної ефективності” проведено розрахунок норм часу на виконання дипломного проекту, витрат на електроенергію, суму амортизаційних відрахувань та ціну дослідження. Також визначено витрати на оплату праці, відрахування на соціальні заходи та економічну ефективність і термін окупності капітальних вкладень.

В розділі “Охорона праці та безпека в надзвичайних ситуаціях” розглянуто питання облаштування робочих місць для користувачів ПК з врахуванням оптимальних параметрів освітлення, мікроклімату, шуму та вібрації, забезпечення

захисту від дії рентгенівського та електромагнітного, ультрафіолетового та інфрачервоного випромінювання. Описано можливості використання комп'ютерної техніки для оцінки можливої обстановки у випадку надзвичайних ситуації та подано функціональні заходи у сфері державного регулювання та контролю захисту населення і територій.

У загальних висновках щодо дипломного проекту описано результати проектування та розробки системи виявлення аритмії серця на записах ЕКГ та проведено їх аналіз, подано рекомендації з подальших напрямків дослідження.

В графічній частині приведено результати проектування та розробки системи автоматичного виявлення аритмії серця на записах ЕКГ.

ВИСНОВКИ

Даний проект присвячено створенню моделі машинного навчання для автоматичної класифікації ЕКГ. Проект розпочався з введення в кардіологію серця, що включає ознайомлення з структурою серця людини та його електричною активністю. Іншим напрямком аналізу був огляд попередніх досліджень в області автоматичного аналізу ЕКГ. В результаті аналізу існуючих рішень було виявлено широкий спектр підходів до вирішення проблеми, визначено їх переваги та недоліки.

Система архітектурно поділена на п'ять компонент: завантаження, візуалізація та попередня обробка даних, модуль моделей машинного навчання та модуль тестування. Проект реалізовано мовою програмування Python. Завантаження даних відбувається з допомогою АПІ бібліотеки SciPy. Графічне представлення здійснюється через методи бібліотеки Matplotlib. Попередній аналіз даних включає обробку сигналу з допомогою дискретного перетворення Фур'є, вейвлетів та частотних фільтрів з застосуванням NumPy та SciPy. В проекті побудовано та здійснено порівняльний аналіз п'яти моделей машинного навчання: багатосаровий перцептрон, повністю згортова, залишкова та рекурентна нейронні мережі, Random Forest. Для оцінки якості моделі використано гармонійну середню (F1) між точністю (precision) та відгуком (recall) системи з застосуванням кросвалідації.

В результаті найкращу оцінку отримала модель Random Forest, що набрала 0.77. Найкращу оцінку серед моделей на основі нейронних мереж досягнуто з допомогою залишкової мережі, який становить 0.7. Результати свідчать про однозначну можливість застосування машинного навчання для автоматичної класифікації ЕКГ. Хоча модель на основі дерев прийняття рішень показала найкращий результат, нейронні мережі теж можуть застосовуватись для вирішення задачі. Необхідні подальші дослідження в цьому напрямку. Іншим напрямком досліджень є аналіз інших відхилень серцевих ритмів, оскільки ця категорія класифікації отримала найнижчий рівень точності. Це пояснюється тим, що ця категорія включає великий список різних відхилень серцевого ритму.

АНОТАЦІЯ

Дипломний проект присвячений розробці моделі для автоматичного виявлення аритмії серця на записах електрокардіограм засобами машинного навчання.

Об'єктом дослідження є використання методів машинного навчання для виявлення аритмій серця на записах ЕКГ.

Предмет дослідження: методи обробки сигналів ЕКГ та засоби машинного навчання, призначені для забезпечення автоматичної класифікації електрокардіограм

Мета роботи: створення моделі для автоматичного виявлення аритмії серця засобами машинного навчання.

Система розроблена з допомогою мови програмування Python та з застосуванням бібліотек NumPy, SciPy, Scikit-Learn, Theano. Система дозволяє проводити автоматичний аналіз коротких записів електрокардіограм з одним відведенням на предмет наявності аритмії серця.

Система з моделлю Random Forest досягла оцінки у 0.77 за критерієм гармонійної середньої між точністю та відкликом (F1). Залишкова нейронна мережа змогла отримати 0.7 за тим же критерієм. Необхідні подальші дослідження застосування нейронних мереж, а також більш детальний аналіз інших відхилень ритмів серця.

Ключові слова: ЕЛЕКТРОКАРДІОГРАМА, АРИТМІЯ СЕРЦЯ, МАШИННЕ НАВЧАННЯ, НЕЙРОННІ МЕРЕЖІ, ДЕРЕВА ПРИЙНЯТТЯ РІШЕНЬ, PHYSIONET.