

Кластеризація медико-біологічних сигналів в розпізнавальних системах, що навчаються з учителем

Лагутін^f В. В., ORCID [0000-0003-3431-1686](https://orcid.org/0000-0003-3431-1686)

Сиротенко^f В. І., ORCID [0000-0002-7542-5885](https://orcid.org/0000-0002-7542-5885)

Шачиков^s А. Д., к.т.н., ORCID [0000-0003-2353-7647](https://orcid.org/0000-0003-2353-7647)

Шуляк^s О. П., к.т.н. доц., ORCID [0000-0002-6920-2212](https://orcid.org/0000-0002-6920-2212)

Кафедра електронної інженерії ee.kpi.ua

Національний технічний університет України

«Київський політехнічний інститут імені Ігоря Сікорського» kpi.ua

Київ, Україна

Анотація—Розглядається розпізнавання медико-біологічних сигналів по характеристикам їх форми за допомогою алгоритмів, яких навчають з учителем, в системах діагностики пацієнтів. Класифікація сигналів вважається відомою, навчальні вибірки – доступними. Передбачається статистична оцінка вірності рішень про типи сигналів на контрольних вибірках за показниками чутливості, специфічності, загальною валідності.

Розкриваються принцип і порядок вдосконалення взятої за основу базової процедури розпізнавання сигналів. Розробляється і досліджується програмний інструментарій для виявлення і реалізації можливих резервів у підвищенні вірності рішень за рахунок більш детального урахування характеру і особливостей імовірнісних розподілів сигналів в фазовому просторі. З цією метою окремо по класам сигналів додатково здійснюється кластеризація навчальних вибірок на локаціях їх скупчення. Формуються прив'язані до цих локацій сімейства кластерних еталонів і вони використовуються при визначенні типів сигналів, що надходять. Навчання розпізнаванню сигналів стає комбінованим. Вихідна їх класифікація, що несе певний медичний сенс, не порушується.

Розробки і дослідження ілюструються прикладами розпізнавання трьох типів QRS-комплексів (N, A і V) в запису електрокардіограми пацієнта. Результативність застосування розробленого інструментарію перевіряється на конкретних навчальних і контрольних вибірках в порівнянні з показниками базового алгоритму.

Ключові слова — діагностичні системи; медико-біологічні сигнали; розпізнавальні процедури; навчання з учителем; кластеризація

1. ВСТУП

Діагностика пацієнтів спирається на аналіз форми медико-біологічних сигналів, які надходять при їх обстеженні [1 – 5]. Прикладами є читання електрокардіограм, аналіз електроенцефалограм, міограм, плетизмограм, інших сигналів [1 – 5].

Форму сигналів часто характеризують наборами коефіцієнтів в розкладах процесів, що спостерігаються, по системах базисних функцій або характеризують просто послідовностями відліків на рівнодискретних сітках часу [1 – 6].

При аналізі сигналів широко використовується накопичений досвід, який виражається в класифікаціях сигналів і в інтерпретації утворених класів відповідними станами досліджуваних систем організму людини [1 – 5]. Від правильності розпізнавання сигналів може безпосередньо залежати правильність виставлених діагнозів [1 – 5].

Підвищення результативності розпізнавання медико-біологічних сигналів може стати сприятливою передумовою підвищення якості діагностики пацієнтів. Звідси – вимога та цінність розробок

щодо підвищення рівня правильності прийнятих рішень про типи сигналів, що розпізнаються.

Досвід розпізнавання сигналів різних типів виражається також в характерній формі сигналів в кожному класі [7 – 13]. Її отримання в процесі роботи діагностичної системи пов'язане як з накопиченням достовірних сигналів кожного класу, так і з узагальненням, часто з усередненням їх форми [7 – 13], а воно може згладжувати, нівелювати не тільки випадкові, а й регулярні елементи форми сигналів, які є важливими для їх розпізнавання.

Використання єдиного зразка на всю навчальну вибірку кожного конкретного класу [7 – 13] може стати перешкодою в урахуванні можливого різноманіття форми сигналів цього класу. Більш детальний розгляд варіабельності форми сигналів в межах класу не виключає можливості підвищення вірності рішень про типи сигналів.

Одна з головних особливостей цієї роботи в цілому полягає в тому, що вихідна класифікація сигналів з її загально визнаною медичною інтерпретацією тут зберігається незмінною, а варіа-



бельність форми сигналів досліджується без виходу за межі кожного класу сигналів.

Навчання з учителем [1 – 5] комбінується, таким з самонавчанням [14 – 16].

Головне питання цієї роботи – отримання програмного інструментарію, який забезпечує розкриття наявності резервів підвищення правильності розпізнавання медико-біологічних сигналів по характеристикам їх форми, оцінку цих резервів і їх реалізацію під час обробки даних в системах діагностики станів організму людини і його захворювань.

Вирішення цього питання в роботі побудовано на більш детальному урахуванні характеру імовірного розподілу сигналів в фазовому просторі, яке з'ясується шляхом додаткової кластеризації сигналів [14 – 16] навчальних вибірок. Завдяки цьому в кожному класі для розпізнавання нових сигналів замість єдиного еталона формуються і використовуються сімейства локальних кластерних еталонів.

Цілеспрямоване групування всіх точок навчальних вибірок в підмножини (кластери) в фазовому просторі в локаціях їх скупчень виконується окремо для різних класів. Об'єднання точок в кластери в цій роботі реалізується за ознакою подібності характеристик форми сигналів [7 – 16] і по близькості точок в евклідовій або іншій метриці [10, 12].

Розташування точок кластерів в такому випадку окреслює реальні імовірнісні розподіли по типам сигналів, що розпізнаються, і виражає кластерними розбивками особливості цих розподілів, передаючи їх сімействами кластерних еталонів. Узагальнені еталони розщеплюються, таким чином, наведеним способом на сімейства локальних еталонів.

При розробці розпізнавальних процедур не виключається, що фазові підпростори, в яких зосереджені вибірки різних класів, мають області перетину (загальні підобласті). Але при розгляді різних модифікацій таких процедур мається на увазі лише роздільна кластеризація по всіх класах сигналів.

Рішення про тип нових сигналів в таких умовах приймається не за узагальненими еталонами різних класів, в яких осереднення може нівелювати корисні для розпізнавання особливості, а за локальними еталонами, що опинилися в безпосередній близькості до цих сигналів.

Найбільш близький локальний еталон в околі реалізації, що надійшла, вказує своїм типом на рішення розпізнавального алгоритму. В цьому полягає принцип прийняття рішень у разі наявності множин локальних еталонів.

Урахування конфігурації розподілів навчальних вибірок в такій спосіб в цілому націлене на підвищення вірності розпізнавання сигналів. У цьому полягають спрямованість і головний зміст роботи, які розкриваються нижче.

Розпізнавання сигналів конкретних типів, що розглядаються в роботі, має чисто ілюстративний характер для розкриття змісту і з'ясування можливостей і особливостей досліджуваних процедур.

Приклад обробки даних конкретного складу можна також розглядати як тест на працездатність модифікацій розпізнавальних алгоритмів.

II. ВИХІДНИЙ АЛГОРИТМ РОЗПИЗНАВАННЯ СИГНАЛІВ ЯК БАЗОВА ПРОЦЕДУРА ДЛЯ КЛАСТЕРИЗАЦІЇ НАВЧАЛЬНИХ ВИБІРОК ЗА ТИПАМИ СИГНАЛІВ

Для конкретики в цій роботі розглядається розпізнавання сигналів з вичерпною класифікацією з трьох класів. Для прикладу розглядається задача розпізнавання N, A, V типів QRS-комплексів [17] в півгодинному запису ЕКГ конкретного пацієнта. Дані такого запису взяті з Internet [17].

Розглядається розпізнавальна система, що навчається з учителем [1 – 5]. Зазначений запис має розмітку положень R-піків кожного кардіоцикла. Для кожного QRS-комплексу вказано його тип [17]. Реалізація QRS-комплексу V типу може не мати R-піку в запису, але його гіпотетичне положення в даних все одно відзначене вчителем.

Для навчання розпізнавального алгоритму складено навчальні вибірки по 62 реалізації для кожного типу сигналів. Аналогічно складені контрольні вибірки для оцінки вірності розпізнавання сигналів процедурою, яка навчена. Використані стандартні показники вірності рішень: чутливість, специфічність, загальна валідність [1 – 5]. Обсяги навчальних і контрольних вибірок однакові для спрощення розрахунків.

Всі вихідні реалізації сигналів розглядаються фрагментарно в єдиному вікні їх аналізу з 128 відліків на рівнодискретній сітці часу з однаковим положенням R-піків. Ширина вікна незмінна.

На етапі попередньої обробки даних запису кожен сигнал перетворюється в характеристику його форми. З нього виключається постійна складова і здійснюється нормування за інтенсивністю [7 – 13], що сприяє підвищенню вірності розпізнавання сигналів [7 – 13].

З навчальних вибірок кожного класу сигналів формується його еталон. Сигнал, який розпізнається, порівнюється з усіма еталонами. Максимум оцінок схожості вказує при цьому на доцільне рішення розпізнавального алгоритму. Подібність сигналу з еталонами оцінюється по кореляції їх характеристик форми, яка розраховується у вигляді відповідних скалярних здобутків.

Таким є алгоритм розпізнавання сигналів, що взятий як вихідний, як базова процедура для подальших модифікацій, що пропонуються, в тому числі, для модифікацій з кластеризацією навчальних вибірок за типами процесів, що аналізуються. Блок-схема базової процедури розпізнавання сигналів (рис. 2.1).



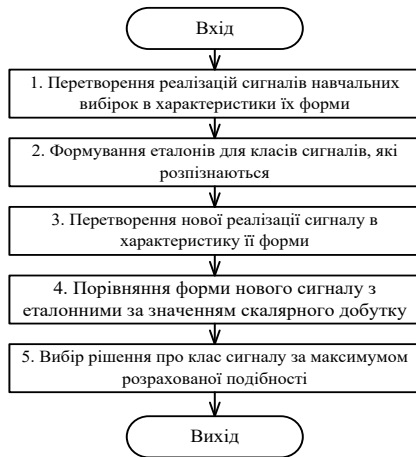


Рис. 2.1. Блок-схема базової процедури розпізнавання сигналів

Навчальна вибірка обробляється на етапі навчання, з неї отримують еталони – середній вигляд сигналів в кожному класі, перерахований в характеристику форми. За контрольною вибіркою складається оцінка рівня вірності розпізнавання сигналів прямим підрахунком кількості правильних і помилкових рішень в роботі алгоритму і оцінкою чутливості, специфічності і загальної валідності рішень.

Такі оцінки будуть отримані для порівняльного аналізу всіх розглянутих модифікацій досліджуваних процедур.

Базова процедура розпізнавання передбачає для кожного типу сигналів єдиний еталон. За допомогою такої процедури отримані еталони N, A, V типів QRS-комплексів, які спільно з сигналами навчальних вибірок зображені на рис. 2.2 – 2.4.

Аналіз графіків показує, що характеристики форми сигналів навчальних вибірок всіх розглянутих типів мають сильний розкид, хоча він значно менший, ніж для сигналів в початковому вигляді, до виключення постійних складових і масштабування. Найменший розкид і найкращі умови розпізнавання в даному випадку мають місце для сигналів N типу.

Процедура розпізнавання таких сигналів реалізована в середовищі MatLab [18 – 20].

Розраховані еталони (середній вид сигналів по їх типам) показані на графіках жирними лініями.

Проведена тестова перевірка результативності розпізнавання сигналів на контрольних вибірках. Результати перевірки представлені в табл. 2.1.

Таблиця 2.1 - Оцінки вірності розпізнавання сигналів базовою процедурою на контрольних вибірках

Тип	A	N	V	Спец.
A	48	2	2	92,31
N	14	60	2	78,95
V	0	0	58	100,00
Чутлив.	77,42	96,77	93,55	89,25%

Таблиця підтверджує працездатність процедури з описом сигналів у вигляді характеристик їх форми і критерієм оцінки їх близькості величинами скалярних добутків векторів сигналу і еталонів.

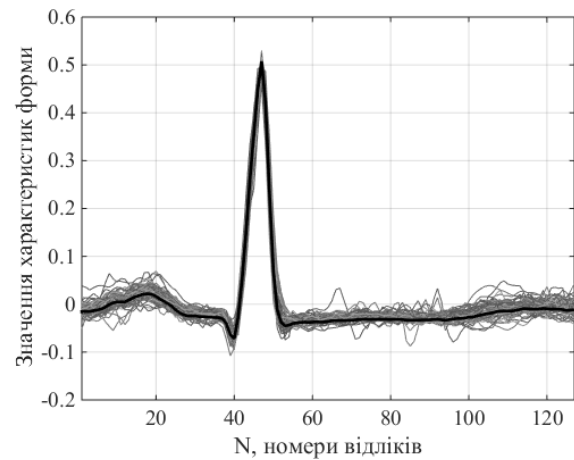


Рис. 2.2. Сигнали навчальної вибірки і еталон (жирна лінія) для QRS-комплексів N типу

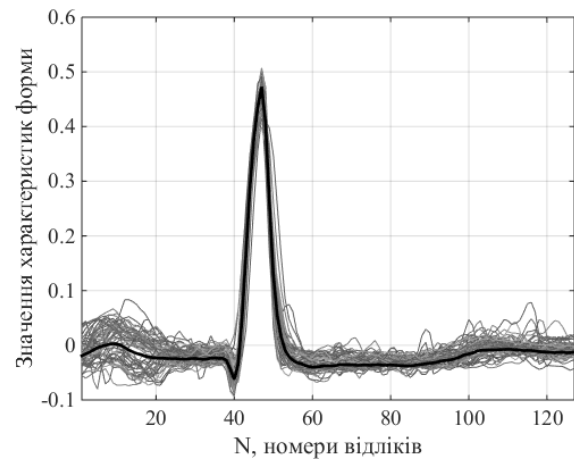


Рис. 2.3. Сигнали навчальної вибірки і еталон (жирна лінія) для QRS-комплексів A типу

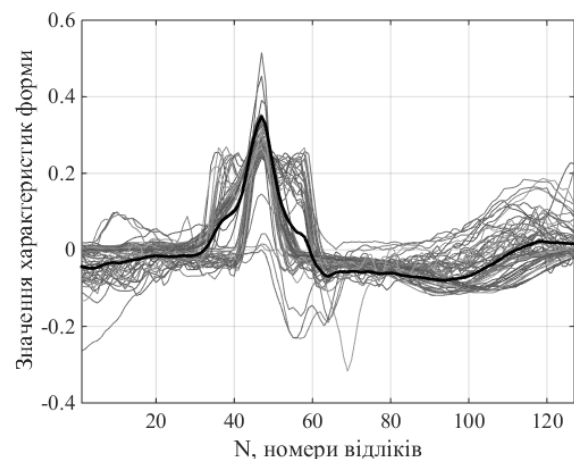


Рис. 2.4. Сигнали навчальної вибірки і еталон (жирна лінія) для QRS-комплексів V типу

Таблиця 2.2 - Оцінки вірності розпізнавання сигналів на контрольних вибірках у разі використання всіх реалізацій навчальних вибірок в якості еталонів

Тип	A	N	V	Спец.
A	56	4	1	91,80
N	6	58	1	89,23
V	0	0	60	100,00
Чутлив.	90,32	93,55	96,77	93,55%

У таблиці рішень, що отримані базовою розпізнає процедурою, відзначена наявність низки помилок. Найменшою є їх кількість для сигналів N типу і це підтверджує наявність для них найкращих умов розпізнавання використовуваною процедурою. Якраз це і було відзначено при аналізі розкиду сигналів навчальних вибірок.

Альтернативою розглянутої базової процедури з найбільш повним узагальненням сигналів навчальних вибірок та отриманням для них єдиних еталонів шляхом усереднення сигналів є процедура розпізнавання сигналів, в якій всі реалізації навчальних вибірок використовуються в якості еталонів і ніякого узагальнення їх форми не здійснюється.

Для такого крайнього альтернативного варіанту, в якому має місце тільки накопичення досвіду (збір достовірно розпізнаних сигналів, що утворюють навчальні вибірки), а його узагальнення не провадиться і всі реалізації вибірок відіграють роль еталонів, був складений окремий алгоритм і проведено додаткове дослідження вірності розпізнавання сигналів. Для тестового прикладу була отримана така таблиця рішень на контрольних вибірках (табл. 2.2).

Дані цієї таблиці розкривають рівень гранично досяжних значень показників вірності в розпізнаванні сигналів для тих процедур, які тим чи іншим способом узагальнюють навчальні вибірки у вигляді додаткових еталонів, формуються, для прийняття рішень, включаючи розглянуту базову процедуру. І дійсно, порівняння таблиць 2.1 і 2.2 показує, що вірність рішень базового алгоритму в цілому нижче.

Отримані в останньому випадку дані є свого роду верхнім орієнтиром по чутливості, специфічності і загальної валідності, обмеженням зверху в оцінці і в порівнянні можливостей модифікацій розпізнавальних процедур, що пропонуються, зокрема, модифікації з роботою над помилками, процедур з кластеризацією навчальних вибірок для різних метрик в фазових просторах сигналів і інших можливих модифікацій.

Далі пропонується розробка такої модифікації розглянутої базової процедури, яка враховує набутий досвід зроблених помилок в розпізнаванні сигналів на етапі її попереднього дослідження. Мається на увазі досвід роботи з навчальними вибірками. Результативність роботи над помилками перевіряється на вибірках контрольних.

III. МОДИФІКАЦІЯ БАЗОВОЇ ПРОЦЕДУРИ РОЗПІЗНАВАННЯ СИГНАЛІВ, ЯКА ВРАХОВУЄ ДОСВІД ПОМИЛКОВИХ РІШЕНЬ

Принцип врахування зроблених помилок може полягати в наступному. За результатами перевірки правильності розпізнавання сигналів різних типів на навчальних вибірках формуються підмножини реалізацій, щодо яких були зроблені однотипні помилкові рішення. Оскільки вибірки навчальні, то правильні рішення для таких підмножин відомі. Їх можна прикріпити до таких підмножин, їх окремим сигналам і еталонам і враховувати надалі при прийнятті рішень. У цьому сенс роботи над помилками.

Під однотипними тут розуміються помилки, в яких має місце одне і те ж поєднання дійсного типу сигналу і конкретного варіанту помилкового рішення. Наприклад, в одну з таких підмножин об'єднуються всі реалізації A типу, щодо яких розпізнавальна процедура прийняла одне і те ж саме рішення, що це сигнали типу N.

Після видалення з навчальних вибірок всіх реалізацій, що призводять до помилок, еталони решти для кожного типу сигналів перераховуються заново. У такому вигляді вони використовуються далі спільно з додатковими еталонами. Якщо помилкових рішень якогось типу немає, то немає і додаткових еталонів.

Таким чином, в загальному випадку для кожного класу сигналів може з'явитися додатковий еталон і, можливо, не один, але декілька, якщо сигнали, що призводять до однотипних помилкових рішень, сильно відрізняються між собою один від одного за їх формою. Додатковим еталоном може стати навіть єдина реалізація сигналу, що призводить до помилкового рішення.

Найпростішим способом роботи над помилками є переведення всіх сигналів, які призвели до помилок, до складу додаткових еталонів відповідних класів. Однак, якщо в зазначених групах сигналів виявилися близькі реалізації, то їх можна зібрати в підгрупи і сформувані для них додаткові еталони. Це знизить трудомісткість прийняття рішень.

Перший блок сформує з навчальних вибірок підмножини сигналів, за якими робляться помилки одного і того ж самого типу. У розглянутому прикладі, коли типів сигналів в їх класифікації всього три, таких підмножин шість. Їм відповідають шість недиагональних елементів таблиці рішень, яка буде отримана для навчальних вибірок. У ній буде підраховано кількість однорідних випадків, що призводять до однотипних помилок. Самі такі сигнали будуть зібрані в зазначені підмножини.

Другий блок в найпростішому варіанті визначить середній вид сигналів в групах з однорідними помилковими рішеннями. При більш детальному аналізі всередині таких однорідних груп будуть скомбіновані близькі реалізації і в таких комбінаціях будуть отримані свої додаткові еталони. Решта окремих поодиноких сигналів також



увійдуть до складу додаткових еталонів і будуть входити в сімейства еталонів своїх класів. У цій роботі блок два є резервним.

Уточнення початкових (основних) еталонів третім блоком в модифікованій процедурі виконується за тим же самим правилом оцінки середнього виду сигналів, що залишилися в групах з безпомилковими рішеннями. Неважко перевірити, що середній вид реалізацій в групі, які є характеристиками форми сигналів, являє собою також характеристику форми за своїми властивостями і вона може використовуватися при прийнятті рішень як еталон без додаткових перетворень.

При таких змінах в базовому алгоритмі і використанні ускладненої системи еталонів (основних і додаткових) процедура розпізнавання сигналів залишається в своїй основі такою ж самою. Черговий вхідний сигнал порівнюється з усіма основними і додатковими еталонами. Близькість їх форми оцінюється значеннями скалярних добутків. Рішення визначається максимумом розрахованої таким чином кореляції

Особливі варіанти форми сигналів, виражені додатковими еталонами, вказують на можливі місця їх появи в фазовому просторі. Мережа додаткових еталонів створюється для того, щоб повторювати особливості розподілу точок у фазовому просторі сигналів.

Перевірка результативності врахування досвіду помилок на контрольних вибірках за допомогою в розробленій програмній процедурі на даному тестовому прикладі відображена в табл. 3.1 - 3.2. Тут представлені кількість вірних і помилкових рішень для першої і десятої ітерацій роботи над помилками і значення відповідних показників чутливості, специфічності і загальної валідності прийнятих рішень.

Спостерігається, що в тестовому прикладі частково збереглися навчальні вибірки для оцінки основних еталонів. Модифікація процедури обробки даних не призвела до її виродження до випадку розпізнавання сигналів без узагальнення досвіду зроблених помилок. Досвід не тільки накопичується, але і узагальнюється. Після десятої ітерації зміна показників не виявлялася. Ітерації набули сталий характер.

Збіжність процесу в цих дослідженнях є гарантованою, оскільки в кожній ітерації оцінка основних еталонів йде з вибуванням зі складу реалізацій, що осереднюються, частки сигналів внаслідок їх переведення до складу додаткових еталонів, а їх число по ітераціях може лише зростати і в граничному разі охопити навчальні вибірки повністю. Ітеративний характер зміни оцінок чутливості, специфічності і валідності рішень в проведених дослідженнях відображений на графіках (рис. 3.1 - 3.3).

Отримані результати підтверджують працездатність розробленої модифікації базової процедури розпізнавання сигналів з таким урахуванням досвіду помилкових рішень. У тестовому прикладі навчальні вибірки для оцінки основних еталонів

частково збереглися. В цілому, робота над помилками за допомогою розробленої процедури на тестовому прикладі виявилася вдалою.

У порівнянні з базовою розпізнавальною процедурою в цьому прикладі підтверджена можливість підвищення вірності прийняття рішень за рахунок врахування досвіду помилок, який отримано на випробуваннях розпізнавальної процедури на навчальних вибірках.

Зростання загальної валідності після проведених ітерацій в порівнянні з базовою процедурою тут склав приблизно 0,5%, хоча в перших ітераціях показники вірності рішень були значно гірше, ніж для базового алгоритму, і додаткові еталони захоплювали «чужі» сигнали при їх розпізнаванні в контрольних вибірках. П'яти ітерацій врахування помилок тут виявилася достатнім, щоб завершити цей процес і вийти на високий рівень показників вірності рішень.

Таблиця 3.1 - Оцінки вірності розпізнавання сигналів на контрольних вибірках з урахуванням досвіду помилкових рішень, в 1-й ітерації

Тип	A	N	V	Спец.
A	34	62	3	34,34
N	9	0	0	0,00
V	19	0	59	75,64
Чутлив.	54,84	0,00	95,16	50,00%

Таблиця 3.2 - Оцінки вірності розпізнавання сигналів на контрольних вибірках з урахуванням досвіду помилкових рішень, в 10-й ітерації

Тип	A	N	V	Спец.
A	45	1	0	97,83
N	17	61	1	77,22
V	0	0	61	100,00
Чутлив.	72,58	98,39	98,39	89,78%

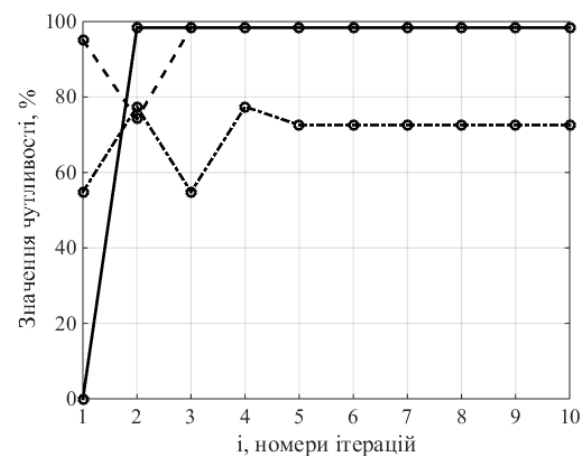


Рис. 3.1. Залежність чутливості розпізнавання сигналів від номеру ітерації (N – безперервна лінія, A – пунктир-крапка, V – пунктир)

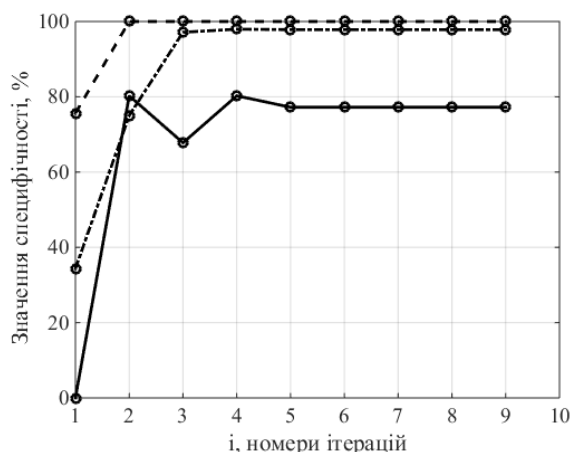


Рис. 3.2. Залежність специфічності розпізнавання сигналів від номеру ітерації (N – безперервна лінія, A – пунктир-крапка, V- пунктир)

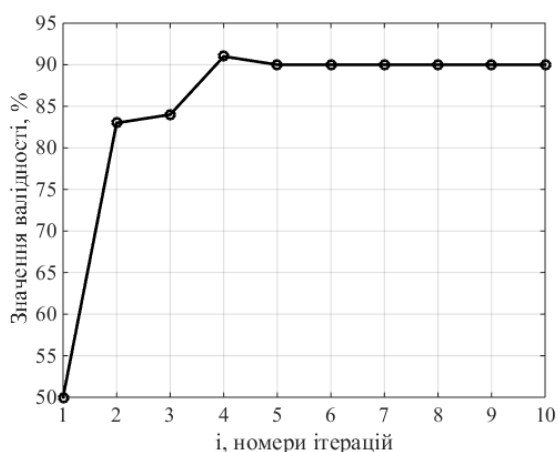


Рис. 3.3. Залежність валідності розпізнавання сигналів від номеру ітерації.

Детальне врахування варіацій форми сигналів в їх класах з метою підвищення вірності рішень може також здійснюватися шляхом навмисного розкладання всіх навчальних вибірок на дрібні групи близьких реалізацій для покриття фазового простору цілими сімействами еталонів в місцях скупчення точок і це може бути реалізовано з використанням відомих процедур кластеризації навчальних вибірок [14 – 16], яка буде вносити в обробку сигналів елементи самонавчання [1 – 5]. Таким способом буде поєднуватися накопичення досвіду розпізнавання сигналів (нарощування навчальних вибірок) з його узагальненням (розкриттям особливостей розподілу достовірно розпізнаних сигналів в фазовому просторі їх форми). Цей додатковий, проміжний варіант узагальнення повинен бути більш детальним і тому більш інформативним, ніж варіант з використанням єдиних узагальнених еталонів, як в базовій розпізнавальній процедурі, і це обговорюється нижче.

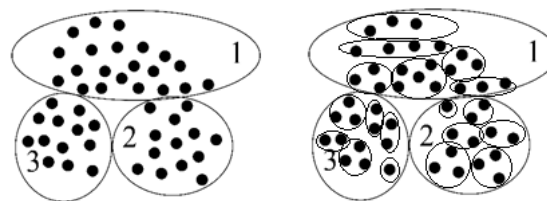


Рис. 4.1. Урахування особливостей локалізації сигналів навчальних вибірок при їх роздільній кластеризації

IV. ПРОЦЕДУРА РОЗПІЗНАВАННЯ СИГНАЛІВ З КЛАСТЕРИЗАЦІЄЮ НАВЧАЛЬНИХ ВИБІРОК ЗА ЕВКЛІДОВОЮ ВІДСТАННЮ ПРИ КОМБІНОВАНОМУ НАВЧАННІ

Під час розкриття принципу обробки даних у запропонованій модифікації розпізнавального алгоритму будемо розглядати кластеризацію навчальних вибірок як дію, яка спрямована на розкриття характеру і особливостей розподілу сигналів кожного типу в просторі опису їх форми шляхом об'єднання реалізацій в підмножини близько розташованих точок з їх прив'язкою до місць скупчень в такому фазовому просторі (рис. 4.1).

Вихідна класифікація сигналів, що задана вчителем і представлена на цьому рисунку зліва типами 1, 2 і 3, тут не порушується кластеризацією (рисунок праворуч). Розбиття на підмножини виконуються всередині кожного класу і без урахування точок інших класів. Це не заважає бачити і враховувати можливі накладення областей кластерів різних класів сигналів один на одного при прийнятті рішень. Локалізація скупчень точок отриманих кластерів в подальшому передається локалізацією кластерних (локальних) еталонів, місце розташування яких узагальнює вміст кластерів. Сигнали, що надійшли знову та розпізнаються, потрапляють в фазовому просторі локально в оточення еталонів з різних класів, які розташовані до них найближчим чином відповідно до реальних імовірнісних розподілів для конкуруючих версій рішень. Розрізненість типів сигналів локально поліпшується. Більш далекі за евклідовою відстанню локальні еталони перестають впливати на формування рішень. Умови та надійність прийняття рішень за обстановкою поблизу стають легшими і сприятливими, ніж у разі роботи з єдиними еталонами в повних навчальних множинах кожного типу. І це спрямовано на поліпшення показників вірності розпізнавання сигналів.

У визначенні типу кожного сигналу беруть участь всі локальні еталони. Їх кількість тут задається числом типів сигналів і обсягом використовуваних для них навчальних вибірок. По відношенню до базової процедури алгоритм стає більш громіздким. Час обробки даних зростає. Проте, якщо еталонів «занадто багато», то можна провести їх кластеризацію, в точності так само, як самих сигналів. Але це може стати на шкоду вірності рішень, тому доцільна додаткова перевірка на контрольних вибірках. Розкритий принцип обробки даних реалізує процедура, блок-схема якої надана на рис. 4.2.

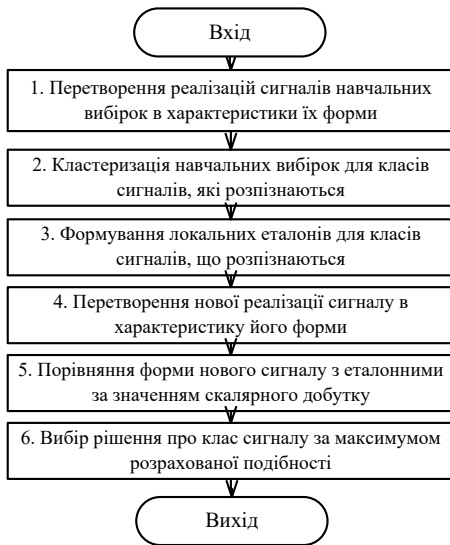


Рис. 4.2. Блок-схема процедури розпізнавання сигналів з кластеризацією навчальних вибірок за евклідовою відстанню

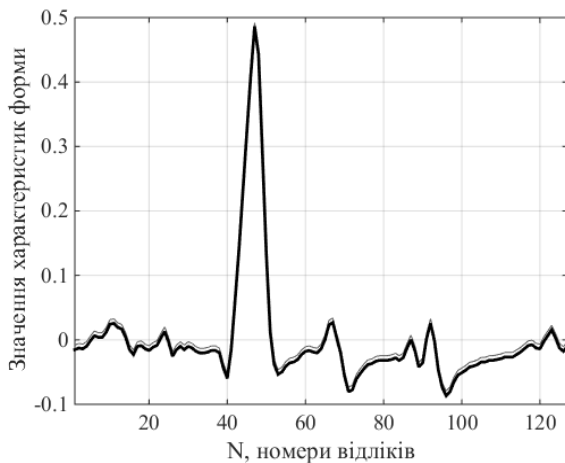


Рис. 4.3. Реалізації сигналів кластера і його локальний еталон (жирна лінія) для QRS-комплексів N типу

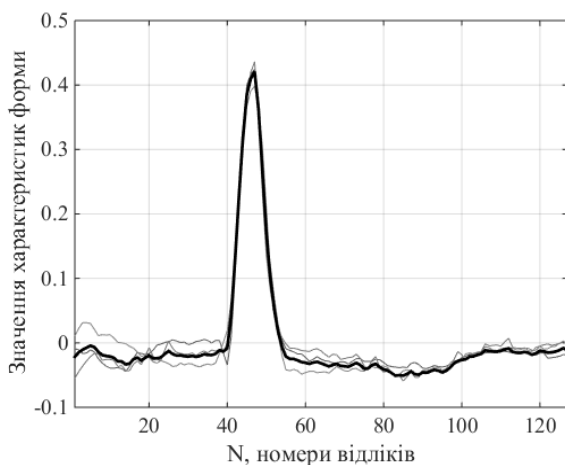


Рис. 4.4. Реалізації сигналів кластера і його локальний еталон (жирна лінія) для QRS-комплексів A типу

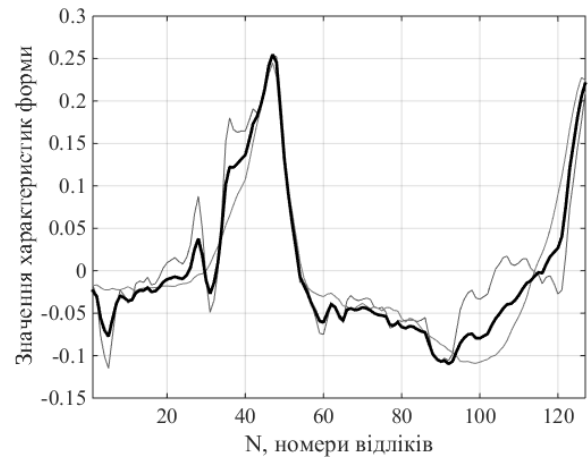


Рис. 4.5. Реалізації сигналів кластера і його локальний еталон (жирна лінія) для QRS-комплексів V типу

Приклади результатів кластеризації (реалізації сигналів кластерах для різних їх типів) і відповідні локальні еталони (показані жирними лініями) зображені на рис. 4.3 – 4.5.

В кластери сигналів різних типів збирається одна і та ж кількість найбільш близьких за формою сигналів навчальних вибірок за допомогою стандартної процедури k-середніх. Відстань між елементами всередині кластерів при цьому мінімізується, міжкластерні відстані збільшуються [14 – 16].

Центроїди кластерів локалізуються в фазовому просторі відповідно до особливостей розподілу сигналів свого типу в кожному класі. Вибірки сигналів різних типів кластеризуються окремо.

У тестовому прикладі найменш варіабельна форма сигналів N типу.

Локальні еталони являють собою середній вид сигналів кластерів і за своїми властивостями є характеристиками форми сигналів.

Отримані для тестового прикладу сімейства локальних (кластерних) еталонів демонструються на рис. 4.6 – 4.8.

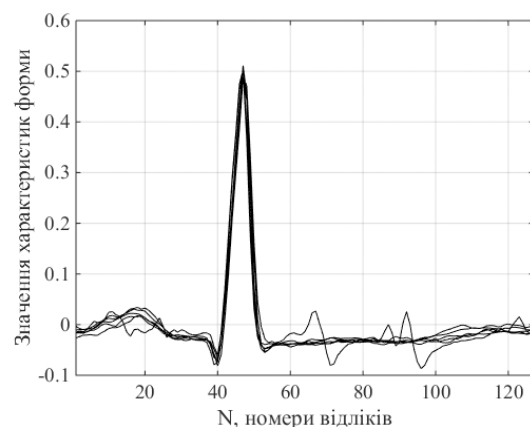


Рис. 4.6. Сімейства локальних еталонів для QRS-комплексів N типу при кластеризації навчальних вибірок за евклідовою відстанню

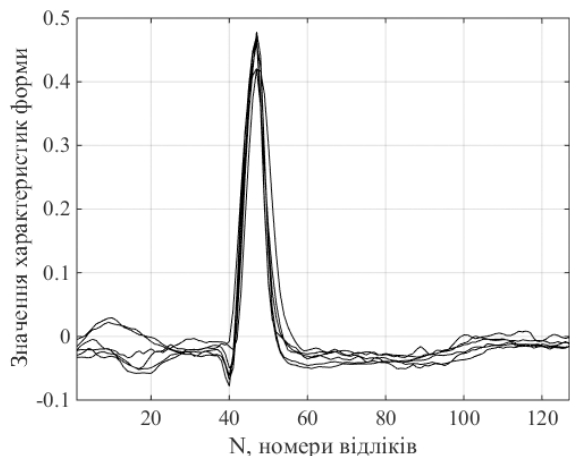


Рис. 4.7. Сімейства локальних еталонів для QRS-комплексів А типу при кластеризації навчальних вибірок за евклідовою відстанню

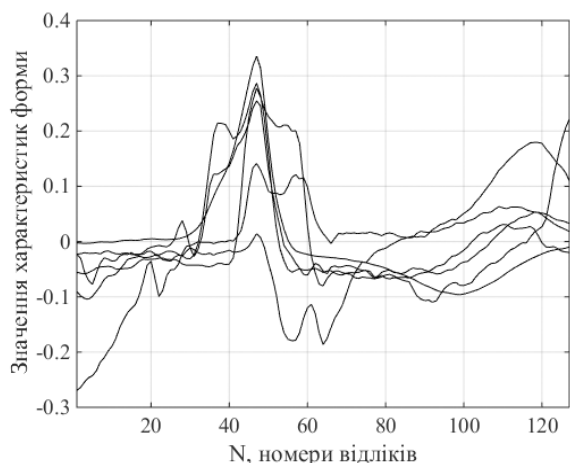


Рис. 4.8. Сімейства локальних еталонів для QRS-комплексів V типу при кластеризації навчальних вибірок за евклідовою відстанню

Локальні еталони N типу найменш відмінні один від одного. Найбільш суттєвий внесок в підвищення вірності рішень за рахунок розшарування еталонів з охопленням значних варіацій форми очікується для сигналів V типу

Результативність роботи розглянутої модифікації розпізнавальної процедури перевірена на контрольних вибірках для тестового прикладу. Результати відображені табл. 3.1.

Рівень вірності рішень в цьому разі є близьким до гранично досяжного.

ТАБЛИЦЯ 3.1 - Оцінки вірності розпізнавання сигналів на контрольних вибірках процедурою з кластеризацією навчальних вибірок за евклідовою відстанню

Тип	A	N	V	Спец.
A	56	4	2	90,32
N	6	58	0	90,63
V	0	0	60	100,00
Чутлив.	90,32	93,55	96,77	93,55%

Кластеризація навчальних вибірок в порівнянні з базовою процедурою з узагальненими еталонами дає помітний позитивний ефект: чутливість, специфічність і загальна валідність розпізнавання зростають.

Для запропонованої модифікації розпізнавальної процедури досліджені також залежності кількості правильних рішень від кількості сигналів в кластерах. Дослідження проводилися на контрольних вибірках тестового прикладу. Графіки представлені на рис. 4.9 - 4.11.

Відсоток правильних рішень, як впливає з графіків, може змінюватися в широких межах. При необхідності можуть бути отримані аналогічні залежності для чутливості і специфічності та зроблений доцільний вибір режиму обробки даних.

Аналогічні дослідження має сенс проводити для діагностичних сигналів будь-якого виду з урахуванням використовуваної їх класифікації.

У разі проведення подібних досліджень треба також брати до уваги обсяги навчальних та контрольних вибірок, які обробляються.

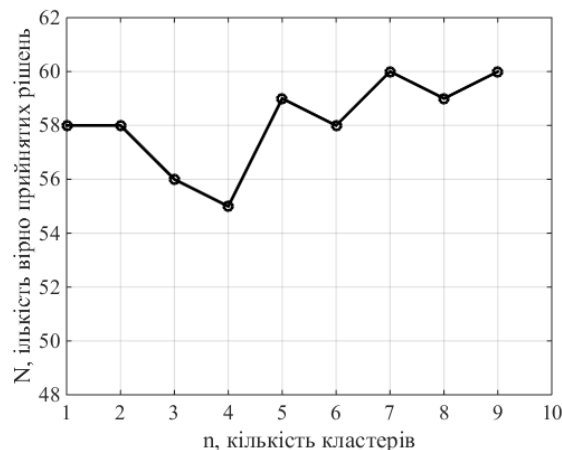


Рис. 4.9. Оцінки показників вірності розпізнавання сигналів типу N в залежності кількості сигналів в кластерах при кластеризації за евклідовою відстанню

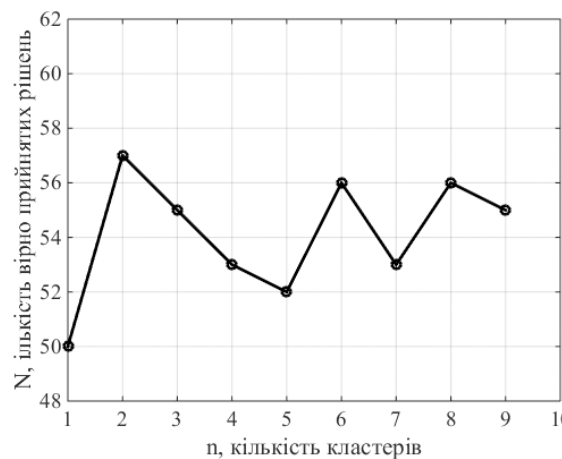


Рис. 4.10. Оцінки показників вірності розпізнавання сигналів типу A в залежності кількості сигналів в кластерах при кластеризації за евклідовою відстанню



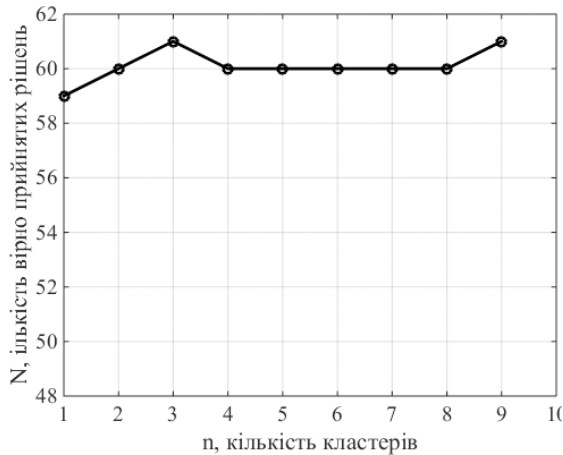


Рис. 4.11. Оцінки показників вірності розпізнавання сигналів типу V в залежності кількості сигналів в кластерах при кластеризації за евклідовою відстанню

Нижче розглядається варіант кластеризації сигналів з метрикою, заснованої на скалярному добутку сигналів, який використовується в критерії прийняття рішень про їхній тип.

V. ПРОЦЕДУРА РОЗПІЗНАВАННЯ СИГНАЛІВ З КЛАСТЕРИЗАЦІЄЮ НАВЧАЛЬНИХ ВИБІРОК ЗА КРИТЕРІЄМ ПОДІБНОСТІ СИГНАЛІВ ПРИ КОМБІНОВАНОМУ НАВЧАННІ

Замість евклідової відстані в стандартній процедурі кластеризації в алгоритмі розпізнавання сигналів використовується критерій α_z .

Відстань між точками в цьому разі характеризується величиною $\alpha_z = 1 - z^*$, де z^* результат перерахунку скалярного добутку z (схожості) векторів, відстань між якими оцінюється: $z^* = (1/2) \cdot (z+1)$

Перерахунок z в z^* є перетворенням шкали $[-1; 1]$ для z в шкалу $[0; 1]$ для z^* .

Таким чином, величина z^* виражає схожість між векторами в межах від 0 до 1, а $1 - z^* = \alpha_z$ набуває сенсу їх відмінності (відстані) між точками в фазовому просторі.

В іншому ця модифікація процедури розпізнавання сигналів повторює попередню процедуру з кластеризацією за евклідовою відстанню.

Метрика простору тут побудована на використанні критерію подібності форми сигналів (взаємної кореляції їх форми, що пояснює узяття назву для цієї модифікації алгоритму.

Приклади сигналів кластерів і їх локальні еталони для комплексів різного типу у разі кластеризації з метрикою α_z наводяться на рис. 5.1 – 5.3.

Графіки показують, що в порівнянні з випадком використання евклідової метрики склад кластерів і локальні еталони неідентичні.

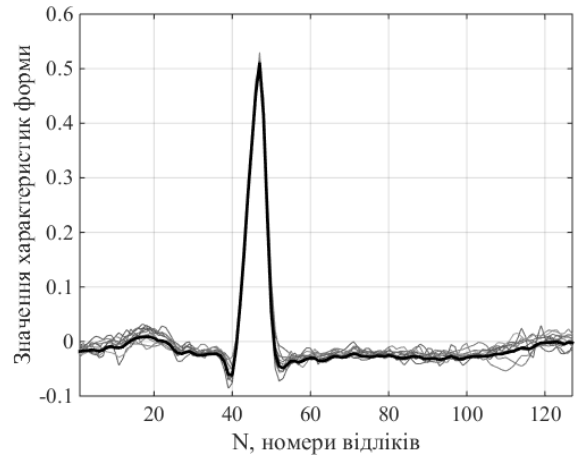


Рис. 5.1. Сигнали кластера і локальний еталон для комплексів N типу при кластеризації з метрикою α_z

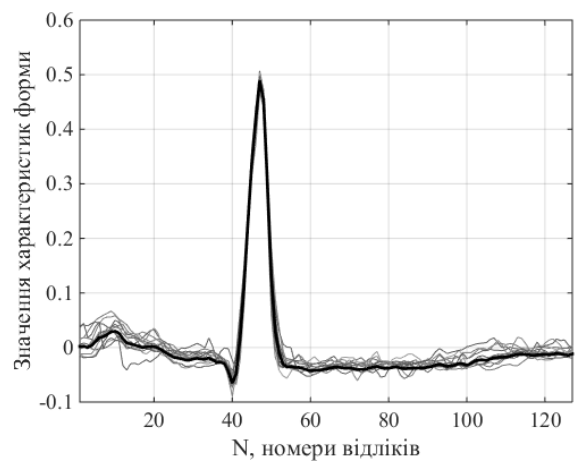


Рис. 5.2. Сигнали кластера і локальний еталон для комплексів A типу при кластеризації з метрикою α_z

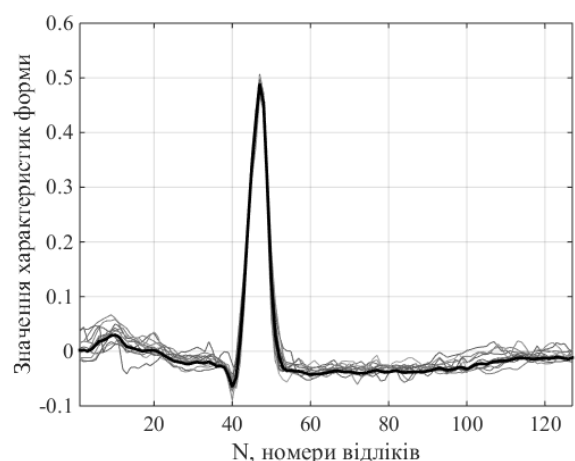


Рис. 5.3. Сигнали кластера і локальний еталон для комплексів V типу при кластеризації з метрикою α_z

Розроблена модифікація розпізнавальної процедури працездатна. Отримані для тестового прикладу сімейства локальних (кластерних) еталонів демонструються на рис. 5.4 - 5.6.

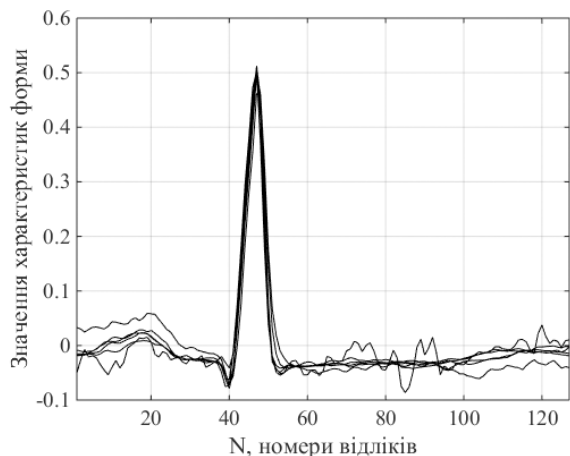


Рис. 5.4. Сімейства локальних еталонів для комплексів N типу при кластеризації з метрикою α_z

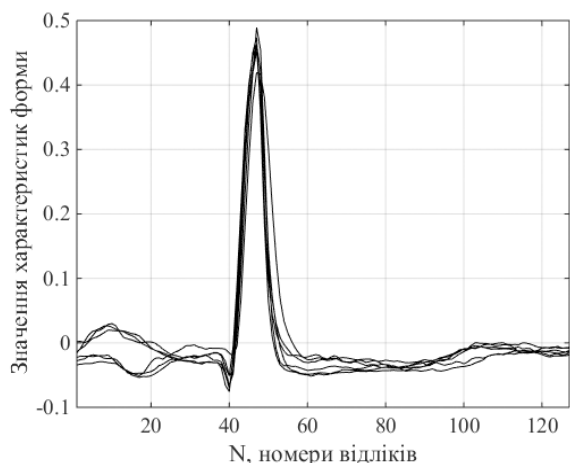


Рис. 5.5. Сімейства локальних еталонів для комплексів A типу при кластеризації з метрикою α_z

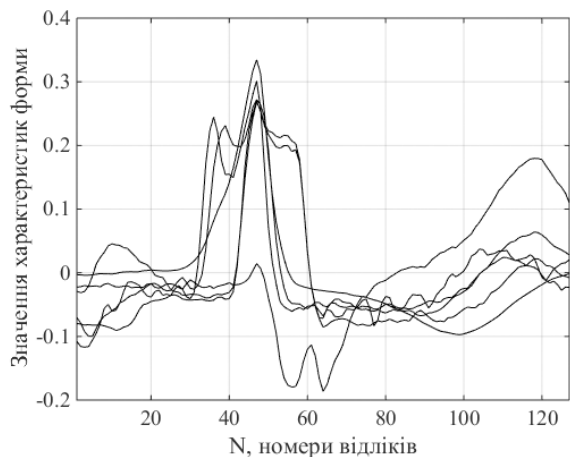


Рис. 5.6. Сімейства локальних еталонів для комплексів V типу при кластеризації з метрикою α_z

Результативність роботи розглянутої модифікації розпізнавальної процедури перевірена на

контрольних вибірках для тестового прикладу. Результати відображені табл. 5.1.

Досягнутий рівень вірності рішень близький до гранично досяжного в цьому разі.

Для цієї модифікації розпізнавальної процедури на тих же самих контрольних вибірках досліджені також залежності показників вірності розпізнавання сигналів (чутливості, специфічності, загальної валідності) від кількості сигналів в кластерах (рис. 5.7 – 5.9).

ТАБЛ. 5.1 - Оцінки вірності розпізнавання сигналів на контрольних вибірках при кластеризації з метрикою α_z

Тип	A	N	V	Спец.
A	56	3	1	93,33
N	6	59	1	89,39
V	0	0	60	100,00
Чутлив.	90,32	95,16	96,77	94,09%

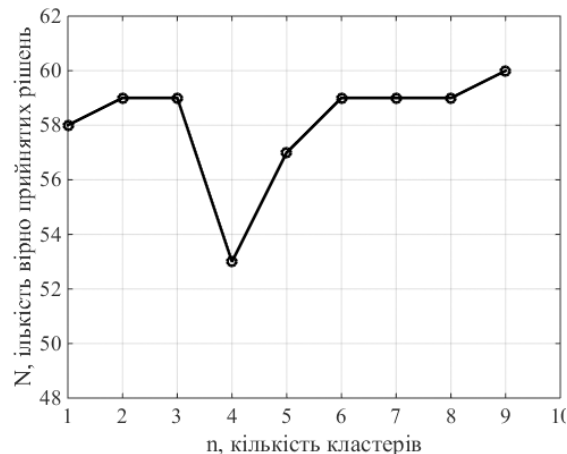


Рис. 5.7. Оцінки показників вірності розпізнавання сигналів N типу в залежності кількості сигналів в кластерах при кластеризації з метрикою α_z

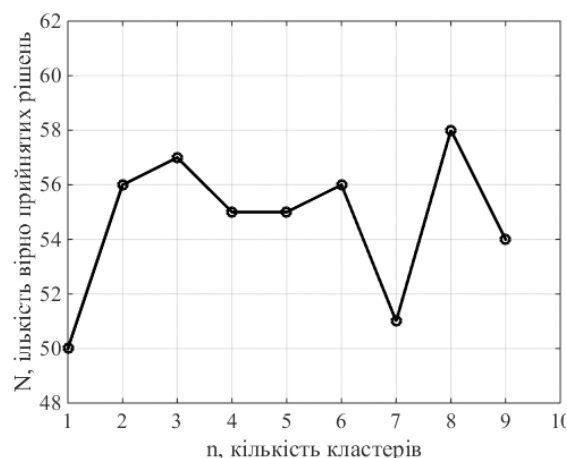


Рис. 5.8. Оцінки показників вірності розпізнавання сигналів A типу в залежності кількості сигналів в кластерах при кластеризації з метрикою α_z



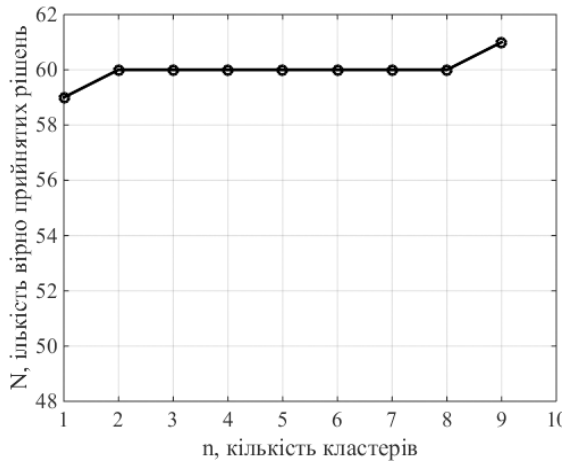


Рис. 5.9. Оцінки показників вірності розпізнавання сигналів V типу в залежності кількості сигналів в кластерах при кластеризації з метрикою α_z

Ці та аналогічні дані по чутливості і специфічності можуть бути використані для вибору умов кластеризації навчальних вибірок.

Отримані результати показують, що представлені модифікації процедур розпізнавання не ідентичні, хоча і близькі (в розглянутому прикладі) за своїми можливостями поліпшення умов прийняття рішень.

Використання критерію подібності форми сигналів в конструюванні метрики для їх кластеризації виявилось продуктивним.

VI. ОЦІНКА РЕЗУЛЬТАТИВНОСТІ ФОРМУВАННЯ ПІДМНОЖИН НАВЧАЛЬНИХ ВИБІРОК В ПОРЯДКУ НАКОПИЧЕННЯ СИГНАЛІВ ДОСТОВІРНИХ ТИПІВ

Це дослідження має на меті підтвердження результативності кластеризації як способу групування сигналів навчальних вибірок в локаціях їх скупчення для виявлення характеру і особливостей розподілу сигналів навчальних вибірок задля підвищення якості розпізнавання сигналів.

Спроможність кластеризації підтверджується методом від протилежного. Як альтернатива кластеризації розглядається досить довільне формування угруповань точок той же кількості з навчальних вибірок без урахування їх взаємного розташування в фазовому просторі, тут – в порядку їх запису в навчальні вибірки з постійним кроком відбору в окремі підмножини.

Таких підмножин буде стільки ж, що і при кластеризації. У кожному класі буде отримано таку ж кількість еталонів. Всі вони будуть використані при розпізнаванні сигналів тих же самих контрольних вибірок. Будуть розраховані ті ж за складом показники вірності рішень. І їх порівняння з показниками при кластеризації допоможе підтвердити її роль в поліпшенні умов для прийняття рішень.

Загальний порядок обробки даних та оцінки результатів залишаємо незмінним. Міняємо тільки порядок формування підмножин в навчальних

вибірках кожного класу сигналів. У зв'язку з цим еталони втрачають прив'язку до скупчень точок. Для оцінки наслідків переглядаємо результати того ж складу, що і у випадку з кластеризацією.

Сигнали і еталон одного з підмножин навчальних вибірок в кожному класі сигналів покажемо на рис. 6.1 - 6.3.

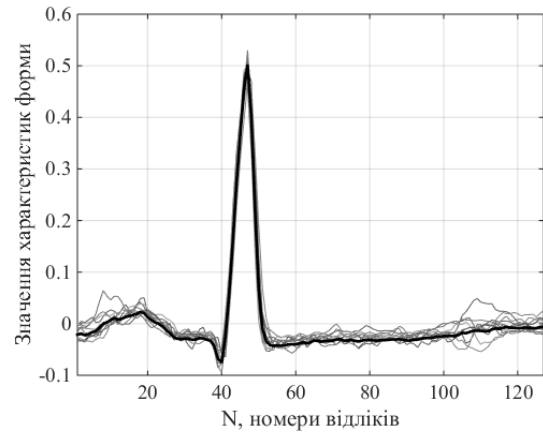


Рис. 6.1. Сигнали підмножини і локальний еталон для комплексів N типу без кластеризації

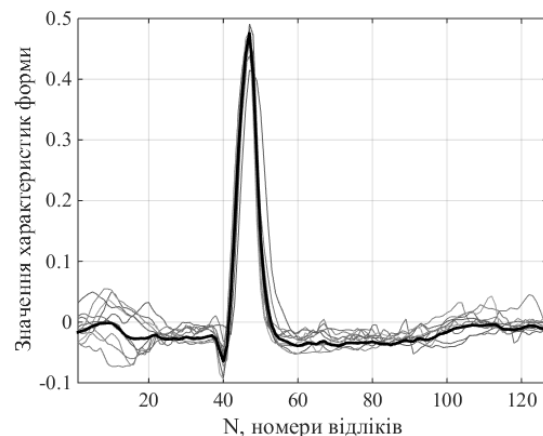


Рис. 6.2. Сигнали підмножини і локальний еталон для комплексів A типу без кластеризації

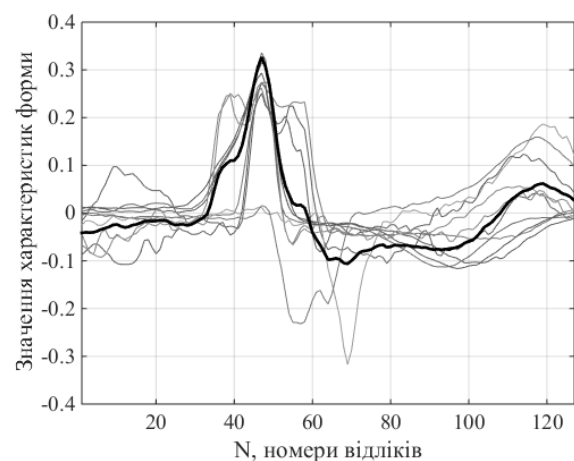


Рис. 6.3. Сигнали підмножини і локальний еталон для комплексів V типу без кластеризації

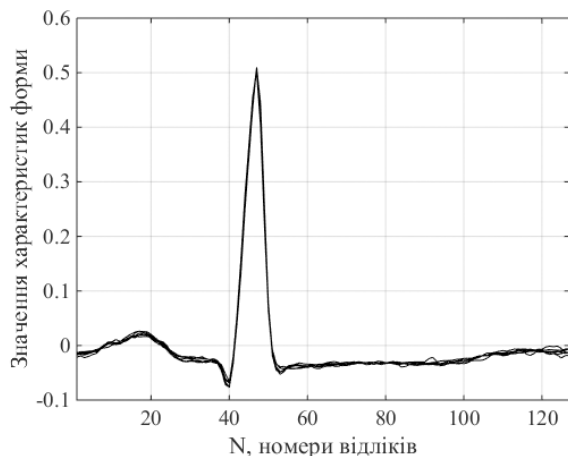


Рис. 6.4. Сімейства локальних еталонів N типу для підмножин без кластеризації

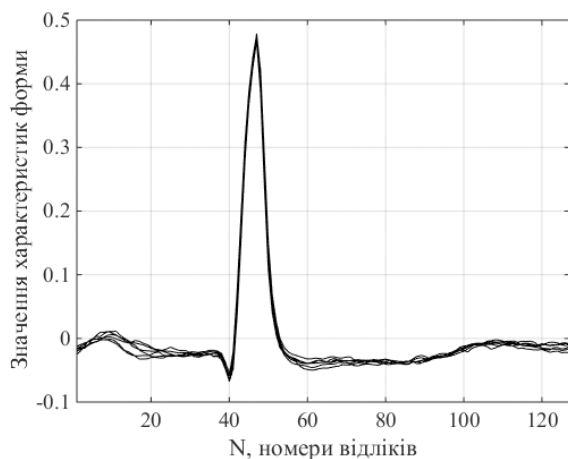


Рис. 6.5. Сімейства локальних еталонів A типу для підмножин без кластеризації

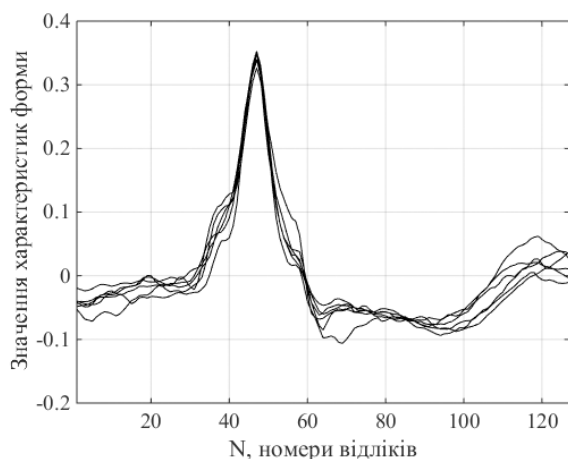


Рис. 6.6. Сімейства локальних еталонів V типу для підмножин без кластеризації

Сімейства еталонів, розрахованих з цих підмножин, зображено на рис. 6.4 – 6.6.

Тут відзначається збільшення різниці реалізацій в групах сигналів N типу. В порівнянні

з випадками з кластеризацією. Розкид сигналів інших типів залишається істотним.

Вірність розпізнавання сигналів при такому їх підборі в підмножини розкриємо в табл. 6.1.

Результуюча вірність рішень в порівняннях з варіантом з кластеризацією тут стає нижче, що і очікувалося і це можна розцінювати як прояв втрати в цьому разі більш систематичного відстеження реальних розподілів сигналів, що мало місце в процедурах з кластеризацією. Очікуваний результат підтверджений даними дослідженням на конкретному прикладі. Кластеризація навчальних вибірок – результативний спосіб забезпечення високого рівня вірності розпізнавання сигналів в системах, які навчають з учителем.

ТАБЛИЦЯ 6.1 - ПОКАЗНИКИ ВІРНОСТІ РОЗПІЗНАВАННЯ СИГНАЛІВ ПРИ ДОВІЛЬНОМУ ЇХ РОЗБИТТІ НА ПІДМНОЖИНИ ДЛЯ ОТРИМАННЯ СІМЕЙСТВ ЕТАЛОНІВ

Тип	A	N	V	Спец.
A	48	2	2	92,31
N	13	60	2	80,00
V	1	0	58	98,31
Чутлив	77,42	96,77	93,55	89,25%

ВИСНОВКИ

Основні результати роботи полягають у наступному.

- Для систем, що навчаються з учителем, побудовані, програмно реалізовані і на тестовому прикладі розпізнавання типів QRS-комплексів в електрокардіограмі пацієнта перевірені модифікації процедур розпізнавання медико-біологічних сигналів по характеристикам їх форми зі скалярним добутком сигналів в оцінці їх близькості по взаємній кореляції.
- Перевірки працездатності розроблених процедур, оцінки результативності використання запропонованих модифікацій базового алгоритму побудовані на статистичному дослідженні показників чутливості, специфічності і загальної валідності прийнятих рішень шляхом прямого підрахунку числа правильних і помилкових рішень з визначенням типів помилок на контрольних вибірках після проведення навчання на вибірках навчальних.

Основні висновки.

- 1) Урахування досвіду помилок базового розпізнавального алгоритму, які знайдені при його випробуваннях на навчальних вибірках, при прийнятті рішень в подальшому шляхом оголошення правильних рішень введенням відповідних додаткових еталонів істотно ускладнює розпізнає процедуру. Доцільність використання такої модифікації залежить від форми сигналів, що розпізнаються, особливостей їх розподілу в фазовому просторі

і повинна встановлюватися в кожному конкретному випадку шляхом проведення випробувань зі статистичною оцінкою чутливості, специфічності та інших показників вірності рішень.

- 2) Комбіноване навчання з роздільною кластеризацією навчальних вибірок є проміжний варіантом, в якому накопичення досвіду розпізнавання сигналів поєднується з його узагальненням. Полярними варіантами в цьому разі є використання всіх навчальних реалізацій як еталонів (тільки накопичення досвіду) і єдиних еталонів в кожній вибірці (повне узагальнення навчальних вибірок).
- 3) Комбіноване навчання поєднує в собі збереження класифікації медико-біологічних сигналів, що має зрозуміле медичне тлумачення, із забезпеченням додаткової деталізації в урахованні розподілу сигналів в фазовому просторі через вибір еталонів і це призводить до підвищення вірності розпізнавання сигналів в порівнянні показниками базової процедури, в якій узагальнення навчальних вибірок в еталонах є повним.
- 4) Запропонований варіант кластеризації вибірок з авторською метрикою α_z за своєю результативністю близький до випадку роботи в евклідовій метриці. Така модифікація процедур обробки даних може розглядатися як додатковий інструмент в розпізнавальних системах, які навчаються з учителем.
- 5) Перехід від кластеризації навчальних вибірок, що об'єднує в групи близькі сигнали, до довільного формування таких груп для формування локальних еталонів знижує в загальному випадку рівень правильності рішень з розпізнавання сигналів, що підтверджує результативність кластеризації і можливість досягнення з її допомогою практично гранично високих показників.
- 6) Запропонований програмний інструментарій при проведенні дослідницьких робіт може сприяти розкриттю і оцінці можливих резервів у підвищенні вірності розпізнавання медико-біологічних сигналів в системах діагностики пацієнтів. Перспективним в розвитку отриманих результатів є дослідження результативності розпізнавальних процедур залежно від режиму кластеризації навчальних вибірок.

ПЕРЕЛІК ПОСИЛАНЬ

- [1] Genkin, A.A. Novaya informacionnaya tekhnologiya analiza medicinskih dannyh (programnyj kompleks OMIS) / A. A. Genkin. – SPb.: Politehnika, 1999. – 191 s.: il.
- [2] V. A. Dyuk, V. L. Emanuel', Informacionnye tekhnologii v mediko-biologicheskikh issledovaniyah, SPb.: Piter, 2003, 528 s.
- [3] Antomonov M. YU. Matematicheskaya obrabotka i analiz mediko-biologicheskikh dannyh [Tekst] / M. YU. Antomonov.- M.: Nauka, 2005. – 558 s.

- [4] Vasil'ev V. I. Raspoznayushchie sistemy. Spravochnik. / V. I. Vasil'ev. – K.: Naukova Dumka, 1983. – 421 s.: il.
- [5] Mincer O. P. Informativni tekhnologii v ohoronі zdorov'ya i praktichnij medicini: navch. posib. / O. P. Mincer, YU. V. Voronenko, V. V. Vlasov. – K.: Vishcha shkola, 2003. – 350 s.: il.
- [6] Korostelev, A. A. Teoreticheskie osnovy radiolokacii: uchebn. posobie dlya vuzov / A. A. Korostelev, N. F. Klyuev, YU. A. Mel'nik, A. A. Vegetyagin, V. A. Gubin, V. E. Dulevich i dr. – M.: Sov. radio, 1978. – 608 s.
- [7] Shulyak A.P., Shachykov A.D., Criteria and procedures for estimation the informativity and feature selection in biomedical signals for the irrecognition / «Visnyk NTUU KPI» Seriya – Radiotekhnika. Radioaparaturbuduvannya. 2016. – Вип. 66. – с. 79 – 86. DOI: [10.20535/RADAP.2016.66.79-86](https://doi.org/10.20535/RADAP.2016.66.79-86)
- [8] Shachykov A. D., Shulyak, A. P. "Analysis of the Biomedical Signals' of the Problem of Recognition", 2014 IEEE XXXIV International Scientific Conference Electronics and Nanotechnology (ELNANO), April 15-18, 2014, Kyiv, Ukraine. p. 281-285. DOI: [10.1109/ELNANO.2014.6873982](https://doi.org/10.1109/ELNANO.2014.6873982)
- [9] Shachykov A. D., Shulyak, A. P. Decomposition of Biomedical Signals on Mutually Orthogonal Components in the Diagnosis of Diseases", 2014 IEEE XXXIV International Scientific Conference Electronics and Nanotechnology (ELNANO), April 15-18, 2014, Kyiv, Ukraine, p. 291-294. DOI: [10.1109/ELNANO.2014.6873914](https://doi.org/10.1109/ELNANO.2014.6873914)
- [10] Shulyak A.P., Shachykov A.D. Otrabotka pryntsyypov analiza struktury tsyklycheskykh medyko-byolohycheskykh syhnalov dlya ykh obnaruzhenyya, raspoznavanyya y klasyfykatsyy [Testing the principles of analysis of the structure of cyclic biomedical signals for their detection, recognition and classification]/ Shachykov A.D., Shulyak A.P., Zbirnyk naukovykh prats' «VISNYK NTUU «KPI». seriya prykladobuduvannya», випуск 49(1), 2015р. – 169 – 179 с. DOI: [10.20535/1970.49\(1\).2015.47101](https://doi.org/10.20535/1970.49(1).2015.47101)
- [11] Shachykov A. D., Shulyak, A. P. "About the Impact of Informative Features Selection in the Mutually Orthogonal Decompositions of Biomedical Signals for their Recognition", 2016 IEEE 36th International Conference ELECTRONICS and NANOTECHNOLOGY (ELNANO), Kyiv, Ukraine, April 19-21, 2016, p. 228-231. DOI: [10.1109/ELNANO.2016.7493054](https://doi.org/10.1109/ELNANO.2016.7493054)
- [12] Shachikov A. D., Shelofast V. A., SHulyak O. P. «Modifikacii procedur othora priznakov ciklicheskih mediko-biologicheskikh signalov dlya ih raspoznavaniya», «Visnik NTUU KPI» seriya priladobuduvannya», vip. 53(1), K.: 2017, s. 103-109. DOI: [10.20535/1970.53\(1\).2017.106810](https://doi.org/10.20535/1970.53(1).2017.106810)
- [13] Shulyak A.P., SHachikov A.D. Osoblivosti vikoristannya harakteristiki formi mediko-biologichnih signaliv pri ih rozpoznavanni/ SHachikov A.D., SHulyak A.P., Zbirnyk naukovykh prac' «VISNYK NTUU «KPI». seriya priladobuduvannya», випуск 51(1), 2016р. – 131 – 139 с. DOI: [10.20535/1970.51\(1\).2016.78217](https://doi.org/10.20535/1970.51(1).2016.78217)
- [14] Mandel I. D. Cluster analysis. – M.: Finance and Statistics, 1988. – 176 p.
- [15] Kotov A., Krasilnikov N. Data clustering. 2006.
- [16] Duran B., Odell P. Cluster analysis – M.: Statistics, 1977. – 128 p.
- [17] PhysioNet. St.-Petersburg Institute of Cardiological Technics 12-lead Arrhythmia Database. <http://www.physionet.org/cgi-bin/atm/ATM>
- [18] Potemkin V.G. Vvedenie v MATLAB. – M.: Dialog-MIFI, 2000
- [19] D'yakonov V. P. MATLAB. Polnyj samouchitel'. – M.: DMK Press, 2012. – 768 s.: il.
- [20] Sergienko A. B. Cifrovaya obrabotka signalov. – SPb, Piter, 2002. — 608 s.:

Надійшла до редакції 17 листопада 2019 р.



УДК 61:681.5

Кластеризация медико-биологических сигналов в распознавательных системах, обучающихся с учителем

Лагутин^f В. В., ORCID [0000-0003-3431-1686](https://orcid.org/0000-0003-3431-1686)Сиротенко^f В. И., ORCID [0000-0002-7542-5885](https://orcid.org/0000-0002-7542-5885)Шачиков^s А. Д., к.т.н., ORCID [0000-0003-2353-7647](https://orcid.org/0000-0003-2353-7647)Шуляк^s А. П., к.т.н. доц., ORCID [0000-0002-6920-2212](https://orcid.org/0000-0002-6920-2212)Кафедра электронной инженерии ee.kpi.ua

Национальный технический университет Украины

«Киевский политехнический институт имени Игоря Сикорского» kpi.ua

Киев, Украина

Аннотация—Рассматривается распознавание медико-биологических сигналов по характеристикам их формы с помощью алгоритмов, обучаемыми с учителем, в системах диагностики пациентов. Классификация сигналов полагается известной, обучающие выборки – доступными. Предполагается статистическая оценка верности решений о типах сигналов на контрольных выборках по показателям чувствительности, специфичности, общей валидности.

Раскрываются принцип и порядок совершенствования взятой за основу базовой процедуры распознавания сигналов. Разрабатывается и исследуется программный инструментарий для выявления и реализации возможных резервов в повышении верности решений за счет более детального учета характера и особенностей вероятностных распределений сигналов в фазовом пространстве. С этой целью отдельно по классам сигналов дополнительно осуществляется кластеризация обучающих выборок на локациях их скопления. Формируются привязанные к этим локациям семейства кластерных эталонов и они используются при определении типов поступающих сигналов. Обучение распознаванию сигналов становится комбинированным. Исходная их классификация, несущая определенный медицинский смысл, не нарушается.

Разработки и исследования иллюстрируются примерами распознавания трех типов QRS-комплексов (N, A и V) в записи электрокардиограммы пациента. Результативность применения разработанного инструментария проверяется на конкретных обучающих и контрольных выборках в сравнении с показателями базового алгоритма.

Ключевые слова - диагностические системы; медико-биологические сигналы; опознавательные процедуры; обучение с учителем; кластеризация.



Clustering of Biomedical Signals in Recognition Systems with Supervised Learning

V. V. Lahutin^f, ORCID [0000-0003-3431-1686](https://orcid.org/0000-0003-3431-1686)

V. I. Sirotenko^f, ORCID [0000-0002-7542-5885](https://orcid.org/0000-0002-7542-5885)

A. D. Shachykov^s, PhD, ORCID [0000-0003-2353-7647](https://orcid.org/0000-0003-2353-7647)

O. P. Shulyak^s, PhD Assoc.Prof., ORCID [0000-0002-6920-2212](https://orcid.org/0000-0002-6920-2212)

Electronic Engineering Department ee.kpi.ua

National technical university of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute» kpi.ua

Kyiv, Ukraine

Abstract—Recognition of biomedical signals by the characteristics of their shape using supervised learning algorithms in patient diagnostic systems is considered. The classification of signals is assumed to be known and training samples are assumed to be available. A statistical assessment of the decision fidelity of the types of signals in the control samples by indicators of sensitivity, specificity, general validity is assumed.

The principle and procedure for improving the basic signal recognition procedure are disclosed. A software toolkit is developed and researched to identify and implement possible reserves in increasing the fidelity of decisions from a more detailed consideration of the nature and characteristics of the probability distributions of signals in their phase space. For this purpose, clustering of training samples is additionally carried out at locations of their aggregation, separately according to signal classes. Families of cluster standards attached to these locations are formed and they are used to determine the types of incoming signals.

The procedures for their processing are getting more complicated. The complexity of training and decision-making is growing. Signal recognition training becomes combined. However, their initial classification, which carries a certain medical meaning, is not violated. The emphasis in decision-making is shifted to comparing recognized signals with nearest standards in their vicinity. Distant standards lose their influence.

The development and research are illustrated by examples of recognition of three types of QRS complexes (N, A, and V) in the patient's electrocardiogram record. The effectiveness of the developed tools is tested on concrete training and control samples in comparison with the basic algorithm.

Keywords — *diagnostic systems; biomedical signals; recognition procedures; supervised learning; clustering.*