

**НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»**

Приладобудівний факультет

Кафедра автоматизації та систем неруйнівного контролю

До захисту допущено:

Завідувач кафедри

_____ Юрій КИРИЧУК

«__» _____ 20__ р.

Дипломна робота

на здобуття ступеня бакалавра

за освітньо-професійною програмою «Комп'ютерно-інтегровані системи та технології в приладобудуванні»

спеціальності 151 «Автоматизація та комп'ютерно-інтегровані технології»

на тему: «Інтелектуальна система виявлення стресу на основі Shimmer-сенсорів»

Виконав:
студент IV курсу, групи ПК-01
Стрельчук Ярослав Вікторович

Керівник:
Доктор технічних наук, професор
Куц Юрій Васильович

Консультант з:
Доктор технічних наук, професор
Нечипоренко Аліна Сергіївна

Консультант з:
Асистент кафедри оториноларингології, PhD ХНМУ
Алексеева Вікторія Вікторівна

Рецензент:
Асистент, кандидат технічних наук
Івасенко Віталій Михайлович

Засвідчую, що у цій дипломній роботі немає
запозичень з праць інших авторів без відповідних
посилань.

Студент (-ка) _____

Київ – 2024 року

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
Приладобудівний факультет
Кафедра автоматизації та систем неруйнівного контролю

Рівень вищої освіти – перший (бакалаврський)

Спеціальність – 151 «Автоматизація та комп'ютерно-інтегровані технології»

Освітньо-професійна програма «Комп'ютерно-інтегровані системи та технології в приладобудуванні»

ЗАТВЕРДЖУЮ

Завідувач кафедри

_____ Юрій КИРИЧУК

«__» _____ 20__ р.

ЗАВДАННЯ

на дипломну роботу студенту

Стрельчуку Ярославу Вікторовичу

1. Тема роботи «Інтелектуальна система виявлення стресу на основі Shimmer-сенсорів», керівник роботи Куц Юрій Васильович, доктор технічних наук, професор, затверджені наказом по університету від «28» травня 2024 р. №2121-с

2. Термін подання студентом роботи: 07.06.2024

3. Вихідні дані до роботи: сукупність аналізованих параметрів: ФПГ та ЕДА.

Час аналізу: не більше 6 хв.

Умови виконання: лабораторні.

Вихідна інформація: наявність/відсутність стресового стану у людини. Досліджувана група учасників експерименту не менше 30.

4. Зміст роботи: Сучасний стан питання ідентифікації стресу та існуючі рішення, Розробка компонентів інтелектуальної системи для виявлення стресу, Виявлення та класифікація стресового стану людини на основі методів машинного навчання,

Розроблення програмного рішення для реалізації розглянутого методу виявлення стресового стану людини.

5. Перелік ілюстративного матеріалу: презентація, 2 плакати.

6. Консультанти розділів роботи

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв
Сучасний стан питання ідентифікації стресу та існуючі рішення	Алексеева В. В.		
Розробка компонентів інтелектуальної системи для виявлення стресу	Нечипоренко А. С.		

7. Дата видачі завдання 12.02.2024

Календарний план

№ з/п	Назва етапів виконання дипломної роботи	Термін виконання етапів роботи	Примітка
1	Отримання завдання	12.02.2024	
2	Ознайомлення з літературою за темою роботи	13.02.2024 – 01.03.2024	
3	Проведення експериментальних досліджень	01.03.2024 – 10.03.2024	
4	Розроблення методики проведення експериментальних досліджень	10.03.2024 – 17.03.2024	
5	Проведення експериментальних досліджень	17.03.2024 – 01.04.2024	
6	Опрацювання результатів лабораторних досліджень	01.04.2024 – 07.04.2024	
7	Розробка програмного рішення	07.04.2024 – 01.05.2024	
8	Оформлення текстової частини дипломної роботи	01.05.2024 – 30.05.2024	
	Підготовка презентації	30.05.2024 – 01.06.2024	
	Представлення роботи до захисту	14.06.2024 – 18.06.2024	

Студент

Ярослав СТРЕЛЬЧУК

Керівник

Юрій КУЦ

Анотація

Дипломна робота на тему «Інтелектуальна система виявлення стресу на основі Shimmer-сенсорів» присвячена створенню автоматизованого рішення для виявлення стресу за допомогою біосигналів. Основною метою роботи було розробити методіку та програмне забезпечення, яке б дозволило ефективно моніторити фізіологічний стан людини та ідентифікувати стресові стани, використовуючи дані, отримані за допомогою сенсорів фотоплетизмографії та електродермальної активності.

У ході дослідження було проведено серію експериментів з 30 учасниками. За стресор було використано сигнал повітряної тривоги.

Для обробки даних було застосовано методи попередньої фільтрації та нормалізації сигналів. Було проведено аналіз варіабельності серцевого ритму та електродермальної активності для визначення основних параметрів, що характеризують стресовий стан.

Для класифікації станів спокою та стресу було використано кілька методів машинного навчання, включаючи K-Nearest Neighbors, Naive Bayes, Random Forest, Decision Tree та Support Vector Regression. На основі отриманих результатів було встановлено, що метод Naive Bayes показав найвищу точність класифікації – 88%. Розроблене програмне забезпечення автоматизує процес обробки та аналізу даних, а також класифікації стану стресу.

Практичне значення роботи полягає в можливості використання створеної системи для моніторингу фізіологічного стану людини, що дозволяє швидко виявляти стресові стани та надавати рекомендації або попередження користувачеві. Це сприяє підвищенню рівня здоров'я та безпеки, а також може бути застосовано у медичній діагностиці та профілактиці захворювань, пов'язаних зі стресом.

Ключові слова: стрес, варіабельність серцевого ритму, фотоплетизмографія, електродермальна активність, біосигнали, Shimmer-сенсори, машинне навчання, класифікація стресу, здоров'я, автоматизована система, фізіологічний моніторинг, ідентифікація стресу, ФПГ сенсори.

Abstract

This thesis titled "Intelligent Stress Detection System Based on Shimmer Sensors" is dedicated to creating an automated solution for detecting stress using biosignals. The main goal of the work was to develop a methodology and software that would allow effective monitoring of a person's physiological state and identifying stress states using data obtained from photoplethysmography and electrodermal activity sensors.

A series of experiments with 30 participants were conducted, using an air raid alarm signal as the stressor.

Methods of filtering and signal normalization were applied for data processing. An analysis of heart rate variability and electrodermal activity was performed to determine the main parameters characterizing the stress state.

Several machine learning methods were used to classify states of calm and stress, including K-Nearest Neighbors, Naive Bayes, Random Forest, Decision Tree, and Support Vector Regression. Based on the obtained results, the Naive Bayes method showed the highest classification accuracy of 88%. The developed software automates the process of data processing, analysis, and stress state classification.

The practical significance of the work lies in the possibility of using the created system to monitor a person's physiological state, allowing for the quick detection of stress states and providing recommendations or warnings to the user. This contributes to improving health and safety levels and can also be applied in medical diagnostics and the prevention of stress-related diseases.

Keywords: stress, heart rate variability, photoplethysmography, electrodermal activity, biosignals, Shimmer sensors, machine learning, stress classification, health, automated system, physiological monitoring, stress identification, PPG sensors.

Зміст

Перелік умовних скорочень.....	7
Вступ	8
Розділ 1. Сучасний стан питання ідентифікації стресу та існуючі рішення.....	10
1.1 Медико-біологічне підґрунтя та визначення стану стресу.....	10
1.2 Аналіз існуючих аналогів	12
1.3 Постановка завдання	16
1.4 Висновки до розділу	17
Розділ 2. Розробка компонентів інтелектуальної системи для виявлення стресу.....	18
2.1 Опис структури інтелектуальної системи	18
2.2 Збір та обробка даних.....	19
2.3 Препроцесинг даних	21
Висновки до розділу	24
Розділ 3. Виявлення та класифікація стресового стану людини на основі методів машинного навчання.....	25
3.1 Методи класифікації для виявлення стресу.....	25
3.1.1 K-Nearest Neighbors	25
3.1.2 Naive Bayes	26
3.1.3 Random Forest.....	27
3.1.4 Decision Tree.....	28
3.1.5 Support Vector Machine.....	29
3.2 Оцінка результатів	30
3.3 Застосування методів машинного навчання	32
Висновки до розділу	35
Розділ 4. Розроблення програмного рішення для реалізації розглянутого методу виявлення стресового стану людини.....	37
4.1 Обґрунтування вибору мови програмування.....	37
4.2 Використані бібліотеки мови програмування Python	37
4.3 Автоматизація процесу прийняття рішення в системі виявлення стресу.....	40
4.4 Тестування моделі та отримані результати.....	43
Висновки до розділу	45
Висновки.....	46
Літературні джерела	48

Перелік умовних скорочень

- BCR – Варіабельність серцевого ритму
- ФПГ – Фотоплетизмографія
- ЕДА – Електродермальна активність
- ЕЕГ – Електроенцефалографія
- ЕКГ – Електрокардіографія
- ЕМГ – Електроміографія
- СКВ – Середнє квадратичне відхилення
- АНС – Автономна нервова система
- IQR –Інтерквартильний діапазон
- ГГН – Гіпоталамо-гіпофізарно-надниркова система
- АД – Артеріальний тиск
- ССС – Серцево-судинна система
- ШГР – Шкірно-гальванічна реакція
- САС – Симпатична адреномедулярна система
- СР - Серцевий ритм
- KNN – K-Nearest Neighbors
- SVM – Support Vector Machine

Вступ

Стрес – це природна фізична і психічна реакція на життєвий досвід. Кожна людина час від часу відчуває стрес. Він може виникати через різні чинники – повсякденні обов'язки, робота і сім'я, серйозні життєві події, виявлений діагноз, війна або смерть близької людини. Організм реагує на стрес вивільненням гормонів, які збільшують частоту серцебиття і дихання та готують м'язи до реакції. У ситуаціях де існує безпосередня загроза це є корисним. Але у випадку якщо така реакція не припиняється і рівень стресу залишається підвищеним набагато довше, ніж це необхідно для виживання, це може негативно позначитися на здоров'ї. Хронічний стрес може сприяти розвиненню фізичних та психічних захворювань, таких як: діабет, артрит, гіпертонія, тривожні розлади, депресія. Також окрім впливу на здоров'я, через тривалий стрес можуть погіршитись когнітивні можливості людини, без розуміння цього самою людиною. Що може нести загрозу для оточуючих, особливо коли через специфіку професії, особи, на яку здійснюється вплив стресорів залежить життя інших.

Вищеописана інформація свідчить про важливість розпізнавання стресу та виконання заходів профілактики. Наразі наявні різні методи якісного виявлення, які ґрунтуються на даних отриманих за допомогою: електрокардіограми, електроенцефалограми, електроміографії. Проте ці методи потребують стаціонарного розташування, освіченого персоналу та суворого дотримання методології вимірювання, що є не завжди доступним як фізично так і фінансово.

Аби спростити процедуру визначення стресу рекомендується використовувати носимі пристрої для вимірювання. Де сигнал буде передаватися на телефон або портативний комп'ютер. Це можна робити швидко, зручно, а головне постійно. Що дає можливість точно відслідковувати зміни та знаходити кореляції між різними видами біологічних сигналів. Одним із таких пристроїв є сенсори фотоплетизмографії та шкірно-гальванічної реакції.

Метою роботи є розробка компонентів інтелектуальної системи виявлення стресу на основі Shimmer-сенсорів на основі аналізу сигналів фотоплетизмографії та

шкірно-гальванічної реакції.

Під час виконання було проведено вимір у стані спокою та під впливом стресу, для того, щоб запропонувати ефективний метод розпізнавання стресу на початковій стадії. За чинник стресу було обрано сигнал повітряної тривоги, який передається за допомогою сирени та досліджено її вплив на осіб з різних культур та різним життєвим досвідом.

Розділ 1. Сучасний стан питання ідентифікації стресу та існуючі рішення

1.1 Медико-біологічне підґрунтя та визначення стану стресу

Стрес є складною фізіологічною та психологічною реакцією на сприйняті загрози, відомі як стресори. Він включає каскад нейроендокринних процесів, які готують організм до реагування на стресор, це ще називають "реакція бий або біжи". Розуміння медичних та біологічних основ стресу є важливим для розробки ефективних стратегій його виявлення та управління.

Коли сприймається стресор, мозок ініціює серію реакцій, що включають гіпоталамус, гіпофіз та наднирники, цю систему називають гіпоталамо-гіпофізарно-наднирковою. Ця система регулює вивільнення кортизолу, основного гормону стресу, який допомагає мобілізувати енергетичні ресурси та підтримувати гомеостаз. Одночасно активується симпатична адреномедулярна система, яка вивільнює адреналін та норадреналін, що збільшують частоту серцевих скорочень, кров'яний тиск та дихання.

Три основні області мозку, критично залучені у стресову реакцію: гіпокамп, мигдалеподібне тіло та префронтальна кора. Мигдалеподібне тіло виявляє емоційні стимули та сигналізує гіпоталамусу про початок стресової реакції. Префронтальна кора, відповідальна за вищі когнітивні функції, оцінює загрозу та модулює реакцію. Хронічний стрес може призвести до змін у цих областях мозку, впливаючи на пам'ять, регуляцію емоцій та прийняття рішень.

Стрес є не тільки фізіологічним феноменом, але й психологічним. Він включає суб'єктивну оцінку стресора та механізми подолання. Сприйняття стресу залежить від індивідуальних відмінностей, таких як особистісні риси, набутий досвід та соціальна підтримка. Поведінкові реакції на стрес можуть включати зміни у виразі обличчя, положенні тіла та мовних патернах, що часто використовуються разом з біосигналами для виявлення стресу.

Хоч гострий стрес і є нормальною адаптивною реакцією, проте хронічний стрес може мати згубний вплив на здоров'я. Тривала активація осі ГГН та підвищені рівні кортизолу можуть призвести до різних проблем зі здоров'ям, включаючи серцево-судинні захворювання, метаболічні розлади та психічні проблеми, такі як тривога та депресія. А порушення пам'яті, регуляції емоцій та прийняття рішень у певної особи можуть нести значну загрозу для оточуючих [1].

Сучасний стан досліджень в області ідентифікації стресу та існуючі рішення включають використання біосигналів для точного виявлення стресу. Біосигнали, такі як електроенцефалографія, електрокардіографія, електроміографія, електро-дермальна активність, частота дихання та температура шкіри, надають важливу інформацію про реакцію автономної нервової системи на стрес, відображаючи зміни у активності мозку, частоті серцевих скорочень, напруженні м'язів, провідності шкіри та дихальних патернах.

Електроенцефалографія вимірює електричну активність мозку і є чутливою до змін у нейрофізіологічних станах, пов'язаних зі стресом. Стрес зазвичай викликає асиметрію в альфа-активності ЕЕГ, з більшою альфа-активністю у правій фронтальній частині порівняно з лівою. Крім того, стрес часто асоціюється зі зменшенням альфа-активності та збільшенням бета-активності, що вказує на підвищений рівень збудження та когнітивного навантаження.

Електрокардіографія використовується для вимірювання частоти серцевих скорочень та їх варіабельності, які є важливими показниками активності АНС. Під час стресу частота серцевих скорочень збільшується через активацію симпатичної нервової системи. Параметри варіабельності серцевого ритму, які відображають баланс між симпатичною і парасимпатичною активністю, також значно змінюються під час стресу. Зменшення ВСР часто спостерігається при стресі, що вказує на зменшення парасимпатичної активності та підвищення активності симпатичної нервової системи.

Фотоплетизмографія також використовується для вимірювання частоти серцевих скорочень та їх варіабельності. Це неінвазивна технологія, яка використовує джерело світла та фотоприймач на поверхні шкіри для вимірювання об'ємних змін

кровообігу. Джерело світла випромінює світло на тканину, а фотоприймач вимірює відбите світло від тканини. Відбите світло пропорційне змінам об'єму крові.

Електродермальна активність, також відома як гальванічна реакція шкіри. Цей метод полягає у вимірюванні зміни провідності шкіри. Провідність змінюється через активність потових залоз, контрольованих симпатичною нервовою системою. Збільшення ЕДА є надійним показником стресу, відображаючи підвищений рівень збудження та різних емоційних станів.

Додаткові біосигнали, що використовуються для виявлення стресу, включають частоту дихання, яка зазвичай збільшується під час стресу, та температуру шкіри, яка може показувати різні патерни в залежності від вимірюваної частини тіла. Напруження м'язів, вимірюване за допомогою ЕМГ, також збільшується під час стресу, особливо у таких м'язах, як трапецієподібний м'яз та м'язи чола.

Сучасні досягнення у виявленні стресу використовують мультимодальні підходи, які об'єднують кілька біосигналів для підвищення точності та надійності. Мультимодальні системи інтегрують дані з різних датчиків, використовуючи алгоритми машинного навчання для класифікації рівнів стресу

Техніки машинного навчання, широко використовуються у виявленні стресу. Алгоритми методів машинного навчання можуть обробляти складні та високорозмірні дані, виявляючи патерни, що корелюють зі станами стресу [2].

У даній ж роботі буде використано ФПГ і ЕДА через їх доступність використання як носимих пристроїв, що дозволяє використовувати їх для постійного аналізу біосигналів. Також наявні напрацювання у цій роботі можна використовувати у майбутньому як систему для сповіщення про наявність стресу для людей з важливими соціальними професіями для забезпечення вищого рівню безпеки, окрім запобігання різного роду хвороб.

1.2 Аналіз існуючих аналогів

Наявними пристроїв для збору даних ФПГ та ЕДА на ринку є три носимі пристрої — Fitbit Sense, Empatica E4 та Shimmer3 GSR+.

Fitbit Sense – це багатофункціональний фітнес-трекер, який вимірює частоту серцевих скорочень і попереджає про аномальні значення. Трекер відстежує стрес за допомогою електродермальної активності та надає детальну інформацію про якість сну на смартфон. Пристрій має відносно невисоку вартість та позиціонується виробником як «персональний менеджер» для стресу, якості сну. Фото пристрою зображено на рисунку 1.1 [3].



Рисунок 1.1 Fitbit Sense

Empratica E4 – це носимий пристрій, який оснащений датчиками для вимірювання електродермальної активності, фотоплетизмографічних сигналів. Пристрій сумісний з мобільними додатками і програмним забезпеченням для збору і аналізу даних у реальному часі, що дозволяє отримувати негайний зворотний зв'язок та проводити глибокий аналіз даних. Empratica E4 має високу вартість та

позиціонується виробником як інструмент для науковців. Фото пристрою зображено на рисунку 1.2 [4].



Рисунок 1.2 Empatica E4

Shimmer3 GSR+ – це носимий пристрій, що дозволяє вимірювати електродермальну активність та отримувати фотоплетизмографічні сигнали. Завдяки високій частоті вибірки і точності, Shimmer3 GSR+ забезпечує детальний аналіз фізіологічних реакцій, що є важливим для досліджень стресу, емоційного стану і поведінкових реакцій. Пристрій сумісний з різними програмними пакетами для збору та аналізу даних, що дозволяє інтегрувати його в існуючі дослідницькі системи і отримувати негайний зворотний зв'язок. Shimmer3 GSR+ має середню вартість в порівнянні зі згаданими вище пристроями та позиціонується виробником як інструмент для науковців. Фото пристрою зображено на рисунку 1.3 [5].



Рисунок 1.3 Shimmer3 GSR+

Користуючись даними дослідження [6], можна сказати, що частота серцевих скорочень, обчислена за допомогою Empatica E4 та Shimmer GSR3+, має значну позитивну кореляцію у більшості учасників дослідження. Fitbit Sense не демонструє значної кореляції з Shimmer GSR3+ для оцінки серцевого ритму.

Вимірювання рівня шкірної провідності, отримані за допомогою пристроїв Empatica E4 та Shimmer GSR3+, показують позитивну і значну кореляцію, що свідчить про, точність вимірювання обох пристроїв для того ж самого параметра. Fitbit Sense також показує значну кореляцію з Shimmer GSR3+, але тільки в окремих випадках і меншою мірою. Це означає, що хоча Fitbit Sense теж може вимірювати рівень шкірної провідності, але його результати менш надійні і точні порівняно з Empatica E4 і Shimmer GSR3+.

Fitbit Sense має обмежений доступ до «сирих даних», що є значною перешкодою, яка ускладнює застосування методів попередньої обробки даних, необхідних для точного тлумачення даних у реальних умовах.

Shimmer 3 пропонує кілька переваг, які роблять його відмінним вибором для досліджень стресу. Він забезпечує точні та надійні вимірювання ЕДА і ФПГ, надає повний доступ до сирих даних, що дозволяє використовувати детальний аналіз та застосовувати спеціальні алгоритми обробки. Пристрій легко інтегрується з іншими системами вимірювання фізіологічних параметрів, що робить його універсальним для комплексних досліджень. Схожими параметрами володіє і Empatica E4, але для цієї

роботи було обрано Shimmer 3 через його фінансову доступність та зручність у роботі з отриманими даними.

1.3 Постановка завдання

Для досягнення мети у роботі було поставлено такі завдання:

1. Проаналізувати існуючі системи та рішення для розпізнавання стресу
2. Зібрати дані для аналізу
3. Розробити метод виявлення стресу на основі методу машинного навчання
4. Розробити автоматизоване програмне рішення

Ця тема є актуальною у галузі медичної діагностики, профілактичної медицини та передчасного виявлення та запобігання загроз викликаних непоміченим стресом.

Стрес впливає як на фізичне, так і на психічне здоров'я людей. Хоч короточасний стрес може мати позитивний ефект, стимулюючи організм до швидкої реакції на загрозу, хронічний стрес може призвести до серйозних захворювань, таких як діабет, артрит, гіпертонія, тривожні розлади та депресія. Також тривалий стрес негативно впливає на когнітивні здібності, що може мати серйозні наслідки для професій, від яких залежить безпека інших людей. Сучасний розвиток технологій носимих пристроїв відкриває нові можливості для спрощення процедури визначення стресу.

Актуальність дослідження полягає у розробці компонентів інтелектуальної системи для розпізнавання стресу на основі аналізу сигналів фотоплетизмографії та шкірно-гальванічної реакції. Використання таких сенсорів дозволяє збирати дані в режимі реального часу та передавати їх на телефон або портативний комп'ютер, що робить можливим постійний моніторинг стану здоров'я людини. Це дає можливість швидко виявляти зміни та знаходити кореляції між різними біологічними сигналами. Дана робота, присвячена визначенню початкових стадій стресу за допомогою носимих пристроїв та сприяє підвищенню загального рівня здоров'я населення та покращенню якості життя.

1.4 Висновки до розділу

У вступному розділі було обґрунтовано актуальність теми, зокрема важливість ідентифікації стресу для підтримки фізіологічного здоров'я та безпеки людей. Стрес є значним фактором ризику для розвитку різних захворювань, таких як діабет, гіпертонія, тривожні розлади та депресія. З огляду на це, виявлення стресових станів на ранніх стадіях стає критично важливим. Була сформульована мета роботи: розробка інтелектуальної системи для виявлення стресу на основі Shimmer-сенсорів, яка б дозволила моніторити фізіологічний стан людини в режимі реального часу.

Основними завданнями роботи стали: аналіз існуючих систем та рішень для розпізнавання стресу, збір даних для аналізу, розробка методу виявлення стресу на основі методів машинного навчання та створення автоматизованого програмного рішення. В ході було виявлено, що носимі пристрої, такі як Shimmer-сенсори, мають значний потенціал для постійного моніторингу стану здоров'я, забезпечуючи зручність і доступність у використанні. Таким чином, у вступному розділі було закладено наукову базу та сформульовано чіткі напрямки для подальшої роботи.

Розділ 2. Розробка компонентів інтелектуальної системи для виявлення стресу

2.1 Опис структури інтелектуальної системи

На рисунку 2.1 зображено структурно-функціональну схему інтелектуальної системи підтримки прийняття рішень. Де, МС – мікросервіси, БВД – блок введення даних, БПД – блок попередньої обробки даних, БСА – блок статистики і агрегації даних, БЗ – база знань, БД – база даних, ППВ – підсистема планування втручань, ОПР – особа, що приймає рішення, БВД – блок виводу даних, БІЗ – блок інтерпретації знань, БПР – блок прийняття рішень, БМПР – блок моделей прийняття рішень, БАД – блок аналізу даних, ІАД – інтелектуальний аналіз даних, СА – статистичний аналіз даних, ПОСТ – постобробка даних [7].

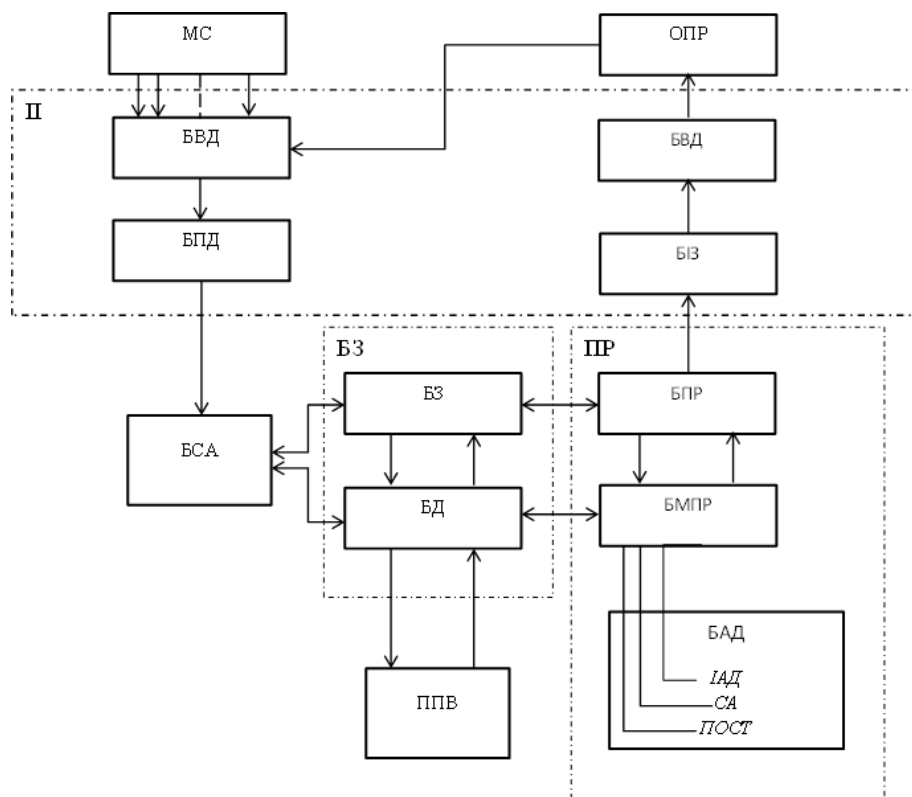


Рисунок 2.1 Структурно-функціональна схема інтелектуальної системи підтримки прийняття рішень

У даній роботі було розроблено програмне забезпечення блоку аналізу даних, який використовує методи машинного навчання для інтелектуального аналізу даних, статистичного аналізу даних та постобробку даних.

2.2 Збір та обробка даних

Збір даних ФПГ та ЕДА стану стресу було проведено у 30 учасників віком від 20 до 30 років, 15 чоловіків, 15 жінок. Дослід полягав в тому щоб зібрати дані біосигналів за відсутності та пізнішою наявністю стресорів. Це було необхідно для того аби знайти кореляції між станами осіб, де заздалегідь відомо про наявність стресорів та часові межі впливу. На рисунку 2.2 зображено фото з наочною демонстрацією проведення дослідження. На зображенні видно розміщення датчиків ФПГ та ЕДА, які з'єднані з Shimmer 3, що закріплені на руці особи, приймаючої участь в досліді. Пристрій Shimmer 3 з'єднано з комп'ютером за допомогою Bluetooth, а дані ФПГ та ЕДА передаються та зберігаються в режимі реального часу.

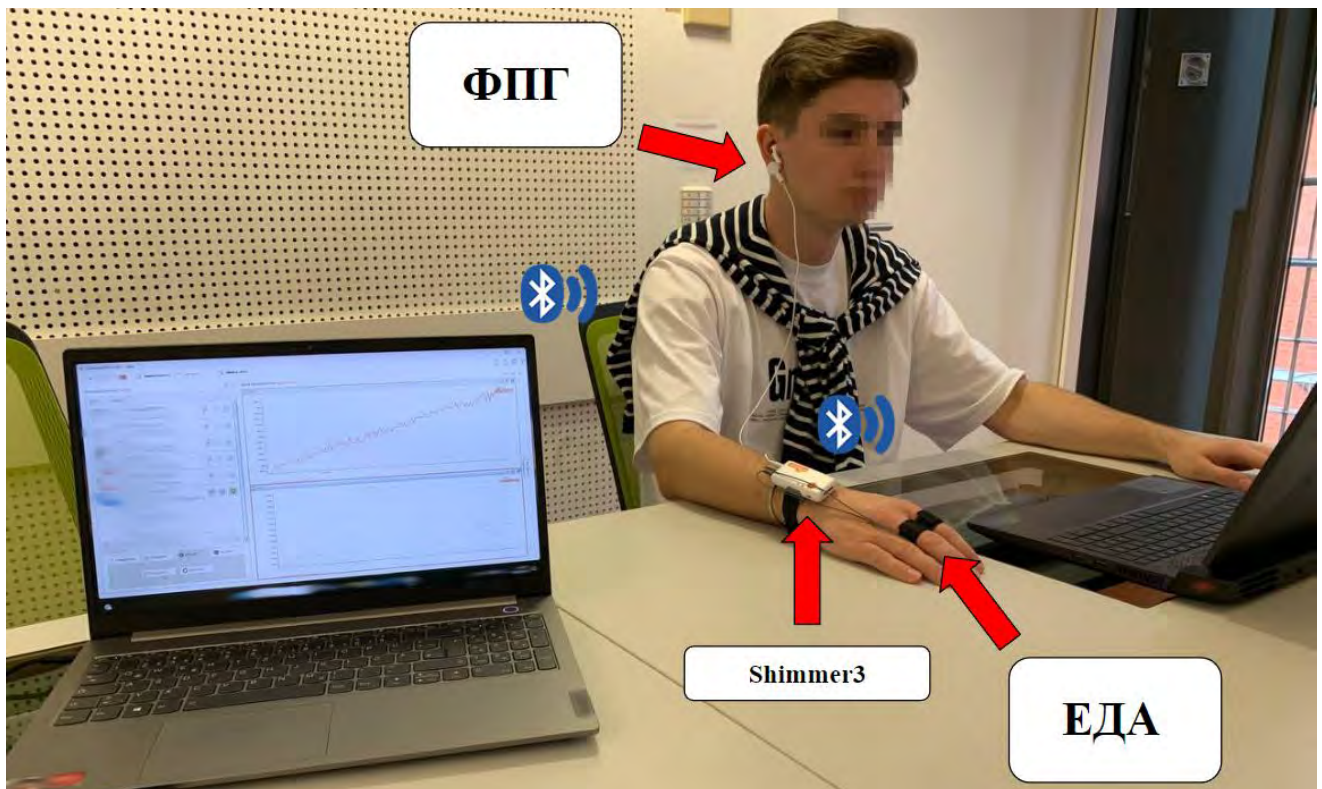


Рисунок 2.2 Ілюстрація проведення дослідження (збору даних)

Для збору даних було розроблено протокол виміру біосигналів, оскільки для усіх учасників умови проведення досліджу маю бути рівними, аби отримати об'єктивні та актуальні дані. Схема протоколу зображена на рисунку 2.3.

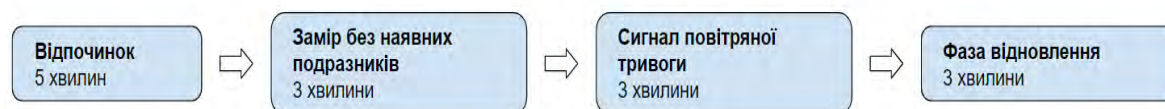


Рисунок 2.3 Протокол вимірювання

Учасники дослідження спостерігали перед собою білий екран, який слугував для уникнення будь-якої додаткової емоційної стимуляції та забезпечення однакових умов для всіх. Це дозволяло учасникам зосередитися, не відволікаючись на зміни візуального контенту, тоді як фактор стресу надходив через навушники.

За фактор стресу було обрано сигнал повітряної тривоги, оскільки він є етично прийнятним стресовим чинником та актуальним для дослідження в сучасних умовах нашої країни. Та це дозволяє аналізувати реакції без необхідності навмисного заподіяння шкоди учасникам.

Учасники тесту 5 хвилин проводили у спокої без подразників, далі проводились заміри тривалістю 3 хвилини, аби отримати результати у стані особи без стресу. Після цього вмикався сигнал повітряної тривоги та проводився замір наступних 3 хвилин. Тоді збиралися дані про відновлення нервової системи, ще 3 хвилини.

Учасники тесту були проінформовані про хід і цілі дослідження та підписали згоду на обробку персональних даних та їх подальше використання в наукових цілях. Також було зібрано додаткові дані, такі як: вік, стать, зріст, вага, для обрахунку індексу маси тіла, наявність захворювань ендокринної та серцево-судинної системи, досвід з сигналом повітряної тривоги, для знаходження додаткових кореляцій.

В результаті збору, отримані біосигнали було збережено у 30 датасетів формату .csv з даними фотоплетизмографії та опором шкіри. Запис даних здійснювався з частотою 128 Гц.

Попередньо було проведено декілька серій вимірювання на тест-групах з різними варіаціями стресорів, а саме сигналами повітряної тривоги різних країн та різної тривалості. Протокол вимірювання описаний вище виявився найбільш релевантним та змістовно значущим для цілей даної роботи. За сигнал повітряної тривоги було обрано поширену в Україні електричну сирену «С-40».

2.3 Препроцесинг даних

Через людський фактор тривалість вимірювань може бути незначною мірою різною. Оскільки початок та кінець запису сигналів починаються вручну. Тому для запобігання похибок, вимірювання заздалегідь проводились на 10 секунд довше ніж це необхідно. Далі програмними методами було автоматизовано видалення даних, записаних пізніше ніж це вимагав протокол. Таким чином було приведено отримані результати до однакової «довжини».

Також через специфіку запису даних пристроєм Shimmer 3, перед дією описаною у попередньому абзаці, було здійснено видалення нерелевантних, цілям даної роботи, даних та відлік часу інтегрований виробником було переведено у зручний для використання з відліком від 0.

Тоді для зручності роботи дані було поділено на дві групи датасетів – біосигнали у стані без наявних подразників та з наявними стресорами.

Отримані датасети містять біосигнали про ЕДА та скорочення серця. Для фільтрації даних скорочення серця було використано методи бібліотек програмних засобів призначених для роботи з ФПГ. Даними ЕДА в цій роботі виступають значення опору шкіри. Цей опір був відфільтрований за допомогою методу інтерквартильного діапазону.

Очищення викидів (аутлаєрів) за допомогою методу інтерквартильного діапазону є популярним підходом у обробці даних для видалення екстремальних значень, які можуть спотворювати результати аналізу. IQR визначається як різниця між третім (Q3) і першим (Q1) квантилями, що охоплює середні 50% даних. Щоб визначити викиди, обчислюють нижню та верхню межі: нижня межа дорівнює Q1 -

1.5IQR, а верхня межа — $Q3 + 1.5IQR$. Значення, що лежать за межами цих діапазонів, вважаються викидами і можуть бути видалені або оброблені окремо, щоб забезпечити точніші та надійніші результати аналізу. Використання IQR методу допомагає зберегти основну структуру даних, видаляючи лише ті значення, які суттєво відрізняються від основної маси [8].

Отримані відфільтровані біосигнали ФПГ було проаналізовано за допомогою методу варіабельності серцевого ритму.

Варіабельність серцевого ритму є показником загального фізіологічного стану організму, який відображає зміну інтервалів між серцевими скороченнями. Регулюється ВСР автономною нервовою системою, що включає симпатичну та парасимпатичну гілки. Симпатична нервова система активується у відповідь на стрес, підвищуючи частоту серцевих скорочень і знижуючи ВСР, тоді як парасимпатична система сприяє релаксації та збільшенню ВСР. Висока ВСР зазвичай свідчить про здоров'я та здатність організму адаптуватися до різних стресових факторів, тоді як низька ВСР може вказувати на проблеми зі здоров'ям і знижений рівень адаптації.

Аналіз ВСР може бути корисним інструментом для оцінки функціонального стану серцево-судинної системи та автономної нервової системи. Зниження ВСР може передувати розвитку серйозних захворювань. ВСР також може бути індикатором психологічної стійкості та здатності до саморегуляції. Було встановлено, що ВСР корелює з емоційною регуляцією, соціальними взаємодіями та когнітивними функціями, такими як увага та прийняття рішень. [9]

ВСР складається з кількох компонентів, які відображають різні аспекти діяльності автономної нервової системи. Сюди належать високочастотні та низькочастотні компоненти, які вказують на активність парасимпатичної та симпатичної нервових систем відповідно. Високочастотна активність пов'язана з дихальними коливаннями, зумовленими парасимпатичною регуляцією серцевого ритму, тобто частота серцевих скорочень збільшується під час вдиху та зменшується під час видиху. Низькочастотна активність є сумішшю симпатичної та парасимпатичної активностей, що відображає загальний баланс між цими двома системами. Також існують дуже низькочастотні компоненти, які можуть бути

пов'язані з терморегуляцією та гормональними впливами. Кожен з цих компонентів дає змогу отримати різноманітну інформацію про стан здоров'я людини, її реакцію на стрес та здатність до адаптації.

Отримані показники ВСР було проаналізовано за допомогою методів машинного навчання, що буде детально описано в наступних розділах. Щоб визначити значущі компоненти для виявлення стану стресу в рамках цієї дипломної роботи, було проведено t-тест для кожного з отриманих компонентів. Відбір компонентів для подальшого аналізу здійснювався на основі р-значення; включалися тільки ті компоненти, для яких р-значення було менше 0.5. Це забезпечило розгляд лише статистично значущих відмінностей, підвищуючи точність і надійність зроблених висновків.

Парний Т-тест — це статистичний метод, який використовується для порівняння середніх значень двох пов'язаних вибірок, зібраних від однієї й тієї ж групи суб'єктів у двох різних умовах [10]. Цей тест дозволяє визначити, чи є статистично значущі відмінності між цими умовами.

Формула для парного t-тесту виглядає так:

$$t = \frac{\bar{d}}{s_d/\sqrt{n}} \quad (2.1)$$

де \bar{d} - середнє значення різниць між парними спостереженнями,

s_d - стандартне відхилення різниць між парними спостереженнями

n - кількість пар спостережень

Р-значення (p-value) є показником у статистиці, який вказує на ймовірність того, що спостережувані результати могли б виникнути випадково, за умови, що нульова гіпотеза є правдивою [11]. Традиційно, рівень значущості встановлюється на рівні 0.05, що означає 5% ймовірності помилково відкинути нульову гіпотезу. Якщо р-значення нижче 0.05, це вказує на те, що спостережувані результати є статистично значущими, і ми можемо з високою ймовірністю стверджувати, що виявлені відмінності не є випадковими.

Для аналізу шкірно-гальванічної реакції було обраховано стандартне відхилення опору шкіри як у стані спокою, так і під час стресу. Виявилося, що за наявності стресорів варіабельність опору шкіри значно зростає. Ця зміна варіабельності опору шкіри є значущою для нашого дослідження, адже вона дозволяє точніше враховувати індивідуальні особливості кожного учасника. Оскільки опір шкіри відрізняється в кожній особі, зростання його варіабельності під час стресу надає нам унікальну можливість краще розуміти, як різні люди реагують на стресові впливи.

Висновки до розділу

У другому розділі було описано структуру інтелектуальної системи для виявлення стресу та протокол збору даних. Проведено вимірювання біосигналів 30 учасників у трьох фазах: стан спокою, стресовий стан, викликаний сигналом повітряної тривоги, та фаза відновлення. Дані були зібрані за допомогою Shimmer-сенсорів з частотою запису 128 Гц.

Для забезпечення якості даних було застосовано методи попередньої фільтрації, нормалізації сигналів та видалення викидів за допомогою методу інтерквартильного діапазону. Аналіз варіабельності серцевого ритму та електродермальної активності дозволить визначити основні параметри, що характеризують стресовий стан.

Отримані результати стали основою для подальшого застосування методів машинного навчання для класифікації станів спокою та стресу. Використання комплексного підходу до збору та обробки даних забезпечило високу точність та надійність аналізу, що є критично важливим для розробки ефективної системи виявлення стресу.

Розділ 3. Виявлення та класифікація стресового стану людини на основі методів машинного навчання

3.1 Методи класифікації для виявлення стресу

У роботі було використано кілька методів машинного навчання для аналізу варіабельності серцевого ритму та електродермальної активності, щоб визначити значущі відмінності між станами стресу та спокою. Основні методи, які були застосовані: K-Nearest Neighbors, Naive Bayes, Random Forest, Decision Tree та Support Vector Regression.

Ці моделі машинного навчання були обрані через їх популярність та ефективність у широкому спектрі застосувань, включаючи біомедичні дослідження. Зокрема, K-Nearest Neighbors, Naive Bayes, Random Forest, Decision Tree та Support Vector Regression відомі своєю здатністю ефективно обробляти складні набори даних, що містять шуми та відхилення, та виявляти складні патерни в даних, що робить їх ідеальними для точного аналізу станів стресу та спокою.

Застосування кількох методів дозволило провести аналіз точності цих методів для системи визначення стану стресу та пізніше вибрати модель з найвищою точністю для виявлення стресу.

Всі зазначені методи належать до категорії методів з наглядом (supervised learning), що означає аналіз наявного навчального набору даних з відомими мітками або результатами, на основі яких модель навчається і потім застосовується для прогнозування нових даних.

3.1.1 K-Nearest Neighbors

Алгоритм K-Nearest Neighbors є одним з основних і широко використовуваних методів машинного навчання для задач класифікації та регресії. Це непараметричний алгоритм, який класифікує або прогнозує значення нових точок даних на основі

схожості з наявними даними. KNN функціонує шляхом обчислення відстаней між новою точкою і всіма точками в наборі даних, після чого обираються k найближчих сусідів, клас яких і визначає клас нової точки [12].

Для вимірювання відстані між точками найчастіше використовують Евклідову метрику, хоча можуть застосовуватися й інші метрики, такі як Манхеттенська та Мінковського. Вибір параметра k є критичним для продуктивності алгоритму: занадто мале значення k може призвести до перенавчання, тоді як занадто велике — до недонавчання. Оптимальне значення k зазвичай вибирається за допомогою методів перехресної перевірки.

Переваги:

- Простота реалізації та інтерпретації.
- Гнучкість в обробці нових даних без необхідності перенавчання моделі.
- Мінімальні вимоги до налаштування гіперпараметрів.

Недоліки:

- Високі вимоги до обчислювальних ресурсів при великих наборах даних.
- Низька масштабованість на великі обсяги даних.
- Схильність до зниження ефективності при високій розмірності даних.

3.1.2 Naive Bayes

Алгоритм Naive Bayes є методом класифікації в машинному навчанні, який базується на теоремі Байєса. Цей алгоритм належить до категорії наглядного навчання і використовується для класифікаційних задач. Naive Bayes моделює розподіл вхідних даних для кожного класу, що дозволяє передбачати ймовірність належності нових даних до певного класу [13].

Теорема Байєса дозволяє оновлювати ймовірності подій на основі нових даних. Вона використовує поняття умовної ймовірності для обчислення ймовірності події з урахуванням наявних доказів. Алгоритм припускає, що всі ознаки є незалежними одна від одної, що значно спрощує обчислення. Незважаючи на те, що це припущення

часто порушується в реальних даних, алгоритм показує хороші результати, особливо на невеликих наборах даних.

В даній роботі використаний класифікатор Gaussian Naive Bayes: Він використовується для безперервних даних, припускаючи нормальний розподіл ознак.

Переваги:

- Простота реалізації та інтерпретації.
- Висока швидкість роботи та ефективність при великих обсягах даних.
- Низькі вимоги до обчислювальних ресурсів.

Недоліки:

- Припущення про незалежність ознак часто не відповідає дійсності, що може призвести до неправильних класифікацій.
- Проблема "нульової частоти", коли певна ознака не зустрічається у тренувальному наборі.

3.1.3 Random Forest

Алгоритм Random Forest є методом машинного навчання, що використовується для задач класифікації та регресії. Він поєднує виходи багатьох дерев рішень для досягнення більш точного результату. Цей метод часто використовується через свою гнучкість та здатність обробляти як числові, так і категорійні дані [14].

Random Forest складається з великої кількості дерев рішень, кожне з яких побудоване на випадковій підмножині даних і ознак. Кожне дерево працює незалежно, і остаточне передбачення робиться шляхом усереднення результатів (у випадку регресії) або більшості голосів (у випадку класифікації). Це дозволяє знизити ризик перенавчання, яке часто спостерігається в індивідуальних деревах рішень.

Переваги:

- Знижений ризик перенавчання: Завдяки об'єднанню результатів багатьох дерев, Random Forest зменшує загальну дисперсію і похибку прогнозу.

- Гнучкість: Алгоритм може ефективно обробляти як задачі класифікації, так і регресії.
- Визначення важливості ознак: Random Forest легко визначає, які ознаки є найбільш впливовими для моделі.

Недоліки:

- Часові витрати: Обчислення для кожного дерева може бути повільним, особливо для великих наборів даних.
- Ресурсоємність: Алгоритм потребує більше пам'яті та обчислювальних ресурсів.

3.1.4 Decision Tree

Алгоритм Decision Tree є непараметричним методом наглядного навчання, який застосовується для задач класифікації та регресії. Цей метод використовує ієрархічну структуру, що складається з кореневого вузла, внутрішніх вузлів (або вузлів рішень), гілок і листових вузлів. Кожен вузол представляє собою тест на певну ознаку, «гілки» — це результати тесту, а «листові вузли» — кінцеві рішення або передбачення [15].

Процес навчання Decision Tree використовує стратегію розділення та завоювання, проводячи пошук для виявлення оптимальних точок розщеплення в дереві. Процес розщеплення повторюється рекурсивно зверху вниз до тих пір, поки всі або більшість записів не будуть класифіковані. Як правило, невеликі «дерева» більш схильні до досягнення «чистих листових вузлів», тоді як великі «дерева» можуть стикатися з проблемою фрагментації даних, що часто призводить до перенавчання.

Переваги:

- Легкість інтерпретації: Логіка булевих операцій та візуальні представлення роблять Decision Trees легкими для розуміння.
- Гнучкість: Decision Trees можуть обробляти різні типи даних, включаючи дискретні та безперервні значення, а також справляються з відсутніми значеннями.
- Універсальність: Використовуються як для класифікації, так і для регресії.

Недоліки:

- Схильність до перенавчання: Складні дерева часто перенавчаються на тренувальних даних і погано узагальнюють нові дані.
- Висока варіативність: Невеликі зміни в даних можуть призвести до значних змін в структурі дерева.
- Побудова дерев є ресурсомістким процесом.

3.1.5 Support Vector Machine

Алгоритм Support Vector Machine (SVM) є методом наглядного навчання, який використовується для задач класифікації та регресії. SVM працює шляхом знаходження оптимальної гіперплощини, яка максимально розділяє дані різних класів у багатовимірному просторі. Розроблений у 1990-х роках, SVM базується на принципі максимізації відстані між класами, що забезпечує високу здатність до узагальнення.

SVM знаходить гіперплощину, яка максимізує відстань між найближчими точками даних різних класів, відомими як опорні вектори. Для нелінійно розділених даних застосовуються ядрові функції, що дозволяють проектувати дані у вищий вимір, де вони можуть бути лінійно розділені [16].

Переваги:

- Висока здатність до узагальнення: Оптимізація маржі дозволяє моделі добре працювати на нових даних.
- Гнучкість: Здатність працювати з лінійно та нелінійно розділеними даними за допомогою ядерних методів.
- Стійкість до перенавчання: Ефективне вирішення задачі класифікації в умовах великої кількості ознак.

Недоліки:

- Обчислювальна складність: Високі витрати обчислювальної здатності комп'ютера на обчислення для великих наборів даних.
- Чутливість до вибору ядра та параметрів: Необхідність налаштування гіперпараметрів для досягнення оптимальної продуктивності.

3.2 Оцінка результатів

Для оцінки отриманих результатів було використано матрицю невідповідності(помилки). Матриця помилок - це таблиця, яка використовується для визначення ефективності алгоритму класифікації. Матриця візуалізує та підсумовує ефективність алгоритму класифікації [17]. На рисунку 3.1 зображено структуру матриці [18].

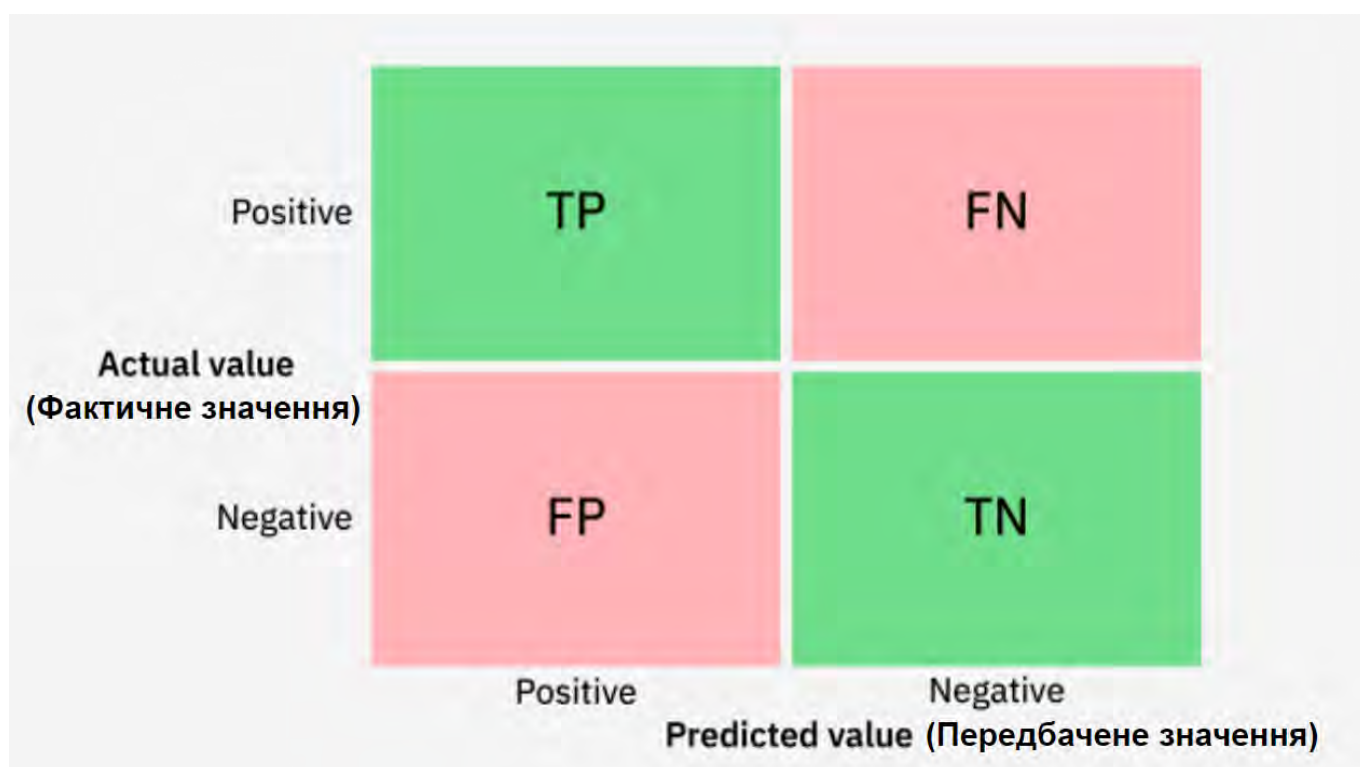


Рисунок 3.1 Матриця помилок

Матриця невідповідності складається з чотирьох основних характеристик (чисел), які використовуються для визначення метрик вимірювання класифікатора.

Цими чотирма числами є

- TP (True Positive): Кількість випадків, коли алгоритм правильно класифікував позитивний стан.
- TN (True Negative): Кількість випадків, коли алгоритм правильно класифікував негативний стан.

- FP (False Positive): Кількість випадків, коли алгоритм неправильно класифікував негативний стан як позитивний. Також відомий як помилка першого роду.

- FN (False Negative): Кількість випадків, коли алгоритм неправильно класифікував позитивний стан як негативний. Також відомий як помилка другого роду.

Метрики ефективності алгоритму: точність, прецизійність, відгук та F1-міру, які розраховуються на основі показників TP, TN, FP та FN.

Точність алгоритму визначається як відношення правильно класифікованих випадків (TP + TN) до загальної кількості випадків (TP + FP + FN + TN)

$$\text{Точність} = \frac{TP + TN}{TP + FP + FN + TN} \quad (3.1)$$

Прецизійність алгоритму визначається як відношення правильно класифікованих позитивних випадків (TP) до загальної кількості передбачених позитивних випадків (TP + FP).

$$\text{Прецизійність} = \frac{TP}{TP + FP} \quad (3.2)$$

Метрика відгуку визначається як відношення правильно класифікованих позитивних випадків (TP) до загальної кількості фактичних позитивних випадків (TP + FN).

$$\text{Відгук} = \frac{TP}{TP + FN} \quad (3.3)$$

Метрика відгуку також називається чутливістю.

F1-міра виражає рівновагу між прецизійністю та відгуком і розраховується за формулою:120

$$F1 \text{ міра} = \frac{\text{прецизійність} * \text{відгук}}{\text{прецизійність} + \text{відгук}} \quad (3.4)$$

3.3 Застосування методів машинного навчання

Після використання: K-Nearest Neighbors, Naive Bayes, Random Forest, Decision Tree та Support Vector Regression для аналізу стану стресу. Отримані результати було проаналізовано за допомогою матриці помилок, для алгоритмів де це можливо. Можемо спостерігати такі значення точності, прецизійності, відгуку та міри F1:

K-Nearest Neighbors:

Точність (Accuracy): 0.5555555555555556

Прецизійність (Precision): 1.0

Повнота (Recall): 0.3333333333333333

F1-міра (F1 Score): 0.5

Naive Bayes:

Точність (Accuracy): 0.8888888888888888

Прецизійність (Precision): 1.0

Повнота (Recall): 0.8333333333333334

F1-міра (F1 Score): 0.9090909090909091

Random Forest:

Точність (Accuracy): 0.5555555555555556

Прецизійність (Precision): 0.75

Повнота (Recall): 0.5

F1-міра (F1 Score): 0.6

Decision Tree:

Точність (Accuracy): 0.5555555555555556

Прецизійність (Precision): 1.0

Повнота (Recall): 0.3333333333333333

F1-міра (F1 Score): 0.5

SVR:

MSE (Середньоквадратична помилка): 0.19079860103983404

На рисунках 3.2, 3.3, 3.4, 3.5 зображені матриці недостовірності для K-Nearest Neighbors, Naive Bayes, Random Forest, Decision Tree.

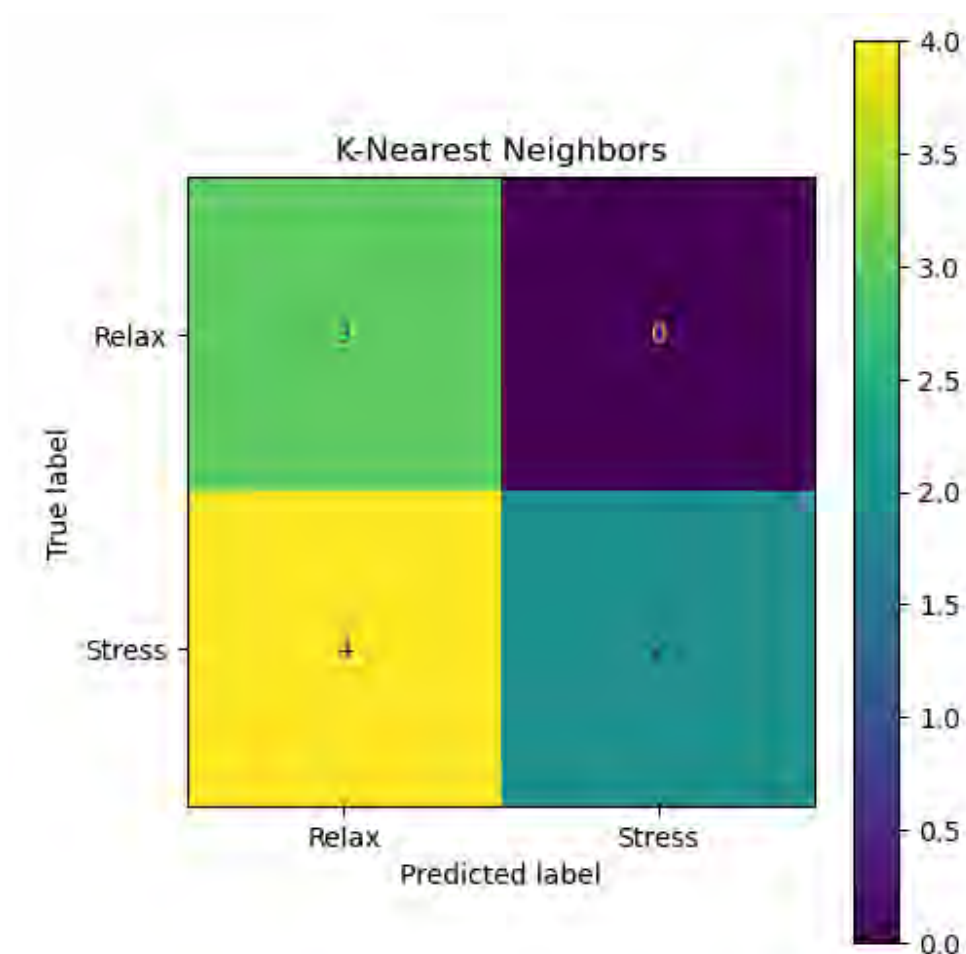


Рисунок 3.2 Матриця недостовірності для K-Nearest Neighbors

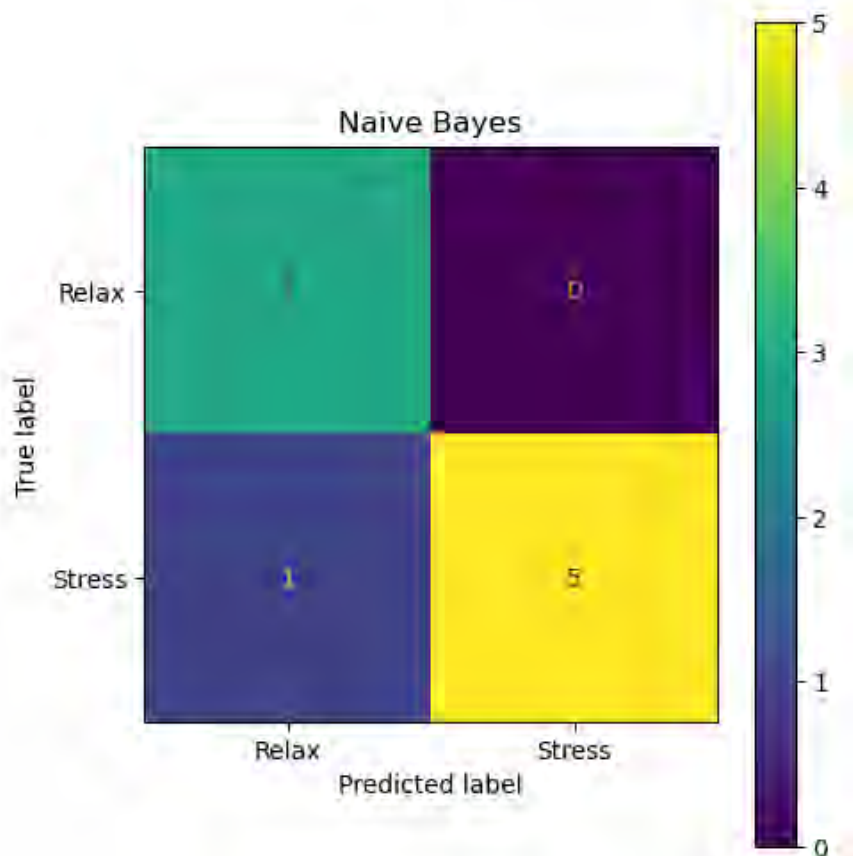


Рисунок 3.3 Матриця недостовірності для Naive Bayes

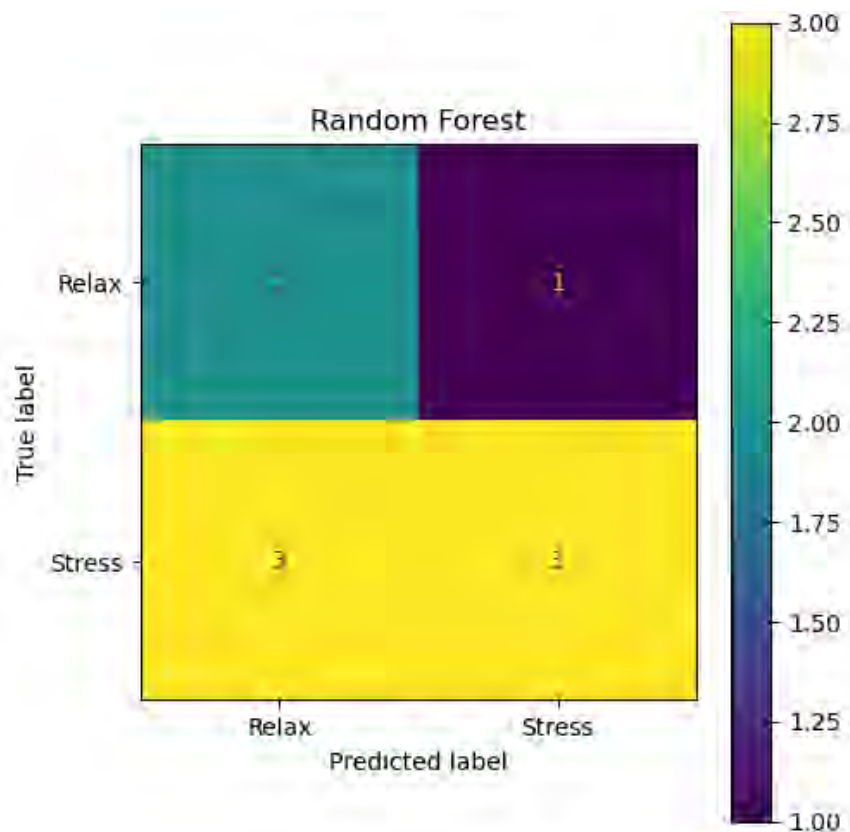


Рисунок 3.4 Матриця недостовірності для Random Forest

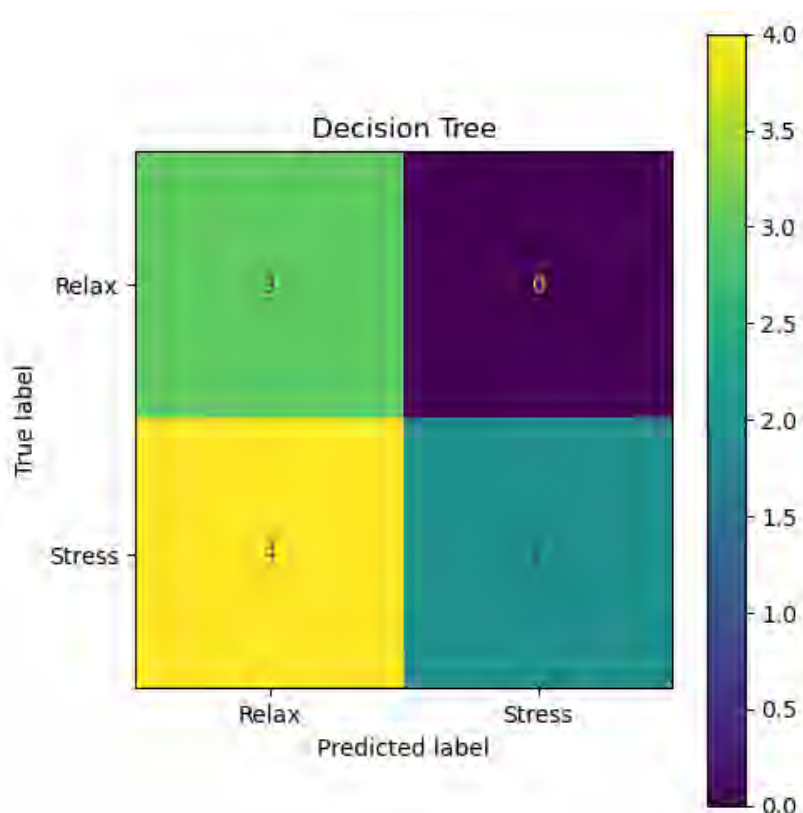


Рисунок 3.5 Матриця недостовірності для Decision Tree

Для SVR відсутня матриця помилок, оскільки вона використовується для оцінки якості класифікації у задачах класифікації. Коли SVR— це метод регресії, а не класифікації. Він прогнозує безперервні числові значення. Натомість, для оцінки якості моделей регресії, таких як SVR, використовуються інші метрики, як у це було зроблено вище з середньоквадратичною помилкою.

Проаналізувавши отримані результати можемо зробити висновок, що для задачі визначення стану стресу у даній роботі найбільше підходять Naive Bayes з точністю 88% та SVR з середньоквадратичною помилкою 19%. Тому Naive Bayes було обрано для аналізу та класифікації стану стресу.

Висновки до розділу

У третьому розділі було застосовано кілька методів машинного навчання для класифікації станів спокою та стресу. До цих методів належать K-Nearest Neighbors,

Naive Bayes, Random Forest, Decision Tree та Support Vector Regression. Кожен з цих методів має свої переваги та недоліки, що були детально проаналізовані у процесі дослідження.

На основі отриманих результатів було встановлено, що метод Naive Bayes показав найвищу точність класифікації – 88%. Інші методи, такі як K-Nearest Neighbors, Random Forest та Decision Tree, показали нижчу точність, але все ж внесли важливий вклад у загальний аналіз даних. Метод Support Vector Regression був використаний для оцінки середньоквадратичної помилки, яка склала 19%. Застосування різних методів машинного навчання дозволило отримати всебічну оцінку даних та вибрати найефективніший підхід для класифікації станів стресу. Це підкреслює важливість використання мульти-модального підходу для підвищення точності та надійності виявлення стресу.

Розділ 4. Розроблення програмного рішення для реалізації розглянутого методу виявлення стресового стану людини

4.1 Обґрунтування вибору мови програмування

У своїй роботі я використав мову програмування Python через низку важливих причин. По-перше, Python має велику кількість бібліотек для роботи з даними, таких як Pandas, Numpy, Scipy та Scikit-learn, які дозволяють ефективно обробляти та аналізувати великі обсяги даних. Ці бібліотеки забезпечують широкий спектр інструментів для автоматизації процесів та зручної роботи з даними, що значно підвищує продуктивність і точність аналізу.

По-друге, Python має спеціалізовані бібліотеки для роботи з біосигналами, зокрема з фотоплетизмографічними даними. Такі бібліотеки надають потужні інструменти для фільтрації, аналізу та візуалізації біосигналів. Використання цієї бібліотеки дозволяє швидко і точно обробляти біосигнали, виділяти необхідні характеристики та проводити детальний аналіз.

Таким чином, Python є ідеальним вибором для моєї роботи завдяки своїй гнучкості, потужності та широкому спектру інструментів, що дозволяють ефективно працювати з даними та біосигналами.

4.2 Використані бібліотеки мови програмування Python

Для практичної реалізації інтелектуальної системи виявлення стресу на основі Shimmer-сенсорів було використано такі бібліотеки мови програмування Python: Pandas, Numpy, Scipy, Scikit-learn та NeuroKit.

Pandas — це бібліотека Python, призначена для роботи з даними в табличній формі. Вона надає високопродуктивні, гнучкі та виразні інструменти для аналізу даних [19]. Основні особливості Pandas включають:

- Структури даних: Pandas забезпечує дві основні структури даних — DataFrame та Series. DataFrame є двовимірною таблицею з рядками та стовпцями, що дозволяє

зручно працювати з табличними даними. `Series` — це одномірний масив, схожий на стовпець у таблиці.

- Обробка даних: `Pandas` надає широкий спектр методів для маніпуляції з даними, включаючи сортування, фільтрацію, групування, злиття та агрегування даних.
- Читання та запис даних: `Pandas` підтримує імпорт та експорт даних у різних форматах, таких як `CSV`, `Excel`, `SQL`, `JSON` та інші.
- Інтеграція з іншими бібліотеками: `Pandas` легко інтегрується з іншими бібліотеками для наукових обчислень, такими як `numpy` та `scipy`.

`NumPy` (`Numerical Python`) — це основна бібліотека `Python` для наукових обчислень. Вона забезпечує підтримку багатовимірних масивів та матриць, а також набір високорівневих математичних функцій для їх обробки [20]. Основні можливості `NumPy` включають:

- Масиви: `NumPy` надає структуру `ndarray` для роботи з багатовимірними масивами. Ці масиви є основою для численних обчислень та аналізу даних у `Python`.
- Векторизовані операції: `NumPy` дозволяє виконувати операції над масивами без використання циклів, що значно підвищує продуктивність обчислень.
- Математичні функції: `NumPy` містить широкий набір математичних функцій, включаючи алгебру, статистику, генерацію випадкових чисел та інші.
- Інтеграція з іншими бібліотеками: `NumPy` є основою для багатьох інших бібліотек `Python`, таких як `pandas`, `scipy` та `scikit-learn`.

`SciPy` (`Scientific Python`) — це бібліотека `Python`, призначена для наукових та технічних обчислень. Вона базується на бібліотеці `NumPy` і надає високорівневі команди для виконання численних задач у наукових дослідженнях та інженерії [21]. Основні можливості `SciPy` включають:

- Лінійна алгебра: `SciPy` містить розширений набір функцій для роботи з лінійною алгеброю, включаючи розв'язання систем лінійних рівнянь, обчислення власних значень та власних векторів, а також матричні операції.
- Оптимізація: Бібліотека надає інструменти для оптимізації, такі як мінімізація функцій, пошук коренів рівнянь та підгонка кривих. Вона також включає алгоритми нелінійного програмування.

- Інтерполяція: Бібліотека містить різні методи інтерполяції даних, такі як лінійна, сплайнова та поліноміальна інтерполяція. Це дозволяє створювати гладкі криві на основі дискретних наборів даних.
- Статистика: SciPy включає широкий спектр статистичних функцій, таких як обчислення розподілів, статистичних тестів, параметричної та непараметричної статистики. Вона також містить функції для генерації випадкових чисел.
- Візуалізація: SciPy інтегрується з іншими бібліотеками для візуалізації, такими як Matplotlib, що дозволяє створювати графіки та візуалізувати дані.
- Інтеграція з іншими бібліотеками: SciPy легко інтегрується з іншими популярними бібліотеками для наукових обчислень, такими як NumPy, Pandas та Scikit-learn, що робить її важливою частиною наукової екосистеми Python.

Scikit-learn — це бібліотека Python для машинного навчання. Вона надає широкий набір інструментів для створення та навчання моделей машинного навчання [22]. Основні можливості Scikit-learn включають:

- Алгоритми машинного навчання: Scikit-learn містить реалізації багатьох алгоритмів машинного навчання, таких як лінійна регресія, логістична регресія, дерева рішень, випадкові ліси, кластеризація (наприклад, K-means) та інші.
- Попередня обробка даних: Бібліотека забезпечує інструменти для попередньої обробки даних, такі як масштабування, нормалізація, кодування категоріальних змінних та розбиття даних на навчальну та тестову вибірки.
- Оцінка моделей: Scikit-learn надає функції для оцінки моделей, такі як крос-валідація, метрики оцінки (точність, прецизійність, повнота, F1-міра) та інші.
- Пайплайни: Бібліотека дозволяє створювати пайплайни для поєднання різних етапів обробки даних та навчання моделей, що робить процес побудови моделей більш організованим та ефективним.
- Інтеграція з іншими бібліотеками: Scikit-learn легко інтегрується з іншими бібліотеками для наукових обчислень та аналізу даних, такими як numpy та pandas.

NeuroKit — це бібліотека Python, яка забезпечує комплексний інструментарій для обробки, аналізу та візуалізації біосигналів. Вона створена для того, щоб

дослідники могли легко і ефективно обробляти фізіологічні дані, такі як ЕКГ, ФПГ, ЕМГ та багато інших [23].

Основні можливості бібліотеки NeuroKit включають:

- Попередня обробка сигналів: Включає фільтрацію, нормалізацію та інтерполяцію сигналів для підготовки їх до подальшого аналізу.
- Виділення характеристик: Виявлення та екстракція важливих особливостей сигналів, таких як піки, частота серцевих скорочень, мінливість серцевого ритму та інші параметри.
- Аналіз сигналів: Містить інструменти для аналізу часових та частотних характеристик сигналів, включаючи розрахунок статистичних показників та використання методів спектрального аналізу.
- Візуалізація даних: Забезпечує можливості для візуалізації сигналів та їх характеристик, що дозволяє дослідникам легко інтерпретувати результати.
- Інтеграція з іншими бібліотеками: NeuroKit інтегрується з іншими популярними бібліотеками для наукових обчислень та візуалізації, такими як numpy, pandas, matplotlib та scipy.

NeuroKit орієнтована на простоту використання, високу гнучкість та надійність, що робить її корисним інструментом для дослідників у галузі нейропсихології, медицини та біомедичних наук. Використання цієї бібліотеки дозволяє знизити час та зусилля, необхідні для обробки та аналізу фізіологічних даних, що сприяє більш ефективному проведенню наукових досліджень.

4.3 Автоматизація процесу прийняття рішення в системі виявлення стресу

Система виявлення стресу складається з таких підпрограм:

- Нормалізація даних
- Розбиття даних на стани "релакс" та "стрес"
- Обробка сигналів ФПГ та розрахунок ВСР
- Т-тест
- Фільтрація даних ЕДА та обрахунок СКВ

- Комбінація даних ВСР та ЕДА
- Машинне навчання

Опис кожної з підпрограм:

- Нормалізація даних:

Для початку усі датасети з біосигналами вимірювань учасників були збережені у файловій директорії та перейменовані у форматі «номер учасника досліджу. csv» для зручності автоматизації.

Тоді, як про це йшлося у пункті 2.3 Препроцесинг даних, було здійснено видалення даних, записаних пізніше ніж це вимагав протокол. Таким чином було приведено отримані результати до однакової «довжини» за допомогою програмного рішення показаному у додатку А. Цей ж програмний текст видаляє нерелевантні, цілям даної роботи, дані та відлік часу інтегрований виробником переводить у зручний для використання з відліком від 0.

- Розбиття даних на стани "релакс" та "стрес":

Для зручності роботи дані було поділено на дві групи датасетів – біосигнали у стані без наявних подразників та з наявними стресорами. Програмний текст для цього продемонстрований у додатку Б задля сепарації біосигналів вимірювання без наявного стресу та у додатку В відповідно з наявними стресорами.

- Обробка сигналів ФПГ та розрахунок ВСР:

Сепаровані датасети з сигналами окремо фільтруються за допомогою влаштованих методів бібліотеки NeuroKit. На даному етапі виконується фільтрування сигналів фотоплетизмографії. До відфільтрованого сигналу ФПГ застосовується метод знаходження R- піків.

R-піки у контексті фотоплетизмографії є точками, що відповідають максимальному об'єму крові в судинах під час систоли, тобто коли серце скорочується і виштовхує кров. Вони використовуються для розрахунку частоти серцевих скорочень та варіабельності серцевого ритму, що допомагає в оцінці фізіологічного стану організму.

За знайденими R- піками розраховуються компоненти VCP. Програмний тест цієї і дій описаних в попередніх двох абзаців надано у додатках Г та Д для даних зібраних без наявного стресу та з.

- T-тест:

Для виявлення значущих параметрів VCP для аналізу стану стресу було проведено T-тест. Отримані компоненти було профільровано за значенням $P < 0.5$ та збережено до готового датасету для аналізу за допомогою методів машинного навчання. Програмний код продемонстровано у додатку Е.

- Фільтрація даних ЕДА та обрахунків СКВ:

Для використання даних ЕДА, було обрано метод аналізу опору шкіри. Опір потрібно відфільтрувати на викиди (аутлаєри) та порахувати середнє квадратичне відхилення до стану стресу та у цьому стані. Пізніше СКВ будуть використовуватися для визначення стану. Програмний текст для фільтрації, обрахунку СКВ надано у додатках Ж та И.

- Комбінація даних VCP та ЕДА:

Отримані значення параметрів VCP та СКВ необхідно з'єднати попарно в залежності від наявності/відсутності стресорів у датасети разом. У додатку К наведено код для цього.

- Машинне навчання:

Коли усі дані отримано тоді до кожного набору додаються мітки: 0 для даних стану релаксу і 1 для стресу. Потім ці два набори даних об'єднуються в один загальний датасет. Для стандартизації ознак було використано `StandardScaler`, що гарантує, що кожна ознака матиме середнє значення 0 і стандартне відхилення 1. Після цього дані було розділено на навчальну і тестову вибірки за допомогою функції `train_test_split`.

Наступним кроком є навчання моделі `Naive Bayes`. Було створено модель `GaussianNB` і навчено її на навчальних даних, а потім використовується модель для передбачення на тестових даних. Для оцінки моделі було обчислено точність передбачень на тестовій вибірці за допомогою методів описаних у пункті 3.2 Оцінка результатів. У додатку Л надано програмний текст для цієї підпрограми, але для усіх

методів машинного навчання описаних у цій роботі для наочного зображення коду виконання усього аналізу.

Після оцінки моделі планується використовувати її для моніторингу фізіологічного стану людини. Модель можна інтегрувати в систему моніторингу здоров'я, яка буде аналізувати дані з біосенсорів і надавати рекомендації або попередження користувачеві в разі виявлення стресу.

4.4 Тестування моделі та отримані результати

Датасет було розділено на дві вибірки з відсотковим відношенням 70%/30% на навчальну та тестову вибірки. Після тренування моделі Naive Bayes було отримано такий результат точності, прецизійності, відгуку, F1-міри:

Точність (Accuracy): 0.8888888888888888

Прецизійність (Precision): 1.0

Повнота (Recall): 0.8333333333333334

F1-міра (F1 Score): 0.9090909090909091

Матрицю невідповідності можна спостерігати на рисунку 4.3

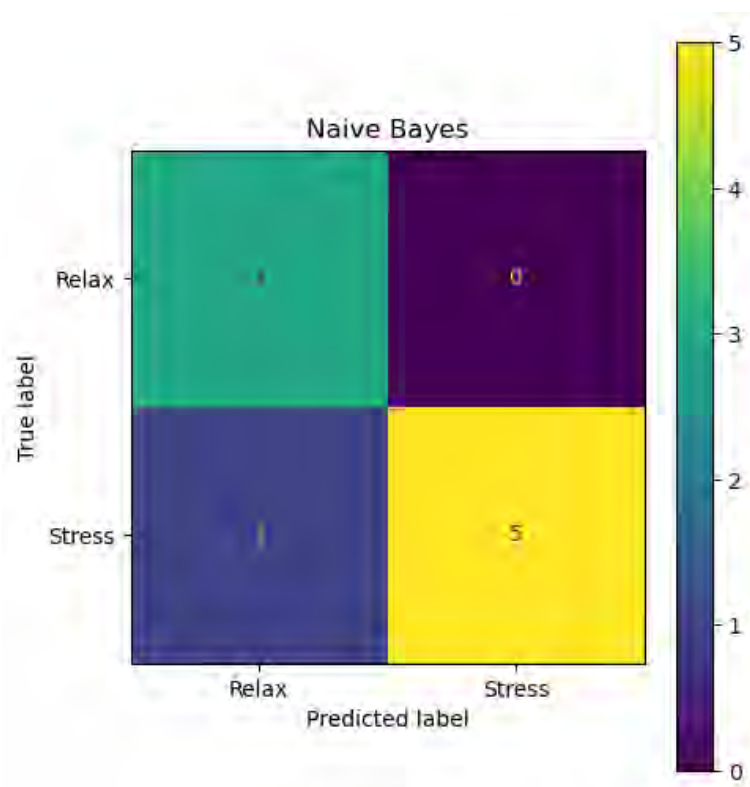


Рисунок 4.3 Матриця невідповідності для Naive Bayes

Для підвищення точності моделі визначення стану стресу людини, планується додати більше даних вимірювань.

Тренована модель готова до використання, необхідно лиш скористатися функцією `predict` для даних де необхідно визначити стан стресу.

На рисунку 4.4 зображено вивід даних після процесу ідентифікації стану людини. Моделі було надано на перевірку уже завчасно відомі метрики виміряні у людини з наявним впливом стресорів. На рисунку можемо спостерігати завантажені метрики та відповідь моделі у правому нижньому куті «Naïve Bayes Prediction». Отримана відповідь моделі була 1, тобто наявність стресу. За відсутності, відповідно отримана відповідь мала би бути 0.

```

    HRV_HF    HRV_VLF    HRV_SD1    HRV_MFDFA_alpha2_Peak    \
0  0.005279  0.001028  34.908922                0.599187

    HRV_MFDFA_alpha2_Fluctuation    HRV_SDANN1    HRV_AI    HRV_MFDFA_alpha1_Max    \
0                0.000498    47.052905    52.419628                0.56902

    HRV_MFDFA_alpha1_Mean    HRV_LZC    ...    HRV_SI    HRV_RMSSD    HRV_DFA_alpha2    \
0                0.810317    0.819251    ...    52.061446    49.270308                0.856286

    HRV_MFDFA_alpha2_Asymmetry    HRV_GI    HRV_MFDFA_alpha2_Increment    \
0                0.0    52.229299                0.015516

    HRV_LF    HRV_SD1d    Resistance_Std    Naive Bayes Prediction
0  0.003217    25.653547                84.063996                1

```

Рисунок 4.4 Результати роботи моделі

Висновки до розділу

У четвертому розділі було здійснено розробку та впровадження програмного рішення для реалізації методу виявлення стресового стану людини на основі даних, отриманих за допомогою Shimmer-сенсорів. Було обґрунтовано вибір мови програмування Python, завдяки її потужним бібліотекам, що дозволяють ефективно обробляти та аналізувати біосигнали.

Було використано бібліотеки Pandas, Numpy, Scipy, Scikit-learn та NeuroKit для реалізації системи. Детально описано процес автоматизації обробки даних, включаючи нормалізацію, розбиття на стани, фільтрацію сигналів, розрахунок параметрів ВСР та ЕДА, проведення Т-тесту, комбінацію даних та застосування методів машинного навчання.

Розроблене програмне забезпечення дозволяє ефективно обробляти біосигнали та автоматизувати процес аналізу даних, що забезпечує високу точність класифікації станів спокою та стресу. На основі отриманих результатів було встановлено, що модель Naive Bayes показала точність класифікації 88%.

Для підвищення точності та надійності системи виявлення стресу, планується розширення обсягу даних та включення додаткових біосигналів у подальших дослідженнях.

Висновки

У ході виконання дипломної роботи «Інтелектуальна система виявлення стресу на основі Shimmer-сенсорів» були досягнуті поставлені завдання та виконані наступні етапи роботи:

Аналіз існуючих систем та рішень для розпізнавання стресу:

Було проаналізовано різні пристрої та методи для виявлення стресу, такі як електрокардіографія, електроенцефалографія, електроміографія та електродермальна активність. Відповідно до проведеного аналізу було обрано фотоплетизмографію та ЕДА через їх доступність для носимих пристроїв, що дозволяють постійний моніторинг стану здоров'я, а також обрано сенсор Shimmer для виконання збору даних.

Збір та обробка даних:

Проведено вимірювання біосигналів 30 учасників у віці від 20 до 30 років під час стану спокою та під впливом стресу, використовуючи сигнал повітряної тривоги як стресор. Дані були збережені у форматі .csv з частотою запису 128 Гц. Проведено попередню обробку даних, включаючи видалення нерелевантних даних та приведення сигналів до однакової довжини.

Розробка методу виявлення стресу на основі методів машинного навчання:

Використано методи K-Nearest Neighbors, Naive Bayes, Random Forest, Decision Tree та Support Vector Regression для аналізу даних.

Проведено t-тест для визначення значущих компонентів варіабельності серцевого ритму.

Розробка автоматизованого програмного рішення:

Використано мову програмування Python та бібліотеки Pandas, NumPy, SciPy, Scikit-learn та NeuroKit для обробки даних та тренування моделей. Створено програмне забезпечення, яке автоматизує процес збирання, обробки та аналізу даних, а також класифікації стану стресу.

Отже, за результатами виконаної роботи можемо бачити, що метод Naive Bayes має найвищу точність у виявленні стану стресу, досягнувши точності 88%. Також

ефективно працювала модель Support Vector Regression із середньоквадратичною помилкою 19%. Застосування методів машинного навчання дозволило досягти поставлених цілей у виявленні стану стресу на основі аналізу ФПГ та ЕДА.

Розроблена інтелектуальна система може бути використана для моніторингу фізіологічного стану людини в режимі реального часу. Це дозволяє швидко виявляти стрес та надавати рекомендації або попередження користувачеві, що сприяє підвищенню рівня здоров'я та безпеки.

Для подальшого підвищення точності моделі та її адаптації до різних умов, необхідно збільшити обсяг даних вимірювань та дослідити додаткові параметри біосигналів. Це дозволить створити більш універсальну та надійну систему для виявлення стресу.

Літературні джерела

1. Definitions and concepts of stress / L. M. Manosso et al. *Glutamate and neuropsychiatric disorders: current and emerging treatments*. 2022. P. 27--63.
2. Review on psychological stress detection using biosignals / G. Giannakakis et al. *IEEE transactions on affective computing*. 2022. Vol. 13, no. 1. P. 440-460.
3. Офіційний сайт Fitbit для трекерів активності та іншого. URL: <https://www.fitbit.com/global/content/dam/fitbit/global/pdp/devices/sense-2/device-360/shadow-grey/prod0.png> (дата звернення: 15.05.2024).
4. Empatica | Medical devices, AI and algorithms for remote patient monitoring. URL: https://www.empatica.com/website/assets/images/e4/2/e4_front-lg-hdpi.jpg (дата звернення: 16.05.2024).
5. Wearable Sensor Technology | Wireless IMU | ECG | EMG | GSR. URL: https://shimmersensing.com/wp-content/uploads/2021/07/gsr_sensor-galvanic_skin_response_op-shimmer-worn.jpg (дата звернення: 16.05.2024).
6. Wearable technologies for electrodermal and cardiac activity measurements: a comparison between fitbit sense, empatica E4 and shimmer GSR3+ / V. Ronca et al. 2023.
7. Нечипоренко А. С. Моделі, методи та інформаційні технології раннього виявлення розладів в нестационарних квазіперіодичних процесах : автореф. дис. ... д-ра техн. наук : 05.13.06. Харків, 2018. 41 с.
8. Interquartile range - Wikipedia. *Wikipedia, the free encyclopedia*. URL: https://en.wikipedia.org/wiki/Interquartile_range (дата звернення: 16.05.2024).
9. ChuDuc H., NguyenPhan K., NguyenViet D. A review of heart rate variability and its applications. *APCBEE procedia*. 2013. Vol. 7. P. 80-85.
10. Paired t-Test. URL: https://www.jmp.com/en_gb/statistics-knowledge-portal/t-test/paired-t-test.html (дата звернення: 16.05.2024).
11. Beers B. P-Value: what it is, how to calculate it, and why it matters. *Investopedia*. URL: <https://www.investopedia.com/terms/p/p-value.asp> (дата звернення: 16.05.2024).
12. What is the k-nearest neighbors algorithm? | IBM. *IBM - United States*. URL: <https://www.ibm.com/topics/knn> (дата звернення: 16.05.2024).
13. What are naïve bayes classifiers? | IBM. *IBM - United States*. URL: <https://www.ibm.com/topics/naive-bayes> (дата звернення: 16.05.2024).
14. What is random forest? | IBM. *IBM - United States*. URL: <https://www.ibm.com/topics/random-forest> (дата звернення: 16.05.2024).
15. What is a decision tree? | IBM. *IBM - United States*. URL: <https://www.ibm.com/topics/decision-trees> (дата звернення: 17.05.2024).
16. What is support vector machine? | IBM. *IBM - United States*. URL: <https://www.ibm.com/topics/support-vector-machine> (дата звернення: 17.05.2024).
17. What is a confusion matrix? | IBM. *IBM - United States*. URL: <https://www.ibm.com/topics/confusion-matrix> (дата звернення: 18.05.2024).
18. *IBM - United States*. URL: https://www.ibm.com/content/dam/connectedassets-adobe-cms/worldwide-content/creative-assets/s-migr/ul/g/c8/a7/binary-matrix.component.complex-narrative-xl.ts=1712087356966.png/content/adobe-cms/us/en/topics/confusion-matrix/jcr:content/root/table_of_contents/body/content_section_styled/content-section-

- [body/complex_narrative_390941229/items/content_group/image](#) (дата звернення: 18.05.2024).
19. Pandas - python data analysis library. *pandas - Python Data Analysis Library*. URL: <https://pandas.pydata.org/> (дата звернення: 18.05.2024).
 20. NumPy -. *NumPy* -. URL: <https://numpy.org/> (дата звернення: 18.05.2024).
 21. SciPy -. *SciPy* -. URL: <https://scipy.org/> (дата звернення: 18.05.2024).
 22. Scikit-learn: machine learning in Python – scikit-learn 1.5.0 documentation. *scikit-learn: machine learning in Python – scikit-learn 0.16.1 documentation*. URL: <https://scikit-learn.org/stable/> (дата звернення: 18.05.2024).
 23. Overview – NeuroKit2 0.2.10 documentation. *GitHub Pages*. URL: <https://neuropsychology.github.io/NeuroKit/introduction.html> (дата звернення: 18.05.2024).
 24. Статистичні методи визначення залежностей між випадковими величинами: навчальний посібник / Ю. В. Куц, Ю. Ю. Лисенко; КПІ ім. Ігоря Сікорського. – Київ: КПІ ім. Ігоря Сікорського, 2023. – 115 с.
 25. Галаган Р.М. Комп'ютерне проектування електронних схем. Комп'ютерний практикум: навчальний посібник. Київ : КПІ ім. Ігоря Сікорського, 2023. – 419 с.

Додатки

ДОДАТОК А
Нормалізація даних

доступ надається за домовленістю з авторами

ДОДАТОК Б
Розбиття даних на стан "релакс"

доступ надається за домовленістю з авторами

ДОДАТОК В
Розбиття даних на стан "стрес"

доступ надається за домовленістю з авторами

ДОДАТОК Г
Обробка сигналів ФПГ та розрахунок ВСР для стану без стресу

доступ надається за домовленістю з авторами

ДОДАТОК Д
Обробка сигналів ФПГ та розрахунок ВСР для стану зі стресом

доступ надається за домовленістю з авторами

ДОДАТОК Е
Т-тест

доступ надається за домовленістю з авторами

ДОДАТОК Ж
Фільтрація даних ЕДА та обрахунок СКВ для стану без стресу

доступ надається за домовленістю з авторами

ДОДАТОК И
Фільтрація даних ЕДА та обрахунків СКВ для стану зі стресом

доступ надається за домовленістю з авторами

ДОДАТОК К
Комбінація даних ВСР та ЕДА

доступ надається за домовленістю з авторами

ДОДАТОК Л
Машинне навчання

доступ надається за домовленістю з авторами