

Міністерство освіти і науки України
Чорноморський національний університет імені Петра Могили
Факультет комп'ютерних наук
Кафедра комп'ютерної інженерії

ДОПУЩЕНО ДО ЗАХИСТУ
Завідувач кафедри,
д-р техн. наук, проф.
_____ І. М. Журавська
« __ « _____ 2023 р.

КВАЛІФІКАЦІЙНА МАГІСТЕРСЬКА РОБОТА
**РОЗРОБКА НОВОГО ДІАГНОСТИЧНОГО ЗНАКУ ЗА
РЕЗУЛЬТАТАМИ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ
АБДОМІНАЛЬНИХ ЕЛЕКТРОМІОГРАМ**

Спеціальність «Комп'ютерна інженерія»
123 – КМР.1 – 605.21710517

Студентка _____ А. І. Мельничук
« __ « _____ 2023 р.

Керівник д-р фіз.-мат.
наук, проф. _____ Г. П. Чуйко
« __ « _____ 2023 р.

Миколаїв 2023

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ	4
ВСТУП	5
1. АНАЛІТИЧНИЙ ОГЛЯД ЛІТЕРАТУРИ ТА ПАТЕНТНОЇ ІНФОРМАЦІЇ	9
1.1 Виявлення сигналу ЕМГ	9
1.2 Основні джерела шуму, які впливають на якість сигналу ЕМГ	12
1.2.1 Електричний шум від ліній електропередач і зовнішніх джерел	12
1.2.2 Артефакт руху	12
1.2.3 Перехресне забруднення	13
1.2.4 Викиди	14
1.2.5 Фізіологічний шум	14
1.3 Забезпечення хорошої якості сигналу ЕМГ	15
1.3.1 Відношення сигнал/шум (SNR)	15
1.3.2 Розташування датчика	18
1.3.3 Підготовка шкіри	19
Висновки до розділу 1	20
2. СПОСІБ ЗБОРУ ДАНИХ ЕЛЕКТРОМІОГРАМ М'ЯЗІВ ЖИВОТА. АЛГОРИТМИ ТА МЕТОДИ ОБРОБКИ ОТРИМАНИХ ДАНИХ	21
2.1 Електроміографи «M-TEST»	22
2.2 Алгоритми та методи обробки отриманих даних	25
2.2.1 Метод експоненційного згладжування	25
2.2.2 Метод аналізу детрендованих флуктуацій (DFA)	30
Висновки до розділу 2	34
3. РЕАЛІЗАЦІЯ АЛГОРИТМІВ ТА МЕТОДІВ ОБРОБКИ ОТРИМАНИХ ДАНИХ У СИСТЕМІ КОМП'ЮТЕРНОЇ МАТЕМАТИКИ MAPLE 2020	35

3.1 Експоненціальна модель згладжування: усунення шумів і зменшення трендів серії ЕМГ	35
3.2 Підмножини патернів дихання	41
3.3 Відсоток викидів	43
3.3.1 Що таке викиди	43
3.3.2 Чотири способи обчислення викидів	44
3.3.2 Розрахунок відсотку викидів	45
3.4 Повнохвильове випрямлення та згладжування ЕМГ методом рухомих медіан	46
Висновок до розділу 3	48
4. ТЕСТУВАННЯ ТА АНАЛІЗ ОБЧИСЛЕНЬ У СИСТЕМІ КОМП'ЮТЕРНОЇ МАТЕМАТИКИ MAPLE 2020	50
4.1 Файл ARFF і крива навчання	50
4.2 Розділ заголовка ARFF	51
4.2.1 Оголошення @relation	51
4.2.2 Оголошення @attribute	51
4.2.3 Рядкові атрибути	52
4.2.4 Атрибути дати	52
4.3 Розділ даних ARFF	52
4.4 Розріджені файли ARFF	54
4.5 Продуктивність класифікаторів Weka	55
4.6 Кластеризація набору даних	59
4.7 Коротке обговорення результатів машинного навчання	60
Висновок до розділу 4	61
ВИСНОВКИ	63
ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ	64

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ

АКФ	–	автокореляційні функції
ЕЕГ	–	електроенцефалографічні
ЕМГ	–	електроміограма
ЕНМГ	–	електронейроміографії
СОАГС	–	синдром обструктивного апное/гіпопное сну
СОАС	–	синдромом обструктивного апное сну

AASM	–	American Academy of Sleep Medicine
ARFF	–	Attribute-Relation File Format
CSA	–	Central sleep apnea
DFA	–	Deterministic finite acceptor
EMG	–	Electromiogram
ETS	–	Errors-Trend-Seasonalities
OHS	–	Hypoventilation syndrome in obesity
OSA	–	Obstructive sleep apnea
RMS	–	Root Mean Square
SNR	–	Signal-to-noise ratio
SSE	–	Sum of squared errors

ВСТУП

Термін «порушення дихання уві сні» охоплює низку станів, що характеризуються ненормальним диханням під час сну; у багатьох випадках це пов'язано зі звуженням або обструкцією верхніх дихальних шляхів (глотки). Порушення дихання варіюється від періодичної, часткової обструкції дихальних шляхів без порушення сну (хропіння), на іншому кінці спектру, частих апное, пов'язаних з повторюваною гіпоксемією та збудженнями, що призводять до порушення сну та сонливості вдень. Термін «обструктивне апное сну» (OSA) відноситься до періодичної обструкції дихальних шляхів, незалежно від наявності денних симптомів. Якщо виникають симптоми, стан називається синдромом обструктивного апное сну (СОАС), також відомим як синдром обструктивного апное/гіпопное сну (СОАГС).

Порушення дихання уві сні також включає:

- центральне апное уві сні (CSA), при якому періодична зупинка дихання відбувається без обструкції дихальних шляхів і яка у дорослих спостерігається в основному при серцевій недостатності;
- синдром гіповентиляції при ожирінні (OHS), при якому дихання уповільнене протягом сну, із супутнім звуженням або обструкцією верхніх дихальних шляхів або без нього.

Порушення дихання уві сні є дуже поширеним явищем у Європі, але систематично не збираються статистичні дані про пов'язану з цим захворюваністю або смертністю [14].

Американська академія медицини сну (AASM) виділяє кілька типів і підтипів розладів дихання, пов'язаних зі сном. Симптоми, тяжкість, причини та лікування розладу дихання уві сні відрізняються залежно від типу. У складних випадках у людини може бути діагностовано більше одного типу.

Основні проблеми:

- Синдром обструктивного апное сну (СОАС) є серйозною проблемою охорони здоров'я в Європі.
- Ожиріння є основним фактором ризику СОАС, але значна меншість хворих не страждає ожирінням.
- Основним методом лікування СОАС є безперервний позитивний тиск у дихальних шляхах – дуже ефективно та економічне втручання.
- Нелікований СОАС значно підвищує ризик нещасних випадків через сонливість.
- Синдром гіповентиляції при ожирінні стає все більш актуальною проблемою охорони здоров'я.

Актуальність магістерської роботи полягає в дослідженні абдомінальних електроміограм для моніторингу процесу дихання пацієнтів уві сні та виявлення залежностей відповідно до стану пацієнта.

Мета магістерської роботи: змодельювати та проаналізувати процес дихання уві сні та виявити залежності відповідно до стану пацієнта.

Об'єкт дослідження магістерської роботи: методи моделювання та аналізу набору даних абдомінальних електроміограм.

Предмет дослідження магістерської роботи: моделювання та аналіз набору даних електроміограм м'язів живота для моніторингу процесу дихання та встановлення залежностей для точного діагнозу пацієнта.

Для досягнення поставленої мети необхідно вирішити такі завдання:

- провести аналіз рішень, що існують для моделювання та аналізу набору даних абдомінальних електроміограм;
- за результатами аналітичного огляду літератури та патентної інформації визначити загальне компонування системи;
- визначити методи та моделі системи;
- реалізувати метод моделювання в одному з програмних пакетів Maple 2020;

- проаналізувати модель та здійснити перевірку отриманих результатів.

Гіпотеза: деякі хвороби призводять до порушення дихання уві сні.

Наукова новизна полягає в використанні методу експоненційного згладжування в обробці даних абдомінальних електроміограм (ЕМГ) для подальшого формування наборів даних ARFF (формат файлу зв'язку атрибутів) і застосування їх в класифікаторі Random Tree програмного пакету Weka 3-9-5. Цей класифікатор дозволяє візуалізувати його у формі «дерева рішень». Такий візуальний класифікатор-діагностик має бути практично зручним і доступним для медичних спеціалістів, які переважно не знайомі з методами Machine Learning.

Практичне значення отриманих результатів: результати дослідження можуть допомогти спеціалістам точніше встановлювати діагноз пацієнтам з проблемами дихальної системи, серцево-судинної та інших захворювань, що викликають порушення дихання під час сну.

Методи дослідження: 25 абдомінальних ЕМГ, що були випадково відібрані на відомому медичному порталі Physionet є набором даних для аналізу та моделювання ЕМГ сплячих дорослих – при виборі теми і постановці мети дослідження та аналізі результатів, аналіз технічної літератури з досліджуваної проблеми, теорії аналізу та моделювання сигналу - для практичної реалізації нових рішень комп'ютерної математики, а також теоремою Вінера-Хінчіна – при проведенні експериментальних досліджень метод експерименту для оцінки кількісних параметрів персистентності абдомінальних ЕМГ.

Робота пройшла апробацію на XXV Всеукраїнської науково-практичної конференції «Могилянські читання» (Миколаїв, 07–11 листопада 2022 р.) [33].

Публікації. Основні положення магістерської роботи опубліковані у збірнику матеріалів XXV Всеукраїнської науково-практичної конференції «Могилянські читання–2022».

1. АНАЛІТИЧНИЙ ОГЛЯД ЛІТЕРАТУРИ ТА ПАТЕНТНОЇ ІНФОРМАЦІЇ

1.1 Виявлення сигналу ЕМГ

Електроміографія (ЕМГ) – об'єктивний метод дослідження нервово-м'язової системи шляхом реєстрації електричних потенціалів м'язів, що дозволяє оцінити їх функціональний стан або використовувати отримані дані при розробці біо контролюємих протезів. Електроміографія унікальний метод вивчення функціонального стану опорно-рухового апарату і завжди заснований на реєстрації електричних процесів різних м'язів шляхом запису, обробки та аналізу потенціалів дії м'язових волокон, які викликають скорочення м'язів, і є перспективним способом вирішувати діагностичні проблеми. виявлення патологічних процесів, біоелектричного протезування. Сигнали електроміограми (ЕМГ) також активно використовуються для відновлення порушених рухових функцій в ортопедії і протезуванні. Наземні реєструючі системи ЕМГ сигналу є невід'ємною частиною сучасної медичної діагностики опорно-рухового апарату. ЕМГ - один з найбільш об'єктивних і інформативних методів вивчення функціонального стану периферичної нервової системи [1].

Точне виявлення дискретних подій у ЕМГ (наприклад, зміни фази в патерні активності, пов'язаного з ініціацією швидкої рухової відповіді) є важливим питанням в аналізі рухової системи. Було запропоновано декілька методів для визначення часу включення та вимкнення м'яза.

Найпоширеніший метод визначення подій, пов'язаних з моторикою, за сигналами ЕМГ полягає у візуальному огляді навченими спостерігачами. «Метод одного порогу», який порівнює сигнал ЕМГ із фіксованим порогом, є найбільш інтуїтивно зрозумілим і поширеним комп'ютерним методом визначення часу початку скорочення м'язів [2]. Цей метод заснований на порівнянні випрямлених необроблених сигналів і амплітудного порогу, значення якого залежить від середньої потужності фонового шуму [3]. Метод

може бути корисним для подолання деяких проблем, пов'язаних з візуальним оглядом. Однак такий підхід, як правило, не є задовільним, оскільки результати вимірювань сильно залежать від вибору порогового значення. Цей вид методу часто покладається на критерії, які є надто евристичними та не дозволяють користувачеві самостійно встановлювати ймовірність виявлення та помилкового сигналу [4]. У «однопороговому методі» співвідношення між ймовірністю виявлення P_{dk} та ймовірністю P_γ того, що зразок шуму перевищує поріг γ , визначається рівнянням 1.1.

$$P_{dk} = \exp\left(\frac{\ln(P_\gamma)}{1+10^{\frac{SNR}{10}}}\right), \quad (1.1)$$

де P_{dk} – ймовірність виявлення сигналу; P_γ – ймовірність, що зразок шуму перевищує поріг γ ; SNR – відношення сигнал/шум.

У 1984 році Вінтер [7] зауважив, що цей підхід загалом є незадовільним, оскільки він сильно залежить від вибору порогу. Щоб подолати проблеми «одного порогу», Vornato et al. [4] представив метод «подвійного порогового виявлення» в 1998 році. Подвійні порогові детектори є кращими від однопорогових, тому що вони дають вищу ймовірність виявлення. Двопорогові сповіщувачі дозволяють користувачеві встановити зв'язок між помилковим сигналом та ймовірністю виявлення з більшим ступенем свободи, ніж однопорогові. Користувач може налаштувати детектор відповідно до різних оптимальних критеріїв, таким чином адаптуючи його характеристики до характеристик кожного конкретного сигналу та програми [4].

Сигнал EMG, зареєстрований під час довільних динамічних скорочень, можна розглядати як гаусівський процес з нульовим середнім $s(t) \in N(0, \sigma_s)$, модульований м'язовою активністю та спотворений незалежним адитивним гаусовим шумом $n(t) \in N$ з нульовим середнім $(0, \sigma_n)$. Якщо ймовірність виявлення дорівнює P_d , то метод подвійного порогу визначається рівнянням 1.2.

$$P_d = \sum_{k=r_o}^m \binom{m}{k} P_{dk}^k (1 - P_{dk})^{m-k}, \quad (1.2)$$

де P_d – ймовірність виявлення сигналу; r_o – поріг; m – довжина вікна спостереження.

Поведінка двопорогового детектора фіксується параметрами: порогом r_o і довжиною вікна спостереження, m . Їх значення вибираються так, щоб мінімізувати значення ймовірності помилкового сигналу та максимізувати P_d для кожного конкретного відношення сигнал/шум (SNR) [4]. У 2004 році Ланьї та Адлер [5] виявили, що метод подвійного порогу, запропонований Борнато, є складним і обчислювально дорогим, що вимагає відбілювання сигналу. Він також не дуже чутливий. Ланьї та Енді запропонували новий алгоритм, заснований на методі подвійного порогу, який є більш чутливим, стабільним і ефективним зі знизеними витратами на обчислення. Для конкретних програм, окрім точності виявлення, важливою мірою може бути швидкість алгоритму. Алгоритми з великим часом обчислення непридатні для онлайн-виявлення. Один специфічний недолік методу Vornato et al. [4] – ймовірність виявлення, яка буде максимальною, коли P_{fa} фіксований, другий поріг має бути обраний рівним «1». Другий поріг фіксується під час виявлення, що означає, що подвійний пороговий детектор фактично стає однопороговим детектором. Цей метод не вимагає етапу відбілювання сигналу, який був необхідний раніше. Процес відбілювання займає багато обчислювального часу. Крім того, процес відбілювання зменшує ймовірність сигналу. Ця функція призведе до того, що виявлення пропустить частину інтервалу активації. Методи, запропоновані Lanyı та Adler [5], забезпечують швидке та більш надійне виявлення м'язів. В таблиці (рис. 1.1) показано порівняння різних, на перший погляд, методів виявлення на основі досліджень Мерло та Фаріни [11] у 2003 році.

Method	SNR(db)								Remark
	2		4		6		8		
	Bias	Std	Bias	Std	Bias	Std	Bias	Std	
Improved method (11)	-39	26	-22	25	-12	22	-3	17	Best
Double threshold (8)	41	68	21	69	12	47	0	53	Good
Single threshold (11)	55	154	67	147	62	135	72	139	Worse

Рисунок 1.1 – Порівняння трьох основних методів виявлення ЕМГ.

1.2 Основні джерела шуму, які впливають на якість сигналу ЕМГ

1.2.1 Електричний шум від ліній електропередач і зовнішніх джерел

Перешкоди від ліній електропередач (50 або 60 Гц), люмінесцентних ламп і електричних пристроїв виникають через електромагнітне випромінювання, яке поширене в усіх середовищах. Хоча це, як правило, не викликає занепокоєння для сучасних технологій, перешкоди в лінії в деяких випадках можуть забруднити записи ЕМГ (рис. 1.2).

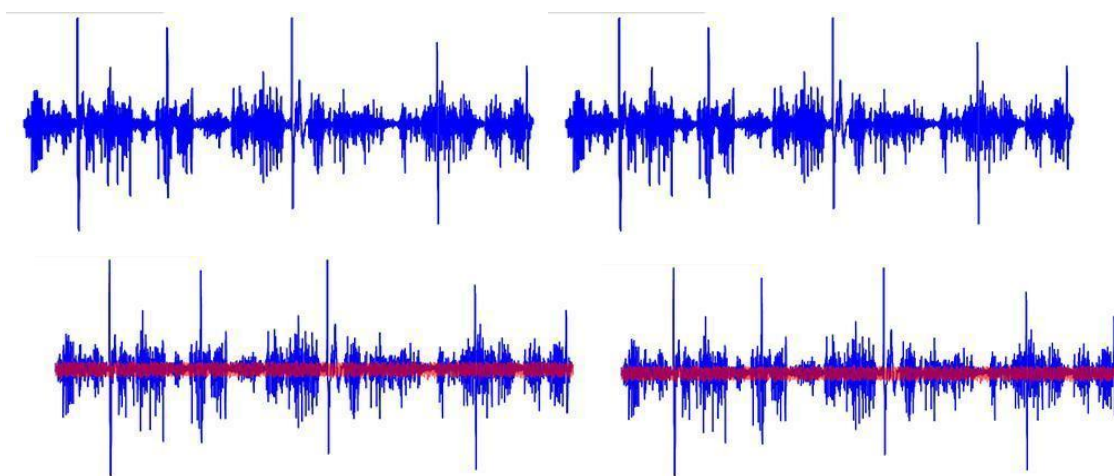


Рисунок 1.2 – Приклад забруднення сигналу ЕМГ.

1.2.2 Артефакт руху

Артефакт руху спричинений відносним переміщенням датчика по відношенню до підлеглої шкіри над потрібним м'язом.

Це може бути наслідком:

- прямого удару по датчику або тілу,
- швидкого переміщення сегмента тіла, до якого прикріплено датчик,

- змін у хімічному балансі поверхні шкіри та електрода через об'ємні зміни при скороченні і розтягуванні під час скорочення м'язів.

Артефакт руху (рис. 1.3) особливо проблематичний під час динамічних скорочень або інтенсивної діяльності.

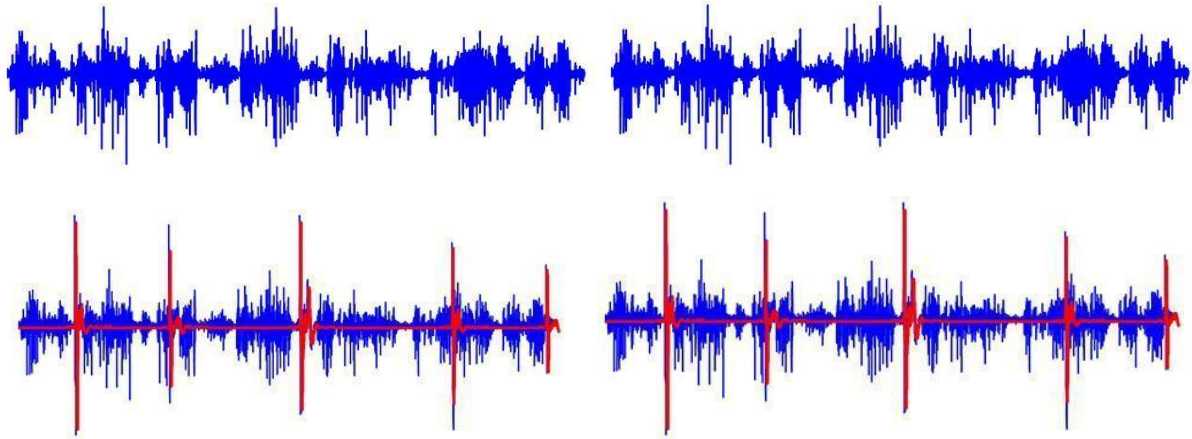


Рисунок 1.3 – Приклад артефакту руху ЕМГ

Високочастотна фільтрація поверхневого сигналу ЕМГ із частотою 20 Гц зменшує артефакт руху, енергія якого здебільшого зосереджена на низьких частотах, зберігаючи при цьому частотний вміст сигналу ЕМГ від досліджуваного м'яза [8].

1.2.3 Перехресне забруднення

Деякі частини ЕМГ-сигналу можуть не походити від потрібного м'яза. Натомість вони можуть бути результатом сигналів, що поширюються на датчик від сусідніх м'язів. Ці «перехресні» сигнали накладаються на ЕМГ-сигнал від досліджуваного м'яза, і їх присутність спотворює амплітуду та час даних ЕМГ (рис. 1.4) [13].

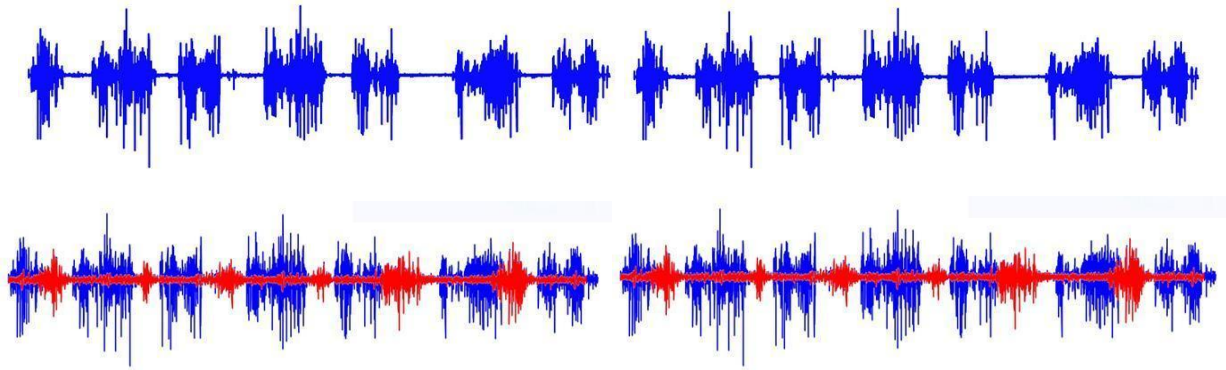


Рисунок 1.4 – «Перехресні» сигнали накладаються на ЕМГ-сигнал від досліджуваного м'яза

1.2.4 Викиди

Від'єднання датчика або надмірна амплітуда ЕМГ-сигналу можуть спричинити насичення, яке зазвичай називають «обрізанням» (рис. 1.5).

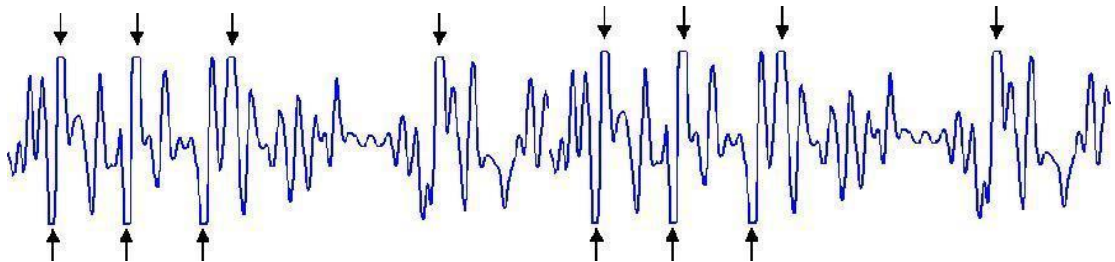


Рисунок 1.5 – Від'єднання датчика або надмірна амплітуда ЕМГ-сигналу можуть спричинити насичення

Якщо це станеться, слід забезпечити контакт між датчиком і шкірою, зменшити посилення підсилювача, якщо це можливо, або змінити розташування датчика на м'язі, щоб зменшити амплітуду сигналу [12].

1.2.5 Фізіологічний шум

Фізіологічний шум походить від тканин, відмінних від м'язів, які генерують електричні сигнали, наприклад сигнал ЕКГ (рис. 1.6).



Рисунок 1.6 – Фізіологічний шум у сигналі ЕМГ

Його можна зменшити, правильно розмістивши датчик ЕМГ подалі від джерела шуму, якщо це можливо. Обертання датчика так, щоб електроди вирівнювалися в еквіпотенціальних площинах, тобто обидва електроди були

на однаковій відстані від джерела, також зменшує фізіологічний шум у сигналі ЕМГ.

1.3 Забезпечення хорошої якості сигналу ЕМГ

1.3.1 Відношення сигнал/шум (SNR)

В аналоговому та цифровому зв'язку відношення сигнал/шум, яке часто називають S/N або SNR, є мірою потужності бажаного сигналу відносно фонового шуму (небажаного сигналу). S/N можна визначити за допомогою фіксованої формули, яка порівнює два рівні та повертає співвідношення, яке показує, чи впливає рівень шуму на бажаний сигнал.

Відношення зазвичай виражається одним числовим значенням у децибелах (дБ). Відношення може бути нулем, додатним чи від'ємним числом. Відношення сигнал/шум понад 0 дБ означає, що рівень сигналу перевищує рівень шуму. Чим вищий коефіцієнт, тим краща якість сигналу.

Наприклад, сигнал Wi-Fi із співвідношенням сигнал/шум 40 дБ забезпечить кращі послуги мережі, ніж сигнал із співвідношенням сигнал/шум 20 дБ. Якщо S/N сигналу Wi-Fi низький, це може вплинути на продуктивність мережі, оскільки пристроям стає важче відрізнити потрібний сигнал від шуму. Це може призвести до скидання пакетів і повторної передачі даних, що призведе до зниження пропускну здатності та більшої затримки.

Шум включає будь-які небажані перешкоди, які погіршують якість бажаного сигналу. Він може включати тепловий, квантовий, електронний, імпульсний або інтермодуляційний шум, а також інші форми шуму. Фактори навколишнього середовища, такі як температура та вологість, також можуть впливати на рівень шуму.

Якщо шум досить значний порівняно з бажаним сигналом, тобто S/N низький, він може порушити передачу широкого діапазону даних, включаючи текстові файли, графіку, телеметрію, програми, аудіо- та відеопотоки.

Інженери зв'язку завжди прагнуть максимізувати S/N. Традиційно це робилося шляхом використання найвужчої можливої смуги пропускання системи прийому відповідно до бажаної швидкості передачі даних. Однак є й інші методи. Наприклад, інженери можуть використовувати методи розширення спектру, щоб покращити продуктивність системи, або вони можуть збільшити вихідну потужність сигналу, щоб збільшити S/N.

У деяких системах високого рівня, таких як радіотелескопи, внутрішній шум мінімізується шляхом зниження температури приймальної схеми майже до абсолютного нуля (-273 градуси Цельсія або -459 градусів Фаренгейта). У бездротових системах завжди важливо оптимізувати роботу передавальної та приймальної антен .

Обчислення відношення сигнал/шум.

Співвідношення сигнал/шум зазвичай вимірюється в децибелах і може бути обчислено за допомогою логарифму за основою 10. Однак точна формула залежить від того, як вимірюються рівні сигналу та шуму.

Наприклад, якщо вони вимірюються в мікровольтах, можна використати таку формулу:

$$S/N = 20 \log_{10}(P_s/P_n) \quad (1.3)$$

P_s — сигнал у мікровольтах, а P_n — шум у мікровольтах.

Однак якщо сигнал і шум вимірюються у ватах, формула дещо відрізняється:

$$S/N = 10 \log_{10}(P_s/P_n) \quad (1.4)$$

Буква P часто використовується в цих формулах для позначення потужності.

Коли P_s дорівнює P_n , S/N буде 0. Співвідношення 0 дБ вказує на те, що сигнал безпосередньо конкурує з рівнем шуму, в результаті чого сигнал межує з нерозбірливим. У цифровому зв'язку це може спричинити зниження швидкості передачі даних через часті помилки, які вимагають від системи передачі повторного надсилання пакетів даних.

Коли P_s більше, ніж P_n , S/N буде додатним. В ідеалі P_s має бути набагато більше, ніж P_n , щоб мінімізувати шумові перешкоди. Як приклад, припустимо, що P_s дорівнює 10 мікрвольтам, а P_n дорівнює 1 мікрвольту. Оскільки 10, поділене на 1, дорівнює 10, для обчислення S/N можна використати наступну формулу:

$$S/N = 20\log_{10}(10) = 20 \text{ дБ} \quad (1.5)$$

Коефіцієнт 20 дБ означає, що сигнал чітко читається. Якщо сигнал набагато слабший, але все ще перевищує рівень шуму - скажімо, 1,3 мікрвольти - S/N набагато нижчий, у цьому випадку лише 2,28 дБ:

$$S/N = 20\log_{10}(1,3) = 2,28 \text{ дБ} \quad (1.6)$$

Це незначна ситуація, яка може вплинути на продуктивність мережі, хоча це не найгірша ситуація. Коли P_s менше, ніж P_n , сигнал/шум негативний, низьке відношення сигнал/шум. У такому типі ситуації надійний зв'язок майже неможливий, тому слід вжити заходів для підвищення рівня сигналу, зниження рівня шуму або застосувати комбінацію обох [24].

SNR є найкращим показником якості сигналу ЕМГ. Він вказує на співвідношення сигналу ЕМГ під час скорочення м'яза до небажаного електричного сигналу, зареєстрованого, коли м'яз перебуває в стані спокою, тобто базового шуму (рис. 1.7).

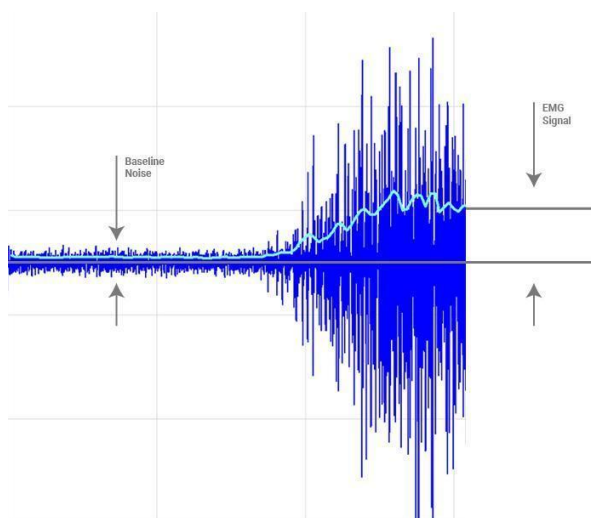


Рисунок 1.7 – Відношення сигнал/шум (SNR)

Величина сигналу ЕМГ залежить головним чином від сили скорочення, розташування та орієнтації датчика на досліджуваному м'язі, конструктивних характеристик датчика та кількості жирової тканини між шкірою та м'язом. Величина базового шуму залежить головним чином від технології запису та якості інтерфейсу шкіра-електрод.

Чим вищий SNR, тим надійніше відрізнити дані ЕМГ від базового шуму.

Селективність детектування сигналу датчика залежить від площі та відстані між поверхнями детектування. Чим більша площа та відстань між електродами, тим більша площа захоплення датчика і, як наслідок, більша амплітуда виявленого сигналу ЕМГ. Однак площа електродів і відстань між електродами не можуть бути надто великими, щоб мінімізувати виявлення небажаних перехресних перешкод від сусідніх м'язів, включаючи м'язи, розташовані глибоко до потрібного [9].

Відстань між електродами є фіксованою, що забезпечує повторюваність і послідовність експериментів. Без фіксованого інтервалу відстань між електродами може змінюватися з рухом, що призводить до небажаних змін у формі сигналу ЕМГ та його частотному вмісті.

1.3.2 Розташування датчика

Розташування датчика на м'язі є найважливішим фактором для хорошого SNR. Розміщення може максимізувати фізіологічний сигнал ЕМГ і мінімізувати перехресні перешкоди від сусідніх м'язів (рис. 1.8).



Рисунок 1.8 – Найкраще розташування датчика на м'язі

Найкраще розташування датчика на м'язі, як правило, на середній лінії м'яза, далеко від витоків сухожилля та зон іннервації, зі стрілкою датчика паралельно напрямку м'язових волокон.

На середній лінії м'яза волокна мають більший діаметр і, як правило, дають більшу амплітуду ЕМГ-сигналів. Навпаки, розташування датчика ЕМГ поблизу сухожилля зменшує амплітуду виявленого ЕМГ-сигналу, оскільки, коли м'язові волокна наближаються до сухожилля, вони стають тоншими та меншими. Волокна меншого діаметру генерують менші амплітуди потенціалів дії. Крім того, у цій області м'яз фізично менший, що ускладнює точне розміщення датчика та уникнення перехресних перешкод від сусідніх м'язів. Якщо можливо, слід уникати зон іннервації, оскільки в цих областях електрична активність поширюється в протилежних напрямках уздовж м'язових волокон і може бути ослаблена [10].

М'язові волокна, як правило, проходять у м'язі поздовжньо. У результаті електрична активність поширюється переважно вздовж довжини м'яза.

1.3.3 Підготовка шкіри

При звичайних нанесеннях поверхню шкіри слід протерти тампоном, змоченим спиртом (рис. 1.9). У більш екстремальних умовах, наприклад, у людей із сухою шкірою, гіпоалергенну стрічку можна кілька разів

наклеювати та відшаровувати до шкіри в місці датчика, щоб видалити зовнішній шар мертвої шкіри [11].

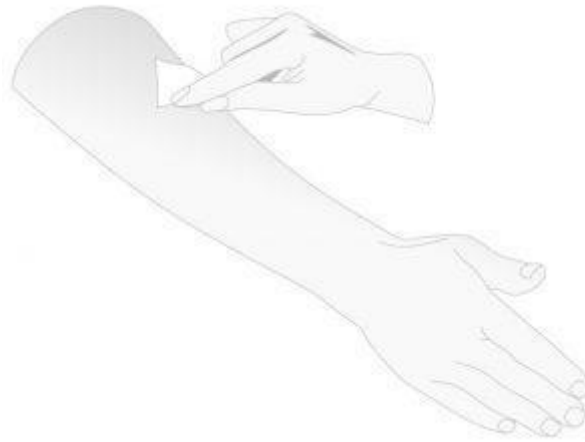


Рисунок 1.9 – Підготовка шкіри до ЕМГ

Висновки до розділу 1

У даному розділі було розглянуто виявлення сигналу, основні джерела шуму та забезпечення хорошої якості сигналу електроміограми. Показано, існуючі різні ефективні методи, що використовуються при отриманні якісного сигналу, це необхідно в даній роботі для подальшого аналізу отриманих даних та моделювання у системі комп'ютерної математики.

На основі розглянутої літератури зроблено висновок, що найнеобхіднішим є розуміння моделі сигналу у вигляді комбінації тренду, сезонних коливань та шумів.

За основу для моделювання та аналізу набору даних абдомінальних електроміограм взято модель Errors-Trend-Seasonalities (ETS), що є життєво важливим для прогнозування та інтерпретації серій ЕМГ.

2. СПОСІБ ЗБОРУ ДАНИХ ЕЛЕКТРОМІОГРАМ М'ЯЗІВ ЖИВОТА. АЛГОРИТМИ ТА МЕТОДИ ОБРОБКИ ОТРИМАНИХ ДАНИХ

Електронеуроміографія об'єднує групу методів функціональної діагностики, які визначають стан периферичних нервів, швидкість провідності імпульсів по нервових волокнах, а також оцінити м'язову збудливість. Методика дозволяє зареєструвати ступінь порушення імпульсного проведення, його пригнічення або відсутність, характер зміни. Є два різновиди ЕНМГ (з поверхневими електродами або використанням голчастих, що вводяться безпосередньо в м'яз). Призначення певної процедури залежить від розташування органів, досліджуваних м'язів або нервів.

Електронеуроміографія (інша назва: електроміографія, а також скорочено ЕНМГ та ЕМГ) – це спосіб функціональної діагностики, що дозволяє виявити порушення в роботі м'язів і периферичних нервів. Завдяки проведенню ЕНМГ верхніх і нижніх кінцівок можна діагностувати неврологічні захворювання на ранніх стадіях, коли у пацієнта відсутні будь-які скарги або симптоми [12].

Основне завдання електронеуроміографії – визначити реакцію периферичних нервів на стимуляцію електричними імпульсами, а також оцінити швидкість і силу реакції, виявити ушкодження та їхній характер.

Медичне обладнання для електронеуроміографії (ЕНМГ) - це комплексне електрофізіологічне дослідження, необхідне для визначення функціонального стану периферичної нервової системи і м'язів. Методика дозволяє виявляти патологічні зміни на самих ранніх стадіях.

Такий медичний апарат називається - електронеуроміограф. Система має всі можливості для проведення всіх рутинних електрофізіологічних тестів [23].

2.1 Електроміографи «M-TEST»

Комп'ютерний електроміограф «M-TEST» реєструє електроміограму по 1, 2 або 4 каналах. Порядок і набір використовуваних методик лікар визначає сам, виходячи з показань до проведення обстеження.

У базі даних зберігаються відомості про пацієнтів, записана електроміограма й результати її обробки. У приладі передбачено повний набір функцій для зручної і швидкої роботи з архівами електроміограм великого обсягу.

Під час запису електроміограми та подальшого її перегляду можна змінювати амплітудне й часове розгорнення і відразу виводити на друк записану ЕМГ. Підсумкові результати аналізу отримуємо в потрібному форматі таблиць, графіків або діаграм, вихідну форму висновку можна сформулювати на власний розсуд.

Функціональні можливості комп'ютерного електроміографа «M-TEST» разом з автоматичним описом результатів дослідження, вбудованою базою медичних норм і функцією її редагування роблять роботу простішою та ефективнішою [13].

Електроміограф «M-TEST» дозволяє реєструвати спонтанну активність м'язів із використанням поверхневих електродів двох видів: із регульованою і постійною межелектродною відстанню.

Система визначає основні параметри інтерференційної ЕМГ: частоту сумарної електричної активності м'язів, максимальну й середню амплітуду сигналу (рис. 2.1).

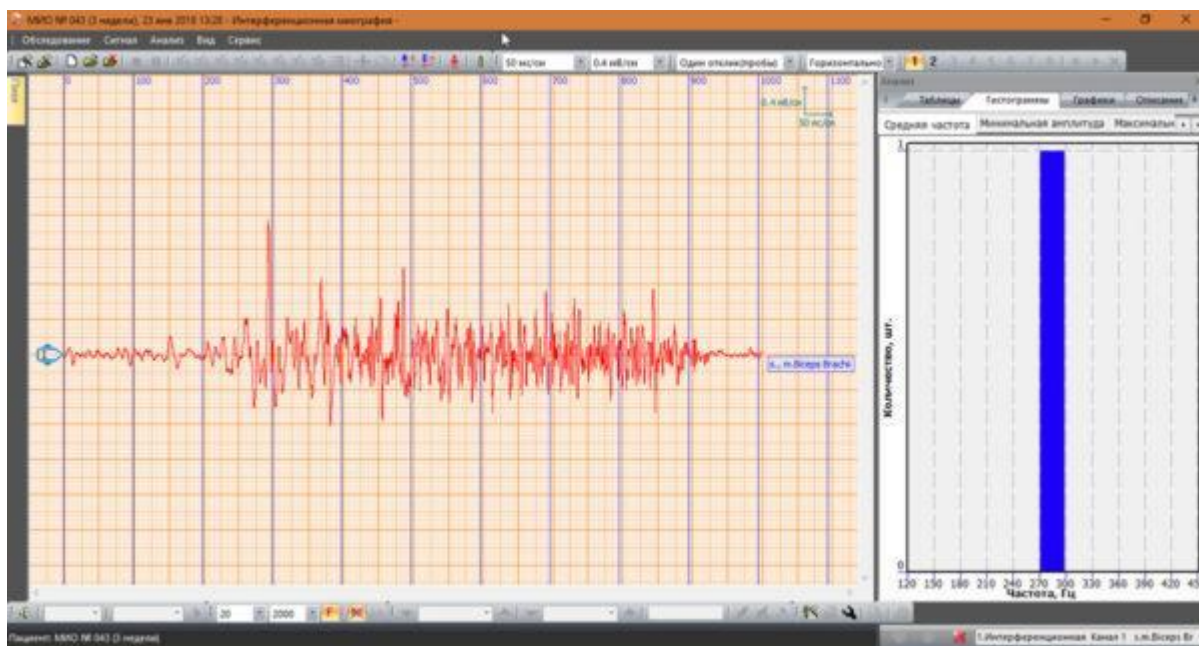


Рисунок 2.1 – Методика інтерференційної міографії

Для ортопедичної стоматології в електроміографі “М-TEST” передбачено додатковий аналіз жувальної проби. Для вивчення хронічного головного болю в приладі є функція дослідження екстероцептивної супресії довільної м’язової активності.

Система “М-TEST” реєструє викликані відповіді м’язів, отримані під час електричної стимуляції периферичного нерва за допомогою стимуляційних електродів. У процесі аналізу характеристик М-відповіді визначаються: поріг генерації, латентний період, амплітуди, тривалості і площі, розраховуються показники резидуальної латентності та швидкості проведення імпульсу по рухових волокнах, визначається динаміка показників.

В електроміографі “М-TEST” здійснюється дослідження потенціалу дії нерва, під час якого розраховується швидкість проведення порушення по чутливих волокнах (СРВс).

У процесі дослідження пізніх відповідей на стимуляцію аналізуються характеристики F-хвилі (латентність, моторна швидкість проведення, хронодисперсія, тахеодисперсія, середня швидкість, амплітуда F-хвилі (рис. 2.2), співвідношення величин F- і М-відповідей) і проводиться аналіз

параметрів Н-рефлекса під час стимуляції одиночними і парними стимулами (латентність, амплітуда, поріг генерації Н-рефлексу, співвідношення динаміки Н-рефлексу і динаміки М-відповіді).

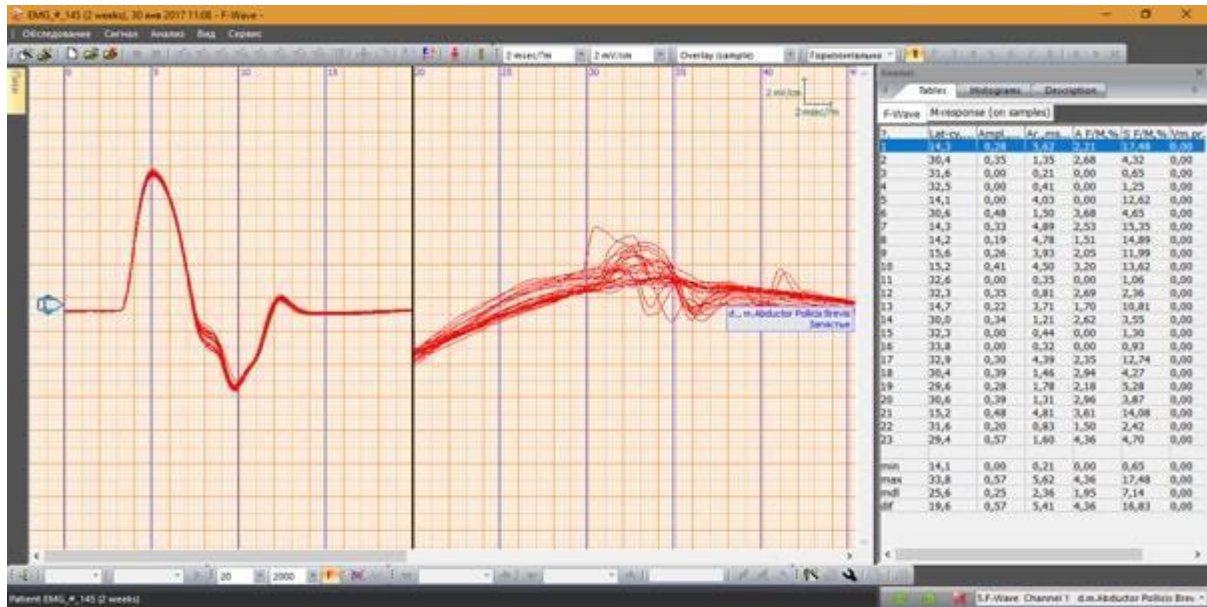


Рисунок 2.2 – Стимуляційна міографія: F-хвиля

“M-TEST” дає можливість реєструвати ЕМГ під час ритмічної стимуляції (із визначенням декременту М-відповіді) і проводити фармакологічні й навантажувальні проби, у тому числі тетанізацію та дослідження посттетанічних феноменів для уточнення характеру порушення нервово-м’язової передачі. Для оцінювання провідних систем рефлекторних дуг у електроміографі “M-TEST” передбачене дослідження мигального рефлексу.

Методика голчастої (локальної) ЕМГ.

Під час дослідження аналізується спонтанна активність, виділяються й аналізуються параметри потенціалів рухових одиниць, визначаються тривалості, амплітуди, площі. Для визначення характеру електроміограми на слух можна ввімкнути озвучування стрічки. У приладі реалізовано методику турно-амплітудного аналізу голчастої ЕМГ (рис. 2.3).

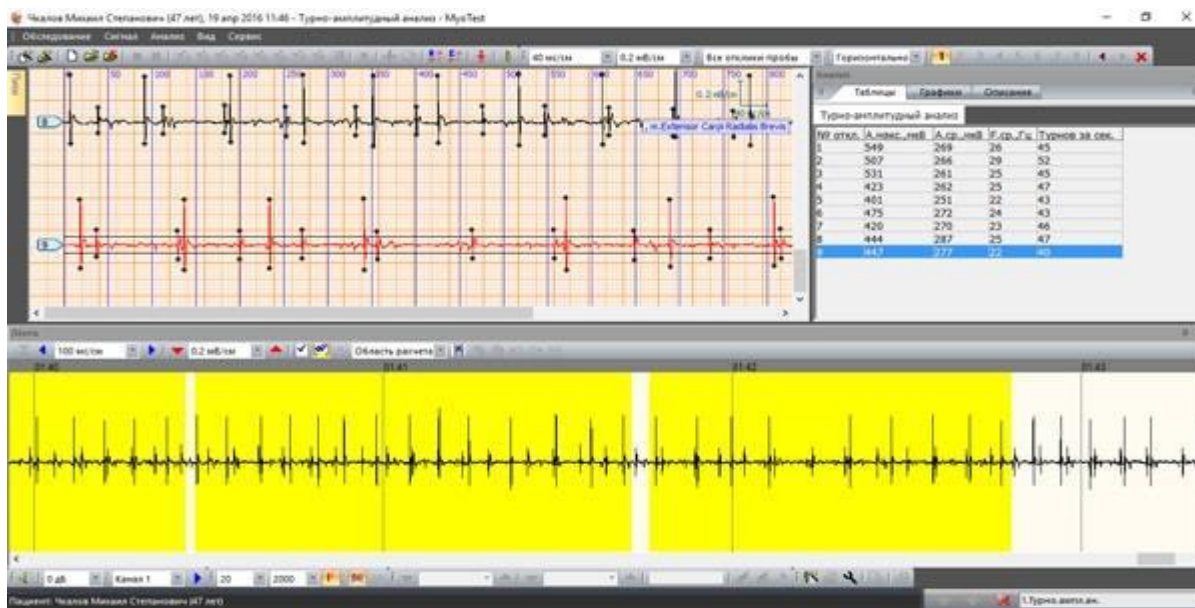


Рисунок 2.3 – Голчаста міографія: турно-амплітудний аналіз

Методики дослідження викликаних потенціалів.

За допомогою електроміографа “M-TEST” можна досліджувати викликані потенціали (ВП):

- реєструвати й аналізувати зорові ВП на спалах;
- зорові ВП на реверсивний шаховий патерн;
- довголатентні слухові ВП;
- когнітивні ВП – умовно-негативну хвилю очікування CNV і феномен P300;
- коротколатентні соматосенсорні ВП [15].

2.2 Алгоритми та методи обробки отриманих даних

2.2.1 Метод експоненційного згладжування

В основі моделей ETS лежить експоненційне згладжування – метод прогнозування, при якому значення змінної за всі попередні періоди входять у прогноз, експоненційно втрачаючи свою вагу з часом. Це дозволяє моделі з достатнім ступенем гнучко реагувати на нові зміни даних, зберігаючи при цьому інформацію про попередню поведінку тимчасового ряду.

ETS розраховує або прогнозує майбутнє значення на основі існуючих значень за допомогою версії AAA алгоритму експоненціального

згладжування. Прогнозована вартість є продовженням історичних значень у вказаній часовій шкалі. Ця модель використовує три параметри – помилку, тенденцію та сезонність [14].

Повна класифікація адитивної моделі ETS за компонентами представлена нижче:

Таблиця 2.1 – Адитивна модель ETS

Trend	Seasonal		
	N	A	M
N	$y_t = \ell_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha \varepsilon_t$	$y_t = \ell_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha \varepsilon_t$ $s_t = s_{t-m} + \gamma \varepsilon_t$	$y_t = \ell_{t-1} s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha \varepsilon_t / s_{t-m}$ $s_t = s_{t-m} + \gamma \varepsilon_t / \ell_{t-1}$
A	$y_t = \ell_{t-1} + b_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha \varepsilon_t$ $b_t = b_{t-1} + \beta \varepsilon_t$	$y_t = \ell_{t-1} + b_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha \varepsilon_t$ $b_t = b_{t-1} + \beta \varepsilon_t$ $s_t = s_{t-m} + \gamma \varepsilon_t$	$y_t = (\ell_{t-1} + b_{t-1}) s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha \varepsilon_t / s_{t-m}$ $b_t = b_{t-1} + \beta \varepsilon_t / s_{t-m}$ $s_t = s_{t-m} + \gamma \varepsilon_t / (\ell_{t-1} + b_{t-1})$
A _d	$y_t = \ell_{t-1} + \phi b_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha \varepsilon_t$ $b_t = \phi b_{t-1} + \beta \varepsilon_t$	$y_t = \ell_{t-1} + \phi b_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha \varepsilon_t$ $b_t = \phi b_{t-1} + \beta \varepsilon_t$ $s_t = s_{t-m} + \gamma \varepsilon_t$	$y_t = (\ell_{t-1} + \phi b_{t-1}) s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha \varepsilon_t / s_{t-m}$ $b_t = \phi b_{t-1} + \beta \varepsilon_t / s_{t-m}$ $s_t = s_{t-m} + \gamma \varepsilon_t / (\ell_{t-1} + \phi b_{t-1})$

Для кожного параметра ми маємо 4 значення:

- A – Адитивний; Для Trend у нас також є Additive Damped
- M – Мультиплікативний
- N – Жодного
- Z – Авто

Отже, ETS (A, M, N) означає – додавання для помилки, мультиплікатив для тренду та жодного для згладжування. Повна класифікація мультиплікативної моделі ETS за компонентами представлена в табл. 2.2.

Експоненційне згладжування — це метод математичного перетворення, який застосовується при прогнозуванні часових рядів. Свою назву він отримав через те, що при кожній наступній ітерації враховуються всі попередні значення ряду, але ступінь врахування зменшується за експоненційним законом.

$$s_t = \{c_1 s_{t-1} + \alpha \cdot (c_t - s_{t-1}) : t = 1 : t > 1 \quad (2.1)$$

де S_t – згладжений ряд; c_t – первинний ряд; a – коефіцієнт згладжування, який обирається апріорі з діапазону ($0 < a < 1$) [15].

Таблиця 2.2 – Мультиплікативна модель ETS

MULTIPLICATIVE ERROR MODELS

Trend	Seasonal		
	N	A	M
N	$y_t = \ell_{t-1}(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}(1 + \alpha\varepsilon_t)$	$y_t = (\ell_{t-1} + s_{t-m})(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} + \alpha(\ell_{t-1} + s_{t-m})\varepsilon_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1} + s_{t-m})\varepsilon_t$	$y_t = \ell_{t-1}s_{t-m}(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}(1 + \alpha\varepsilon_t)$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$
A	$y_t = (\ell_{t-1} + b_{t-1})(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1})\varepsilon_t$	$y_t = (\ell_{t-1} + b_{t-1} + s_{t-m})(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$ $b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$	$y_t = (\ell_{t-1} + b_{t-1})s_{t-m}(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1})\varepsilon_t$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$
Ad	$y_t = (\ell_{t-1} + \phi b_{t-1})(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + \phi b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1})\varepsilon_t$	$y_t = (\ell_{t-1} + \phi b_{t-1} + s_{t-m})(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$ $b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$	$y_t = (\ell_{t-1} + \phi b_{t-1})s_{t-m}(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + \phi b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1})\varepsilon_t$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$

Експоненціальний метод згладжування використовує історичні середні значення змінної за період, щоб спробувати передбачити її майбутню поведінку [16].

Модель припускає, що майбутнє буде дещо таким же, як недавнє минуле. Єдина модель, яку експоненціальне згладжування вивчає з історії попиту, це його рівень – середнє значення, навколо якого попит змінюється з часом.

Експоненціальне згладжування зазвичай використовується для прогнозування даних часових рядів на основі попередніх припущень користувача, таких як сезонність або систематичні тенденції.

Експоненціальне згладжування є широко точним методом прогнозування для короткострокових прогнозів. Техніка призначає більші ваги більш пізнім спостереженням, призначаючи експоненціально зменшуючі ваги, оскільки спостереження стають все більш віддаленими. Цей метод дає трохи ненадійні довгострокові прогнози.

Експоненціальне згладжування може бути найбільш ефективним, коли параметри часового ряду повільно змінюються з часом.

Основними типами методів прогнозування експоненціального згладжування є:

1) Просте або одинарне експоненціальне згладжування

Просте або одинарне експоненціальне згладжування (SES) — це метод прогнозування часових рядів, який використовується з одновимірними даними без тенденції та сезонної моделі. Для цього потрібен єдиний параметр, який називається альфа (α), також відомий як коефіцієнт згладжування. Альфа контролює швидкість експоненціального зменшення впливу минулих спостережень. Параметр часто має значення від 0 до 1.

Проста формула експоненціального згладжування визначається так:

$$S_t = \alpha x_t + (1 - \alpha)S_{t-1} = S_{t-1} + \alpha(x_t - S_{t-1}) \quad (2.2)$$

де S_t — згладжена статистика (просте середньозважене поточного спостереження x_t); S_{t-1} — попередня згладжена статистика; α — коефіцієнт згладжування даних, $0 < \alpha < 1$; t — період часу.

2) Подвійне експоненціальне згладжування

Цей метод відомий як модель тренду Холта або експоненціальне згладжування другого порядку. Подвійне експоненціальне згладжування використовується в прогнозуванні часових рядів, коли дані мають лінійний тренд, але не мають сезонної моделі. Основна ідея тут полягає в тому, щоб ввести термін, який може розглядати можливість демонстрації певної тенденції в серії.

Окрім параметра альфа, подвійне експоненціальне згладжування потребує іншого згладжування

фактор під назвою бета (b), який контролює спад впливу зміни тренду. Метод підтримує тренди, які змінюються адитивно (згладжування з лінійним трендом), і тренди, які змінюються мультиплікативними способами (згладжування з експоненціальним трендом).

Формули подвійного експоненціального згладжування:

$$S_1 = x_1$$

$$B_1 = x_1 - x_0$$

Для $t > 1$

$$S_t = \alpha x_t + (1 - \alpha)(S_{t-1} + b_{t-1}) \quad (2.3)$$

$$\beta_t = \beta(S_t - S_{t-1}) + (1 - \beta)b_{t-1} \quad (2.4)$$

де S_t – згладжена статистика (просте середньозважене поточного спостереження x_t);

b_t – найкраща оцінка тренду в момент часу t ;

β – коефіцієнт згладжування тренду, $0 < \beta < 1$.

3) Потрійне експоненціальне згладжування

Цей метод є найдосконалішим варіантом експоненціального згладжування, який використовується для прогнозування часових рядів, коли дані мають лінійні тренди та сезонні моделі. Техніка застосовує експоненціальне згладжування три рази – згладжування рівня, згладжування трендів і сезонне згладжування. Для контролю впливу сезонної складової додано новий параметр згладжування під назвою гамма (g) [17].

Метод потрійного експоненціального згладжування називається експоненціальним згладжуванням Холта-Вінтерса, названим на честь Чарльза Холта та Пітера Вінтерса.

Експоненціальне згладжування Холта-Вінтерса має дві категорії залежно від характеру сезонної складової:

- **Аддитивний метод Холта-Вінтера – для сезонності, яка викликає залежність.**
- Мультиплікативний метод Холта-Вінтера – для мультиплікативної сезонності.

Щоб налаштувати експоненціальне згладжування, аналітикам потрібно явно вказати всі гіперпараметри моделі. Однак це може бути складно як для початківців, так і для експертів.

Натомість чисельна оптимізація зазвичай використовується для пошуку та фінансування коефіцієнтів згладжування (альфа, бета, гамма, фі) для моделі, що призводить до найменшої помилки.

Експоненціальний метод згладжування може отримати значення для невідомих параметрів, оцінюючи їх на основі даних спостереження. Початкові значення та невідомі параметри можна оцінити шляхом мінімізації суми квадратів помилок (SSE).

Параметри, які вказують на тип зміни тенденції чи сезонності (наприклад, чи є вони адитивними чи мультиплікативними, чи їх слід пом'якшувати), потрібно вказати явно.

2.2.2 Метод аналізу детрендованих флуктуацій (DFA)

Наявність моделі ETS дозволяє прибрати шуми та тренди необроблених серії. Іншими словами, можна виділити шуми та тренди із сигналів шляхом їх розкладання. Далі можна знайти автокореляційні функції (АКФ, корелограми) зі ступеневим спадом для таких оброблених рядів без шумів і трендів. Все це схоже на відомий метод аналізу флуктуації з детрендуванням (DFA).

Клінічний діагноз і базові дослідження критично залежать від здатності записувати та аналізувати фізіологічні сигнали. Приклади включають записи частоти серцевих скорочень пацієнтів із високим ризиком раптової смерті, електроенцефалографічні (ЕЕГ) записи при епілепсії та інших розладах, а також коливання гормонів та інших молекулярних сигнальних месенджерів у нейроендокринній динаміці. Однак традиційний аналіз цих сигналів у лікарні та лабораторний аналіз не встигають за великими досягненнями в технології, які дозволяють записувати та зберігати масивні набори даних сигналів, що постійно коливаються. Дивно, але хоча нещодавно було показано, що ці типово складні сигнали представляють процеси, які є нелінійними, нестационарними та нерівноважними за своєю природою, інструменти для аналізу таких даних часто все ще припускають лінійність, стаціонарність та

умови, подібні до рівноваги. Такі звичайні методи включають аналіз середніх значень, стандартних відхилень та інших характеристик гістограм разом із класичним аналізом спектру потужності [18].

Відкриття полягає в тому, що такі складні набори даних можуть містити приховану інформацію, визначену тут як інформацію, яку неможливо отримати звичайними методами аналізу. Така інформація обіцяє бути клінічною цінністю, а також стосуватися основних механізмів здорової та патологічної функції. Фрактальний аналіз є одним із найбільш багатообіцяючих нових підходів для вилучення такої прихованої інформації з фізіологічних часових рядів. Частково це пов'язано з тим, що відсутність характерних часових (або просторових) масштабів – відмінної риси фрактальної поведінки – може надавати важливі біологічні переваги, пов'язані з адаптивністю реакції [16].

Спрощене та загальне визначення характеризує часовий ряд як стаціонарний, якщо його середнє значення, стандартне відхилення та старші моменти, а також кореляційні функції є інваріантними щодо часового зсуву. Сигнали, які не підкоряються цим умовам, є нестаціонарними. Як обговорювалося вище, обмежений часовий ряд можна відобразити на самоподібний процес шляхом інтегрування; наприклад, таким чином послідовність підкидання монети може бути відображена в одновимірному випадковому блуканні, стаціонарний інтегрований часовий ряд. Однак ще одна проблема, з якою стикаються дослідники, які застосовують цей тип фрактального аналізу до фізіологічних часових рядів, полягає в тому, що вони часто нестаціонарні. Процедура інтеграції зробить нестаціонарність вихідних даних ще більш очевидною.

Щоб подолати цю складність, запроваджено модифікований середньоквадратичний аналіз випадкового блукання, який називається аналізом флуктуації з виключеним трендом (DFA) [17, 18], який можна застосовувати до аналізу біологічних даних. Серед переваг DFA над

звичайними методами (наприклад, спектральним аналізом і аналізом Херста) є те, що він дозволяє виявити внутрішню самоподібність, вбудовану в, здавалося б, нестационарний часовий ряд, а також уникає фальшивого виявлення очевидної самоподібності, яка може бути артефактом зовнішніх тенденцій. Цей метод був успішно застосований до широкого діапазону змодельованих і фізіологічних часових рядів в останні роки [19].

Будь ласка, зверніть увагу, що хоча алгоритм DFA добре працює для певних типів нестационарних часових рядів (особливо повільно змінних тенденцій), він не призначений для обробки всіх можливих нестационарностей у даних реального світу.

Щоб проілюструвати алгоритм DFA, використано інтербітальний часовий ряд, показаний на рис. 2.4 як приклад. Спочатку інтегрується часовий ряд міжтактового інтервалу (загальної довжини N)

$$y(k) = \sum_{i=1}^k [B(i) - B_{ave}], \quad (2.5)$$

де $B(i)$ є i -м міжтактальним інтервалом і B_{ave} є середнім міжтактальним інтервалом.

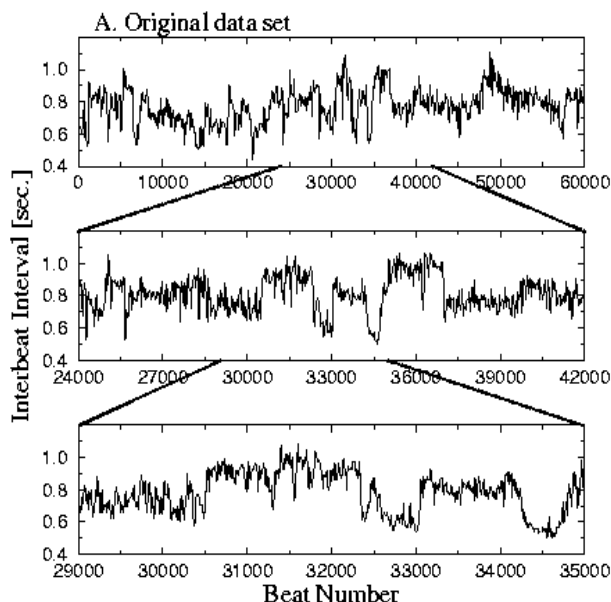


Рисунок 2.4 – Часові ряди серцевого інтервалу між серцевими скороченнями (обернені частоті серцевих скорочень)

Як показано вище, цей крок інтеграції відображає вихідний часовий ряд на самоподібний процес. Далі ми вимірюємо вертикальний характерний

масштаб інтегрованого часового ряду. Для цього інтегрований часовий ряд ділиться на блоки однакової довжини, n . У кожному прямокутнику довжини n лінія найменших квадратів відповідає даним (що представляє тенденцію в цьому полі) (рис. 2.5). Координата y відрізків прямих позначається через $y_n(k)$. Далі ми знімаємо тренд інтегрованого часового ряду, $y(k)$, віднімаючи локальний тренд, $y_n(k)$, у кожному полі. Для заданого розміру прямокутника n характерний розмір коливання для цього інтегрованого та детрендованого часового ряду обчислюється за формулою 2.6.

$$F(n) = \sqrt{\frac{1}{N} \sum_{k=1}^N [y(k) - y_n(k)]^2} \quad (2.6)$$

Local Detrending in the DFA Algorithm

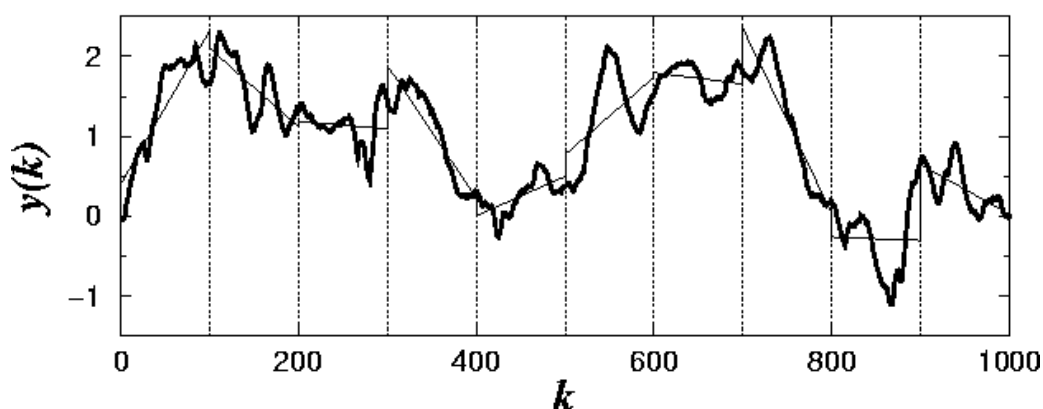


Рисунок 2.5 – Інтегрований часовий ряд формула 3. Вертикальні пунктирні лінії позначають блоки розміром $n = 100$, а суцільні відрізки прямої лінії представляють «тенденцію», оцінену в кожному квадраті лінійним методом найменших квадратів.

Це обчислення повторюється протягом усіх часових масштабів (розмірів сегменту), щоб забезпечити зв'язок між $F(n)$ і розміром сегменту n . Як правило, $F(n)$ буде збільшуватися з розміром сегменту n . Лінійний зв'язок на подвійному логарифмічному графіку вказує на наявність масштабування (самоподібності) – флуктуації в малих прямокутниках пов'язані з флуктуаціями у більших прямокутниках за степеневим законом. Нахил лінії, відношення $\log F(n)$ до $\log n$, визначає експоненту масштабування (параметр самоподібності) α , як зазначено раніше [20].

Висновки до розділу 2

У даному розділі було розглянуто принципи роботи обладнання необхідного для збору абдомінальних електроміограм. Розглянуто роботу електроміографа «M-TEST» та його методики. Розглянуто метод експоненційного згладжування та метод аналізу детрендованих флуктуацій для обробки отриманих електроміограм.

3. РЕАЛІЗАЦІЯ АЛГОРИТМІВ ТА МЕТОДІВ ОБРОБКИ ОТРИМАНИХ ДАНИХ У СИСТЕМІ КОМП'ЮТЕРНОЇ МАТЕМАТИКИ MAPLE 2020

3.1 Експоненціальна модель згладжування: усунення шумів і зменшення трендів серії ЕМГ

Обробка даних почалася з їх моделювання в рамках методу експоненціального згладжування [8]. Цей метод дозволяє представити ряд як набір: Помилка-Тренд-Сезонність (ETS) плюс сам сигнал. Таким чином, за потреби ці компоненти можна розглядати разом. Існує кілька десятків різноманітних математичних моделей для рядів у цьому методі. Отже, вибір однієї з моделей є задачею оптимізації за критерієм Байєса.

Крім того, очищені сигнали, знешумлені та усунені тренди, дозволяють отримати високоякісні корелограми, які називаються автокореляційними функціями. Перетворення Фур'є функції автокореляції є спектром потужності сигналу. Так стверджується відома в прикладній математиці теорема Вінера-Хінчина [9]. Спектри потужності (періодограмми), отримані, як описано тут, визначили патерн дихання. Цей спосіб був схожий на [10], де він використовувався в менших масштабах. Метод відрізняється від звичного для багатьох авторів, які віддають перевагу часовій області (див., наприклад, [11]).

Як рекомендовано в [4, 5], огинаючи ЕМГ були отримані шляхом згладжування ковзними (рухомими) медіанами. Такий спосіб є надійнішим, ніж звичайний середньоквадратичний (RMS) через велику кількість викидів, які зазвичай зустрічаються в більшості серій ЕМГ [4]. Таким чином, початкова задача виглядає двовимірною: медіани та пікові значення огинаючих ЕМГ формують її вхідний вектор (набір атрибутів) [5].

Хороша практика машинного навчання вимагає нормалізації атрибутів вхідного вектора. Існують різні способи такого нормування. Обрано так

званий модифікований z-скоринг, який використовує медіани та медіанні відхилення (MAD [12]):

$$z = 0.6745 \left(\frac{(x - \text{Median}(x))}{\text{MAD}(x)} \right), \quad (3.1)$$

де x тут є атрибутом. Причини віддавати перевагу медіанам і середнім відхиленням замість середніх значень і стандартних відхилень очевидні з попередніх зауважень щодо величезної кількості викидів.

Незважаючи на те, що z-показники безрозмірні та сумісні за шкалами, вони все одно можуть бути досить сильно корельованими, щоб бути статистично незалежними. Кореляційний аналіз і аналіз основних компонентів є обов'язковими етапами, якщо нам потрібні некорельовані, статистично незалежні атрибути [13].

Належним чином попередньо оброблені атрибути та шаблони дихання (класи) утворюють набір вхідних даних для машинного навчання. У цьому дослідженні ми працювали з Weka 3-9-5, популярним програмним інструментом серед багатьох практиків машинного навчання. Це програмне забезпечення було створено в Новій Зеландії (Університет Вайкато). Weka віддає перевагу ARFF (Attribute-Relation File Format), текстовому файлу ASCII, який описує список екземплярів, які спільно використовують набір атрибутів [14]. Наприклад, текстові файли ASCII можна легко редагувати в будь-якому текстовому редакторі, і Блокнот є одним із них.

Команда TimeSeries створює об'єкт часового ряду для використання з пакетом TimeSeriesAnalysis. Об'єкт TimeSeries містить один або кілька наборів даних, виміряних в один набір часових моментів.

Аргумент даних є обов'язковим. Він містить фактичні числа, що утворюють набори даних. Якщо об'єкт має представляти єдиний набір даних, то дані можна надати як список або вектор чисел. Якщо об'єкт для представлення k наборів даних, його можна надати як список k однаково довгі списки чисел, або Матриця чисел з k колонками. Крім того, можна

використати вже існуючий об'єкт набору даних `TimeSeries` для надання даних.

У обчисленнях пакета `TimeSeriesAnalysis` усі часові інтервали розглядаються як однакові, тобто, якщо ви вказуєте часові точки, розташовані нерівномірно, ця інформація використовується лише для таких цілей, як надання відповідних позначок на осях для графіків.

Оптимальна ESM модель

Функція `seq` використовується для побудови послідовності значень. Найпростішою формою є `seq(m..n)`, яка генерує послідовність числових значень $m, m+1, \dots, M$. Якщо nm є цілим числом, то M дорівнює n , інакше $n-1 < M < n$.

Більш типовим є виклик `seq(f(i), i = m..n)`, який генерує послідовність $f(m), f(m+1), \dots, f(M)$, де M є таким, як описано вище. Зверніть увагу, що `seq(i, i = m..n)` еквівалентний простішій формі `seq(m..n)`.

Якщо вираз f не містить ітераційної змінної i , то створюється послідовність ідентичних значень. У цьому випадку змінну ітерації також можна опустити в аргументі діапазону. Тобто `seq(y, i=m..n)` еквівалентний `seq(y, m..n)` і створює послідовність із $\text{floor}(nm)+1$ повторень y .

Усі наведені вище виклики `seq` можуть приймати необов'язковий останній аргумент `step`, який може мати одну з двох форм:

- Якщо `крок` є числовим значенням, послідовність генерується за допомогою значень $m, m+\text{крок}, \dots, M$, де $n\text{-крок} < M \leq n$.
- Якщо `крок` має форму `numelems=N`, де N — ціле число, послідовність генерується з використанням значень $m, m+q, \dots, n$, де $q = (n - m) / (N - 1)$.

Форма `seq(f(i), i = x)`, де x не є діапазоном, генерує послідовність шляхом застосування f до кожного операнда або запису x . Тут x найчастіше буде набором або списком, але це може бути будь-яка інша структура даних, до якої можна застосувати `op`, наприклад `sum` або `product`. Для таблиць

Arrays, Vectors і Matrices скануються записи x , а не операнди. Значення x також може бути рядком, у цьому випадку seq повторює окремі символи в ньому.

Команда ExponentialSmoothingModel створює модель експоненціального згладжування або модель ETS для часового ряду (рис. 3.1). Ці моделі є узагальненням ідеї використання експоненціального згладжування попередніх даних як прогнозу. Параметри моделі можуть бути автоматично оптимізовані відповідно до заданого часового ряду, або параметри можна встановити вручну.

```
> esm:=seq(ExponentialSmoothingModel(ts[q],damping=false),q=1..N2);# Optimal ESM
models
esm := < an ETS(A,A,N) model >, < an ETS(A,A,N) model >, < an ETS(A,A,N) model >,
< an ETS(A,A,N) model >, < an ETS(A,A,N) model >, < an ETS(A,A,N) model >,
< an ETS(A,A,N) model >, < an ETS(A,A,N) model >, < an ETS(A,A,N) model >,
< an ETS(A,A,N) model >, < an ETS(A,A,N) model >, < an ETS(A,A,N) model >,
< an ETS(A,A,N) model >, < an ETS(A,A,N) model >, < an ETS(A,A,N) model >,
< an ETS(A,A,N) model >, < an ETS(A,A,N) model >, < an ETS(A,A,N) model >,
< an ETS(A,A,N) model >, < an ETS(A,A,N) model >, < an ETS(A,A,N) model >,
< an ETS(A,A,N) model >, < an ETS(A,A,N) model >, < an ETS(A,A,N) model >,
< an ETS(A,A,N) model >, < an ETS(A,A,N) model >, < an ETS(A,A,N) model >,
< an ETS(A,A,N) model >, < an ETS(A,A,N) model >, < an ETS(A,A,N) model >,
< an ETS(A,A,N) model >, < an ETS(A,A,N) model >, < an ETS(A,A,N) model >
```

Рисунок 3.1 – Створення моделі ETS для часового ряду за допомогою ExponentialSmoothingModel

Це означає, що адитивні помилки та адитивні тенденції без будь-яких сезонних змін для всіх серій ЕМГ.

Існує 30 варіантів цієї моделі. Їх відрізняє:

- чи є помилки адитивними чи мультиплікативними;
- чи є тренд, і якщо він є, то чи є він адитивним чи мультиплікативним і чи є він демпфованим чи ні;
- чи є сезонний вплив, і якщо є, чи є він адитивним чи мультиплікативним.

За замовчуванням Maple створює модель, яка потенційно представляє будь-який із 30 варіантів. Maple може автоматично вибрати той, який підходить для часового ряду, або ви можете обмежити вибір вручну.

Обмеження вибору вручну можна виконати за допомогою параметрів E, T і S або за допомогою названих параметрів errors, trend і seasonal (рис. 3.2). Якщо використовується E, T і S, потрібно вказати всі три параметри; використовуючи названі параметри, можна використовувати будь-яку підмножину. Значення наведені в наступній таблиці.

```
> dc := [seq(Decomposition(esm[q], ts[q]), q=1..N2)]:whattype(%);# Decompositions
  oa signals on Errors, trends and levels (detrended and denoised signals)
                                     list
```

Рисунок 3.2 – Декомпозиція сигналів на помилки, тренди та рівні (сигнали з усуненням тренду та шумом).

Детрендування та знешумлення сигналу у вигляді матриці-стовпця наведено на рис. 3.3. Графік відношення сигналу до шуму наведено на рис. 3.4.

```
> ls:=Matrix([seq(GetData(dc[q][2]), q=1..N2)]):n1,n2:=Dimensions(ls);# Detrended
  and denoised signals as a column matrix
                                     n1, n2 := 32768, 40
```

Рисунок 3.3 – Детрендування та знешумлення сигналів як матриці-стовпця
Графік відношення сигналу до шуму наведено на рис. 3.4.

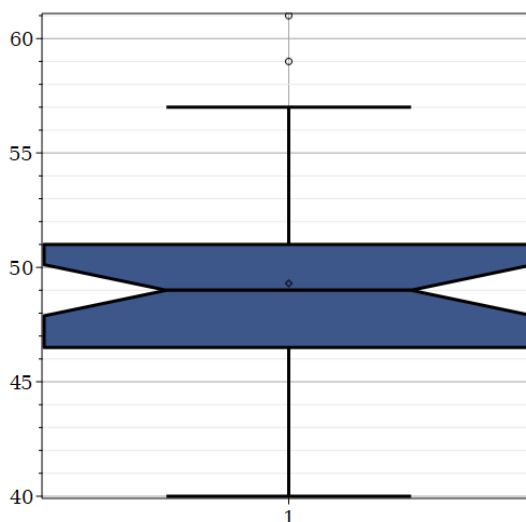


Рисунок 3.4 – Графік відношення сигналу до шуму

Пакет Statistics – це набір інструментів для математичної статистики та аналізу даних. Пакет підтримує широкий спектр типових статистичних завдань, таких як кількісний і графічний аналіз даних, моделювання та підгонка кривої. Окрім стандартних інструментів аналізу даних, пакет

Statistics надає широкий спектр символічних і числових інструментів для обчислень із випадковими величинами. Пакет підтримує понад 35 основних розподілів ймовірностей і надає можливості для визначення нових розподілів.

Команда NormalPlot створює ймовірний графік нормального розподілу (на горизонтальній осі) проти значень відповіді (на вертикальній осі). Значення відповіді відсортовані, якщо не вказано інше. Випадкова вибірка теоретичного розподілу, розмір якої дорівнює розміру даних, генерується, сортується за зростанням і наноситься на графік відносно (впорядкованих) значень відповіді. Перший параметр, X , є значеннями відповіді, які відповідають у-координатам точок на графіку. Копія значень відповіді сортується у порядку зростання у функції (тобто, якщо сортування є хибним, як і за замовчуванням), і наноситься на графік (рис. 3.5) із відсортованою випадковою вибіркою заданого теоретичного розподілу, що відповідає х-координатам графіку.

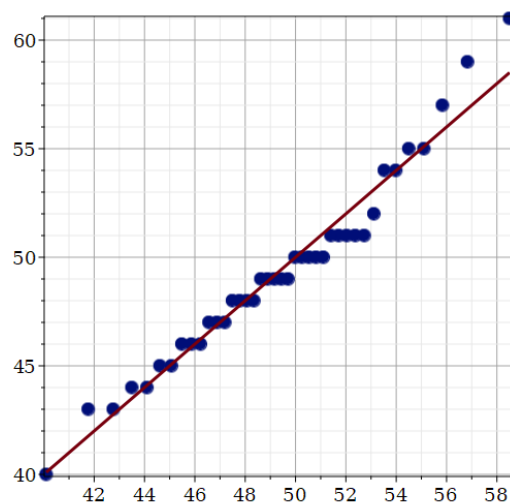


Рисунок 3.5 – Графік нормального розподілу сигналу

Стандартний графік показує, що фактичний розподіл SNR досить близький до розподілу Гаусса, хоча деякі тенденції до формування невеликих «плато» привертають увагу.

Метод експоненціального згладжування служить головним чином як інструмент прогнозування часових рядів. Основна ідея полягає в тому, що

старі дані набагато менш достовірні для прогнозу, ніж свіжі [8, 15]. Отже, старіші дані мають ваги з експоненціальним спаданням.

Вирішення оптимізаційної задачі показало, що оптимальна модель ETS (Error-Trend-Seasonality) виявилася однаковою для всіх ЕМГ набору даних. Ця універсальна модель прогнозує:

- 1) адитивні помилки;
- 2) адитивні та незатухаючі тренди;
- 3) відсутність сезонності.

Крім того, отримана модель ETS відокремлює шум (помилки) і тренди від сигналів. Можна знайти автокореляційні функції для оброблених рядів без шумів і трендів, а потім високоякісні спектри потужності (періодограми).

Варто відзначити відносно високе відношення сигнал/шум (SNR), що означає досить невеликі шуми. Це співвідношення було в діапазоні (40 - 69) дБ для наших ЕМГ. Розподіл SNR досить близький до гауссового. Для порівняння, SNR для ЕМГ [4] не перевищував 20 дБ. Це виглядає дивно, якщо згадати величезну кількість викидів на рис.1. Можливо, ці викиди не є стандартними помилками чи шумом.

3.2 Підмножини патернів дихання

Команда `time` повертає загальний час процесора, використаний з початку сеансу Maple. Одиниці вимірюються в секундах, а повернуте значення є числом із плаваючою комою. Послідовність викликів `time(x)` повертає час, витрачений на обчислення виразу `x`. Команда `time[real]` повертає реальний час, що минув з моменту запуску ядра, у секундах. Послідовність викликів `time[real](x)` повертає реальний час, використаний для обчислення виразу `x`.

Треба зауважити, що під час побудови певних структур даних Maple робить деякі автоматичні спрощення. Наприклад, Maple визнає, що вираз 3^2 може бути точно представлений у його розгорнутій формі, і таким чином

автоматично та негайно виконує розширення. Оскільки це обчислення виконується під час спрощення, і оскільки команда `time()` реєструє лише час ЦП, витрачений під час оцінки (а не спрощення), розширення не реєструється як таке, що зайняло будь-який час [5].

Послідовність викликів `Display(maplet)` відображає (запускає) програму `Maplet`. Повернутий результат може бути `NULL`, рядком, списком рядків або рядком і списком рядків, залежно від параметрів, указаних для елемента команди `Shutdown`, який закриває програму `Maplet` (рис. 3.6).

Щоразу, коли програма `Maplet` відображається, вона призначається змінній `lastmaplet`. До нього можна отримати доступ для налагодження або для повторного запуску програми `Maplet`.

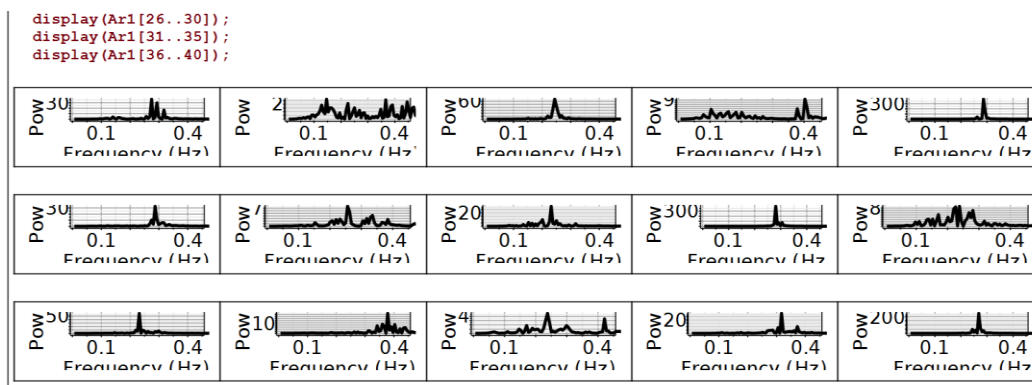


Рисунок 3.6 – Графік спектрів потужності різних моделей дихання пацієнтів

На рис. 3.7 можна виявити розбіжності між спектрами потужності представників різних моделей дихання. Високі та вузькі піки, можливо, з кількома незначними сателітами, характерні для регулярного дихання (ліва сторона рис. 3.7). Нерегулярна картина показує набагато більше низьких вершин у набагато ширшому діапазоні частот. Таким чином, розпізнавання патернів дихання є візуально доступним для клініцистів, якщо вони мають спектри потужності ЕМГ у частотній області [6].

Розробка нового діагностичного знаку за результатами інтелектуального аналізу абдомінальних електроміограм

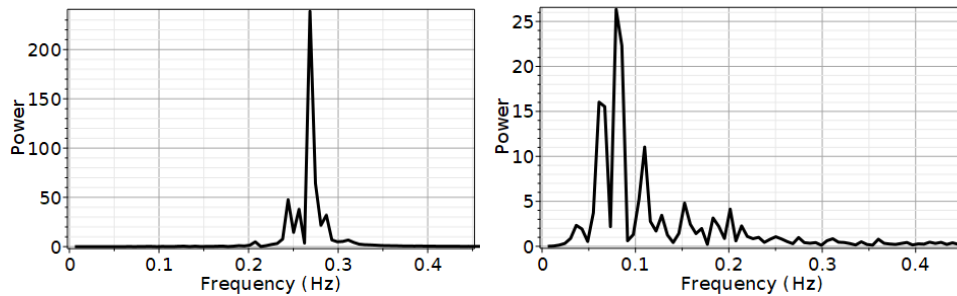


Рисунок 3.7 – Спектри потужності двох суб'єктів: один із регулярним диханням (лівосторонній графік), інший (правий) із нерегулярним диханням.

ЕМГ, звільнені від шумів і трендів, використовувалися для побудови спектрів потужності (періодограмм) відповідно до згаданої вище теореми Вінера-Хінчина. Було проаналізовано ці спектри і виявлено десять суб'єктів із регулярним і тридцять із нерегулярним диханням. Таким чином, середня частота регулярного дихання у дорослих під час сну була приблизно дорівнює одній чверті (0,25) від усього набору даних [22].

3.3 Відсоток викидів

3.3.1 Що таке викиди

Під час дослідження даних викиди – це крайні значення в наборі даних. Це означає, що викидні точки даних сильно відрізняються від очікуваних значень – вони або значно більші, або значно менші. Для даних, які відповідають нормальному розподілу, значення, які падають більше ніж на три стандартні відхилення від середнього, зазвичай вважаються викидами.

Викиди можуть потрапляти в набір даних природним шляхом через мінливість або вони можуть бути результатом таких проблем, як людська помилка, несправне обладнання або погана вибірка. Незалежно від того, як вони потрапляють у дані, викиди можуть мати великий вплив на статистичний аналіз і машинне навчання, оскільки вони впливають на такі обчислення, як середнє та стандартне відхилення, і можуть спотворити перевірку гіпотез. Аналітик даних повинен використовувати різні методи візуалізації та ідентифікації викидів, перш ніж вирішити, чи слід їх видалити, зберегти або змінити [23].

3.3.2 Чотири способи обчислення викидів

Можна сортувати кількісні змінні від низького до високого та шукати надзвичайно низькі чи надзвичайно високі значення. Позначте будь-які екстремальні значення, які знайдете.

Це простий спосіб перевірити, чи потрібно вам досліджувати певні точки даних перед використанням більш складних методів.

Набір даних для пілотного експерименту складається з 8 значень.

180 156 9 176 163 1827 рік 166 171.

Ви сортуєте значення від низького до високого та шукаєте крайні значення.

9 156 163 166 171 176 180 1872 рік.

Використання візуалізацій

Ви можете використовувати програмне забезпечення для візуалізації своїх даних за допомогою прямокутної діаграми або діаграми коробки та вусів, щоб ви могли одразу побачити розподіл даних. Цей тип діаграми виділяє мінімальні та максимальні значення (діапазон), медіану та інтерквартильний діапазон для ваших даних.

У багатьох комп'ютерних програмах викиди на діаграмі виділяються зірочкою, і вони лежатимуть поза межами графіка [25].

Виявлення статистичних викидів

Виявлення статистичних викидів передбачає застосування статистичних тестів або процедур для визначення екстремальних значень.

Ви можете конвертувати екстремальні точки даних у z-рахунки, які покажуть, на скільки стандартних відхилень вони від середнього значення.

Якщо значення має досить високий або досить низький z-оцінку, його можна вважати викидом. Як правило, значення з оцінкою az більше 3 або менше -3 часто визначаються як викиди.

Інтерквартильний діапазон (IQR) повідомляє про діапазон середньої половини вашого набору даних (рис. 3.8). Ви можете використовувати IQR,

щоб створити «паркани» навколо своїх даних, а потім визначити викиди як будь-які значення, які виходять за межі цих парканів.

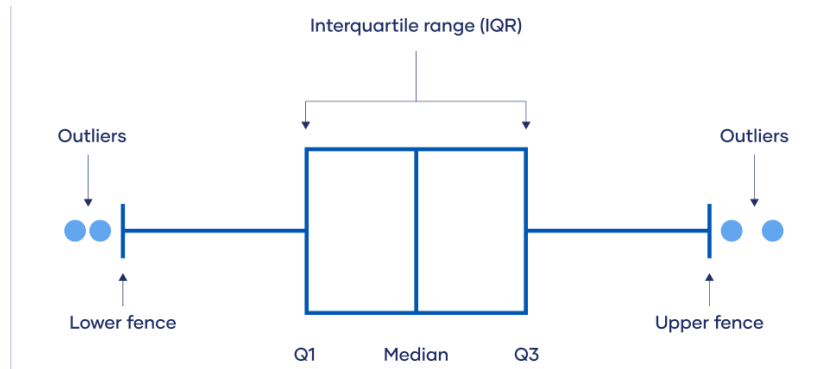


Рисунок 3.8 – Візуалізація інтерквартильного діапазону

Цей метод корисний, якщо у вас є кілька значень на крайніх кінцях набору даних, але ви не впевнені, чи може якесь із них вважатися викидом.

Метод міжквартильного діапазону:

1. Відсортувати дані від низького до високого
2. Визначити перший квартиль (Q1), медіану та третій квартиль (Q3).
3. Обчислити свій $IQR = Q3 - Q1$
4. Обчислити верхню огорожу $= Q3 + (1,5 * IQR)$
5. Обчислити нижню огорожу $= Q1 - (1,5 * IQR)$
6. Використати свої обмеження, щоб виділити будь-які викиди, усі значення, які виходять за межі ваших парканів [24].

3.3.2 Розрахунок відсотку викидів

Для підрахунку кількості викидних частин (частки, відсотки) використано пакунки: `Statistics`, `LinearAlgebra`, `SignalProcessing`.

Пакет `LinearAlgebra` пропонує підпрограми для побудови та обробки матриць і векторів, обчислення стандартних операцій, результатів запитів і вирішення задач лінійної алгебри.

Матриця, вектор або послідовність виразів обох. Матриці 10 x 10 або менше та вектори 10 x 1 і менше, відображають відповідну матрицю або вектор на аркуші Maple (рис. 3.9). Матриці та вектори, більші за цей, відображають заповнювач як вихідні дані.

Пакет SignalProcessing містить команди для маніпулювання сигналами. Зокрема, ці команди працюють із зразками сигналів, що зберігаються у векторах із типом даних float або complex.

```
> No:=nout(dtm);numelems(No);whattype(No);max(No);min(No);# Outliers percentage (fractions)
No := [0.019135, 0.24564, 0.0033569, 0.038177, 0.00076294, 0.015228, 0.082153, 0.0092773, 0., 0.014923, 0.013397, 0.0075378, 0.10327, 0.0079041, 0.,
0.061829, 0., 0.049896, 0.16071, 0.056030, 0.016388, 0.020935, 0.10693, 0.0097351, 0.019470, 0.035309, 0., 0.070801, 0.061340, 0.0074158, 0., 0.19293,
0.097565, 0.072632, 0.044464, 0.031189, 0.00048828, 0.0032959, 0.031647, 0.013702]
40
list
0.24564
0.
```

Рисунок 3.9 – Викиди у відсотках (дробки)

Функція MedianDeviation обчислює середнє абсолютне відхилення від медіани вказаної випадкової змінної або набору даних (рис. 3.10).

```
> bw1:=0.9*min(MedianDeviation(z1), InterquartileRange(z1)/1.34)*N2^(-0.2);
bw1 := 0.29029
```

Рисунок 3.10 – Обчислення середнього абсолютного відхилення від медіани

Зобразимо отримані результати статистичних викидів у вигляді гістограми (рис. 3.11):

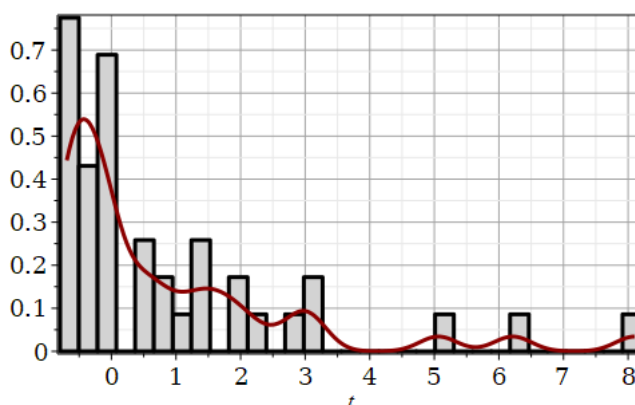


Рисунок 3.11 – Гістограма статистичних викидів

3.4 Повнохвильове випрямлення та згладжування ЕМГ методом рухомих медіан

Основним методом фіксації обвідної ЕМГ є обчислення середньоквадратичного (RMS) значення сигналу в межах «ковзного вікна» з вибраним розміром. Середньоквадратичне значення сигналу у вікні наноситься в центрі вікна, щоб уникнути часових зсувів огинаючої відносно руху. Тим часом RMS-віконне виявляється недостатньо надійним під

впливом численних викидів у серії ЕМГ [26,27]. Тому застосовано метод [27] робастної обробки серії ЕМГ із вікном рухомих медіани.

Важливе значення тут має правильний вибір розміру вікна. Вікно має бути не дуже великим і не дуже маленьким. Нашим фаворитом виявилася кількість точок (вибірок) на інтервалі усереднення (вікні), що дорівнює 1024 вибіркам. Таке рухоме вікно відповідає часовому інтервалу довжиною приблизно 5,1 секунди, що становить $1/32$ загальної тривалості кожної серії.

Частка викидів у наборі даних була в діапазоні (0–0,274), і лише дві ЕМГ не мали викидів. Медіани огинаючих були в діапазоні (39 - 1868) мкВ; пікові значення були в діапазоні (від 198 до 3180) мкВ. Виявилось, що медіани за z-балами та пікові значення огинаючої ЕМГ досить сильно корелюють: коефіцієнт кореляції дорівнював 0,7245. Критичне значення коефіцієнта кореляції, який розділяє значущі і незначущі кореляції, було нижчим: 0,2635 при заданому розмірі популяції (40) і рівні довіри 0,95. Отже, нам потрібні були основні компоненти z-показників, які були б статистично незалежними. Методи аналізу основних компонентів підходять для таких компонентів (c1, c2).

Гістограми та оцінки щільності ймовірності відповідають [28,29] і показані на рис.3.12 Обидва мають мультимодальні розподіли з первинними (вищими) модами в додатних діапазонах головних компонент. Другорядні (нижчі) моди займають негативні діапазони, як показано на рис.3.

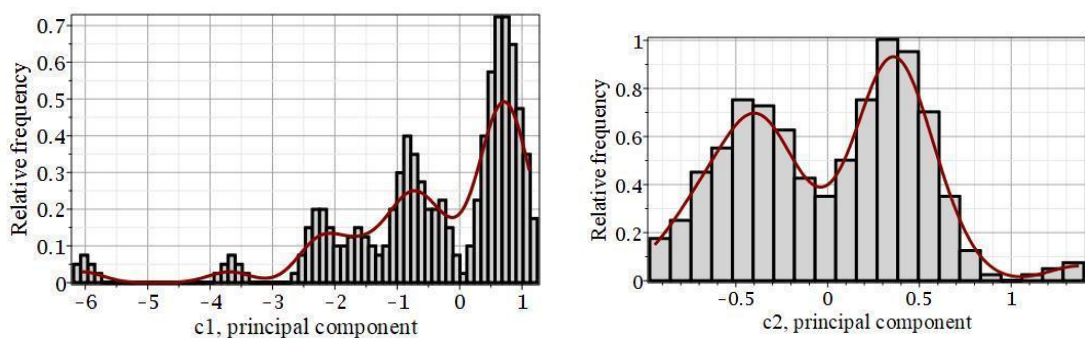


Рисунок 3.12 – Мультимодальні розподіли головних компонент (c1 і c2); криві відображають оцінки функцій густини ймовірностей згідно з [30].

Такі особливості гістограм спонукають нас розділити випадки з обома позитивними числовими атрибутами ($c_1, c_2 > 0$ означає випадок А, інакше В) на дві окремі категорії. Вони можуть бути додатковим номінальним атрибутом із двома можливими станами (А та В або навіть двійковими: 0 та 1). Таким чином, ми маємо в результаті набір даних із 40 екземплярів. Кожен з них описується чотирма атрибутами: двома числовими (c_1, c_2) і двома номінальними: {А, В} і атрибутом класу ({irregular, regular}).

Візьмемо абсолютне значення сигналу (рис. 3.13). Це також називається повнохвильовим випрямленням (W.Rose).

```
> st1:=time() :  
fwr:=Matrix([seq(map(abs,Column(ls,q)),q=1..N2)]) :m1,m2:=Dimensions(fwr) ;  
t_fwr:=time()-st1;  
#Take the absolute value of the signal. This is also called full wave rectification (W.Rose)  
|  
  
m1,m2 := 32768,40  
t_fwr := 59.67
```

Рисунок 3.13 – Реалізація повнохвильового випрямлення

Нехай кількість точок (вибірок) на інтервалі усереднення дорівнює $2^8=256$ вибірок. Такий рухомий інтервал (вікно) збігається з часом тривалістю приблизно 1,28 с, що становить $1/128$ частину загальної тривалості сигналу (3.14).

```
> st2:=time() :  
env:=Matrix([seq(Statistics:-MovingStatistic(Column(fwr,q),2^10,Median),q=1..N2)]) :  
Dimensions(env) ;|  
t_s:=time()-st2;# Matrix of envelope  
  
31745,40  
t_s := 197.01
```

Рисунок 3.14 – Матриця згортання

Висновок до розділу 3

У даному розділі детально описано експоненціальну модель згладжування: усунення шумів і зменшення трендів серії ЕМГ. Детально розглядалися принципи розрахунку викидів тощо.

У даному розділі мова йшла про конкретний алгоритм застосування методів. Оглянуто було різноманітні патерни дихання, котрі були виявлені у

досліджуваних пацієнтів. Окрім того обирався спосіб обрахунку відсотків викидів. Обране було виявлення статистичних викидів.

Також описано повнохвильове випрямлення та згладжування ЕМГ методом рухомих медіан. Обрано розмір вікна, з урахуванням викидів. Розглянуто розділення випадків з обома позитивними числовими атрибутами та обрано абсолютне значення сигналу.

4. ТЕСТУВАННЯ ТА АНАЛІЗ ОБЧИСЛЕНЬ У СИСТЕМІ КОМП'ЮТЕРНОЇ МАТЕМАТИКИ MAPLE 2020

4.1 Файл ARFF і крива навчання

ARFF означає Attribute-Relation File Format. Це текстовий файл ASCII, який описує список екземплярів, які мають спільний набір атрибутів. Файли ARFF були розроблені Проектом машинного навчання на кафедрі комп'ютерних наук Університету Вайкато для використання з програмним забезпеченням машинного навчання Weka. Цей документ описує версію ARFF, яка використовується з версіями Weka від 3.2 до 3.3; це розширення формату ARFF, як описано в книзі з інтелектуального аналізу даних, написаній Яном Х. Віттенем і Ейбе Франком (новими доповненнями є атрибути рядка, атрибути дати та розріджені екземпляри).

Файли ARFF мають два окремі розділи. Перший розділ — це інформація заголовка, за якою йде інформація про дані.

Заголовок файлу ARFF містить назву відношення, список атрибутів (стовпців у даних) і їх типи. Приклад заголовка стандартного набору даних IRIS виглядає так:

```
% 1. Title: Iris Plants Database
%
% 2. Sources:
%   (a) Creator: R.A. Fisher
%   (b) Donor: Michael Marshall (MARSHALL%PLU@io.arc.nasa.gov)
%   (c) Date: July, 1988
%
@RELATION iris

@ATTRIBUTE sepallength NUMERIC
@ATTRIBUTE sepalwidth NUMERIC
@ATTRIBUTE petallength NUMERIC
@ATTRIBUTE petalwidth NUMERIC
@ATTRIBUTE class {Iris-setosa,Iris-versicolor,Iris-virginica}
```

Дані файлу ARFF виглядають так:

```
@DATA
5.1,3.5,1.4,0.2,Iris-setosa
4.9,3.0,1.4,0.2,Iris-setosa
4.7,3.2,1.3,0.2,Iris-setosa
4.6,3.1,1.5,0.2,Iris-setosa
5.0,3.6,1.4,0.2,Iris-setosa
5.4,3.9,1.7,0.4,Iris-setosa
4.6,3.4,1.4,0.3,Iris-setosa
5.0,3.4,1.5,0.2,Iris-setosa
```

```
4.4,2.9,1.4,0.2,Iris-setosa
```

```
4.9,3.1,1.5,0.1,Iris-setosa
```

Рядки, які починаються з %, є коментарями. Оголошення @RELATION, @ATTRIBUTE і @DATA не чутливі до регістру.

Кілька відомих наборів даних машинного навчання поширюються разом із Weka у каталозі \$WEKAHOME/data у вигляді файлів ARFF [26].

4.2 Розділ заголовка ARFF

Розділ заголовка ARFF файлу містить оголошення відношення та оголошення атрибутів.

4.2.1 Оголошення @relation

Ім'я відношення визначається як перший рядок у файлі ARFF. Формат: @relation <ім'я-зв'язку>, де <ім'я-відношення> є рядком. Рядок має бути взято в лапки, якщо ім'я містить проміжки.

4.2.2 Оголошення @attribute

Оголошення атрибутів мають форму впорядкованої послідовності операторів @attribute . Кожен атрибут у наборі даних має власний оператор @attribute , який унікально визначає назву цього атрибута та його тип даних. Порядок оголошення атрибутів вказує на позицію стовпця в розділі даних файлу. Наприклад, якщо атрибут є третім оголошеним, Weka очікує, що всі значення атрибутів будуть знайдені в третьому стовпці, розділеному комами.

Формат оператора @attribute такий: @attribute <назва-атрибуту> <тип даних>, де <назва-атрибуту> повинна починатися з букви. Якщо в назві мають бути пробіли, то ім'я повністю має бути взято в лапки [27].

<datatype> може бути будь-яким із чотирьох типів, які зараз (версія 3.2.1) підтримуються Weka:

- числові;
- <номінальна специфікація>;
- рядок;
- дата [<формат-дати>];

де <номінальна специфікація> і <формат дати> визначені нижче. Ключові слова numeric, string і date не чутливі до регістру.

Числовими атрибутами можуть бути дійсні чи цілі числа.

Номінальні значення визначаються шляхом надання <nominal-specification> із переліком можливих значень: {<nominal-name1>, <nominal-name2>, <nominal-name3>, ...}

Наприклад, значення класу набору даних Iris можна визначити таким чином: @ATTRIBUTE class {Iris-setosa,Iris-versicolor,Iris-virginica}.

Значення, які містять проміжки, повинні бути взяті в лапки.

4.2.3 Рядкові атрибути

Рядкові атрибути дозволяють нам створювати атрибути, що містять довільні текстові значення. Це дуже корисно в програмах для видобутку тексту, оскільки ми можемо створювати набори даних із рядковими атрибутами, а потім писати фільтри Weka для маніпулювання рядками (наприклад, StringToWordVectorFilter). Рядкові атрибути оголошуються наступним чином: @ATTRIBUTE LCC string.

4.2.4 Атрибути дати

Оголошення атрибутів дати мають такий вигляд: @attribute <name> date [<date-format>], де <name> – це ім'я атрибута, а <date-format> – необов'язковий рядок, що вказує, як значення дати мають аналізуватися та друкуватися (це той самий формат, який використовується SimpleDateFormat). Рядок формату за замовчуванням приймає комбінований формат дати й часу ISO-8601: «rrrr-MM-дд'T'ГГ:хм:сс» [28, 29].

Дати мають бути вказані в розділі даних як відповідні рядкові представлення дати/часу (див. приклад нижче).

4.3 Розділ даних ARFF

Розділ даних ARFF у файлі містить рядок декларації даних і фактичні рядки екземпляра.

Декларація @data

Оголошення @data — це один рядок, який позначає початок сегмента даних у файлі. Формат: @data.

Кожен екземпляр представлено в одному рядку, а символи повернення каретки позначають кінець екземпляра.

Значення атрибутів для кожного екземпляра розділені комами. Вони мають з'являтися в тому порядку, в якому вони були оголошені в розділі заголовка (тобто дані, що відповідають n-му оголошенню @attribute , завжди є n-м полем атрибута).

Відсутні значення позначаються одним знаком питання, наприклад:

@data

4.4,?,1.5,?,Iris-setosa

Значення рядкових і номінальних атрибутів чутливі до регістру, і всі, що містять пробіли, повинні бути взяті в лапки, як показано нижче:

@relation LCCvsLCSH

@attribute LCC string

@attribute LCSH string

@data

AG5, 'Encyclopedias and dictionaries.;Twentieth century.'

AS262, 'Science -- Soviet Union -- History.'

AE5, 'Encyclopedias and dictionaries.'

AS281, 'Astronomy, Assyro-Babylonian.;Moon -- Phases.'

AS281, 'Astronomy, Assyro-Babylonian.;Moon -- Tables.'

Дати мають бути вказані в розділі даних за допомогою рядкового представлення, указанного в декларації атрибута. Наприклад:

Позначки часу @RELATION

@ATTRIBUTE мітка часу ДАТА «pppp-ММ-дд ГГ:xx:cc»

@data

«2001-04-03 12:12:12»

«2001-05-03 12:59:55»

4.4 Розріджені файли ARFF

Розріджені файли ARFF дуже схожі на файли ARFF, але дані зі значенням 0 не представлені явно.

Розріджені файли ARFF мають однаковий заголовок (тобто теги @relation і @attribute), але розділ даних відрізняється. Замість представлення кожного значення в порядку, як це:

@data

0, X, 0, Y, «клас А»

0, 0, W, 0, «клас В»

Ненульові атрибути явно ідентифікуються номером атрибута та їхнім значенням, як це:

@data

{1 X, 3 Y, 4 «клас А»}

{2 Вт, 4 «клас В»}

Кожен екземпляр оточений фігурними дужками, а формат кожного запису: <індекс> <пробіл> <значення>, де індекс — індекс атрибута (починаючи з 0).

Зауважимо, що пропущені значення в розрідженому екземплярі дорівнюють 0, це не «пропущені» значення! Якщо значення невідоме, маємо явно представити його знаком питання (?).

Існує відома проблема збереження об'єктів SparseInstance із наборів даних, які мають рядкові атрибути. У Weka значення рядків і номінальних даних зберігаються як числа; ці числа діють як індекси в масиві можливих значень атрибутів (це дуже ефективно). Однак першому значенню рядка присвоюється індекс 0: це означає, що внутрішньо це значення зберігається

як 0. Коли записується SparseInstance, екземпляри рядка з внутрішнім значенням 0 не виводяться, тому їх значення рядка втрачається (і коли файл arff читається знову, значення за замовчуванням 0 є індексом іншого значення рядка, тому значення атрибута виглядає зміненим). Щоб вирішити цю проблему, додайте фіктивне значення рядка в індексі 0, яке ніколи не використовується, коли ви оголошуєте атрибути рядка, які, імовірно, використовуватимуться в об'єктах SparseInstance і зберігаються як файли Sparse ARFF [30].

Далі наведено лістинг коду розробленого для роботи з файлом ARFF та кривою навчання:

```
Mt:=Tr(Matrix([seq(<z_1[q], z_2[q], zab[q], z4[q]>, q=1..N2)]));
      _rtable[18446745645995885382]
seq((convert(Row(Mt,p),list),p=40)):op(%);
      0.116791271649549, -1.24600085042767, B, irregular
with(RandomTools);
[AddFlavor, BlumBlumShub, Generate, GenerateSimilar,
  GenerateSimilarODE, GetFlavor, GetFlavors, GetState, HasFlavor,
  LinearCongruence, MersenneTwister, QuadraticCongruence,
  RandomExpand, RemoveFlavor, SetState, genIntegrationExpression]
Generate(set(posint(range = 40), 20));nops(%);
      {1, 2, 5, 13, 19, 21, 22, 29, 32, 33, 34, 35, 36, 37, 40}
```

15

4.5 Продуктивність класифікаторів Weka

WEKA – програмне забезпечення з відкритим вихідним кодом надає інструменти для попередньої обробки даних, впровадження кількох алгоритмів машинного навчання та інструменти візуалізації, щоб розробляти методи машинного навчання та застосовувати їх до реальних проблем інтелектуального аналізу даних. Те, що пропонує WEKA, узагальнено на наступній діаграмі (рис. 4.1).

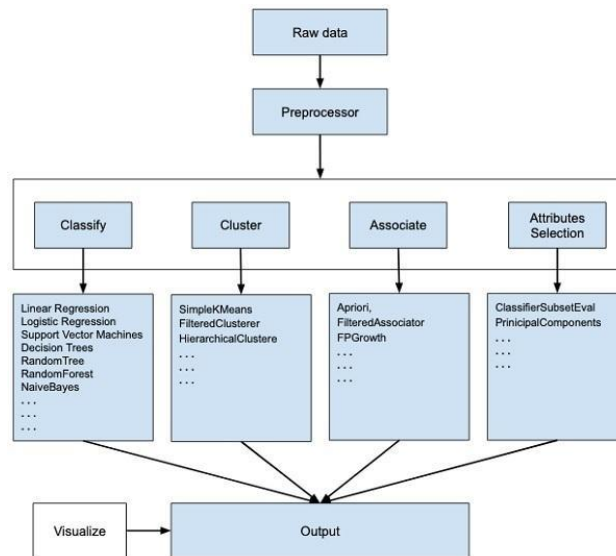


Рисунок 4.1 – Діаграма роботи програмного забезпечення WEKA

Якщо спостерігати за початком потоку зображення, зрозуміло, що є багато етапів роботи з великими даними, щоб зробити їх придатними для машинного навчання.

По-перше, почнемо з зібраних необроблених даних. Ці дані можуть містити кілька нульових значень і нерелевантних полів. Для очищення даних використовуємо інструменти попередньої обробки даних, надані в WEKA.

Потім збережемо попередньо оброблені дані у локальному сховищі для застосування алгоритмів ML.

Далі, залежно від типу моделі ML, яку намагаємося розробити, можна вибрати один із варіантів, наприклад Classify, Cluster або Associate. Вибір атрибутів дозволяє автоматично вибирати об'єкти для створення скороченого набору даних [30].

Звернімо увагу, що в кожній категорії WEKA забезпечує реалізацію кількох алгоритмів. Необхідно вибрати алгоритм, встановити потрібні параметри та запустити його на наборі даних.

Тоді WEKA надасть статистичні результати обробки моделі. Він надає інструмент візуалізації для перевірки даних.

Різні моделі можна застосовувати до одного набору даних. Потім можна порівняти результати різних моделей і вибрати найкращу, яка відповідає вашим цілям.

Таким чином, використання WEKA призводить до більш швидкого розвитку моделей машинного навчання в цілому.

Інтерфейс Explorer містить наступні панелі:

- Панель попереднього опрацювання уможливорює імпорт даних з бази даних, текстових файлів у форматі CSV, а також попереднє опрацювання цих даних за допомогою різноманітних алгоритмів (фільтрів). Ці фільтри використовуються для трансформування даних, а також для видалення певних атрибутів.
- Панель класифікації надає можливість застосувати алгоритми класифікації та регресійного аналізу до обраного набору даних, візуалізувати та оцінити результати, відобразити ROC криві тощо.
- Панель асоціації надає доступ до методів, які дозволяють оцінити взаємозв'язки між атрибутами.
- Панель кластеризації містить різноманітні методи кластеризації, наприклад метод кластеризації методом k-середніх, EM-алгоритм тощо.
- Панель вибору атрибутів дозволяє ідентифікувати атрибути, які найбільш впливають на якість прогнозування.
- Панель візуалізації відображає точкові діаграми.

Weka 3-9-5 [31] пропонує багато алгоритмів класифікації та чотири режими тестування. Для прогнозування вихідних даних WEKA містить класифікатори. Доступні для навчання алгоритми класифікації - це дерева рішень, векторні машини підтримки, класифікатори на основі екземплярів, логістична регресія та байєсівські мережі. Залежно від вимоги, використовуючи пробне і тестове дослідження, користувач може знайти відповідний алгоритм для аналізу даних. Класифікатори використовуються

для класифікації наборів даних на основі характеристик атрибутів. Перехресна перевірка є однією із цих чотирьох. Класифікатор оцінюється шляхом перехресної перевірки, використовуючи кількість складок, введена в текстове поле Folds (у нашому випадку 4, 5 або 8). Таблиця 4.1 містить результати набору даних, класифікованих кількома класифікаторами Weka, і основні оцінки їх ефективності. Елементи матриці плутанини визначають більшість таких індексів ефективності.

Таблиця 4.1 – Продуктивність класифікаторів Weka для набору даних

Класифікатор	Складки	Матриця плутанини	Точність	Чутливість	МСС	Площа РПЦ	Каппа
Випадкове дерево	8	$\begin{pmatrix} 26 & 5 \\ 2 & 7 \end{pmatrix}$	0,851	0,825	0,562	0,808	0,551
Багатошаровий перцептрон	5	$\begin{pmatrix} 27 & 4 \\ 3 & 6 \end{pmatrix}$	0,833	0,825	0,518	0,763	0,517
LWL (локально зважене навчання)	8	$\begin{pmatrix} 29 & 2 \\ 4 & 5 \end{pmatrix}$	0,842	0,850	0,540	0,854	0,533
Логістика	8	$\begin{pmatrix} 28 & 3 \\ 3 & 6 \end{pmatrix}$	0,850	0,850	0,570	0,681	0,570
LMT (дерево логістичної моделі)	4	$\begin{pmatrix} 28 & 3 \\ 3 & 6 \end{pmatrix}$	0,850	0,850	0,570	0,782	0,570
AdaBoostM1	8	$\begin{pmatrix} 29 & 2 \\ 4 & 5 \end{pmatrix}$	0,842	0,850	0,540	0,837	0,533

Зокрема:

- 1) Точність - це частка правильно класифікованих екземплярів, які відповідають класу .
- 2) Чутливість (відкликання): частка екземплярів, класифікованих як даний клас, поділена на фактичну загальну кількість у цьому класі; може бути показано або для кожного класу, або як середньозважене для набору даних (як у таблиці 4.1).

- 3) MCC (коефіцієнт кореляції Метьюза) оцінює кореляцію між спостережуваною та прогнозованою бінарними класифікаціями; ця кореляція краща, якщо цей індекс ближче до 1.
- 4) ROC area: цей показник дає нам уявлення про ефективність класифікаторів у цілому; продуктивність дуже хороша, якщо цей індекс перевищує 0,85.
- 5) Статистика Каппа є майже синонімом надійності між оцінювачами; узгодження є помірним для діапазону (0,41-0,60) і суттєвим для каппа з діапазону (0,61-0,8).

4.6 Кластеризація набору даних

Давайте оцінимо, наскільки добре попередньо призначені класи в даних збігаються з вибраними кластерами. Це можливо в режимі «Оцінка класів для кластерів», доступному в Weka у вікні «Кластеризація». Таблиця 4.2 демонструє ступінь узгодженості між класами та реальними кластерами.

Таблиця 4.2 – Відповідність класів кластерам

Метод	Кластерні екземпляри	Призначено для кластерів	Центроїд кластерів	Неправильно згруповані екземпляри
EM (максимізація очікування)	0 – 11 (28%)	0 1 4 ; 27	0,5031; 0,6251; A	6 (15%)
	1 – 29 (72%)	нерегулярний 7 ; 2 reg	-1,3415; - 0,3826; B	
FartgestFirst	0 – 9 (23%)	0 1 3 ; 28	0,7037; 0,1647; A	6 (15%)
	1 – 31 (77%)	нерегулярний 6 ; 3 reg	-0,5374 0,5458; B	
Ієрархічний	0 – 9 (23%)	0 1		6 (15%)
	1 – 31 (77%)	3 ; 28 нерегулярний 6 ; 3 reg		

Зауважте, що кластери позначені 0 і 1 у таблиці 4.2. Результати виглядають досить задовільними: три різні методи дають приблизно однакову кластеризацію. Два кластери, які містять 85% правильно кластеризованих екземплярів, непогано відповідають двом попередньо

призначеним класам. Відповідність гірша, якщо розглядати регулярний патерн дихання окремо. Він не перевищує 66%.

4.7 Коротке обговорення результатів машинного навчання

Здається корисним порівняти результати таблиць 4.1 і 4.2 з аналогічними результатами для інших наборів медичних даних, щоб оцінити наші результати. На щастя, у Weka 3-9-5 є набір медичних даних під назвою «рак молочної залози.arff». Цей набір даних складається з 286 екземплярів, кожен з яких має десять атрибутів, включаючи атрибут класу: no-recurrence-Events (201) або recurrence-Events (85). Отже, цей набір даних виглядає більш повним, ніж наш (40 екземплярів і чотири атрибути).

Класифікація реального набору даних «рак молочної залози» за допомогою одного з методів таблиці 1 (LWL, локально зважене навчання) у тому ж режимі перехресної перевірки (з 28 згортками) дає нам досить скромну продуктивність. Точність дорівнює 0,717, чутливість - 0,738, МСС - 0,305, площа ROC - 0,643, а каппа-статистика дорівнює 0,288. Неправильно класифіковані екземпляри 75 (20 і 55 для обох класів відповідно).

Кластеризація набору даних «рак молочної залози» за методом найдалшого першого кластеризатора в тому самому режимі «оцінка класів у кластери» дає нам два кластери з 221 і 65 екземплярами відповідно. Однак кількість неправильно кластеризованих екземплярів досить велика: 98 (39 і 59 відповідно).

Порівнюючи ці оцінки з результатами попереднього розділу (таблиці 4.1 і 4.2), вони є перспективними для такого невеликого набору даних. Дійсно, успіх машинного навчання часто залежить від вибору атрибутів, а не від кількості екземплярів і атрибутів. Таким чином, пропозиції [32] були успішними з цієї точки зору.

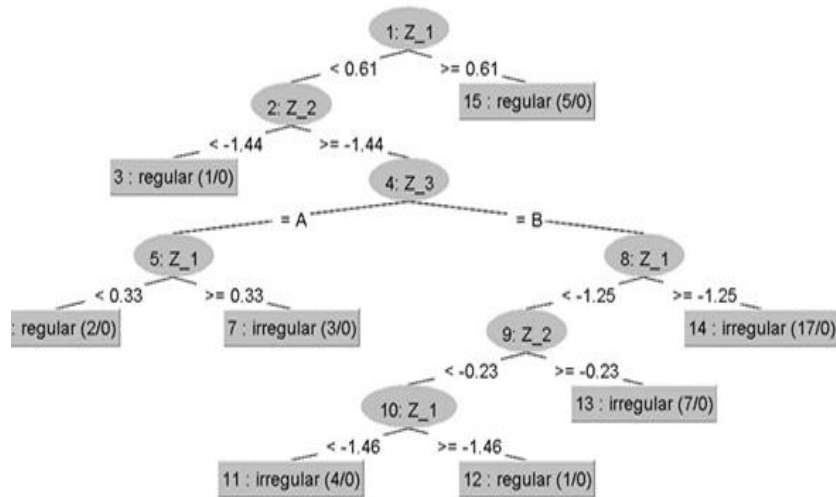


Рисунок 4.2 – «Дерево рішень» у класифікаторі Random Tree; овали – це вузли, а прямокутники – листя.

Варто зазначити, що результати таблиці 4.1 є більш-менш однорідними. Вони не надто залежать від класифікаторів. Зокрема, досить помірні статистичні значення МСС і карра з таблиці 4.1 проявляються в таблиці 4.2 як неправильне призначення класів кластерам. Тому збіг регулярних патернів дихання з кластерами 0 знаходиться лише в межах (60-66)%.

Рис. 4.2 візуалізує дерево рішень у класифікаторі Random Tree, яке показало «середню продуктивність» у таблиці 4.1. Класифікуємо екземпляр з атрибутами, згенерованими Weka (-0,870, -0,125, B, ?) за цим візуальним класифікатором. Можна переконатися, використовуючи «Дерево» на рис. 4.2, що цей «екземпляр» повинен належати до класу нерегулярного дихання (як п'ятий екземпляр листа №11). Таким чином, візуальний класифікатор на рис. 4.2 доступний навіть людям, які не знайомі з машинним навчанням.

Висновок до розділу 4

У даному розділі описано формат файлу ARFF, його атрибути та роботу з ним.

Було оцінено класифікатор шляхом перехресної перевірки та оцінено, наскільки добре попередньо призначені класи в даних збігаються з вибраними кластерами.

Продемонстровано ступінь узгодженості між класами та реальними кластерами. Порівняно таблиці значень продуктивності класифікаторів Weka для набору даних та відповідність класів кластерам.

Побудовано «Дерево рішень» у класифікаторі Random Tree.

У результаті проведення тестування було показано, що дана система моделювання даних працює.

У наведених етапах демонстрації усі елементи виконані правильно.

ВИСНОВКИ

Під час виконання дипломної роботи було виконано аналіз рішень, що існують для збору та моделювання набору даних абдомінальних електроміограм. За результатами аналітичного огляду літератури та патентної інформації визначено загальне компонування системи, обрані найбільш оптимальні методи та алгоритми між обробки отриманих даних, розроблено алгоритм роботи системи. Проаналізовано модель та здійснено перевірку отриманих результатів.

За розробленим алгоритмом змодельовано набір даних абдомінальних електроміограм у системі комп'ютерної математики Maple 2020 та протестовано його реалізацію.

Використання методу експоненційного згладжування в обробці даних абдомінальних електроміограм (ЕМГ) для подальшого формування наборів даних ARFF (формат файлу зв'язку атрибутів) і застосування їх в класифікаторі Random Tree програмного пакету Weka 3-9-5. Цей класифікатор дозволяє візуалізувати його у формі «дерева рішень». Такий візуальний класифікатор-діагностик має бути практично зручним і доступним для медичних спеціалістів, які переважно не знайомі з методами Machine Learning.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

- [1] W. Rose, “KAAP686 Mathematics and Signal Processing for Biomechanics Electromyogram analysis,” 2011.
- [2] R. RAILIANU, “The effect on the electroactivity of the abdominal muscles of the level of connective tissue dysplasia and the method of combined hernioplasty of the inguinal hernia,” *Journal of Experimental and Clinical Surgery*, vol. 13, no. 2, pp. 86–92, Jun. 2020, doi: 10.18499/2070-478X-2020-13-2-86-92.
- [3] R. I. Railianu, G. I. Podolinyi, and A. v. Marshaluk, “EFFECT OF CONNECTIVE INSUFFICIENCY AND SIZES OF MEDIAN HERNIAS ON BEFORE AND POSTOPERATIVE FUNCTION OF THE ABDOMINAL MUSCLE,” *RUDN Journal of Medicine*, vol. 23, no. 1, pp. 40–53, Dec. 2019, doi: 10.22363/2313-0245-2019-23-1-40-53.
- [4] A. NAKAZAKI, S. YAMAGUCHI, H. MOTEGI, Y. ISHI, M. OKAMOTO, and M. FUJIMURA, “Detection of the Moment of Nerve Decompression Using Continuous Monitoring of Evoked Facial Nerve Electromyograms in a Patient with Facial Nerve Schwannoma: Case Report,” *NMC Case Rep J*, vol. 9, no. 0, pp. 2022–0117, Dec. 2022, doi: 10.2176/jns-nmc.2022-0117.
- [5] M. N. Musselwhite *et al.*, “Differential effects of acute cerebellectomy on cough in spontaneously breathing cats,” *PLoS One*, vol. 16, no. 6, p. e0253060, Jun. 2021, doi: 10.1371/journal.pone.0253060.
- [6] R. Kohiruimaki, S. Maeo, and H. Kanehisa, “Suspended Push-up Training Augments Size of not only Upper Limb but also Abdominal Muscles,” *Int J Sports Med*, vol. 40, no. 12, pp. 789–795, Nov. 2019, doi: 10.1055/a-0989-2482.
- [7] R. Kawama, A. Ike, A. Soma, T. Hojo, and T. Wakahara, “Side-To-Side Difference in Electromyographic Activity of Abdominal Muscles during Asymmetric Exercises,” *J Sports Sci Med*, pp. 493–503, Oct. 2022, doi: 10.52082/jssm.2022.493.

- [8] Y. Zhou *et al.*, “Overexpressed DDX3x promotes abdominal aortic aneurysm formation and activates AKT in ApoE knockout mice,” *Biochem Biophys Res Commun*, vol. 634, pp. 138–144, Dec. 2022, doi: 10.1016/j.bbrc.2022.09.077.
- [9] Y. Y. Xie and D. Djurdjanovic, “Monitoring of human neuromusculoskeletal system performance through model-based fusion of electromyogram signals and kinematic/dynamic variables,” *Struct Health Monit*, vol. 20, no. 3, pp. 804–817, May 2021, doi: 10.1177/1475921719848006.
- [10] H. Ou, D. Gates, S. Johnson, and D. Djurdjanovic, “Model-based Fusion of Surface Electromyography with Kinematic and Kinetic Measurements for Monitoring of Muscle Fatigue,” *Int J Progn Health Manag*, vol. 13, no. 2, Jul. 2022, doi: 10.36001/ijphm.2022.v13i2.3132.
- [11] J. Lockstone, S. M. Parry, L. Denehy, I. K. Robertson, D. Story, and I. Boden, “Non-Invasive Positive airway Pressure therapy to Reduce Postoperative Lung complications following Upper abdominal Surgery (NIPPER PLUS): a pilot randomised control trial,” *Physiotherapy*, vol. 117, pp. 25–34, Dec. 2022, doi: 10.1016/j.physio.2022.06.001.
- [12] T. Kapelner, M. Sartori, F. Negro, and D. Farina, “Neuro-Musculoskeletal Mapping for Man-Machine Interfacing,” *Sci Rep*, vol. 10, no. 1, p. 5834, Apr. 2020, doi: 10.1038/s41598-020-62773-7.
- [13] S. R. Devasahayam, “The Electromyogram: Modeling and Analysis,” in *Signals and Systems in Biomedical Engineering: Physiological Systems Modeling and Signal Processing*, Singapore: Springer Singapore, 2019, pp. 353–385. doi: 10.1007/978-981-13-3531-0_12.
- [14] V. A. Staude, I. V. Kotulskyi, D. R. Duplii, and O. D. Karpynska, “Особливості функціонування м’язів — глобальних стабілізаторів в осіб з асиметричним розташуванням таза та крижів у фронтальній площині
Повідомлення 2. Дослідження активності m. gluteus medius, m. obliquus

- externus, m. biceps femoris, m. rectus femoris,” *TRAUMA*, vol. 18, no. 5, pp. 64–72, Jan. 2022, doi: 10.22141/1608-1706.5.18.2017.114120.
- [15] O. H. Shadrin, A. A. Kovalchuk, S. V. Diukareva, V. M. Fysun, and N. M. Basaraba, “Шляхи корекції порушень функціонального стану шлунково-кишкового тракту при гельмінтозах у дітей,” *CHILD`S HEALTH*, no. 7.75, pp. 27–32, Sep. 2021, doi: 10.22141/2224-0551.7.75.2016.86720.
- [16] L. Lievi and O. Zuma, “СУЧАСНІ ІНТЕЛЕКТУАЛЬНІ МЕТОДИ МОДЕЛЮВАННЯ СКЛАДНИХ ТЕХНОЛОГІЧНИХ ОБ’ЄКТІВ,” *Системи управління, навігації та зв’язку. Збірник наукових праць*, vol. 1, no. 63, pp. 49–53, Feb. 2021, doi: 10.26906/SUNZ.2021.1.049.
- [17] S. V. Kosynska, “Ефективність лікування тривожно-депресивних розладів у хворих на гастроентерологічну патологію,” *Gastroenterology*, no. 1.55, pp. 11–14, Jan. 2022, doi: 10.22141/2308-2097.1.55.2015.80340.
- [18] O. Korzhyk, O. Pavlovych, S. Braniuk, and A. Morenko, “Особливості електроміографічної активності дистальних м’язів кисті в жінок із різною модальною α -частотою,” *Notes in Current Biology*, no. 4(377), pp. 115–121, Jun. 2018, doi: 10.29038/2617-4723-2018-377-115-121.
- [19] A. A. Khyzhnyak, A. S. Sokolov, S. S. Dubovska, O. V. Kudinova, Ye. A. Vausov, and N. D. Bitchuk, “Роль комбінованої метаболічної терапії у відновленні післяопераційної когнітивної дисфункції у геронтологічних хворих, які перенесли невідкладні абдомінальні операції,” *Emerg Med (N Y)*, no. 4.75, pp. 84–88, Jan. 2022, doi: 10.22141/2224-0586.4.75.2016.75822.
- [20] V. v. Grubnik *et al.*, “Функціональний стан м’язів черевної стінки при післяопераційних вентральних грижах: сучасні клінічні інструменти діагностики та прогнозу,” *Klin Khir*, no. 1, pp. 59–62, Jan. 2018, doi: 10.26779/2522-1396.2018.01.59.
- [21] R. Zhuang *et al.*, “Age-Related Differences in Stepping Reactions to a Balance Perturbation: A Functional Near-Infrared Spectroscopy and Surface

- Electromyography Study,” *Brain Sci*, vol. 12, no. 11, p. 1479, Oct. 2022, doi: 10.3390/brainsci12111479.
- [22] H. Song *et al.*, “Grouped-sampling technique to deal with unbalance in Raman spectral data modeling,” *Photodiagnosis Photodyn Ther*, vol. 40, p. 103059, Dec. 2022, doi: 10.1016/j.pdpdt.2022.103059.
- [23] G. Chuiko, Y. Darnapuk, O. Dvornik, and Y. Krainyk, “Improved robust handling of electromyograms with mining of new diagnostic signs,” 2021.
- [24] Y. Li and B. C. Lee, “Phronesis: Efficient Performance Modeling for High-dimensional Configuration Tuning,” *ACM Transactions on Architecture and Code Optimization*, vol. 19, no. 4, pp. 1–26, Dec. 2022, doi: 10.1145/3546868.
- [25] V. Gupta, S. Bedathur, S. Bhattacharya, and A. De, “Modeling Continuous Time Sequences with Intermittent Observations using Marked Temporal Point Processes,” *ACM Trans Intell Syst Technol*, vol. 13, no. 6, pp. 1–26, Dec. 2022, doi: 10.1145/3545118.
- [26] K. Madle *et al.*, “Abdominal wall tension increases using Dynamic Neuromuscular Stabilization principles in different postural positions,” *Musculoskelet Sci Pract*, vol. 62, p. 102655, Dec. 2022, doi: 10.1016/j.msksp.2022.102655.
- [27] R. Malátová, P. Bahenský, M. Rost, and D. Marko, “Breathing pattern and its evaluation by muscle dynamometer md03,” *Proc Inst Mech Eng H*, vol. 236, no. 12, pp. 1777–1782, Dec. 2022, doi: 10.1177/09544119221129324.
- [28] K. Alzyoud, S. Al-Murshedi, K. Abualkhael, F. Alqawasmeh, and A. England, “Supine and erect abdominal radiography: A comparison of radiation dose and image quality,” *Applied Radiation and Isotopes*, vol. 190, p. 110477, Dec. 2022, doi: 10.1016/j.apradiso.2022.110477.
- [29] S. Węglarczyk, “Kernel density estimation and its application,” *ITM Web of Conferences*, vol. 23, p. 00037, Nov. 2018, doi: 10.1051/itmconf/20182300037.

- [30] A. Schertel, M. Funke-Chambour, T. Geiser, and A.-K. Brill, “P231 Respiratory breathing patterns and cough in idiopathic pulmonary fibrosis: awake, asleep and over time,” *Chest*, vol. 151, no. 5, p. A131, May 2017, doi: 10.1016/j.chest.2017.04.138.
- [31] S. Shastri, A. Sharma, V. Mansotra, A. Sharma, A. S. Bhadwal, and M. Kumari, “A Study on Exponential Smoothing Method for Forecasting,” *International Journal of Computer Sciences and Engineering*, vol. 6, no. 4, pp. 482–485, Apr. 2018, doi: 10.26438/ijcse/v6i4.482485.
- [32] T. M. Fidale *et al.*, “Eletromyography of abdominal muscles in different physical exercises,” *Medicine*, vol. 97, no. 17, p. e0395, Apr. 2018, doi: 10.1097/MD.00000000000010395.
- [33] Чуйко Г. П., Дарнапук Є. С., Мельничук А. І. Моделювання набору даних абдомінальних електроміограм у системі комп'ютерної математики Maple 2020. Могилянські читання – 2022 : тези доп. XXV Всеукр. наук.-метод. конф. Миколаїв, 07–11 листоп. 2022 р. Миколаїв : Чорном. нац. ун-т ім. Петра Могили, 2022. С. 104–106.

Додаток А

Програмний код

Файл «Modeling_EMG.mw»

```
restart: Digits:=5:
with(LinearAlgebra,Rank,Dimension,Dimensions,Row,Column,SubMatrix,Norm);
with(Statistics,LineChart,Correlation,AutoCorrelation,ColumnGraph,Mean,StandardDeviation,ScatterPlot,BoxPlot,Histogram,KernelDensity,KernelDensityPlot,Quantile,InterquartileRange,Variance,Mode,Median,ViolinPlot,Tally,Range,Difference,MedianDeviation,RousseeuwCrouxSn,RousseeuwCrouxQn,HodgesLehmann,FivePointSummary,Detrend,ScreePlot,PCA);
with(plots,listplot,pointplot,display,setoptions,matrixplot);
plots:-setoptions(axes = boxed, gridlines,font=[Times,14],thickness=3);
with(SignalProcessing,DWT,Periodogram,SignalPlot,DownSample,Threshold,RootMeanSquare);
with(TimeSeriesAnalysis,TimeSeries,TimeSeriesPlot,ExponentialSmoothingModel,Decomposition,GetData,Forecast);
with(Optimization,Maximize,Minimize,LPSolve,NLPSolve);

1-(1-0.36)^8;
alias( pgm=SignalProcessing:-Periodogram, mad=Statistics:-MedianDeviation,RMS=SignalProcessing:-RootMeanSquare,norm=LinearAlgebra:-Norm,Tr=LinearAlgebra:-Transpose, rop=LinearAlgebra:-RowOperation,acf=Statistics:-AutoCorrelation);

dt35:=ExcelTools:-Import("C:\\Users\\Melnychuk\\Documents\\Data\\EMG_data35.xlsx",
"abdom", "A5:AI32802");# Dataset of 35 abdominal EMGs

dt1:=SubMatrix(dt35,1..2^15,1..35):Dimensions(dt1);# This submatrix is a bit reduced
by rows, in comparison with the matrix (1.3.1), but a handier to processing the
dataset
dt5:=ExcelTools:-Import("C:\\Users\\Melnychuk\\Documents\\Data\\ENG_data5.xlsx",
"Аркуш1", "A3:E32802"):dt_5:
whattype(dt5);Dimensions(dt5);# Additional 5 data points
dt2:=SubMatrix(dt5,1..2^15,1..5):Dimensions(dt2);# This submatrix is a bit reduced by
rows, in comparison with the matrix (1.3.1), but a handier to processing the dataset
dtm:=<seq(Column(dt1,q),q=1..Dimensions(dt1)[2])|seq(Column(dt2,q),q=1..Dimensions(dt
5)[2])>:
```

```

N1,N2:=Dimensions(dtm);# Data set finally
ts:=TimeSeries(dtm,frequency=1);
esm:=seq(ExponentialSmoothingModel(ts[q],damping=false),q=1..N2);# Optimal ESM models

dc := [seq(Decomposition(esm[q], ts[q]),q=1..N2)]:whattype(%)# Decompositions oa
signals on Errors, trends and levels (detrendend and denoised signals)
list

ls:=Matrix([seq(GetData(dc[q][2]),q=1..N2)]:n1,n2:=Dimensions(ls);# Detrended and
denoised signals as a column matrix
n1, n2 := 32768, 40

SNR:=seq(round(20*log10(RootMeanSquare(GetData(dc[q][2]))/RootMeanSquare(GetData(dc[q]
][1])))),q=1..N2);
BoxPlot([SNR],gridlines,notched=true,outliers=true,deciles=false); # Signals to noise
ratios ( in db) and their box-plot

Statistics:-
NormalPlot([SNR],sorted=false,gridlines,symbol=solidcircle,symbolsize=18);Range([SNR]
);
tm1:=time():
acs:=[seq(acf(Column(ls,q)),q=1..N2)]:
t__acs:=time()-tm1# Aotocorrelation functions of cleaned signals

seq(Periodogram(acs[q],samplerate=200,powerscale="absolute",view=[0..0.45,default],th
ickness=3),q=1 ..N2):Ar1:=Array([%]);numelems(Ar1);

display(Ar1[1..5]);
display(Ar1[6..10]);
display(Ar1[11..15]);
display(Ar1[16..20]);
display(Ar1[21..25]);
display(Ar1[26..30]);
display(Ar1[31..35]);
display(Ar1[36..40]);

re:={3,5,6,9,11,15,20,27,31,37};numelems(re);

irre:={seq(i,i=1..40)} minus re;nops(%);
irre := {1, 2, 4, 7, 8, 10, 12, 13, 14, 16, 17, 18, 19, 21, 22,

```

23, 24, 25, 26, 28, 29, 30, 32, 33, 34, 35, 36, 38, 39, 40}

```

z4:=Array(1..N2,fill='irregular'):
for i in re do z4[i]:='regular'od:
z_4:=convert(z4,Vector[column]):# Classes vector column

display(Array(1..2,[Ar1[15],Ar1[23]]),size=[500,"golden"]);

nout:=proc(dt::Matrix)::list;
local tr1,tr2,N1,N2,i,j,th,Nout;
option remember;
description "Count of outliers parts (fractions, percentage)";
uses Statistics,LinearAlgebra,SignalProcessing;
N1:=Dimension(dt)[1];N2:=Dimension(dt)[2];
Nout:=Array(1..N2);
#for i to N2 do tr1[i]:=Statistics:-Quartile(Column(dt,i),1)-1.5*Statistics:-
    InterquartileRange(Column(dt,i));tr2[i]:=Statistics:-
    Quartile(Column(dt,i),3)+1.5*Statistics:-
    InterquartileRange(Column(dt,i));# Tukey's method
for i to N2 do tr1[i]:=Statistics:-Median(Column(dt,i))-2.3*Statistics:-
    InterquartileRange(Column(dt,i));tr2[i]:=Statistics:-
    Median(Column(dt,i))+2.3*Statistics:-InterquartileRange(Column(dt,i));#
    Median rule
Threshold(Column(dt,i),[tr1[i], tr2[i]], 'value' = [infinity, infinity]):
th[i]:=Tally(%,output=list):
for j to numelems(th[i]) do if lhs((th[i])[j])=Float(infinity) then
    Nout[i]:=rhs((th[i])[j]) else next end if end do;
end do;
return convert(evalf(Nout/N1),list);
end proc;

No:=nout(dtm);numelems(No);whattype(No);max(No);min(No);# Outliers percentage
(fractions)

bw0:=0.9*min(MedianDeviation(No),InterquartileRange(No)/1.34)*N2^(-0.2);

Statistics:-Histogram(No,binwidth=bw0,
    averageshifted=1,color="LightGrey",range=min(No)..max(No), size=[500,
    "golden"],gridlines,frequencyscale = absolute); ris1:=%:

```

```

z1:= [seq(0.6745*(No[q]-Median(No))/MedianDeviation(No),q=1..N2)]:max(z1);min(z1);# z1
      is No modified z-scores [4,6]

c1:=Array(1..N2,0):alpha:=0.025:
for i from 1 to N2 do if No[i]<=alpha then c1[i]:=i: else next end if end do;
convert (c1,set);c1:=%[2..numelems(%)];nops(c1);

c2:=Array(1..N2,0):
for i from 1 to N2 do if No[i]<=0.065 and No[i]>=0.03 then c2[i]:=i: else next end
      if end do;
convert (c2,set);c2:=%[2..numelems(%)];nops(c2);

c3:=Array(1..N2,0):
for i from 1 to N2 do if No[i]<=0.11 and No[i]>=0.07 then c3[i]:=i: else next end if
      end do;
convert (c3,set);c3:=%[2..numelems(%)];nops(c3);

c4:=Array(1..N2,0):
for i from 1 to N2 do if No[i]>=0.15 then c4[i]:=i: else next end if end do;
convert (c4,set);c4:=%[2..numelems(%)];nops(c4);

bw1:=0.9*min(MedianDeviation(z1),InterquartileRange(z1)/1.34)*N2^(-0.2);

Statistics:-Histogram(z1,binwidth=bw1,
      averageshifted=1,color="LightGrey",range=min(z1)..max(z1), size=[500,
      "golden"],gridlines): gr1:=%:

Statistics:-KernelDensityPlot(z1, bandwidth=bw1, kernel=gaussian,
      linestyle=1,range=min(z1)..max(z1), color="DarkRed", thickness=2,
      size=[500, "golden"]):gra1:=%:
display(gr1,gra1);

c1:=Array(1..N2,0):alpha:=0.05:
for i from 1 to N2 do if z1[i]<=alpha then c1[i]:=i: else next end if end do;
convert (c1,set);low_g:=%[2..numelems(%)];nops(low_g);#

st1:=time():
fwr:=Matrix([seq(map(abs,Column(ls,q)),q=1..N2)]:m1,m2:=Dimensions(fwr);
t__fwr:=time()-st1;
#Take the absolute value of the signal. This is also called full wave rectification
      (W.Rose]

```



```

st2:=time():
env:=Matrix([seq(Statistics:-MovingStatistic(Column(fwr,q),2^10,Median),q=1..N2)]):
Dimensions(env);
t__s:=time()-st2;# Matrix of envelope
          31745, 40

          t__s := 197.01

maxt:=[seq(max(Column(env,q)),q=1..n2):max(maxt);min(maxt);# Peak values of
          envelopes
medt:=[seq(Median(Column(env,q)),q=1..n2):max(medt);min(medt);# Medians of envelopes
[seq(maxt[q]/medt[q],q=1..n2)]:
mint:=[seq(min(Column(env,q)),q=1..n2):max(mint);min(mint);# Peak values of
          envelopes

BoxPlot([medt,maxt],
          outliers=true,deciles=false,gridlines,notched=true,datasetlabels=[
          "medians","peak_values"],font=[TIMES,14],size=[400,"golden"]);

1:=[seq(0.6745*(medt[q]-
          Median(medt))/MedianDeviation(medt),q=1..N2):max(z1);min(z1);# z1 is
          medt modified z-scores [4,6]

z2:=[seq(0.6745*(maxt[q]-
          Median(maxt))/MedianDeviation(maxt),q=1..N2):max(z2);min(z2);# z1 is
          maxt modified z-scores [4,6]

Statistics:-CorrelationMatrix(<<z1>|<z2>>);

with(Student:-Statistics):
ct:=CriticalTable('StudentT', alpha = 0.95, parameter1 = N2);

Corr_critical:=t/sqrt(N2-2+t^2);
cc_crit:=eval(Corr_critical,t=ct[2,2]);# with 0.95 level
# critical value of corr coeffs essential non-zero

dm:=Matrix([convert(z1,Vector[column]),convert(z2,Vector[column])]);# Data matrix
          (z1,z2)

pca:=PCA(dm,center=false,correlation=false,method=eigenvector,summarize=true);

```

```

dmm:=-pca:-principalcomponents;
-<x1|x2> .pca:-rotation;

pca:-varianceproportion;ColumnGraph(% ,gridlines,color=gray);

z_1:=Column(dmm,1);z_2:=Column(dmm,2);

BoxPlot([z_1,z_2], outliers=true,deciles=false,gridlines,notched=true,datasetlabels=[
    "c1","c2","c3"],font=[TIMES,14],size=[400,"golden"]);

bw_1:=0.9*min(MedianDeviation(z_1),InterquartileRange(z_1)/1.34)*N2^(-0.2);

Statistics:-Histogram(z_1,binwidth=bw_1,
    averageshifted=3,color="LightGrey",range=min(z_1)..max(z_1), size=[500,
    "golden"],gridlines,labeldirections=[horizontal,vertical],labels=["c1,
    principal component","Relative frequency"],labelfont=[TIMES,16]):
    r1:=%:

Statistics:-KernelDensityPlot(z_1, bandwidth=bw_1, kernel=gaussian,
    linestyle=1,range=min(z_1)..max(z_1), color="DarkRed", thickness=3,
    size=[500, "golden"]):ri1:=%:
display(r1,ri1);

bw_2:=0.9*min(MedianDeviation(z_2),InterquartileRange(z_2)/1.34)*N2^(-0.2);

Statistics:-Histogram(z_2,binwidth=bw_1,
    averageshifted=3,color="LightGrey",range=min(z_2)..max(z_2), size=[500,
    "golden"],gridlines,labeldirections=[horizontal,vertical],labels=["c2,
    principal component","Relative frequency"],labelfont=[TIMES,16]);
    r_1:=%:

Statistics:-KernelDensityPlot(z_2, bandwidth=bw_2, kernel=gaussian,
    linestyle=1,range=min(z_2)..max(z_2), color="DarkRed", thickness=3,
    size=[500, "golden"]):ri_2:=%:
display(r_1,ri_2);

envre:=SubMatrix(env,1..Dimension(env)[1],[op(re)]:
ar2:=Array(1..Dimension(envre)[2],[seq(SignalPlot(Column(envre,q)),q=1..Dimension(env
re)[2]))]);
[Length of output exceeds limit of 5000000]

```

```

display(ar2[1..5],view=[default,default]);
display(ar2[6..Dimension(envre)[2]],view=[default,default]);re;

envirre:=SubMatrix(env,1..Dimension(env)[1],[op(irre)]):
ar3:=Array(1..Dimension(envirre)[2],[seq(SignalPlot(Column(envirre,q)),q=1..Dimension
    (envirre)[2]))];
    [Length of output exceeds limit of 5000000]

display(ar3[1..5],view=[default,default]);
display(ar3[6..10],view=[default,default]);
display(ar3[11..15],view=[default,default]);
display(ar3[16..20],view=[default,default]);
display(ar3[21..25],view=[default,default]);
display(ar3[26..30],view=[default,default]);irre;

zab:=Array(1..N2,fill='B');numelems(zab);

for j to N2 do if z_1[j]>0.1 and z_2[j]>0.1 and z_2[j]<0.75 then zab[j]:='A' else
    next fi od; zab;

zab[37];
Digits=5;

Mt:=Tr(Matrix([seq(<z_1[q],z_2[q],zab[q],z4[q]>,q=1..N2)]));
seq((convert(Row(Mt,p),list),p=40)):op(%);

with(RandomTools);
Generate(set(posint(range = 40), 20));nops(%);

```