
Визначення типу біологічної тканини за допомогою машинного навчання

Віктор Нікітін

кафедра біомедичної кібернетики, Національний технічний університет «Київський політехнічний інститут імені Ігоря Сікорського», м. Київ, Україна
ORCID 0000-0003-1640-8986

Андрій Дубко

інститут електрозварювання ім. Є.О. Патона НАН України; кафедра біомедичної інженерії; Національний технічний університет «Київський політехнічний інститут імені Ігоря Сікорського», м. Київ, Україна
ORCID 0000-0001-6070-3945

Для цитування цієї статті:

Віктор Нікітін, Андрій Дубко. Визначення типу біологічної тканини за допомогою машинного навчання. International Science Journal of Engineering & Agriculture. Vol. 2, No. 6, 2023, pp. 1-11. doi: 10.46299/j.isjea.20230206.01

Надійшла до редакції: 30 вересня 2023 р.; **Схвалено:** 01 листопада 2023 р.;

Опубліковано: 01 грудня 2023 р.

Анотація: Розвиток інформаційних технологій розпізнавання та класифікації даних вимірювань частотних характеристик імпедансу біологічних тканин заснованих на технологіях машинного навчання, є актуальним завданням для реалізації теоретичного та прикладного дослідження питання визначення типу біологічних тканин. Метою роботи є розробка та навчання моделі для класифікації типу біологічної тканини. За допомогою оціночної плати Eval AD5933EBZ виміряно амплітудно-частотні характеристики імпедансу біологічних тканин (яблука, ківі та цибулі), після чого, за допомогою машинного навчання, проведено їх оптимальна класифікація. Для проведення моделювання застосовано систему GMDH Shell, яка дозволяє вирішувати такі завдання моделювання: класифікацію (передбачення категорії); багатовимірне прогнозування часових рядів; регресію (безперервне прогнозоване значення); поліноміальна апроксимацію кривих. У статті використано такі методи класифікації: комбінаторний, покроковий з додаванням, покроковий змішаний. Як можна побачити з результатів класифікації, та графіків амплітудно-частотних характеристик живих біологічних тканин, результат класифікації напряму залежить від якості даних вимірювань, прикладом цього є яблуко. На графіку амплітудно-частотної характеристики яблука відсутні шуми, що дозволило правильно класифікувати тканину на 100%.

Ключові слова: машинне навчання, штучний інтелект, інформаційні технології, інтелектуальний аналіз даних, оціночна плата EVAL-AD5933EBZ, імпеданс біологічних тканин, амплітудно-частотна характеристика.

1. Вступ

Аналіз частотних характеристик імпедансу біологічних тканин є інформативним методом для отримання даних про стан цих тканин [1]. Обробка даних вимірювання таких характеристик біологічних матеріалів з використанням машинного навчання має велике значення для медицини. Особливістю методів машинного навчання є не прямий розв'язок задачі, а навчання на множині подібних прикладів, що дозволяє використовувати ці методи

для обробки великих обсягів даних та виявляти в них нові, нетривіальні, корисні та доступні для інтерпретації знання [2, 3]. Розвиток інформаційних технологій (ІТ) розпізнавання та класифікації даних вимірювань частотних характеристик імпедансу біологічних тканин заснованих на технологіях машинного навчання, є актуальним завданням для реалізації теоретичного та прикладного дослідження питання визначення типу біологічних тканин.

2. Об'єкт і предмет дослідження

Об'єктом дослідження є дані вимірювань імпедансу біологічних тканин в широкому частотному діапазоні (5-100 кГц).

Предметом дослідження є оцінка оптимальних алгоритмів машинного навчання в задачах класифікації типу біологічних тканин.

3. Мета та задачі дослідження

Метою роботи є розробка та навчання моделі для класифікації типу біологічної тканини. Для досягнення поставленої мети необхідно: провести аналіз науково-технічної літератури з теоретичних основ машинного навчання, обрати типи класифікації; провести вимірювання імпедансу біологічних тканин; зберегти дані вимірювань; провести класифікацію біологічних тканин різних типів за допомогою методів машинного навчання.

4. Аналіз літератури

Машинне навчання (Machine Learning – ML) є дуже популярним у багатьох галузях, таких як охорона здоров'я, фінанси, електронна комерція, тощо. Його техніки та алгоритми весь час розвиваються, а сам напрямок, безперечно, має великі перспективи [4, 5]. Машинне навчання – великий підрозділ штучного інтелекту (Artificial Intelligence -AI), що вивчає методи побудови алгоритмів, здатних навчатися (рис.1) [6].



Рис. 1. Штучний інтелект і його складові.

В середині 20 століття концепцію штучного інтелекту та машинного навчання «машина, яка може навчатися на досвіді» запропонував створити математик Алан Тюрінг [7]. Области практичного застосування штучного інтелекту в сучасних умовах наведено у таблиці 1 [8].

Таблиця 1. Области практичного застосування штучного інтелекту в сучасних умовах

| Область застосування | Характеристика |
|---|---|
| Машинне навчання | Здійснює автоматизацію побудови аналітичної моделі, збирає, аналізує та використовує статистичні дані. Таким чином, формує уявлення стосовно певних ситуацій та як їх вирішувати у різних сферах діяльності людини |
| Нейронна мережа | Один із типів машинного навчання, необхідного для встановлення потрібної зв'язки для корекції виконання поставлених задач або прийняття заздалегідь правильних рішень у відповідних ситуаціях |
| Глибоке навчання | Має змогу формувати багаторівневі нейронні мережі, що дає можливість використовувати переваги обчислювальних потужностей та вдосконалені методи навчання для обробки більш складних моделей з більшими масивами даних. |
| Когнітивні обчислення | Когнітивні обчислення використовуються для імітації процесів. На прикладі людини, котра спочатку інтерпретує зображення та мову, а потім уже самостійно може говорити та виконувати певні дії. |
| Комп'ютерне бачення | Машини здатні розпізнавати образи та вивчати, що відбувається на зображенні чи відео. Таке опція дозволяє машинам самостійно обробляти та аналізувати відео чи зображення і пропонувати свої рішення щодо обробки та використання матеріалу |
| Доказ теорем | У процесі розвитку штучного інтелекту важливу роль відіграло вивчення прийомів доказу теорем. Багато різноманітних задач використовують ті ж методичні підходи, що використовуються під час доказу теорем. При цьому, доведення теореми включає не лише проведення дедукції на основі гіпотез, але й створення інтуїтивних припущень, про те, що необхідно довести, щоб підтвердити теорему |
| Розпізнавання зображень | Розробник системи формує список ознак, від котрого багато залежить якість розпізнавання. Суть розпізнавання полягає в апріорному отриманні вектору ознак для виділеного окремого об'єкта, і потім на основі списку ознак, визначення котрій з фігур відповідає даний вектор ознак |
| Машинний переклад і розуміння людської мови | На основі семантичної моделі представлення текстів створено мову для внутрішнього представлення знань. Тому сьогодні системи здійснюють аналіз фраз та текстів у наступні етапи: морфологічний, синтаксичний, семантичний та прагматичний аналіз |
| Ігрові програми | Одним з прикладів є навчання системи гри в шахи. При цьому в шахах є декілька рівнів складності, що відображають якість гри системи та ідентифікують чіткі критерії оцінки інтелектуального зростання системи. |
| Машинна творчість | Програмні системи, здатні самостійно створювати музику, вірші, оповідання, статті, дипломи і навіть дисертації. Додатково створено безліч музичних додатків: системи обробки звуку, синтезу звуку, системи інтерактивної композиції, програми алгоритмічної композиції. |
| Експертні системи | Використовуються в науці, бізнесі, техніці, виробництві та інших сферах, де існує цілком визначена предметна область. Умовою ефективної роботи такої системи є існування алгоритму у визначеній предметній області. |

Штучний інтелект спроможний також вирішити дуже розповсюджену сьогодні для молоді проблему діагностування очних хвороб. За допомогою використання штучного інтелекту «Deep Mind» від «Google» та британського центру з боротьби з захворюванням

очей «Moorfields Eye Hospital» розроблена спеціальна нейромережа, яка з використанням усього одного знімку сітківки спроможна діагностувати понад 50 хвороб очей [9].

Останніми роками проведено велику кількість досліджень, що демонструють упровадження методів машинного навчання в різні сфери [10].

ML перебуває на стику математичної статистики, методів оптимізації та класичних математичних дисциплін, але має також і власну специфіку, що пов'язана з проблемами обчислювальної ефективності та перенавчання (рис. 2). Багато методів тісно пов'язано з витягуванням інформації й інтелектуальним аналізом даних (Data Mining) [6].



Рис. 2. Місце машинного навчання серед інших дисциплін.

Машинне навчання ML має вчитися та вдосконалюватися на наданих даних і застосовувати їх у динамічному середовищі. Основні етапи побудови моделі ML включають дизайн дослідження, збір даних, підготовку даних, навчання моделі, оцінку моделі та покращення продуктивності (рис. 3) [11].

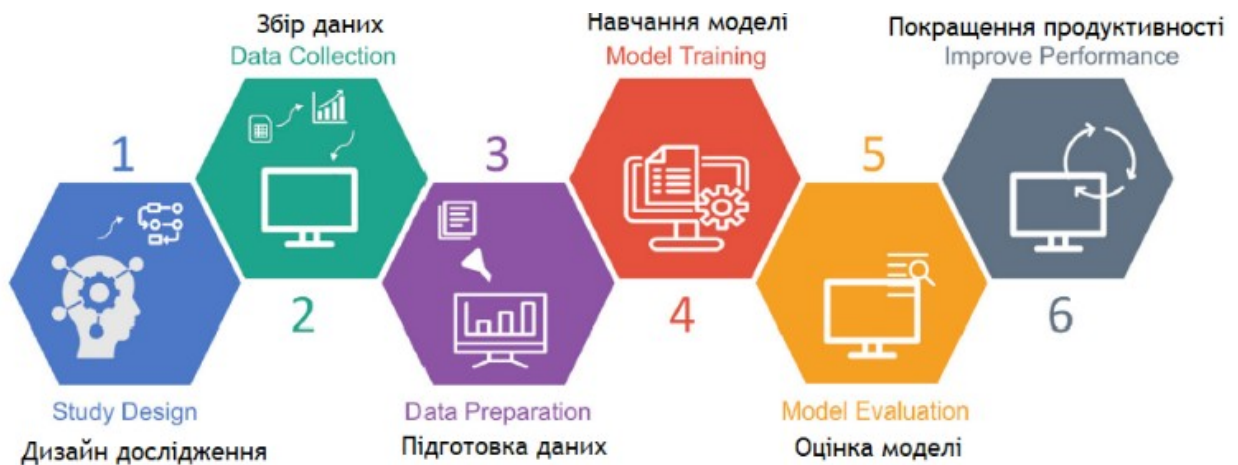


Рис. 3. Основні етапи побудови моделі машинного навчання.

Усі завдання, які вирішуються за допомогою машинного навчання, належать до таких категорій (таблиця 2) [12].

Таблиця 2. Категорії завдань

| Завдання | Призначення |
|-----------------------|--|
| Класифікації | Отримання категоріальної відповіді на основі набору ознак. Має кінцеве кількість відповідей (як правило, у форматі «так» або «ні»): чи є на фотографії кіт, чи є зображення людським обличчям, хворий пацієнт раком. Застосовується в маркетингу під час оцінювання кредитоспроможності позичальників, визначенні лояльності клієнтів, розпізнаванні образів, медичній діагностиці та в багатьох інших сферах. |
| Регресії | Прогнозування на основі вибірки об'єктів з різними ознаками. На виході має вийти дійсне число (2, 35, 76,454 тощо). Наприклад, ціна квартири, вартість цінного паперу після півроку, очікуваний дохід магазину на наступний місяць, якість вина при сліпому тестуванні. |
| Кластеризації | Розподіл даних на групи: поділ усіх клієнтів мобільного оператора за рівнем платоспроможності, зарахування космічних об'єктів до тієї чи іншої категорії (планета, зірка, чорна діра тощо). |
| Зменшення розмірності | Зведення великої кількості ознак до меншого (зазвичай 2–3) для зручності їх подальшої візуалізації (наприклад, стиснення даних). |
| Виявлення аномалій | Виявлення аномалій від стандартних випадків. На практиці таким завданням є, наприклад, виявлення шахрайських дій з банківськими картами. |

На рисунку 4 показані основні види машинного навчання [11].

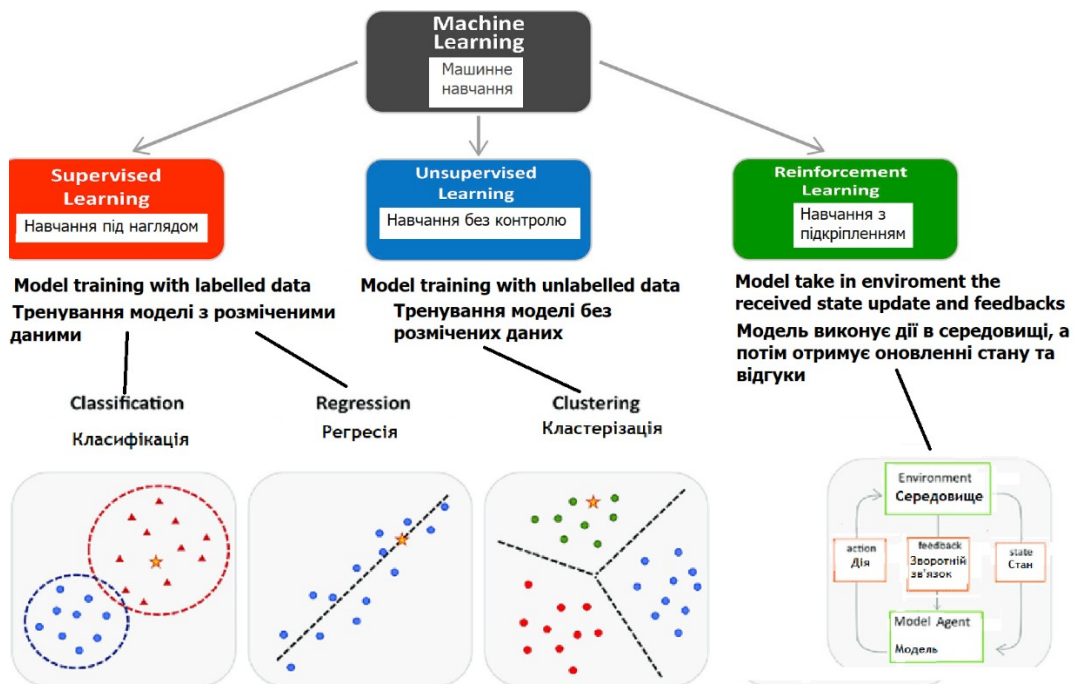


Рис. 4. Основні види машинного навчання.

Основні види машинного навчання включають такі категорії, як класифікацію, регресію при навчанні під наглядом та кластеризацію в навчанні без контролю. Навчання з підкріпленням покращує продуктивність моделі шляхом взаємодії з середовищем. Кольорові крапки та трикутники представляють навчальні дані. Жовті зірочки представляють нові дані, які може передбачити навчена модель.

У випадку задач класифікації процес навчання включає послідовні етапи, такі як підготовка даних, створення навчального набору, генерація класифікатора, навчання

класифікатора, прогнозування, оцінка продуктивності класифікатора та налаштування параметрів.

Ринок програмного забезпечення, що підтримує методи й алгоритми інтелектуального аналізу даних, досить широкий і представлений, наприклад, такими продуктами: R, Python, SPSS, Statsoft Statistica, SAS Enterprise Miner, Oracle DM, MATLAB, Viscovery, Orange, WEKA, Deductor, GMDH Shell.

5. Методи досліджень

Для вирішення наукових завдань, поставлених у статті використано такі методи класифікації: комбінаторний, покроковий з додаванням, покроковий змішаний. Моделювання виконувалось методом групового урахування аргументів (МГУА) в програмному забезпеченні GMDH Shell (Group Method of Data Handling). Для отримання даних використано метод вимірювання імпедансу за допомогою оціночної плати Eval AD5933EBZ [13, 14] на базі інтегральної мікросхеми AD5933 [15], також використані методи цифрової обробки та статистичної обробки сигналів вимірювань.

6. Результати досліджень

Для проведення дослідження застосовуємо систему GMDH Shell, яка є інтелектуальним інструментом моделювання, формує математичні моделі та використовується для вирішення задач інтелектуального аналізу даних, таких як класифікація, прогнозування, аналіз часових рядів [16].

Вона розроблена компанією Geos Research Group (www.gmdhshell.com), яка заснована у 2009 році з метою побудувати найкраще програмне забезпечення для прогнозування. Обчислювальний процес у цій програмній системі заснований на методі групового урахування аргументів МГУА, за допомогою якого відбувається структурно-параметрична оптимізація математичних моделей, що відображають закономірності багатопараметричних даних об'єкта, що досліджується. Програмна система GMDH Shell надає два паралельні алгоритми інтелектуального аналізу даних - нейронні мережі, що самоорганізуються, і комбінаторну структурну оптимізацію моделей. Середовище GMDH Shell дозволяє вирішувати такі завдання моделювання: багатовимірне прогнозування часових рядів; регресія (безперервне прогнозоване значення); класифікація (передбачення категорії); поліноміальна апроксимація кривих [17].

У наших експериментальних дослідженнях ми використовуємо алгоритми платформи GMDH Shell. Ця платформа пропонує наступні 4 алгоритми для прогнозування часових рядів [18, 19]:

1. Комбінаторний. Це класичний алгоритм сортування на основі GMDH, який розглядає всі можливі комбінації змінних;

2. Нейронний. Це також заснований на GMDH алгоритм нейроподібної релаксації, де згенеровані змінні використовуються разом із початковими;

3. Покроковий з додаванням. Це подібно до покрокової регресії, де процедура додає нового члена до поточної моделі, перевірявши її відповідно до принципів GMDH;

4. Покроковий змішаний. Це подібно до поетапної регресії, де процедура може додавати успішних членів до поточної моделі, а також видаляти неуспішні з поточної моделі, перевірявши їх відповідно до принципів GMDH.

На рисунках 5 – 7 показані амплітудно-частотні характеристики імпедансу живих рослинних тканин, в нашому випадку це яблуко, цибуля та ківі. Для кожної тканини було проведено 10 експериментів. Кількість точок вимірювання АЧХ дорівнює 101. На цих рисунках спостерігається дисперсія імпедансу, яка може бути лише в живих тканинах.

Цифрова та статистична обробка даних проводилась за допомогою мови програмування Python та бібліотеки Pandas.

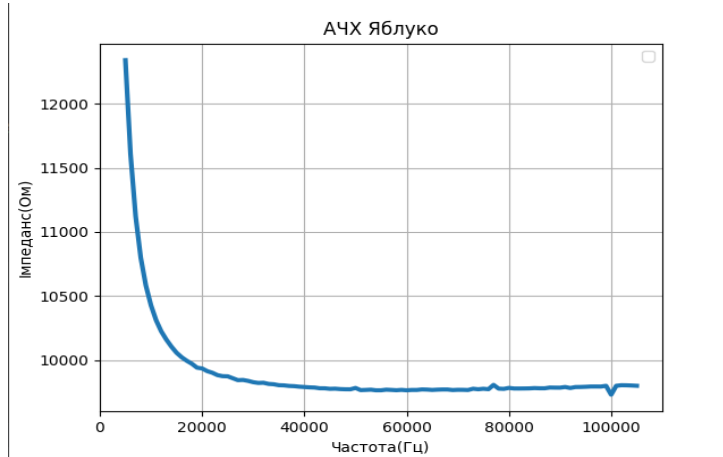


Рис. 5. Амплітудно-частотна характеристика яблука.

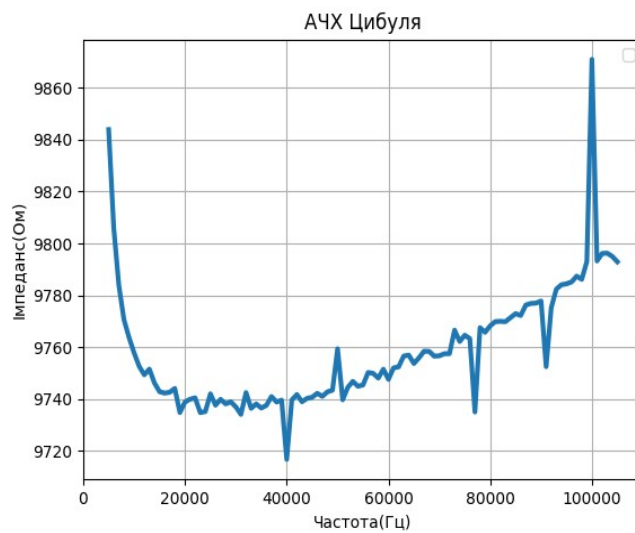


Рис. 6. Амплітудно-частотна характеристика цибулі.

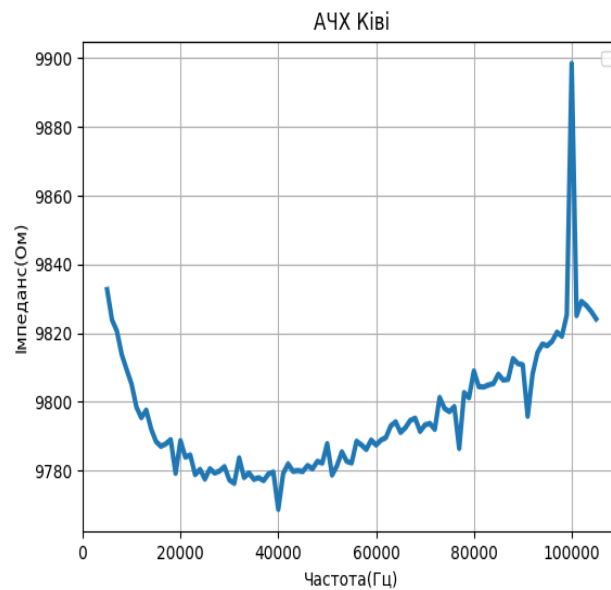


Рис. 7. Амплітудно-частотна характеристика ківі.

Набір даних складає собою 3030 вимірювань АЧХ імпедансу рослинних тканин платою EVAL AD5933EBZ. Класи одразу збалансовані по кількості вимірювань. Ці тканини були закодовані наступним чином: яблуко – 1; цибуля – 2; ківі – 3 (рис. 8).

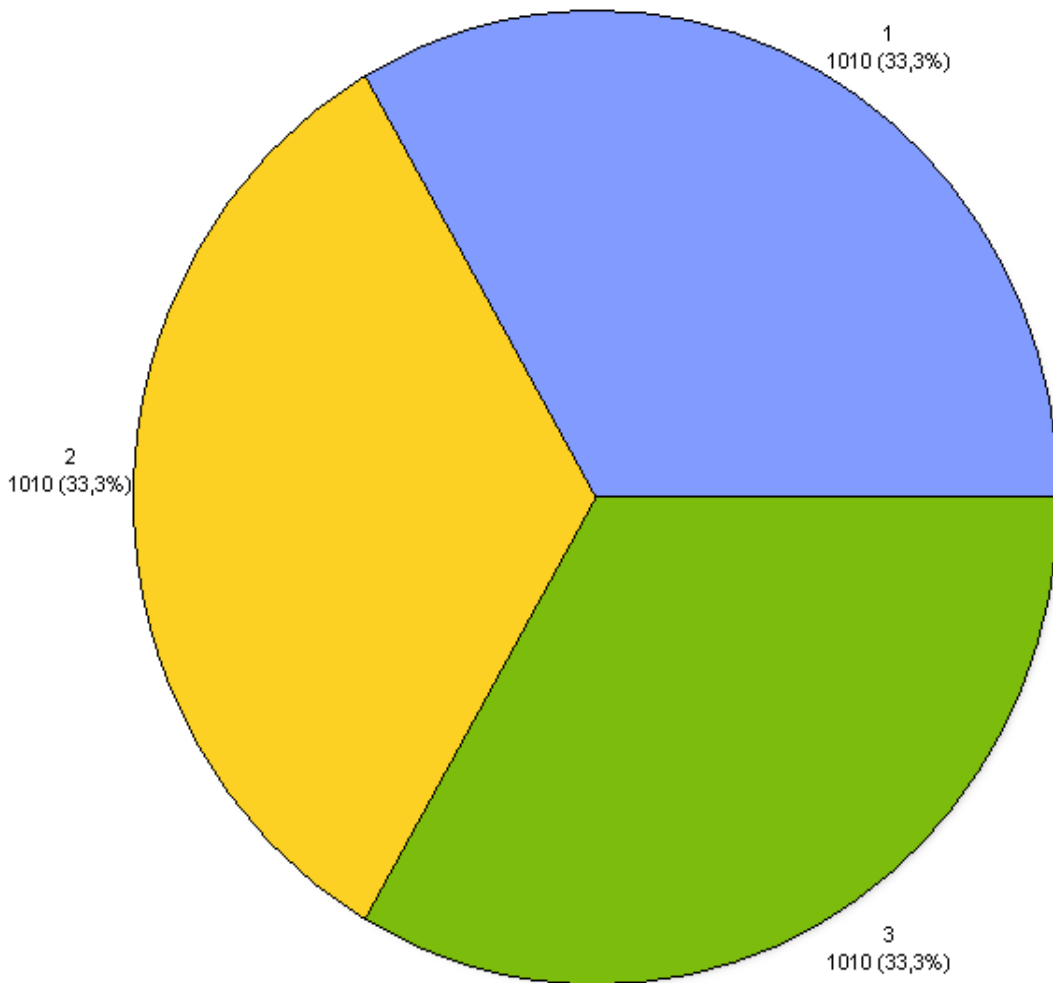


Рис. 8. Кругова діаграма набору даних, де 1-яблуко, 2-цибуля, 3-ківі.

Моделі будувалися в співвідношенні розбиття набору даних: 20% на тестову вибірку це 606 точок, та 80% на навчальну вибірку це 2424 точок відповідно.

Був обраний метод покроковий з додаванням, так як він показав найвищі результати, точність на навчанні склала 93,8%, 2274 вимірювань було класифіковано вірно, 150 невірно. На екзамені - 94.1%, 570 - вірно, 36 - невірно. У таблиці 3 показано точність моделей класифікації на екзамені.

Таблиця 3. Точність моделей класифікації на екзамені

| Метод | Точність |
|-------------------------|----------|
| Комбінаторний | 92,4% |
| Покроковий з додаванням | 94.1% |
| Покроковий змішаний | 92,2% |

На рис. 9 наведені точності класифікації на навчальній виборці.

| | | Predicted class | | | Total | Recall |
|------------------|---|-----------------|-------|-------|-------|--------|
| | | 1 | 2 | 3 | | |
| Actual class | 1 | 808 | 0 | 0 | 808 | 1,000 |
| | 2 | 0 | 790 | 18 | 808 | 0,978 |
| | 3 | 0 | 109 | 699 | 808 | 0,865 |
| Total | | 808 | 899 | 717 | 2424 | |
| Precision | | 1,000 | 0,879 | 0,975 | | |
| F-measure | | 1,000 | 0,926 | 0,917 | | |
| Baseline | | 0,667 | 0,667 | 0,667 | 0,333 | |
| Accuracy | | 1,000 | 0,948 | 0,948 | 0,948 | |

Рис. 9. Точність класифікації на навчанні.

На рис. 10 наведені точності класифікації на екзаменаційній виборці.

| | | Predicted class | | | Total | Recall |
|------------------|---|-----------------|-------|-------|-------|--------|
| | | 1 | 2 | 3 | | |
| Actual class | 1 | 202 | 0 | 0 | 202 | 1,000 |
| | 2 | 0 | 199 | 3 | 202 | 0,985 |
| | 3 | 0 | 33 | 169 | 202 | 0,837 |
| Total | | 202 | 232 | 172 | 606 | |
| Precision | | 1,000 | 0,858 | 0,983 | | |
| F-measure | | 1,000 | 0,917 | 0,904 | | |
| Baseline | | 0,667 | 0,667 | 0,667 | 0,333 | |
| Accuracy | | 1,000 | 0,941 | 0,941 | 0,941 | |

Рис. 10. Точність класифікації на іспиті.

З рис. 10 можна побачити, що найкраще всього було класифіковано 1-ий клас - яблуко, всі 202 вимірювання правильно класифіковані (точність 1). В той час як 2-ий клас правильно класифікував 199 вимірювань вірно з 202, 3-ий клас класифікував 169 вимірювань вірно. 2 та 3 клас мають точність 0,941.

7. Перспективи подальшого розвитку досліджень

У подальших роботах будуть проведені експерименти з допомогою оціночної плати EVAL-AD5933EBZ для прикладного дослідження не тільки типу а й стану живих біологічних тканин.

8. Висновки

За допомогою оціночної плати Eval AD5933EBZ виміряно амплітудно-частотні характеристики імпедансу біологічних тканин (яблука, ківі та цибулі), після чого, за допомогою машинного навчання, проведено їх оптимальна класифікація. Як можна побачити з результатів класифікації, та графіків амплітудно-частотних характеристик живих біологічних тканин, результат класифікації напряму залежить від якості даних вимірювань, прикладом цього є яблуко. На графіку АЧХ яблука відсутні шуми, що дозволило правильно класифікувати тканину на 100%.

Список літератури:

1) Нікітін, В., Дубко, А. (2023). Частотні залежності імпедансу еквівалентних електричних схем заміщення живих біологічних тканин, які моделюють норму та патологічний стан. *International Science Journal of Engineering & Agriculture*, 2(2), 10-20. doi: <https://doi.org/10.46299/j.isjea.20230202.02>.

2) Кононова, К. (2020). Машинне навчання: методи та моделі. Харків: ХНУ імені В. Н. Каразіна, 301.

3) Савченко, Є. (2019). Задача багатозадачного навчання як одна з важливих задач машинного навчання. *Індуктивне моделювання складних систем*, 11. 103-111.

4) Що таке machine learning? Як працює машинне навчання та де воно використовується. Available at: <https://www.telegraf.in.ua/advertisement/10119869-scho-take-machine-learning-jak-pracjuje-mashinne-navchannja-ta-de-vono-vikoristovuyetsja.html>.

5) Rao, C., Govindaraju, V. (2013). *Machine Learning: Theory and Applications*. Handbook of Statistics, 31. 552.

6) Могильний, С. (2019). Машинне навчання з використанням мікрокомп'ютерів: навч.-метод. посіб. За ред. О. В. Лісового та ін. К., 224.

7) Turing, A. (1995). Lecture to the London Mathematical Society on 20 February 1947. 1986. *MD. Comput.* 12, 390–397.

8) Погореленко, А. (2018). Штучний інтелект: сутність, аналіз застосування, перспективи розвитку. *Науковий вісник Херсонського державного університету*, 32. 22-27.

9) Штучний інтелект навчився діагностувати очні хвороби. Available at: <http://bukovina.biz.ua/news/48130>.

10) Lytvyn, V., Vysotska, V. (2015). Designing Architecture of Electronic Content Commerce System. *Computer Science and Information Technologies: Proc. of the Xth Int. Conf. CSIT'2015*, 14–17 September, 2015, Lviv, Ukraine. Publishing Lviv Polytechnic. 115 – 119.

11) Junjie, P., Elizabeth, C., Pierre, D. and Coziana, C. (2018). Machine Learning Techniques for Personalised Medicine Approaches in Immune-Mediated Chronic Inflammatory Diseases: Applications and Challenges. *Frontiers in Pharmacology. Machine Learning in Autoimmune Diseases 2021*, 12. Article 720694. 1-18. doi: 10.3389/fphar.2021.720694.

12) Кравченко, С., Гришкун, Є., Власенко, О. (2020). Методи класифікації машинного навчання з використанням бібліотеки SCIKIT-LEARN. *Вчені записки ТНУ імені В.І. Вернадського. Серія: технічні науки*, 31(70). 121-125. doi: <https://doi.org/10.32838/TNU-2663-5941/2020.3-1/19>.

13) Нікітін, В., Дубко, А. (2022). Вимірювання імпедансу біологічних тканин та інших середовищ за допомогою оціночної плати EVAL AD5933EBZ. *International Science Journal of Engineering & Agriculture*, 1(4), 80–91. doi: <https://doi.org/10.46299/j.isjea.20220104.04>.

14) EVAL-AD5933. Available at: <https://www.analog.com/en/design-center/evaluation-hardware-and-software/evaluation-boards-kits/eval-ad5933.html#eb-documentation>.

15) Базаев, Н., Пржиялговская А., Руденко П. (2016). Особенности использования микросхемы AD5933 в качестве измерителя импеданса при проектировании малогабаритных систем. *Известия вузов. Электроника. Том 21 № 3*, 279-285.

16) GMDH Shell for Data Science uses the power of the GMDH (Group Method of Data Handling). Available at: <http://www.gmdh.net/>.

17) Рябцев, В., Шевченко, Н. (2014) Прогнозирование социальных показателей при помощи системы GMDH SHELL. *Вісник КНУТД 2014, №2*. 162-168.

18) Mogilev, P., Alexandrov, M., Boldyreva, A., Cardiff, J. (2020). GMDH-based Models for Mid-term Forecast of Cryptocurrencies (on example of Waves). 2020 IEEE 15th International Conference on Computer Sciences and Information Technologies (CSIT). 23-26 September, 2020, Zbarazh-Lviv, Ukraine, 2. 13-16.

19) Learning algorithms. Available at: https://gmdhsoftware.com/docs/learning_algorithms.

Determination of the type of biological tissue using machine learning

Viktor Nikitin

Department of Biomedical Cybernetics National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Kiev, Ukraine

Anrii Dubko

Department of Welding and Related Technologies in Medicine and Ecology E.O. Paton Electric Welding Institute; Department of Biomedical Engineering National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Kiev, Ukraine

Abstract: The development of information technologies for the recognition and classification of measurement data of the frequency characteristics of the impedance of biological tissues based on machine learning technologies is an urgent task for the implementation of theoretical and applied research on the issue of determining the type of biological tissues. The aim of the work is to develop and train a model for classifying the type of biological tissue. Using the evaluation board Eval AD5933EBZ, the amplitude-frequency characteristics of the impedance of biological tissues (apples, kiwi and onions) were measured, after which, with the help of machine learning, their optimal classification was carried out. For modeling, the GMDH Shell system is used, which allows solving the following modeling tasks: classification (category prediction); multidimensional forecasting of time series; regression (continuous predicted value); polynomial approximation of curves. The following classification methods are used in the article: combinatorial, step-by-step with addition, step-by-step mixed. As can be seen from the classification results and graphs of amplitude-frequency characteristics of living biological tissues, the classification result directly depends on the quality of the measurement data, an example of this is an apple. There is no noise on the graph of the amplitude-frequency characteristic of the apple, which made it possible to correctly classify the tissue by 100%.

Keywords: machine learning, artificial intelligence, information technology, intelligent data analysis, evaluation board EVAL-AD5933EBZ, impedance of biological tissues, amplitude-frequency characteristic.
