

УДК 004.9

ЕФЕКТИВНІСТЬ АЛГОРИТМІВ ПОПЕРЕДНЬОЇ ОБРОБКИ ДАНИХ В ЗАДАЧІ КЛАСИФІКАЦІЇ ВИСОКОЧАСТОТНИХ БІОЛОГІЧНИХ СИГНАЛІВ

А.Л. Бойко, Є.А. Настенко, В.А. Павлов, О.О. Коновал, М.А.Осадчий,
А.В. Дементьев

Національний технічний університет України «Київський політехнічний інститут
імені Ігоря Сікорського»

*anna_anna2612@mail.ru, nastenko.e@gmail.com, alexandra.konoval@gmail.com,
pavlov.vladimir264@gmail.com, m.osadchyy@gmail.com, liken2134@gmail.com*

В роботі наведено результати досліджень, що дозволили оцінити ефективність застосування різних алгоритмів згладжування сигналів пульсометрії в задачі класифікації здоров'я людини з точки зору стану серцево-судинної системи. Моделювання згладжених сигналів відбувалося за рекурентним адитивно-мультиплікативного багатоетапним алгоритмом (РАМБА) методу групового урахування аргументів (МГУА), проведення класифікації отриманих моделей реалізовано засобами програми GMDH Shell.

Ключові слова: алгоритми згладжування, метод групового урахування аргументів, метод ковзного середнього, експоненціальне згладжування, медіанний фільтр, фільтр Савицького-Голая.

This paper presents the results of effectiveness research on various algorithms for smoothing pulsometry signals in the classification problem of human health from the view point of cardiovascular system state. The modeling of smoothed signals are realized using the Recurrent Aditive-Multiplicative Multistage Algorithm (RAMMA) of Group Method of Data Handling (GMDH).

Keywords: smoothing algorithms, Group Method of Data Handling, Moving Average Method, exponential smoothing, median filter, Savitzky-Golay filter.

В данной работе приведены результаты исследования эффективности различных алгоритмов сглаживания сигналов пульсометрии в задаче классификации здоровья человека с точки зрения состояния сердечнососудистой системы. Моделирование сглаженных сигналов реализовано с помощью рекуррентного адитивно-мультипликативного многоэтапного алгоритма РАМБА метода группового учета аргументов.

Ключевые слова: алгоритмы сглаживания, метод группового учета аргументов, метод скользящего среднего, экспоненциальное сглаживание, медианный фильтр, фильтр Савицкого-Голая.

Вступ. Відомим способом зменшення впливу артефактів різної природи на біомедичні сигнали, що реєструються, є використання методів згладжування та цифрової фільтрації, до переваг яких відносять високу точність, гнучкість налаштування, стабільність функціонування [1]. Проте при застосування в подальшому перетворених даних у задачі класифікації, принципово не втратити характерні для класів ознаки сигналу. В даній роботі вирішується проблема вибору способу первинної обробки даних саме з такої точки зору.

1. Постановка задачі.

Досліджується можливість покращення результатів класифікації біомедичних сигналів пульсометрії шляхом вибору інструменту первинної обробки даних. Приклад сигналу для одного з об'єктів наведено на рис.1. Сигнал відображає реакцію серцево-судинної системи студента на тестове навантаження у 20 присідань.

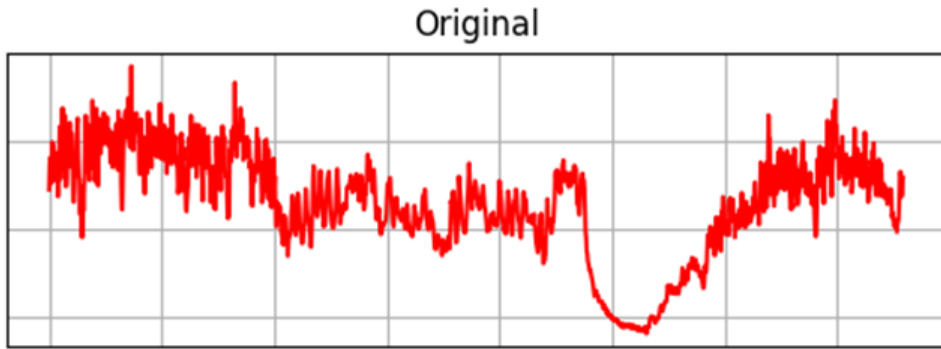


Рис.1. Графік оригінального сигналу пульсометрії при тестовому навантаженні

Серед алгоритмів, що досліджуються - ковзного середнього, експоненційного згладжування, медіанний фільтр та фільтр Савицького-Голая. Класифікація сигналів реалізується у просторі параметрів структури їх моделей, що визначаються за допомогою рекурентного адитивно-мультиплікативного багатоетапного алгоритму (РАМБА) [2] методу групового урахування аргументів (МГУА) [3]. Метою роботи є визначення серед механізмів первинної обробки даних алгоритму, що може забезпечить вищу якість класифікації класів загального здоров'я людини з точки зору стану серцево-судинної системи.

2. Основна частина.

Для проведення досліджень використано базу даних із записами пульсометрії 92-х студентів (у вигляді .hgv-файлів), а також загальну оцінку здоров'я кожного студента в діапазоні шкали від -8 до 14. Шкала відповідає 5-ти групам здоров'я – «низький» (від -8 до -5), «нижче середнього» (від -4 до 0), «середній» (від 1 до 5), «вище середнього» (від 6 до 10), «високий» (від 11 до 14). Для обробки запису сигналів пульсометрії було застосовано 4 алгоритми: метод ковзного середнього, експоненціальне згладжування, медіанний фільтр та фільтр Савицького-Голая. Коротко наведемо дані про алгоритми.

Метод ковзних середніх (1) нівелює випадкові відхилення значень сигналу. При згладжуванні, фактичні значення ряду у часі замінюються середніми значеннями, що характеризують серединну точку вікна ковзання. Згладжування ґрунтується на складанні нового ряду з середніх арифметичних, обчислених для проміжків часу довжиною q :

$$x(k) = \sum_{t=k}^{q+k} \frac{x(t)}{q}, (k = 1, 2, \dots, n - q + 1) \tag{1}$$

де q - довжина періоду згладжування, що залежить від характеру часового ряду, мети згладжування та вибирається дослідником, k - порядковий номер середньої.

Метод експоненційного згладжування застосовується для прогнозування нестаціонарних часових рядів, що мають випадкові зміни рівня і кута нахилу. В якості основної моделі ряду розглядається його локальна апроксимація у вигляді полінома невисокого ступеня p :

$$x(t) = a_0(t) + a_1(t) \cdot t + a_2(t) \cdot t^2 + \dots + a_p(t) \cdot t^p + h \tag{2}$$

коефіцієнти якого a_i повільно змінюються з часом.

В основі медіанного фільтру лежить обчислення ковзної медіани. Щоб одержати розрахунок значення ковзної медіани в точці t , обчислюється медіана значень ряду в часовому інтервалі $[t-q, t+q]$. Відповідне значення є $(2q+1)$ -точкова ковзна медіана. Основна перевага медіанного згладжування - стійкість до наявності викидів [4].

У фільтрі Савицького-Голая замість лінійного наближення сигналу в околі кожної точки вимірювання будується за методом найменших квадратів поліном n -го порядку, що апроксимує первинний ряд (3). Значення чутливості при заданому хвильовому числі k замінюється значенням полінома в цій точці [5]:

$$g(k_i) = a_0 + a_1 SX(k_i) + a_2 SX(k_{i+1}) + \dots + a_n SX(k_{i+n-1}) \quad (3)$$

На рис.2, 3, 4, 5 наведено графіки сигналу (рис.1), що згладжено чотирма згаданими алгоритмами та їх кореляція Пірсона з оригінальним сигналом

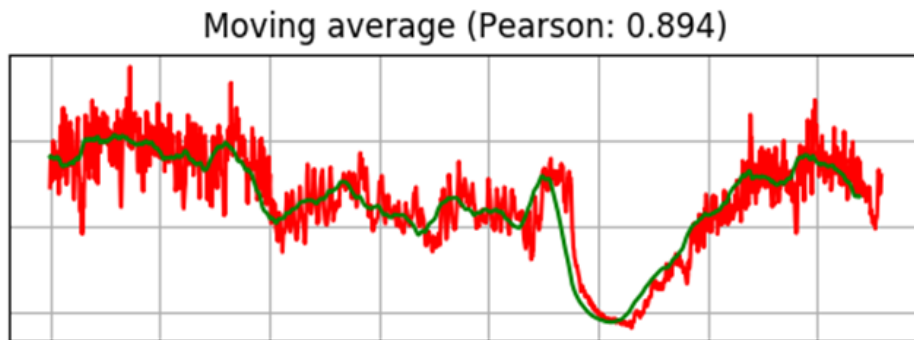


Рис. 2. Графік сигналу пульсометрії, що згладжено за алгоритмом ковзних середніх

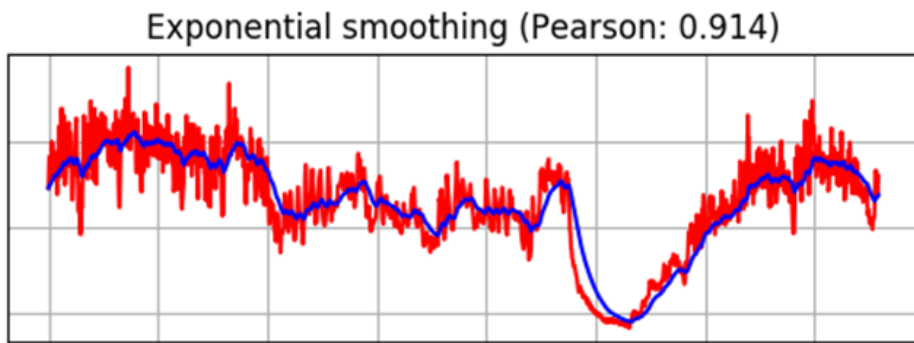


Рис. 3. Графік сигналу пульсометрії, що згладжено за експоненційним алгоритмом

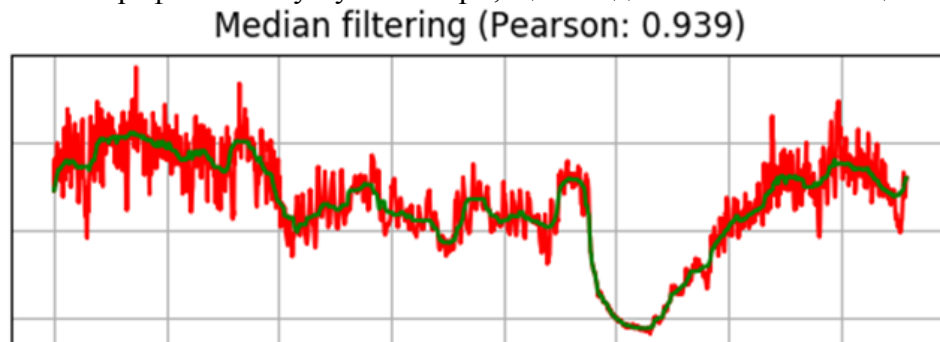


Рис. 4. Графік сигналу пульсометрії, що оброблено медіанним фільтром

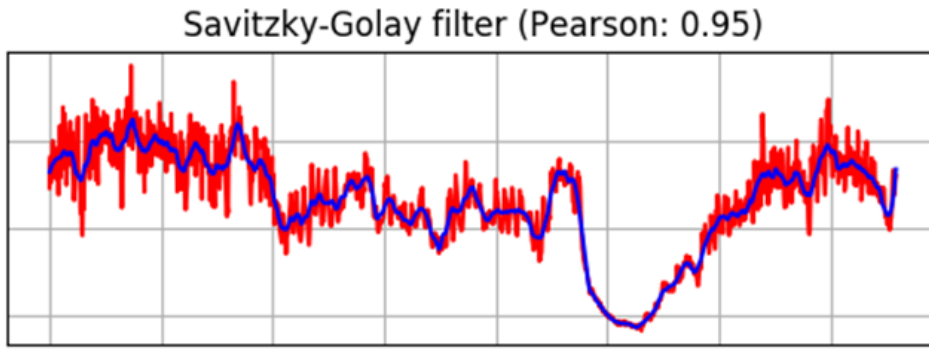


Рис. 5. Графік сигналу пульсометрії, що оброблено фільтром Савицького-Голая

Оцінюючи результат візуально, можна спостерігати деякі зсуви у результатах роботи алгоритмів ковзного середнього та експоненціального згладжування, що притаманно цим алгоритмам при високому значенні ширини вікна та малому значенні параметру α відповідно. Фільтри Савицького-Голая та медіанний не мають помітних відхилень, хоча видно, що медіанний фільтр значно менше реагує на викиди сигналу.

Для попередньої оцінки обробки даних використано коефіцієнт кореляції Пірсона між оригінальним сигналом та його згладжених версій. Результати оцінки дозволяють зробити висновок, що фільтр Савицького-Голая найменше міняє сигнал, при цьому достатньо добре згладжує його. Медіанний фільтр має менший коефіцієнт кореляції, саме через свій опір до викидів. Алгоритми ковзного середнього та експоненційний мають найменші коефіцієнти кореляції через те, що вони мають властивість дещо зсувати вихідний сигнал. Тож оцінюючи згладжуючи властивості алгоритмів з урахуванням значення коефіцієнту кореляції можна зробити попередній висновок, що з формальної точки зору (без врахування кінцевої мети використання даних) доцільним є використання алгоритму Савицького-Голая.

Для оригінального сигналу та сигналів, отриманих в результаті фільтрації, було розраховано оптимальні структури моделей за допомогою алгоритму РАМБА [2] із застосуванням комбінованого критерію точності.

Оптимальна структура для оригінального сигналу:

$$x(t) = a_0 + a_1t + a_2t^{-2} + a_3t^{\frac{5}{3}} \quad (4)$$

Оптимальна структура для сигналу, згладженого методом ковзного середнього:

$$x(t) = a_0 + a_1t \quad (5)$$

Оптимальна структура для сигналу, одержаного експоненціальним згладжуванням:

$$x(t) = a_0 + a_1t + a_2t^{-1} \quad (6)$$

Оптимальна структура для сигналу, згладженого медіанним фільтром:

$$x(t) = a_0 + a_1t + a_2t^{-1} + a_3t^{\frac{4}{3}} + a_4t^{-2} \quad (7)$$

Оптимальна структура для сигналу, згладженого фільтром Савицького-Голая:

$$a_0 + a_1t + a_2t^{\frac{4}{3}} + a_3t^{-\frac{4}{3}} + a_4t^{-1} + a_5t^{-2} \quad (8)$$

П'ять структур, що було одержано, застосовано для розрахунку п'яти наборів коефіцієнтів моделей для кожного з об'єктів. Значення кожного з наборів параметрів формують для класів, що відображають оцінку здоров'я студента, матрицю об'єкт-ознаки, для свого алгоритму обробки сигналів.

Для оцінки впливу алгоритмів згладжування на класифікацію сигналів (належність об'єкту до відповідної групи здоров'я) у просторі параметрів кожної з одержаних структур розраховано класифікатори засобами програми GMDH Shell [6] (табл. 1-2).

Таблиця 1

Класифікатори, що одержано за умов застосування різних алгоритмів обробки сигналу

Фільтр	Клас	Рівняння класифікатора
Експоненційний	1	$Y_1 = 85.9498 - 28.9329 \cdot \sqrt[3]{a_0} + 2.45312 \cdot \sqrt[3]{a_0^2} - 4.37549E(-05) \cdot a_0^2 + 32.5758 \cdot a_1 + 0.411465 \cdot a_1^2 - 0.400891 \cdot \sqrt[3]{a_2} - 0.0108346 \cdot \sqrt[3]{a_2^2} - 0.03971 \cdot a_0 a_1 + 0.000303368 \cdot a_0 \sqrt[3]{a_2} + 0.663463 \cdot \sqrt[3]{a_1 a_2} + 8.55082E(-05) \cdot a_2 \sqrt[3]{a_2}$
	2	$Y_2 = 0.114344 + 0.0396022 \cdot \sqrt[3]{a_2^2} + 5.10941 E(-06) \cdot a_2^2 - 0.000244258 \cdot a_0 \sqrt[3]{a_2} - 0.000901045 \cdot a_2 \sqrt[3]{a_2} + 0.0208818 \cdot \sqrt[3]{a_0 a_2}$
	3	$Y_3 = 0.496144 + 1.11292 \cdot \sqrt[3]{a_2} + 0.000758409 \cdot a_0 \sqrt[3]{a_2} - 0.187104 \cdot \sqrt[3]{a_0 a_2}$
	4	$Y_4 = -0.804524 + 0.706461 \cdot \sqrt[3]{a_1} + 15.165 \cdot \sqrt[3]{a_1^2} + 26.6045 \cdot a_1 - 5.29275 E(-07) \cdot a_2^2 + 0.0239248 \cdot a_1 a_2 + 5.06587 E(-06) \cdot a_0 a_2 - 0.956397 \cdot a_1 \sqrt[3]{a_2} - 0.000155316 \cdot a_0 \sqrt[3]{a_2} - 1.001 \cdot a_1 \sqrt[3]{a_0}$
	5	$Y_5 = 11.9718 + 95.0727 \cdot \sqrt[3]{a_1} + 0.0311318 \cdot \sqrt[3]{a_2} + 286.411 \cdot \sqrt[3]{a_1^2} - 0.036608 \cdot \sqrt[3]{a_2^2} + 361.279 \cdot a_1 - 1.51828 E(-06) \cdot a_2^2 + 0.000398182 \cdot a_2 \sqrt[3]{a_2} + 162.937 \cdot a_1 \sqrt[3]{a_1} + 0.0677184 \cdot a_1 \sqrt[3]{a_2}$
Медіанний	1	$Y_1 = 0.741744 + 1.60241 E(-13) \cdot \sqrt[3]{a_2^2} - 1.58198 E(-12) \cdot \sqrt[3]{a_4^2} + 19222 E(-08) \cdot \sqrt[3]{a_0 a_3} - 988016 E(-09) \cdot \sqrt[3]{a_2 a_3}$
	2	$Y_2 = 0.571778 + 0.0127192 \cdot \sqrt[3]{a_3} - 1.03392 \cdot \sqrt[3]{a_1^2} - 0.00228492 \cdot \sqrt[3]{a_3^2} - 1.78975 \cdot a_1 + 6.43588 E(-07) \cdot a_1 \sqrt[3]{a_2} + 1.45081 E(-27) \cdot a_2 \sqrt[3]{a_2} - 1.02225 E(-07) \cdot \sqrt[3]{a_1 a_2}$
	3	$Y_3 = 0.444732 - 1.18156 E(-12) \cdot \sqrt[3]{a_4^2} + 0.172468 \cdot \sqrt[3]{a_1 a_3} + 2.15965 E(-06) \cdot \sqrt[3]{a_1 a_4} - 9.31727 E(-08) \cdot \sqrt[3]{a_3 a_4}$
	4	$Y_4 = 0.303899 + 1.80097 \cdot \sqrt[3]{a_1} - 9.99114 E(-13) \cdot \sqrt[3]{a_0^2} + 1.47099 \cdot \sqrt[3]{a_1^2} - 4.78248 E(-36) \cdot a_4^2 - 6.78777 E(-24) \cdot a_0 \sqrt[3]{a_4} + 0.0937847 \cdot a_1 \sqrt[3]{a_3} - 3.09858 E(-07) \cdot a_1 \sqrt[3]{a_2} + 3.69134 E(-11) \cdot a_3 \sqrt[3]{a_4}$
	5	$Y_5 = 0.609231 + 3.31378 E(-07) \cdot \sqrt[3]{a_2} + 2.75875 \cdot a_1 - 5.18013 E(-20) \cdot a_2$
Ковзн ого середн	1	$Y_2 = -232.565 + 52.8631 \cdot \sqrt[3]{a_0} - 2.99203 \cdot \sqrt[3]{a_0^2} + 118.65 \cdot a_1 - 4.33829 \cdot a_1^2 - 13.0054 \cdot a_1 \sqrt[3]{a_0}$

	2	$Y_2 = 0.08675 - 8.33299 \cdot a_1 + 1.69187 E(-06) \cdot a_0^2 +$ $+ 20.9968 \cdot a_1^2 + 0.0200941 \cdot a_0 a_1$
	3	$Y_3 = 28.9039 - 4.41808 \cdot \sqrt[3]{a_0} + 0.0141182 \cdot a_0 - 2.92093 \cdot a_1 \sqrt[3]{a_0} +$ $+ 34.7073 \cdot a_1 \sqrt[3]{a_1}$
	4	$Y_4 = 8.27433 + 0.9754912 \cdot a_0 - 16.9353356 \cdot a_1 + 1.89183 E(-05) \cdot a_0^2 +$ $+ 2.24184 \cdot a_1^2$
	5	$Y_5 = 898.866 - 289.559 \cdot \sqrt[3]{a_0} + 3.42617 \cdot \sqrt[3]{a_1} + 31.1029 \cdot \sqrt[3]{a_0^2} +$ $+ 3.77553 \cdot \sqrt[3]{a_1^2} - 1.11262 \cdot a_0$
Оригінал	1	$Y_1 = -0.488232 + 16.6605 \cdot \sqrt[3]{a_3} + 2.18836 E(-06) \cdot a_0^2 + 13127.8 \cdot a_3^2 +$ $+ 11.0288 \cdot a_1 \sqrt[3]{a_3} - 5.2516 \cdot a_3 \sqrt[3]{a_0} - 1.45833 \cdot \sqrt[3]{a_0 a_3}$
	2	$Y_2 = -17.3922 + 3.87846 \cdot \sqrt[3]{a_0} - 0.541205 \cdot \sqrt[3]{a_3} - 0.211772 \cdot \sqrt[3]{a_0^2} +$ $+ 0.407841 \cdot \sqrt[3]{a_1^2}$
	3	$Y_3 = 1.25369 + 0.0564133 \cdot \sqrt[3]{a_2} - 0.00170536 \cdot a_0 + 45.0748 \cdot a_3$
	4	$Y_4 = -47.1998 + 10.1221 \cdot \sqrt[3]{a_0} + 10.2116 \cdot \sqrt[3]{a_3} + 2.56121 \cdot \sqrt[3]{a_1} - 0.534345 \cdot \sqrt[3]{a_0^2} +$ $+ 160.76 \cdot \sqrt[3]{a_3^2} + 34.8069 \cdot \sqrt[3]{a_1 a_3} + 0.755639 \cdot \sqrt[3]{a_2 a_3}$
	5	$Y_5 = 1.94998 - 0.00204567 \cdot a_0 - 0.549044 \cdot a_1 - 0.00014447 \cdot a_2$
Савицького-Голая	1	$Y_1 = 0.0285225 + 0.0465457 \cdot \sqrt[3]{a_0} + 1.01742 \cdot \sqrt[3]{a_3} + 0.00817724 \cdot \sqrt[3]{a_2 a_5}$
	2	$Y_2 = 0.617803 + \sqrt[3]{a_1} - 0.0103558 \cdot \sqrt[3]{a_4} + 0.00479187 \cdot \sqrt[3]{a_5} - 0.0134424 \cdot \sqrt[3]{a_6} -$ $- 0.000452644 \cdot \sqrt[3]{a_4^2} - 0.426867 \cdot \sqrt[3]{a_1 a_3} - 0.0200976 \cdot \sqrt[3]{a_0 a_1} +$ $+ 0.000819375 \cdot \sqrt[3]{a_5 a_7} + 0.000375239 \cdot \sqrt[3]{a_4 a_7} + 0.0053635 \cdot \sqrt[3]{a_2 a_5} -$ $- 0.00354493 \cdot \sqrt[3]{a_2 a_7} - 0.0275363 \cdot \sqrt[3]{a_1 a_2} -$ $- 2.09735 E(-05) \cdot \sqrt[3]{a_6 a_7}$
	3	$Y_3 = 0.513531 - 1.86628 E(-06) \cdot a_0 + a_0 + 4.2759 E(-10) \cdot a_5 +$ $+ 4.31222 E(-12) \cdot a_6^2 + 2.09872 E(-06) \cdot a_1 a_5 - 4.05468 E(-10) \cdot a_5 a_7$
	4	$Y_4 = 0.594861 - 0.0538055 \cdot \sqrt[3]{a_2^2} - 0.00847078 \cdot \sqrt[3]{a_1 a_5} -$ $- 0.000161437 \cdot \sqrt[3]{a_5 a_6} + 0.00380038 \cdot \sqrt[3]{a_3 a_5} + 0.00401101 \cdot \sqrt[3]{a_2 a_4} +$ $+ 0.002785 \cdot \sqrt[3]{a_3 a_6} + 0.000190281 \cdot \sqrt[3]{a_5 a_7} - 2.46439 E(-07) \cdot a_5 \sqrt[3]{a_4} +$ $+ 6.63505 E(-07) \cdot a_1 a_5 - 1.12355 E(-06) \cdot a_5 \sqrt[3]{a_1}$
	5	$Y_5 = 0.379021 + 7.16243 E(-07) \cdot a_0 - 8.98043 E(-06) \cdot a_4 +$ $+ -2.0146 E(-06) \cdot a_7$

Показники якості класифікації кожного варіанту класифікаторів наведено у таблиці 2.

Таблиця 2

Результати класифікації

Тип обробки	Міра класифікації	1	2	3	4	5
Оригінал	Відсоток правильно класифікованих	89,6	76,6	60,7	93,7	98,5
	Середній відсоток	83,82				
	Правильно класифіковані у «своєму» класі	87,18				
	Середнє значення F-міри	0,874				
Експонентний	Відсоток правильно класифікованих	88,7	83,1	81,5	92,1	83,5
	Середній відсоток	85,78				
	Правильно класифіковані у «своєму» класі	94,91				
	Середнє значення F-міри	0,88				
Ковзного середнього	Відсоток правильно класифікованих	72,9	90,1	66,7	89,0	94,1
	Середній відсоток	82,56				
	Правильно класифіковані у «своєму» класі	86,78				
	Середнє значення F-міри	0,855				
Медіанний	Відсоток правильно класифікованих	81,5	85,6	76,6	87,4	76,3
	Середній відсоток	81,48				
	Правильно класифіковані у «своєму» класі	89,5				
	Середнє значення F-міри	0,857				
Савицького-Голая	Відсоток правильно класифікованих	98,5	97,9	65,2	90,1	69,7
	Середній відсоток	84,28				
	Правильно класифіковані у «своєму» класі	86,46				
	Середнє значення F-міри	0,86				

Кращі результати класифікації сигналу пульсометрії дали моделі, одержані при умові використання алгоритму експоненційного згладжування. Середній відсоток правильно класифікованих об'єктів і F-міра вищі, ніж в при використанні сигналів, що згладжені іншими алгоритмами чи оригінального сигналу.

Висновки.

Результати класифікації дали підстави визначити алгоритм експоненційного згладжування як найбільш ефективний для попередньої обробки даних сигналу пульсометрії з точки зору збереження істотних для класифікації сигналів властивостей. Для вирішення задач класифікації даного виду сигналів алгоритм експоненційного згладжування буде інтегровано в єдину систему з програмною реалізацією алгоритму РАМБА.

Література

1. Федотов А.А., Измерительные преобразователи биомедицинских сигналов систем клинического мониторинга /Федотов А.А., Акулов С.А.;М.:Радио и связь, 2013. –250 с.
2. Павлов В.А. Коновал А.О. Рекуррентный аддитивно-мультипликативный многоэтапный алгоритм МГУА для задач классификации объектов, заданных множествами наблюдений. Індуктивне моделювання складних систем. Збірник наук. праць. // К.: МННЦТС, 2015. – Вип.7. – С. 220-232.
3. Ивахненко А.Г. Помехоустойчивость моделирования / А.Г. Ивахненко, В.С. Степашко // Киев: «Наук.думка». – 1985.– 216 с.
4. Розенберг Г.С. Экологическое прогнозирование (Функциональные предикторы временных рядов) / Розенберг Г.С., Шитиков В.К., Брусиловский П.М.; Тольятти, 1994. - 182 с.
5. Никонов А. В. Фильтрация методом Савицкого-Голея спектральных характеристик чувствительности матричных фотоприемных устройств / Никонов А. В., Давлетшин Р. В., Яковлева Н. И., Лазарев П. С.; Успехи прикладной физики, 2016 – том 4, № 2 – С. 198-205
6. GMDH Shell Documentation [Електронний ресурс] – Режим доступу: <http://d.gmdhshell.com/docs/> (дата звернення: 22.04.2017)