

УДК 519.25:616.12-073.7**DOI: 10.31866/2617-796X.6.2.2023.293593****Анатолій Пашко,**

доктор фізико-математичних наук, професор,
професор кафедри теоретичної кібернетики,
Київський національний університет
імені Тараса Шевченка,
Київ, Україна
aapashko@gmail.com
<https://orcid.org/0000-0001-6944-8477>

Олена Чайковська,

кандидат педагогічних наук,
доцент кафедри комп'ютерних наук,
Київський національний університет
культури і мистецтв,
Київ, Україна
oachaikovska@gmail.com
<https://orcid.org/0000-0001-7769-1004>

Юрій Харченко,

магістр кафедри комп'ютерних наук,
Київський національний університет
культури і мистецтв,
Київ, Україна
java.kharchenko@gmail.com
<https://orcid.org/0009-0002-5275-5686>

ВИКОРИСТАННЯ ІНСТРУМЕНТІВ СТАТИСТИЧНОГО АНАЛІЗУ ДЛЯ ОБРОБКИ ЕКГ

Мета статті – дослідження алгоритмів статистичного та інтелектуального аналізу даних та їхнього використання для обробки і аналізу електрокардіограм (ЕКГ). Розглянуто методи та алгоритми, які є основою статистичної обробки і аналізу даних.

Методи дослідження ґрунтуються на застосуванні статистичних методів та алгоритмів для аналізу та попередньої обробки медичних даних. Вона є необхідним кроком в обробці даних, що дає можливість більш ефективно аналізувати, будувати більш точні моделі та зменшити їхню розмірність.

Наукова новизна. У статті проаналізовано класичні статистичні методи, що використовують на першому етапі обробки даних. Доведена їхня ефективність та необхідність під час аналізу медичних даних. Результати роботи продемонстровані на обробці реальних даних, а саме електрокардіограм.

Висновки. Застосування методів статистичного аналізу для попередньої обробки медичних даних розглянуто на прикладі обробки і аналізу ЕКГ. Були обчислені основні статистичні характеристики: середнє, дисперсія, середнє квадратичне відхилення, мода, медіана, коефіцієнт асиметрії, коефіцієнт ексцесу та коефіцієнт варіації. Отримані дані використовують для оцінки закону розподілу даних, перевірки гіпотез про закони розподілу, нормалізації даних.

Для спостережень ЕКГ побудовано кореляційну матрицю, обчислені власні числа та власні вектори, на їхній основі визначено головні компоненти. Використання головних компонент дає можливість зменшити розмірність даних для більш глибокого аналізу. У цьому дослідженні об'єм даних було зменшено в чотири рази.

Виконано дискретне перетворення Фур'є. Аналіз його результатів дав можливість виокремити високочастотні електромагнітні перешкоди, що передаються по кабелю від мережі електроживлення до приладу, визначена частота перешкоди. Були виявлені артефакти руху, пов'язані з диханням хворого. Частота таких перешкод перебуває у діапазоні від 1 до 4 Гц.

Була проведена класифікація точок ЕКГ, побудований модуль ЕКГ для подальшого аналізу. Результати класифікації дали змогу виокремити множину точок, що містяться в околі R-піків. Це дає можливість, не використовуючи складний математичний апарат, локалізувати QRS комплекси.

Результати попередньої обробки медичних даних дають змогу зменшити їх розмірність, виявити наявність або відсутність лінійних зв'язків, оцінити частотні характеристики даних. На основі попередньої обробки можливо планувати подальші дослідження і більш якісно будувати моделі для інтелектуального аналізу даних.

Ключові слова: електрокардіограма; статистичний аналіз; кореляційний аналіз; головні компоненти; класифікація.

Вступ. Істотність та точність аналізу медичних даних залежить від здатності здійснювати прогнози, діагностику та планування лікування на основі точних даних. Інтелектуальний аналіз медичних даних, а саме штучний інтелект в цьому контексті, стає потужним інструментом для розуміння характеристик захворювань та прогнозування їхнього розвитку. Це відкриває нові перспективи для забезпечення ефективного та персоналізованого підходу до кожного пацієнта.

Розуміння значущості штучного інтелекту у сфері аналізу медичних даних відкриває шлях до революційних змін у вигляді організації та надання медичних послуг. Сучасні медичні системи здатні зібрати величезний обсяг різноманітних відомостей про пацієнтів: клінічні дані, дані досліджень, образи, генетичні дані та дані стосовно способу життя.

Однак важливим аспектом є попередня обробка цих даних для покращення якості діагностики, лікування та прогнозування захворювань.

Для аналізу медичних даних, а особливо ЕКГ, використовують різні підходи. Вченими досліджувалися аналіз, розпізнавання та класифікація фізіологічних сигналів за допомогою алгоритмів машинного навчання (Krak, Pashko, Khorozov and Stelia, 2020), обробка сигналів ЕКГ та визначення QRS комплексів (Krak, Pashko, Stelia, Barmak and Pavlov, 2020), виділення інформативних ознак для аналізу комплексу QRS в ЕКГ-сигналах (Pashko et al., 2021), корекція базової лінії сигналів електрокардіограми для ефективною попередньої обробки (Pashko et al., 2022).

Нині штучний інтелект розвивається з кожним днем, також проводиться багато досліджень для класифікації сигналів ЕКГ та інших біомедичних сигналів з використанням архітектур CNN (Khorrami and Moavenian, 2010). Автори застосували неперервне та дискретне вейвлет-перетворення (DWT) і дискретне косинусне перетворення (DCT) для сигналів ЕКГ.

Мохаммад М. Аль-Рахаль, Якуб Базі, Мансур Аль-Зуайр, Есам Отман, Білель Бен-Дждіра (Al Rahhal et al., 2018) трансформували сигнали з різних наборів даних за допомогою CWT для ідентифікації аритмій на сигналах ЕКГ.

Цзін-Шан Хуан, Ван-Шан Лю, Бін Яо, Чжан-Сян Ван, Сі-Фан Чен, Вей-Фан Сунь (Huang et al., 2019) трансформували ЕКГ-сигнали за допомогою STFT і у своєму дослідженні отримали двовимірні скалограми.

Улас Баран Балоглу, Мухаммед Тало, Озал Йилдирим, Ру Сан Тан, У. Раджендра Ачар'я (Baloglu et al., 2019) розробили 10-рівневу наскрізну архітектуру CNN для класифікації одновимірних багатокласових даних ЕКГ.

Тобто у сучасних дослідженнях інструменти штучного інтелекту, такі як машинне навчання, глибоке навчання та нейронні мережі, багато важать у цьому процесі. Вони можуть виявляти складні зв'язки в даних та використовувати цю інформацію для передбачення ризиків захворювань, визначення найефективніших методів лікування та розробки індивідуальних планів лікування для кожного пацієнта.

Але для ефективного використання сучасних методів аналізу необхідна якісна попередня обробка даних.

Результати дослідження. Електрокардіограма – це запис електричної активності серця. ЕКГ оцінює ритм та провідність серця, дає можливість оцінити кровопостачання серцевого м'яза в стані спокою, а також виявити збільшення камер серця (передсердь та шлуночків) (Хемптон та Едлем, 2020).

Кожен такий відрізок (рис. 1) відповідає одному серцевому циклу, тобто одному повноцінному скороченню серцевого м'яза.

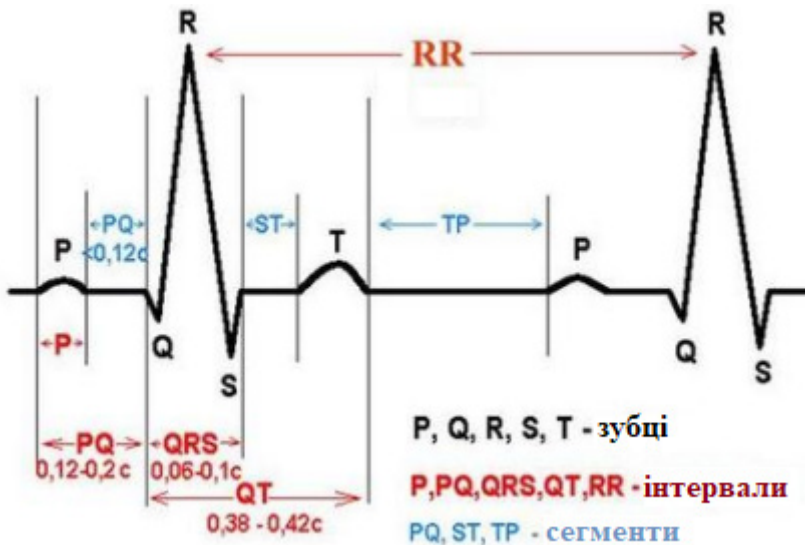


Рис. 1. Основні елементи ЕКГ

Людське серце складається з чотирьох камер (відділів): двох передсердь та шлуночків, послідовне скорочення кожної з них фіксується на ЕКГ у вигляді відхилень від ізолінії. Такі відхилення називають «зубцями» і називають латинськими літерами *P, Q, R, S, T*. Зубці, розташовані вище ізолінії, називаються позитивними, нижче – негативними. Розрізняють зубці: *P, Q, R, S, T*. Зубці *P, R, T* – спрямовані вгору, а зубці *Q, S* – вниз (Хемптон та Едлем, 2020).

Загалом процес формування електрокардіограми можливо описати так:

- формування імпульсу збудження в синусовому вузлі;
- збудження передсердь, формування *R*-зубця;
- реєстрація інтервалу *P-Q*;
- збудження міокарда шлуночків, запис *QRS*-комплексу;
- згасання збудження та відновлення вихідного стану, поява сегмента *S-T* і *T*-зубця.

Електрокардіографічне відведення – це сукупність двох електродів, що розташовані на поверхні тіла. Вісь відведення – це уявна пряма, що з'єднує між собою два електроди. Відведення можуть бути двополюсними або однополюсними.

Для повного опису стану серця зазвичай використовують дванадцять відведень. Вони об'єднані у дві групи: шість відведень від кінцівок та шість грудних. Відведення від кінцівок поділяють на стандартні відведення I, II, III та посилені відведення *aVR, aVL, aVF* (Хемптон та Едлем, 2020). Водночас стандартні відведення є двополюсними, а посилені та грудні – однополюсними.

Літера «*V*» у назвах цих відведень означає, що вимірюють значення потенціалу (*voltage*) під активним електродом, літера «*a*» – цей потенціал посилений (*augmented*).

ЕКГ відображає сумарні електричні струми, що виникають у численних волокнах міокарда під час збудження. Водночас сумарна електрорушійна сила серця змінює величину та напрямок, вона є векторною величиною.

Крім того, осі відведень характеризуються не лише напрямом, але й полярністю: один електрод приєднується до позитивного полюса електрокардіографа, інший – до негативного. Якщо в певний проміжок часу сумарний вектор спрямований у бік позитивного полюса, то крива ЕКГ зміщується вгору, а якщо в бік негативного – вниз. Якщо сумарний вектор спрямований під прямим кутом до осі цього відведення, записується ізолінія.

За умови нормального стану серцево-судинної системи проміжок часу між двома сусідніми серцевими скороченнями (*R*-піками) змінюється від скорочення до скорочення. Ця мінливість називається варіабельністю серцевого ритму (BCP) (Файнзильберг, 2013).

Аналіз варіабельності серцевого ритму є методом дослідження статистичних властивостей сигналу. Активного свого розвитку він сягнув завдяки успіхам у космічній медицині на рубежі 60-х років ХХ ст.

Інформація про варіабельність ритму серця отримується за рахунок дослідження ритмограм – числових послідовностей, елементами яких є проміжки часу між двома сусідніми серцевими скороченнями (Файнзильберг, 2013).

Ще одним методом аналізу ВСР є побудова скатерограм. Скатерограма є графічним відображенням послідовних пар серцевих інтервалів (попереднього та наступного) на координатній площині.

Аналізуючи вигляд ритмограми або скатерограми можливо судити про порушення серцевого ритму або аномальних викидах, виконувати підрахунок частоти серцевих скорочень (ЧСС), а також визначати джерело збудження (місце розташування водія ритму), що викликає скорочення передсердь та шлуночків (Файнзильберг, 2013).

Процес аналізу ЕКГ-сигналу можливо розділити на 2 стадії:

- стадія попередньої обробки;
- виокремлення ознак.

Розглянемо детально стадію попередньої обробки.

1. Перша стадія, тобто стадія попередньої обробки, полягає у видаленні шумів. Шум у цьому випадку – високочастотні компоненти кардіосигналу, а також низькочастотний тренд сигналу. Коли шум видаляється, відбувається згладжування та стиснення ЕКГ-сигналу.

Біопотенціали серця є слабкими сигналами в оточенні великої кількості інших сигналів. Будь-який інший сигнал, крім досліджуваного ЕКГ, є перешкодою (артефактом).

Перешкоди, що виникають під час отримання ЕКГ:

- дрейф ізоляції;
- мережева перешкода;
- м'язовий тремор;
- артефакти руху.

Дрейф ізоляції є низькочастотною перешкодою з частотою менше 1 Гц, обумовлений впливом дихання та високого шкірного опору. Непостійність ізоляції впливає на точність виміру амплітудних параметрів електрокардіосигналу, оскільки саме від неї ведеться відлік.

Мережева перешкода – високочастотна електромагнітна перешкода, що передається по кабелю від мережі електроживлення до приладу. Частота цієї перешкоди становить 50 чи 60 Гц.

М'язовий тремор виникає внаслідок високої електричної активності тканин і є хаотичним коливанням ізоелектричної лінії в частотному діапазоні від 30 до 100 Гц. Накладається на ЕКС у пацієнтів із тремтливим паралічем, паркінсонізмом тощо.

Артефакти руху виявляються як поодинокі або циклічні сплески, спектр яких нагадує спектр комплексів. Частота таких перешкод перебуває у діапазоні від 1 до 40 Гц. Виникають вони у результаті зміни становища пацієнта чи електрода, гикавки, кашлю тощо.

Частотні характеристики ЕКГ досліджують за допомогою перетворення Фур'є.

2. Друга стадія є процесом вилучення потрібної інформації, тобто виокремлення *QRS*-комплексів та всіх зубців.

Зараз є безліч методик виділення *QRS*-комплексів, які, залежно від використовуваного підходу, можливо поділити на три основні групи:

- алгоритми аналізу ЕКГ у часовій області;
- алгоритми частотно-часової області;
- алгоритми класифікації елементів ЕКГ.

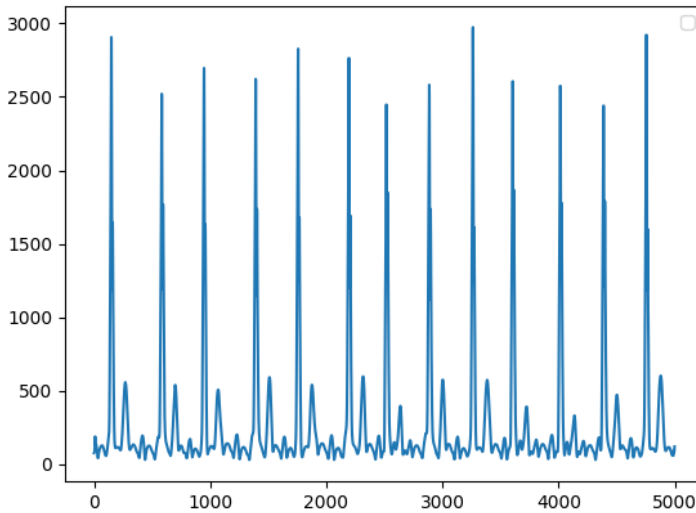


Рис. 2. Модуль сигналу ЕКГ по 12 каналах

Алгоритми першої групи ґрунтуються на застосуванні до вхідного сигналу, крім процедур лінійної фільтрації, деякого нелінійного перетворення. Алгоритми другої групи засновані на застосуванні після ряду процедур попередньої обробки ЕКГ різних частотно-тимчасових перетворень. Алгоритми третьої групи використовують методи машинного навчання.

Для проведення попередньої обробки аналізу сигналу ЕКГ використано запис електрокардіограми в текстовому файлі. Файл містить запис електрокардіограми людини по 12 каналах. Час запису – 10 с. Дискретність: 500 точок за 1 с.

Довжина запису $N=5000$, $\Delta t = 1/500 = 0.002$.

Для візуалізації даних ЕКГ будуюмо графіки по всіх каналах ЕКГ. Знайдено модуль сигналу по 12 каналах (рис. 2), використовуючи метрику Евкліда.

Для кожного каналу ЕКГ було оцінено основні статистичні параметри: середнє, мода, медіана, дисперсія, середнє квадратичне відхилення (табл. 1).

Таблиця 1.

Основні статистичні параметри

Номер каналу	середнє	мода	медіана	дисперсія	СКВ
0 :	25.772	-2.489	-37.103	12709.055	112.734
1 :	18.664	-21.0205	115.66	17853.5098	133.617
2 :	0.4266	-8.9739	-17.109	2627.76232	51.2617
3 :	-19.449	11.835	-29.394	14438.9437	120.162
4 :	20.887	0.27605	150.18	3409.29527	58.3891
5 :	11.101	-15.879	-32.343	7162.04938	84.6288
6 :	-39.168	-8.0782	48.889	21926.6145	148.076
7 :	-28.683	-1.5364	-12.152	35245.4304	187.737

Продовження табл. 1

8 :	11.539	-0.5165	17.115	26102.4330	161.562
9 :	31.864	-0.773	-24.629	39532.8088	198.828
10 :	43.142	-4.6887	-10.911	48439.8703	220.090
11 :	41.785	-5.7603	-24.124	36925.0916	192.159

Для подальшого аналізу початкові дані були нормалізовані по кожному стовпчику (математичне сподівання рівне нулю, дисперсія рівна 1).

Перевірено гіпотезу про нормальний закон розподілу, для цього використовується тест Шапіро-Уїлка.

Тест засвідчив, що гіпотезу про нормальний закон розподілу потрібно відхилити.

Наприклад, для першого каналу

data: Stat\$V1 W = 0.67013, p-value < 2.2e-16

Наступним кроком було проведення кореляційного аналізу. Для нормалізованих змінних, обчислюємо кореляційну матрицю (рис. 3.)

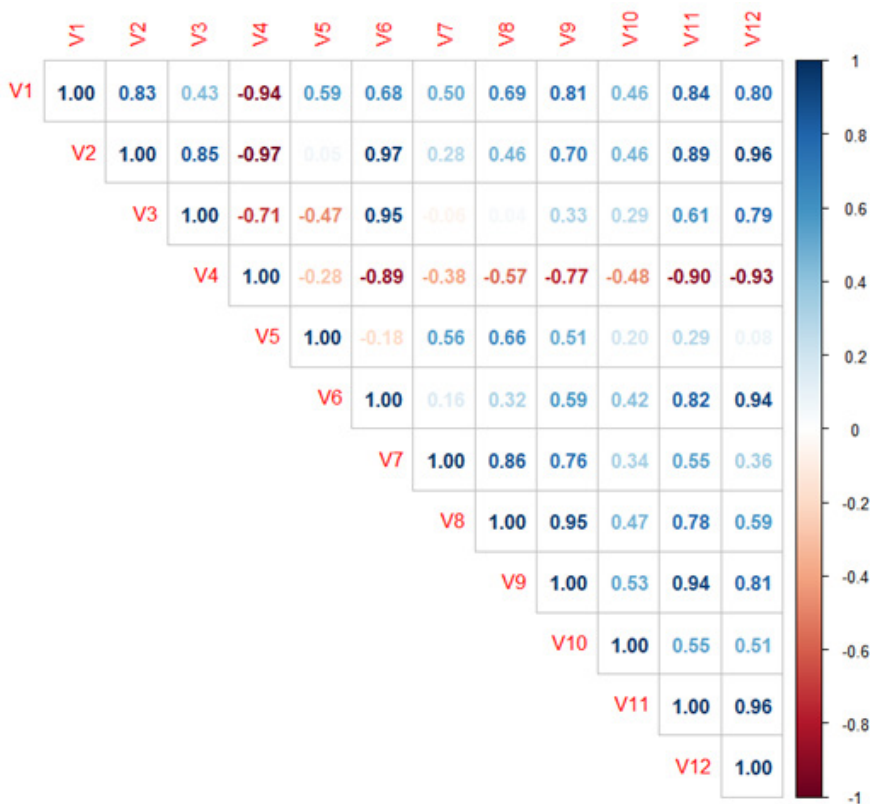


Рис. 3. Кореляційна матриця спостережень по 12 каналах

Проводимо аналіз кореляційної матриці – виокремили групу із чотирьох параметрів, парна кореляція між якими велика $a=1$, $b=2$, $c=4$, $d=11$ (табл. 2).

Таблиця 2.

Найбільш корельовані параметри

	V1	V2	V4	V11
V1	1.00	0.82	-0.94	0.84
V2		1.00	-0.97	0.89
V4			1.00	-0.90
V11				1.00

Подальший крок – факторний аналіз. Знаходимо головні компоненти.

Для початку знаходимо власні числа кореляційної матриці з рівняння

$$|r - \lambda E| = 0, k = \overline{1, m},$$

де E – одинична матриця розміром $m \times m$.

Власні числа кореляційної матриці – [0.05573696, 0.48140159, 0.49675618, 0.21999798, 0.29976447, 0.47899287, 0.2931390, 0.03406193, 0.12536206, 0.12361535, 0.15055222, 0.09967].

Побудуємо графік власних чисел (алгоритм кам'янистого осипу) (рис. 4).

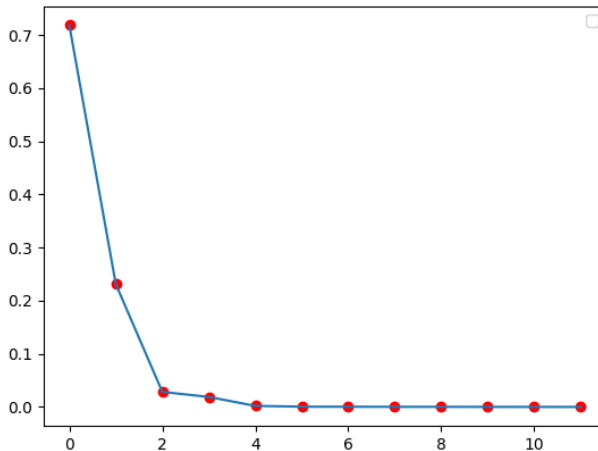


Рис. 4. Графік власних чисел

Вибираємо три найбільших власних числа [0.71862116, 0.2310473, 0.02829887]. Сума трьох найбільших власних чисел дорівнює 0.977967338504827. Отже, вони пояснюють майже 98 % дисперсії.

Власні вектори головних компонент

[[0.24470513 0.28539232 0.04682806 -0.26433968 0.10226645 0.16527131 -0.14415832 0.00571964 0.29688983 0.44857927 0.4995489 0.43485502]

[0.00242543 0.05098572 0.04057141 -0.0270346 -0.02119657 0.04882629
-0.51654628 -0.75117064 -0.38355531 -0.08106214 0.02962585 0.08089215]

[0.05573696 -0.48140159 -0.49675618 0.21999798 0.29976447 -0.47899287
-0.29313903 0.03406193 0.12536206 0.12361535 0.15055222 0.09967]]

Знаходимо модуль трьох головних компонентів (рис. 5) та різницю між модулем 12 компонентів і модулем 3 головних компонентів (рис. 6).

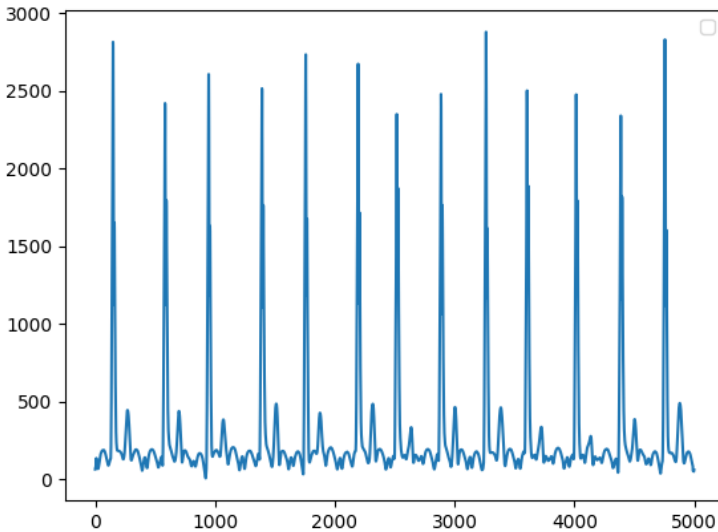


Рис. 5. Графік модульів 3 головних компонентів

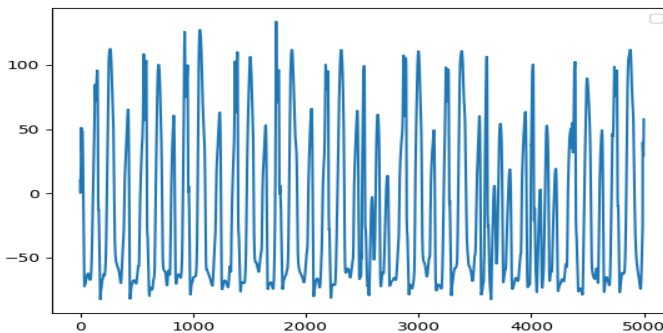


Рис. 6. Різниця модульів

Отже, різниця не суттєва. Тобто замість 12 каналів можливо досліджувати три канали, або один модуль.

Для аналізу частотних характеристик сигналу ЕКГ використовуємо дискретне перетворення Фур'є. На рис. 7 показано спектр модульів сигналу 12 і 3 каналів та їхню різницю.

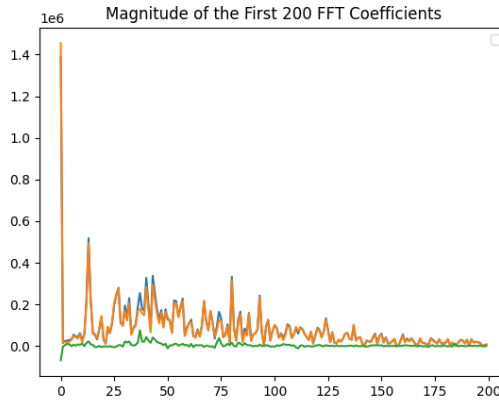


Рис. 7. Спектри модулів сигналів (зеленим кольором показана різниця)

Кластерний аналіз. Для проведення кластерного аналізу використовувався метод k-means.

Будемо вважати, що записана кардіограма (12 каналів) є множиною багатовимірних точок деякого евклідового простору. Розмірність точок – 12, кількість точок – 5000.

Використовуючи алгоритми кластеризації, а саме k-means, проводимо розбиття множини точок на k підмножин. Розглянуто варіанти k=7, k=11. Можемо побачити, що результати кластеризації є схожими. Розміри кластерів майже ідентичні.

На рис. 8 зображено результат кластеризації на 11 класів та центри класів.

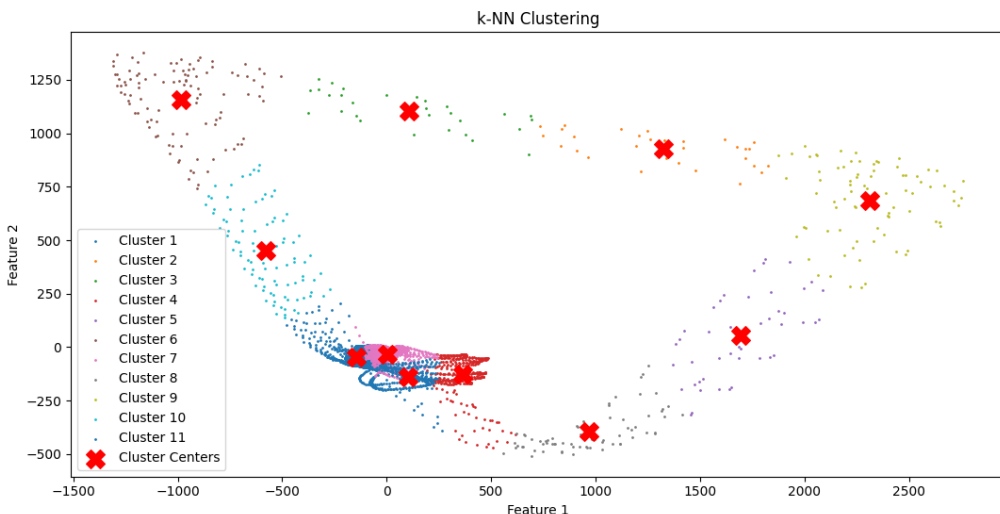


Рис. 8. Кластеризація на 11 класів головних компонентів

Кількість елементів у кластерах

Cluster 1: 2612 елементів

Cluster 2: 84 елементи

Cluster 3: 104 елементи

Cluster 4: 359 елементів

Cluster 5: 1275 елементів

Cluster 6: 32 елементи

Cluster 7: 92 елементи

Cluster 8: 50 елементів

Cluster 9: 295 елементів

Cluster 10: 67 елементів

Cluster 11: 30 елементів

У кластер 10 входять точки, що містяться в околі R-піків.

Висновки. Отримані результати попередньої обробки дають можливість зробити перші висновки.

Основні статистичні характеристики, а саме: середнє, дисперсія, середнє квадратичне відхилення, мода, медіана, коефіцієнт асиметрії, коефіцієнт ексцесу та коефіцієнт варіації, використовують для перевірки гіпотез про закони розподілу, нормалізації даних.

Аналіз кореляційної матриці дав змогу встановити наявність зв'язку між каналами, виокремити групу найбільш корельованих параметрів та допомогти лікарю встановити діагноз.

Власні числа та власні вектори кореляційної матриці використовують для визначення головних компонент. Використання головних компонент дає можливість зменшити розмірність даних для більш глибокого аналізу. Для такого аналізу ЕКГ можливо використовувати не всі 12 каналів, а три головні компоненти, або модуль сигналу. Отже, об'єм даних було зменшено в чотири рази.

Для уточнення і виявлення специфічних властивостей діагнозу хворого потрібно використовувати кожен канал окремо.

Дискретне перетворення Фур'є дало можливість виокремити високочастотні електромагнітні перешкоди, що передаються по кабелю від мережі електроживлення до приладу, визначена частота перешкоди. На основі дискретного перетворення Фур'є були виявлені артефакти руху, що пов'язані з диханням хворого. Частота таких перешкод перебуває у діапазоні від 1 до 4 Гц.

Була проведена класифікація точок ЕКГ, побудований модуль ЕКГ для подальшого аналізу. Результати класифікації дали змогу виокремити множину точок, що містяться в околі R-піків. Це дає можливість, не використовуючи складний математичний апарат, локалізувати QRS-комплекси.

Використання методів статистичного аналізу для попередньої обробки медичних даних дає змогу зменшити їх розмірність, виокремити наявність або відсутність лінійних зв'язків, оцінити частотні характеристики даних. На основі попередньої обробки можливо планувати подальші дослідження і більш якісно будувати моделі для інтелектуального аналізу даних.

Отримані результати на прикладі ЕКГ доводять необхідність і доцільність попередньої обробки даних.

СПИСОК ПОСИЛАНЬ

- Файнзильберг, Л.С., 2013. *Компьютерная диагностика по фазовому портрету электрокардиограммы*. Киев: Освіта України.
- Хемптон, Д. та Едлем, Д., 2020. *ЕКГ у практиці*. Переклад з англійської. 7-е вид. Київ: Медицина.
- Al Rahhal, M.M., Bazi, Y., Al Zuair, M., Othman, E. and BenJdira, B., 2018. Convolutional Neural Networks for Electrocardiogram Classification. *Journal of Medical and Biological Engineering*, [e-journal] 38 (6), pp.1014-1025. <https://doi.org/10.1007/s40846-018-0389-7>
- Baloglu, U.B., Talo, M., Yildirim, O., Tan, R.S. and Acharya, U.R., 2019. Classification of myocardial infarction with multi-lead ECG signals and deep CNN. *Pattern Recognition Letters*, [e-journal] 122, pp.23-30. <https://doi.org/10.1016/j.patrec.2019.02.016>
- Huang, J., Chen, B., Yao, B. and He, W., 2019. ECG Arrhythmia Classification Using STFT-Based Spectrogram and Convolutional Neural Network. *IEEE Access*, 7, pp. 92871-92880.
- Khorrani, H. and Moavenian, M., 2010. A comparative study of DWT, CWT and DCT transformations in ECG arrhythmias classification. *Elsevier Expert Systems with Applications*, [e-journal] 37(8), pp.5751-5757. <https://doi.org/10.1016/j.eswa.2010.02.033>
- Krak, I., Pashko, A., Khorozov, O. and Stelia, O., 2020. Physiological Signals Analysis, Recognition and Classification Using Machine Learning Algorithms. In: S. Subbotin, ed. *Computer Modeling and Intelligent Systems*, Proceedings of The Third International Workshop CMIS-2020, Zaporizhzhia, Ukraine, April 27- May 1, 2020, [e-Book] National University "Zaporizhzhia Polytechnic", pp.955-965. Available at: <<https://ceur-ws.org/Vol-2608/paper71.pdf>> [Accessed 15 July 2023].
- Krak, I., Pashko, A., Stelia, O. Barmak, O. and Pavlov, S., 2020. Selection Parameters in the ECG Signals for Analysis of QRS Complexes. In: T. Hovorushchenko, O. Savenko, P. Popov and S. Lysenko, eds. *Intelligent Information Technologies & Systems of information Security*, Proceedings of the 1st International Workshop, Khmelnytskyi, Ukraine, June 10-12, 2020, [e-Book] Khmelnytskyi National University, pp.1-13. Available at: <<http://ceur-ws.org/Vol-2623/paper1.pdf>> [Accessed 15 July 2023].
- Pashko, A., Krak, I., Stelia, O. and Khorozov, O., 2021. Isolation of Informative Features for the Analysis of QRS Complex in ECG Signals. In: S. Babichev, V. Lytvynenko, W. Wójcik and S. Vyshemyrskaya, eds. *ISDMCI 2020: Lecture Notes in Computational Intelligence and Decision Making*, International Scientific Conference «Intellectual Systems of Decision Making and Problem of Computational Intelligence», Kherson, Ukraine, May 25-29, 2020, [e-Book] Cham: Springer, vol. 1246, pp.409-422. https://doi.org/10.1007/978-3-030-54215-3_26
- Pashko, A., Krak, I., Stelia, O. and Wojcik, W., 2022. Baseline Wander Correction of the Electrocardiogram Signals for Effective Preprocessing. In: S. Babichev, and V. Lytvynenko, eds. *ISDMCI 2021: Lecture Notes in Computational Intelligence and Decision Making*, International Scientific Conference «Intellectual Systems of Decision Making and Problem of Computational Intelligence», Kherson, Ukraine, 2021 [e-Book] Cham: Springer, vol. 77, pp.507-518. https://doi.org/10.1007/978-3-030-82014-5_34

REFERENCES

- Al Rahhal, M.M., Bazi, Y., Al Zuair, M., Othman, E. and BenJdira, B., 2018. Convolutional Neural Networks for Electrocardiogram Classification. *Journal of Medical and Biological Engineering*, [e-journal] 38 (6), pp.1014-1025. <https://doi.org/10.1007/s40846-018-0389-7>
- Baloglu, U.B., Talo, M., Yildirim, O., Tan, R.S. and Acharya, UR., 2019. Classification of myocardial infarction with multi-lead ECG signals and deep CNN. *Pattern Recognition Letters*, [e-journal] 122, pp.23-30. <https://doi.org/10.1016/j.patrec.2019.02.016>
- Fainzilberg, L.S., 2013. *Kompiuternaia diagnostika po fazovomu portretu elektrokardiogrammy* [Computer diagnostics using the phase portrait of an electrocardiogram]. Kyiv: Osvita Ukrainy
- Huang J., Chen, B., Yao, B. and He, W., 2019. ECG Arrhythmia Classification Using STFT-Based Spectrogram and Convolutional Neural Network. *IEEE Access*, 7, pp. 92871-92880.
- Khempton, D. and Edlem, D., 2020. *EKH u praktytsi* [The ECG Made Practical]. Translation from English. 7th ed. Kyiv: Medytsyna.
- Khorrani, H. and Moavenian, M., 2010. A comparative study of DWT, CWT and DCT transformations in ECG arrhythmias classification. *Elsevier Expert Systems with Applications*, [e-journal] 37(8), pp.5751-5757. <https://doi.org/10.1016/j.eswa.2010.02.033>
- Krak, I., Pashko, A., Khorozov, O. and Stelia, O., 2020. Physiological Signals Analysis, Recognition and Classification Using Machine Learning Algorithms. In: S. Subbotin, ed. *Computer Modeling and Intelligent Systems*, Proceedings of The Third International Workshop CMIS-2020, Zaporizhzhia, Ukraine, April 27- May 1, 2020, [e-Book] National University "Zaporizhzhia Polytechnic", pp.955-965. Available at: <<https://ceur-ws.org/Vol-2608/paper71.pdf>> [Accessed 15 July 2023].
- Krak, I., Pashko, A., Stelia, O. Barmak, O. and Pavlov, S., 2020. Selection Parameters in the ECG Signals for Analysis of QRS Complexes. In: T. Hovorushchenko, O. Savenko, P. Popov and S. Lysenko, eds. *Intelligent Information Technologies & Systems of information Security*, Proceedings of the 1st International Workshop, Khmelnytskyi, Ukraine, June 10-12, 2020, [e-Book] Khmelnytskyi National University, pp.1-13. Available at: <<http://ceur-ws.org/Vol-2623/paper1.pdf>> [Accessed 15 July 2023].
- Pashko, A., Krak, I., Stelia, O. and Khorozov, O., 2021. Isolation of Informative Features for the Analysis of QRS Complex in ECG Signals. In: S. Babichev, V. Lytvynenko, W. Wójcik and S. Vyshemyskaya, eds. *ISDMCI 2020: Lecture Notes in Computational Intelligence and Decision Making*, International Scientific Conference "Intellectual Systems of Decision Making and Problem of Computational Intelligence", Kherson, Ukraine, May 25-29, 2020, [e-Book] Cham: Springer, vol. 1246, pp.409-422. https://doi.org/10.1007/978-3-030-54215-3_26
- Pashko, A., Krak, I., Stelia, O. and Wojcik, W., 2022. Baseline Wander Correction of the Electrocardiogram Signals for Effective Preprocessing. In: S. Babichev, and V. Lytvynenko, eds. *ISDMCI 2021: Lecture Notes in Computational Intelligence and Decision Making*, International Scientific Conference "Intellectual Systems of Decision Making and Problem of Computational Intelligence", Kherson, Ukraine, 2021 [e-Book] Cham: Springer, vol. 77, pp.507-518. https://doi.org/10.1007/978-3-030-82014-5_34

UDC 519.25:616.12-073.7**Anatolii Pashko,**

*Doctor of Science in Physics and Mathematics, Professor,
Professor at the Department of Theoretical Cybernetics,
Taras Shevchenko National University of Kyiv,
Kyiv, Ukraine
aapashko@gmail.com
<https://orcid.org/0000-0001-6944-8477>*

Olena Chaikovska,

*PhD in Pedagogy, Associate Professor at the Computer Science Department,
Kyiv National University of Culture and Arts,
Kyiv, Ukraine
oachaikovska@gmail.com
<https://orcid.org/0000-0001-7769-1004>*

Yurii Kharchenko,

*Master at the Computer Science Department,
Kyiv National University
of Culture and Arts
Kyiv, Ukraine
java.kharchenko@gmail.com
<https://orcid.org/0009-0002-5275-5686>*

USE OF STATISTICAL ANALYSIS TOOLS FOR ECG PROCESSING

The purpose of the article is to study the algorithms of statistical and intellectual data analysis and their use for processing and analysing electrocardiograms (ECG). The methods and algorithms that form the basis of statistical data processing and analysis are considered.

The research methods are based on the application of statistical methods and algorithms for the analysis and pre-processing of medical data. Pre-processing is a necessary step in data processing, which makes it possible to analyze more efficiently, build more accurate models and reduce their dimensionality.

Scientific novelty. The article analyses classical statistical methods used at the first stage of data processing. Their effectiveness and necessity in medical data analysis are proved. The results of the work are demonstrated on real data processing, namely, on electrocardiogram processing.

Conclusions. The application of statistical analysis methods for pre-processing medical data is considered on the example of ECG processing and analysis. The main statistical characteristics were calculated: mean, variance, standard deviation, mode, median, skewness coefficient, kurtosis coefficient, and coefficient of variation. The obtained data is used to estimate the law of data distribution, test hypotheses about the laws of distribution, and normalize data.

A correlation matrix is constructed for the ECG observation matrix, eigenvalues and eigenvectors are calculated, and principal components are determined on their basis. The use of principal components makes it possible to reduce the dimensionality of the data for a deeper analysis. In this study, the amount of data was reduced by a factor of four.

A discrete Fourier transform was performed. The analysis of the Fourier's transform results made it possible to isolate high-frequency electromagnetic interference transmitted through the cable from the power supply network to the device, and the frequency of the interference was determined. Motion artefacts associated with the patient's breathing were detected. The frequency of such interference is in the range from 1 to 4 Hz.

The ECG points were classified and an ECG module was built for further analysis. The classification results made it possible to identify a set of points in the vicinity of R-peaks. This makes it possible to localize QRS complexes without using a complex mathematical apparatus.

The results of medical data pre-processing allow to reduce the data dimensionality, identify the presence or absence of linear relationships, and evaluate the frequency characteristics of the data. Based on the pre-processing, it is possible to plan further research and build better models for data mining.

Keywords: electrocardiogram; statistical analysis; correlation analysis; principal components; classification.

15.08.2023