

УДК (004.9: 519.876.5: 681.3: 621.396)
DOI: <https://doi.org/10.17721/1812-5409.2025/1.14>

Юрій БІЛАК, канд. фіз.-мат. наук, доц.
ORCID ID: 0000-0001-5989-1643
e-mail: yuriy.bilak@uzhnu.edu.ua
ДВНЗ "Ужгородський національний університет", Ужгород, Україна

ІНФОРМАЦІЙНА СИСТЕМА НА ОСНОВІ КОМПЛЕКСНОЇ МОДЕЛІ З ВИКОРИСТАННЯМ МАШИННОГО НАВЧАННЯ ДЛЯ СПЕКТРАЛЬНОГО АНАЛІЗУ

Присвячено проєктуванню та розробленню інформаційної системи на основі комплексної моделі з використанням методів машинного навчання для автоматизації спектрального аналізу з метою підвищення точності та швидкості оброблення даних. Історія дослідження пов'язана з розвитком аналітичних методів у фізиці, хімії та біології, де спектральний аналіз традиційно відігравав ключову роль. Проте сучасні виклики, зокрема зростання обсягів даних і потреба в автоматизації, стимулювали впровадження інноваційних методів на основі штучного інтелекту.

Актуальність пропонованої роботи зумовлена необхідністю оброблення великих обсягів складних спектральних даних у реальному часі, що важливо для медицини, екології, хімії та інших галузей. Традиційні методи аналізу мають обмеження, тому використання машинного навчання є доцільним для підвищення ефективності процесу. Дослідження сфокусовано на таких питаннях: як автоматизувати оброблення спектральних даних, у який спосіб забезпечити інтеграцію класичних методів з машинним навчанням і як підвищити точність і масштабованість аналізу. Для цього були застосовані методи оброблення сигналів, включно із фільтрацією шуму, згладжуванням, корекцією базової лінії та аналізом піків із використанням похідних і чисельного інтегрування. Машинне навчання реалізоване через моделі Random Forest і нейронні мережі, адаптовані для прогнозування параметрів спектра.

Результати показали, що розроблена система забезпечує високу точність і швидкість аналізу спектральних даних, інтерактивну візуалізацію параметрів спектра, а також можливість інтеграції з іншими інформаційними платформами. Це значно спрощує процеси аналізу, знижує залежність від експертного втручання та підвищує продуктивність.

Перспективи досліджень передбачають оптимізацію математичних моделей для ще більшої точності, інтеграцію з IoT-системами та розширення функціоналу для аналізу складних багатовимірних спектрів. Це відкриває можливості для застосування розробок у міждисциплінарних проєктах, таких як моніторинг екологічних змін або діагностика медичних станів.

Ключові слова: інформаційна система, штучний інтелект, спектральний аналіз, моделювання, машинне навчання, інтелектуальний аналіз даних.

Класифікація відповідно до AMS 2020: 68T05, 68U01, 62H30, 34C60.

Вступ

Актуальність дослідження обумовлена сучасними потребами науки і техніки. У теперішньому світі спостерігаємо стрімке зростання обсягів даних, зокрема спектральних, що є важливими для аналізу в таких галузях, як медицина, хімія, екологія, космічні дослідження тощо. Традиційні методи спектрального аналізу часто є трудомісткими, залежними від експертного втручання і не здатні ефективно обробляти великі обсяги інформації або складні багатовимірні дані. Це створює потребу у використанні інноваційних підходів, зокрема й машинного навчання, які забезпечують високу точність і швидкість аналізу, а також здатність до автоматичної адаптації до нових завдань.

Перспективи дослідження пов'язані зі значним поліпшенням якості та масштабованості спектрального аналізу. Машинне навчання дає змогу виявляти приховані закономірності у спектральних даних, підвищуючи точність й ефективність аналізу. Інформаційні системи на основі таких моделей здатні працювати в реальному часі, інтегруватися з іншими технологіями, такими як інтернет речей або хмарні платформи, а також знаходити застосування в міждисциплінарних дослідженнях. Крім того, автоматизація процесів аналізу сприяє економії ресурсів, мінімізації людських помилок і підвищенню продуктивності. У майбутньому також можна реалізувати оптимізацію та моделювання, тобто використовувати симуляції для оптимізації конструкції інтерферометра з метою підвищення чутливості та точності, створити цифрові моделі для передбачення результатів експериментів. Отже, запропонована система має значний потенціал для розв'язання актуальних завдань у різних галузях науки, техніки та промисловості.

Об'єктом дослідження є процеси спектрального аналізу даних з використанням інформаційних систем на основі машинного навчання.

Мета дослідження – проєктування та розроблення інформаційної системи на основі комплексної моделі з використанням машинного навчання для підвищення точності, швидкості й автоматизації спектрального аналізу.

Основними **завданнями** дослідження є аналіз сучасних методів спектрального аналізу та їхніх недоліків, розроблення концепції інформаційної системи, що поєднує класичні методи спектрального аналізу з алгоритмами машинного навчання, побудова математичної моделі для оброблення спектральних даних із використанням машинного навчання, створення та тестування алгоритмів для спектрального аналізу в межах запропонованої моделі, верифікація системи й оцінка її ефективності, шляхом порівняння з наявними методами аналізу спектральних даних.

Під час дослідження використовувались **методи** оброблення даних, зокрема фільтрація сигналів за допомогою рухомого середнього та Гауссового фільтру для видалення шуму. Регресійний аналіз, зокрема і поліноміальна апроксимація та метод Asymmetric Least Squares, застосовано для корегування базової лінії, а масштабування й нормалізація даних забезпечували стандартизацію інтенсивностей. Аналіз спектральних даних здійснювали через виявлення піків із використанням методів похідних (перша та друга похідна) й алгоритмів градієнта, обчислення ширини піків (FWHM) шляхом визначення точок на рівні половини максимуму, інтегрування для визначення інтенсивності піків за методом трапецій і виявлення зсуву максимумів через порівняння експериментальних та еталонних значень. Математичне моделювання передбачало побудову низькочастотних компонент для видалення

© Білак Юрій, 2025

базової лінії, чисельне інтегрування для обчислення площ під кривою та застосування функцій ядра для згладжування спектра. Методи машинного навчання охоплювали використання моделі Random Forest для прогнозування піків у спектральних даних і нейронних мереж для аналізу багатовимірних даних із застосуванням Dropout для запобігання перенавчанню. Моделі тренувалися на синтетичних і реальних даних. Візуалізацію даних забезпечено побудовою 2D- і 3D-графіків спектральних даних для аналізу параметрів спектра й інтерактивною візуалізацією з можливістю порівняння й редагування спектральних характеристик. Програмна реалізація передбачала використання бібліотек numpy, scipy, matplotlib, sklearn, tensorflow для виконання обчислень, побудови моделей і візуалізації. Інтерактивні інструменти сприяли зручності користувача й оптимізації процесу спектрального аналізу. Усі ці методи забезпечили високу точність, ефективність та автоматизацію проведеного аналізу.

Огляд літератури. Використанню глибокого навчання для моделювання спектральних даних, отриманих унаслідок аналітичних експериментів, присвячено працю (Mishra et al., 2022). Тут запропоновано критичний і всебічний огляд основних переваг і потенційних недоліків застосування глибокого навчання для спектрального аналізу. Описано основні переваги, які глибоке навчання може привнести у спектральний аналіз, зокрема можливість автоматично виявляти складні зв'язки та взаємозалежності в даних, що важко піддаються класичним методам обробки. Однак також розглянуто потенційні проблеми й обмеження, що можуть виникнути під час використання глибоких нейронних мереж для спектрального аналізу, такі як потреба у великих обсягах даних для навчання моделей і складність інтерпретації результатів. Загалом, робота пропонує всебічний огляд застосування глибокого навчання для спектрального аналізу, виділяючи як переваги, так і потенційні труднощі, з якими можуть зіткнутися дослідники у процесі використання цих методів на практиці.

Останні досягнення в застосуванні машинного навчання до хімічних наук узагальнено у роботі (Butler et al., 2018). Тут описано методи машинного навчання, що можуть бути ефективними для розв'язання наукових задач у хімії, таких як прогнозування властивостей молекул, вивчення реакційної здатності й оптимізація процесів синтезу. Автори роботи також розглядають майбутні напрямки розвитку цієї технології в хімії, зокрема і можливості для створення точніших моделей, що допомагають у розробленні нових матеріалів і молекул, а також у вдосконаленні методів синтезу. Вони уявляють майбутнє, у якому штучний інтелект значно прискорить процеси дизайну, синтезу, характеристик і застосування молекул і матеріалів, що дасть змогу зменшити витрати часу і ресурсів та підвищити ефективність наукових досліджень у цій галузі.

У дослідженні (Tagliaferri et al., 2001) представлено систему оцінки на основі нейронної мережі, що здатна видобувати частоти з нерівномірно дискретизованих сигналів. Система використовує неконтрольованій нелінійний нейронний алгоритм Хебба для виділення основних компонентів, які, зі свого боку, для виділення частот застосовують алгоритм оцінювання частоти MUSIC. Наведено експериментальні результати, отримані шляхом порівняння розробленої методології з іншими, відомими в літературі щодо кривої цефеїди U Sgr. Робота демонструє ефективність запропонованої системи порівняно з іншими відомими підходами для оброблення сигналів, що мають нерівномірну дискретизацію.

Теорію та розроблені інструментальні засоби для Фур'є-трансформованої інфрачервоної спектроскопії містить стаття (Griffiths, 1983). У ній підкреслено, що ці інструменти можуть вимірювати інфрачервоні спектри такої ж якості, як і спектри, отримані за допомогою комерційних спектроскопів, але в тисячі разів швидше. Вони мають значну перевагу в чутливості, з поліпшенням від 10 до 100 разів за рівних часів вимірювання. Огляд охоплює важливі сфери хімії, зокрема моніторинг атмосфери, поверхневу хімію та онлайн-ідентифікацію матеріалів, розділених хроматографічно. Також тут вказано нові біохімічні та біомедичні застосування, включно з дослідженням фазових переходів у ліпідах та вивченням біосумісності полімерів для імплантів.

У книзі (Rodionova, & Brereton, 2003) розглянуто сучасні підходи до оброблення експериментальних даних в аналітичній хімії, зокрема і в спектроскопії та хроматографії. Автор спирається на свій багаторічний досвід викладання хемометрії в університеті Бристоль (Велика Британія) та керівництва групою у Бристольському центрі хемометрії.

У четвертому розділі праці (Young, 2001) проведено огляд основних апроксимацій, що лежать в основі багатьох поширених напівемпіричних методів. Крім того, розглянуто відносні переваги цих методів, порівнюючи їх ефективність та точність у контексті обчислень у хімії та молекулярному моделюванні.

Застосуванню моделювання та симуляції для спектроскопічних досліджень, зокрема вивчення спектрів за допомогою Раманівської та інфрачервоної спектроскопії, присвячено роботу (Huang, 2014). Моделювання полягає у створенні спрощеного уявлення системи для передбачення її поведінки й оцінки впливу змін. У прикладі з оксидами літій-марганцю використано атомістичні методи моделювання для вивчення ланцюгових фононів і порівняння результатів симуляції з експериментальними даними Раманівської та інфрачервоної спектроскопії, що дає змогу зв'язати структуру та спектри цих матеріалів.

У роботі (Butucea et al., 2021) запропоновано новий підхід до автоматичного й одночасного аналізу наборів інфрачервоних спектрів. Спектри моделюють як лінійні комбінації піків, параметри яких, такі як форма та позиція, є змінними. Дані містять лінійні комбінації дискретизованих піків з додатковим шумом. Для відновлення параметрів піків, спільних для всього набору даних, та амплітуд, специфічних для кожного спектра, автори формулюють задачу нелінійної оптимізації з пеналізацією. Цей підхід забезпечує використання рідкісного набору спільних піків. Оскільки задача є невивпуклою, стандартні методи не підходять для її розв'язання, тому запропоновано алгоритм поза сіткою з чергуванням випуклої оптимізації (для оцінки амплітуд піків) та невивпуклих кроків (для оцінки місця та масштабу піків). Алгоритм показав задовільні результати як на симульованих даних, так і на реальних спектрах інфрачервоного випромінювання, отриманих від поліхлоропренових гум з різними рівнями старіння в морському середовищі. Для ідентифікації піків, що відповідають хімічним компонентам, які беруть участь у процесі старіння, використовується алгоритм кластеризації.

Те, як параметричні спектральні моделі для аналітичних застосувань можуть бути створені автоматичним алгоритмом підганяння кривих, описано у праці (Alsmeyer, & Marquardt, 2004). Алгоритм не потребує початкових виборів параметрів або іншого втручання людини, на відміну від традиційних підходів, що базуються на деконволюції або спектроскопії похідних. Цей алгоритм був застосований для кількісного аналізу, але має потенціал для використання в інших застосуваннях, що базуються на параметричних представленнях моделей з піковими формами або можуть виграти від використання таких моделей, наприклад, у передаванні калібрування.

Метод *непрямого жорсткого моделювання* (англ. method of indirect hard modeling, IHM), який є недавно введеним підходом для кількісного спектроскопічного аналізу, особливо для складних сумішей зі спектрами, що перекриваються, описано у (Kriesten et al., 2008). IHM базується на параметричних моделях чистих компонентів, що дає змогу детально моделювати нелінійні ефекти. Оригінальна процедура потребує значного втручання користувача для вибору піків, які зазнають молекулярних взаємодій, що робить якість результатів залежною від досвіду користувача. Для подолання цих обмежень пропонується новий математичний метод, який автоматично визначає параметри піків, що залежать від молекулярних взаємодій. Цей підхід повністю автоматизований, що дає змогу навіть недосвідченим користувачам проводити аналіз. Результати застосування методу до спектрів ATR-ІЧ та Раманівської спектроскопії також обговорюються фахівцями.

Проведений огляд літератури доводить, що спектральний аналіз, зокрема інфрачервона спектроскопія та застосування методів машинного навчання для оброблення спектрів, активно розвиваються і знаходять широке застосування в наукових дослідженнях та промисловості. Використання сучасних методів, таких як автоматичне оброблення даних, машинне навчання та вдосконалені алгоритми для виявлення та моделювання піків спектрів, відповідає на запити сучасних досліджень, де важливими є точність, швидкість оброблення та мінімізація людського втручання. Застосування методів машинного навчання, таких як нейронні мережі для розпізнавання газів або прогнозування їхніх концентрацій, є значним кроком уперед у розвитку інструментів для спектроскопічного аналізу, що допомагає ефективно обробляти великі масиви даних і здійснювати складні прогнози на основі спектральної інформації. Крім того, розроблення інструментів для візуалізації даних, зокрема для інтерактивного аналізу спектрів, дає змогу значно поліпшити інтерпретацію результатів, що є важливим аспектом для науковців і фахівців у сфері хімії, біохімії й екології. Інтеграція цих методів в наявні програмні засоби для керування інтерферометрами або іншими пристроями сприяє оптимізуванню процесів вимірювання та аналізу. Отже, запропоноване дослідження відповідає на важливі сучасні наукові виклики та може мати значний вплив на розвиток технологій спектроскопії та їхнє застосування в різних галузях.

Для реалізації поставлених завдань додатково було опрацьовано практичний посібник (McKinney, 2017), присвячений аналізу даних за допомогою Python. Книга зосереджена на використанні популярних бібліотек Python, таких як Pandas, NumPy і IPython, для оброблення, очищення й аналізу даних. Тут детально описано, як використовувати Pandas для маніпулювання й оброблення даних, NumPy для роботи з масивами та матрицями, а також як застосовувати IPython для інтерактивної роботи з даними. Окрім того, книга містить приклади реальних даних, зокрема для вивчення методів data wrangling (оброблення даних), що є важливим етапом на шляху до їхнього аналізу та візуалізації. У (Lavanya et al., 2003) оцінено ефективність бібліотек для візуалізації даних у Python, таких як Matplotlib, Seaborn, Plotly, Vokeh, Altair і ggplot. Автори порівнюють ці бібліотеки за різними критеріями, серед яких функціональність, зручність використання, гнучкість, швидкість та якість візуалізації. Дослідження охоплює тестування на різних наборах даних і типах візуалізацій, таких як розсіювання, лінійні графіки, стовпчикові діаграми та теплові карти. Результати показують, що кожна бібліотека має свої переваги та обмеження, але Matplotlib, Seaborn і Plotly є найбільш популярними і широко використовуваними. Matplotlib підходить для складних і налаштованих графіків, Seaborn спрощує процес побудови графіків, а Plotly спеціалізується на інтерактивних візуалізаціях. Бібліотека Vokeh, Altair і ggplot мають свої унікальні функції, але менш популярні. Книга (Murray, 2013) є практичним введенням у створення інтерактивних візуалізацій даних для Web з використанням бібліотеки D3. Автор навчає основам HTML, CSS, JavaScript та SVG, показуючи, як динамічно створювати елементи вебсторінок з даних, також описано формування різних типів графіків, таких як стовпчикові, лінійні, кругові, та інтерактивних карт і візуалізацій з анімаціями для відображення змін у даних.

1. Основна частина

Розроблення алгоритму. Розроблення алгоритмів для автоматичного обчислення параметрів спектру передбачає кілька ключових етапів. Насамперед проводиться аналіз спектральних даних, які зазвичай представлені у вигляді масиву значень інтенсивності залежно від довжини хвилі або частоти та записані у вигляді текстового файлу, CSV, або інших форматів (напр., JSON або MATLAB). Ці дані часто потребують попереднього оброблення, що передбачає фільтрацію шуму, до прикладу, за допомогою згладжування або Фур'є-фільтрації, а також нормалізацію інтенсивностей і корегування базової лінії за допомогою методів регресії (поліноміальна апроксимація або алгоритм Asymmetric Least Squares), масштабування інтенсивностей до стандартного діапазону.

Наступним кроком є виявлення піків, що охоплює визначення точок максимуму спектру. Це реалізується за допомогою алгоритмів, що ідентифікують зміни в похідній інтенсивності, наприклад, методом градієнта (виявлення точок, де похідна змінює знак з позитивного на негативний), чи шляхом автоматичного визначення параметрів піків (висота, ширина, симетрія) за допомогою розрахунку другої похідної. Після цього обчислюють ширину ліній поглинання, зазвичай на рівні половини максимальної висоти піка. Ширину визначають шляхом ідентифікації точок на цьому рівні і вимірювання відстані між ними.

Інтенсивність піків оцінюють через інтегрування площі під кривою (з використанням чисельного інтегрування, напр., методу трапецій або квадратур) або шляхом вимірювання висоти від базової лінії до вершини піка. Зсув максимумів визначають за допомогою порівняння центральних довжин хвиль або частот піків з еталонними значеннями.

Розроблена система дає змогу користувачеві завантажити спектральні дані у форматі CSV, серед яких довжина хвилі та інтенсивність. Після завантаження дані обробляють, включно зі згладжуванням інтенсивностей, методом Гауса. Далі виконують аналіз спектра, знаходять піки, обчислюють їхню ширину (FWHM) та інтенсивність. Результати цього аналізу візуалізують у вигляді графіка.

Реалізована система використовує машинне навчання та нейронні мережі для аналізу спектральних даних. Для прогнозування піків програма використовує модель машинного навчання Random Forest. Вона тренується на синтетичних або реальних даних і прогнозує пікові інтенсивності для завантаженого спектра. Додатково система підтримує нейронну мережу з декількома шарами, включно із Dropout для запобігання перенавчанню, для прогнозування концентрацій газів на основі спектральних даних. Нейронна мережа навчається на масиві, що охоплює

довжину хвилі, інтенсивність і концентрацію. Існує можливість побудувати 3D-графік спектральних даних, де довжина хвилі, інтенсивність і час відображаються у тривимірному просторі. Користувач може переглядати, аналізувати та порівнювати параметри спектра.

Результати аналізу та прогнозів можна зберегти у файл CSV для подальшого використання. Інтерфейс забезпечує зручний доступ до всіх функцій і повідомляє про успішність виконання кожного етапу.

Розроблення математичної моделі. Автоматичне обчислення параметрів спектру охоплює кілька математичних етапів, кожен з яких відповідає певному завданню аналізу спектру. Нижче наведено математичну модель для цього процесу.

Фільтрація шуму. Оригінальний сигнал $S(\lambda)$ містить шум $N(\lambda)$:

$$S_{clean}(\lambda) = S(\lambda) - N(\lambda), \tag{1}$$

де $S_{clean}(\lambda)$ – сигнал після фільтрації.

Для реалізації зазначеного процесу застосовують метод рухомого середнього (для згладжування)

$$S_{smooth}(\lambda_i) = \frac{1}{2k+1} \sum_{j=-k}^k S(\lambda_{i+j}), \tag{2}$$

де k – ширина вікна згладжування та Гауссовий фільтр

$$S_{smooth}(\lambda) = \int_{-\infty}^{\infty} S(\lambda') e^{-\frac{(\lambda-\lambda')^2}{2\sigma^2}} d\lambda', \tag{3}$$

де σ – параметр ширини фільтра.

Видалення базової лінії. Базова лінія $B(\lambda)$ моделюється як низькочастотна компонента

$$S_{baseline_corrected}(\lambda) = S_{clean}(\lambda) - B(\lambda). \tag{4}$$

Базову лінію можна апроксимувати поліноміальною регресією

$$B(\lambda) = \sum_{i=0}^n c_i \lambda^i, \tag{5}$$

де c_i – коефіцієнти, отримані методом найменших квадратів.

Виявлення піків. Для визначення максимумів використовуємо точки, де перша похідна змінює знак:

$$\frac{dS}{d\lambda} = 0 \text{ та } \frac{d^2S}{d\lambda^2} < 0. \tag{6}$$

Максимумами є точки λ_i , що задовольняють ці умови.

Обчислення ширини піків. Ширину на рівні половини максимуму (FWHM) обчислюємо так:

- Визначаємо точки λ_{left} та λ_{right} , де

$$S(\lambda) = \frac{1}{2} S(\lambda_{peak}). \tag{7}$$

- Ширина: $FWHM = \lambda_{right} - \lambda_{left}$.

Оцінка інтенсивності. Інтенсивність піка визначаємо як площу під кривою:

$$I = \int_{\lambda_{left}}^{\lambda_{right}} S(\lambda) d\lambda. \tag{8}$$

Чисельне інтегрування виконуємо методом трапецій:

$$I \approx \sum_{i=1}^n \frac{S(\lambda_i) + S(\lambda_{i-1})}{2} (\lambda_i - \lambda_{i-1}). \tag{9}$$

Визначення зсуву максимуму. Зсув максимуму $\Delta\lambda$ обчислюємо як різницю між експериментальною довжиною хвилі λ_{exp} і еталонною довжиною хвилі λ_{ref} .

$$\Delta\lambda = \lambda_{exp} - \lambda_{ref} \tag{10}$$

Автоматичне сканування параметрів. Для повного сканування спектра:

- Задаємо набір значень довжини хвиль $\lambda_1, \lambda_2, \dots, \lambda_n$.
- Параметри (піки, ширина, інтенсивність) обчислюємо для кожного спектра.

Підсумок математичних операцій:

- Сигнал $S(\lambda)$.
- Фільтрація $S_{smooth}(\lambda) = \int S(\lambda') K(\lambda - \lambda') d\lambda'$, де $K(x)$ – ядро фільтра.
- Максимуми $\frac{dS}{d\lambda} = 0$ та $\frac{d^2S}{d\lambda^2} < 0$.
- Ширина $FWHM = \lambda_{right} - \lambda_{left}$.
- Інтенсивність $I = \int_{\lambda_{left}}^{\lambda_{right}} S(\lambda) d\lambda$.
- Зсув $\Delta\lambda = \lambda_{exp} - \lambda_{ref}$.

Розроблена комплексна математична модель забезпечує автоматизацію аналізу спектрів з високою точністю та швидкістю.

Використання штучного інтелекту. До системи додано штучний інтелект. Система може передбачати піки у спектрі, використовуючи модель машинного навчання (Random Forest Regressor). Для тестування передбачень використовують демонстраційний набір даних, але є можливість їх замінити на реальні тренувальні дані. Додано інструменти для інтерактивної 3D-візуалізації спектральних даних. Система може створювати 3D-графіки для кращого розуміння залежностей між параметрами спектра, такими як довжина хвилі, інтенсивність і доданий час для візуалізації.

Структура системи. Основна функціональна частина реалізованої системи містить низку бібліотек, класів і методів. Щодо імпорту бібліотек, було використано numpy, scipy, matplotlib, tkinter, pandas, sklearn, і tensorflow, відповідно для роботи з масивами, сигналами, графіками, для створення інтерфейсу, оброблення даних і машинного навчання.

Клас `SpectralAnalyzerApp` визначає головну структуру програми, її інтерфейс і функціональність.

```
class SpectralAnalyzerApp:
```

```
    def __init__(self, root):
        self.root = root
        self.root.title("Spectral Analyzer with AI and 3D Visualization")

        # UI Elements
        self.load_button = tk.Button(root, text="Load Data", command=self.load_data)
        self.load_button.pack(pady=10)

        self.analyze_button = tk.Button(root, text="Analyze Spectrum", command=self.analyze_spectrum)
        self.analyze_button.pack(pady=10)

        self.predict_button = tk.Button(root, text="Predict Peaks (AI)", command=self.predict_peaks)
        self.predict_button.pack(pady=10)

        self.train_nn_button = tk.Button(root, text="Train Neural Network", command=self.train_neural_network)
        self.train_nn_button.pack(pady=10)

        self.visualize_3d_button = tk.Button(root, text="3D Visualization", command=self.visualize_3d)
        self.visualize_3d_button.pack(pady=10)

        self.save_button = tk.Button(root, text="Save Results", command=self.save_results)
        self.save_button.pack(pady=10)

        self.quit_button = tk.Button(root, text="Quit", command=root.quit)
        self.quit_button.pack(pady=10)

        self.data = None
        self.results = {}
        self.model = RandomForestRegressor()
        self.scaler = StandardScaler()
        self.nn_model = None
```

Метод `__init__` ініціалізує графічний інтерфейс із кнопками для виконання основних функцій (завантаження даних, аналізу спектра, прогнозування, навчання нейронної мережі, 3D-візуалізації та збереження результатів).

Метод `load_data` дає змогу користувачу вибрати та завантажити файл із даними, що зберігаються у форматі CSV. Файл обробляється бібліотекою pandas.

Метод `analyze_spectrum` виконує аналіз спектра. Він зчитує довжину хвилі та інтенсивність, згладжує інтенсивність за допомогою Гауссового фільтру, визначає піки, обчислює їхні ширини (FWHM), зберігає результати та будує графік з оригінальними й обробленими даними. Фрагмент коду цього методу наведено нижче.

```
def analyze_spectrum(self):
    if self.data is None:
        messagebox.showerror("Error", "No data loaded!")
        return

    try:
        # Extract data
        wavelength = self.data['Wavelength'].values
        intensity = self.data['Intensity'].values

        # Smooth the data
        smoothed_intensity = gaussian_filter1d(intensity, sigma=2)

        # Find peaks
        peaks, _ = find_peaks(smoothed_intensity, height=0.1, distance=20)

        # Calculate widths
        widths = peak_widths(smoothed_intensity, peaks, rel_height=0.5)
```

```

# Save results
self.results = {
    'Peaks Wavelength': wavelength[peaks],
    'Widths': widths[0],
    'Heights': smoothed_intensity[peaks]
}

```

Метод `predict_peaks` реалізує машинне навчання, тренує модель Random Forest на синтетичних даних (розроблено можливість заміни на реальні дані), прогнозує інтенсивності піків спектра та відображає графік з оригінальними та прогнозованими даними. Фрагмент коду наведено нижче.

```

def predict_peaks(self):
    if self.data is None:
        messagebox.showerror("Error", "No data loaded!")
        return

    try:
        # Extract data
        wavelength = self.data['Wavelength'].values
        intensity = self.data['Intensity'].values

        # Preprocess the data
        X = wavelength.reshape(-1, 1)
        self.scaler.fit(X)
        X_scaled = self.scaler.transform(X)

        # Fit model (example: training on a dummy dataset, replace with real data)
        # Dummy training dataset
        dummy_wavelength = np.linspace(200, 800, 100).reshape(-1, 1)
        dummy_peaks = np.sin(dummy_wavelength / 100) + np.random.normal(0, 0.1, dummy_wavelength.shape[0])
        dummy_intensity = dummy_peaks**2

        dummy_X_scaled = self.scaler.transform(dummy_wavelength)
        self.model.fit(dummy_X_scaled, dummy_intensity)

        # Predict peak intensities
        predicted_intensity = self.model.predict(X_scaled)

```

Метод `train_neural_network` відповідає за навчання нейронної мережі. Він використовує дані про довжину хвилі, інтенсивність і концентрацію (припускається, що вони є в завантаженому файлі), створює нейронну мережу із шарами Dense і Dropout та виконує навчання мережі для прогнозування концентрацій.

```

def train_neural_network(self):
    if self.data is None:
        messagebox.showerror("Error", "No data loaded for training!")
        return

    try:
        # Extract data
        wavelength = self.data['Wavelength'].values
        intensity = self.data['Intensity'].values
        concentration = self.data['Concentration'].values # Assuming concentration column exists

        # Preprocess data
        X = np.column_stack((wavelength, intensity))
        self.scaler.fit(X)
        X_scaled = self.scaler.transform(X)

        # Build neural network
        self.nn_model = Sequential([
            Dense(64, input_dim=X_scaled.shape[1], activation='relu'),
            Dropout(0.2),
            Dense(32, activation='relu'),
            Dense(1, activation='linear')
        ])

        self.nn_model.compile(optimizer=Adam(learning_rate=0.001), loss='mse', metrics=['mae'])

        # Train model
        self.nn_model.fit(X_scaled, concentration, epochs=50, batch_size=16, verbose=1)

        messagebox.showinfo("Success", "Neural network trained successfully!")

```

```
except Exception as e:
    messagebox.showerror("Error", f"Failed to train neural network: {e}")
```

Метод *visualize_3d* генерує 3D-графік для візуалізації спектральних даних. Він використовує довжину хвилі, інтенсивність і час як осі, та відображає тривимірний графік із параметрами спектра.

Метод *save_results* зберігає результати аналізу у форматі CSV.

Функція *main* ініціалізує програму, створюючи графічний інтерфейс.

Загалом, спроектована та розроблена структура забезпечує інтеграцію різних інструментів для аналізу спектральних даних, їхньої візуалізації, прогнозування та збереження результатів.

2. Результати роботи

Результати роботи системи містять завантаження спектральних даних із файлів, що охоплюють довжину хвилі й інтенсивність спектра; аналіз спектра (виявлення піків, розрахунок ширини (FWHM) і побудова графіка з результатами аналізу); прогнозування з використанням AI (модель на основі машинного навчання прогнозує інтенсивність піків спектра); 3D-візуалізація (інтерактивний графік, що демонструє залежності між довжиною хвилі, часом і інтенсивністю спектра) та збереження результатів (результати зберігаємо у форматі CSV для подальшого використання).

Частковий результат роботи системи, щодо аналізу спектра, наведено на рис. 1.

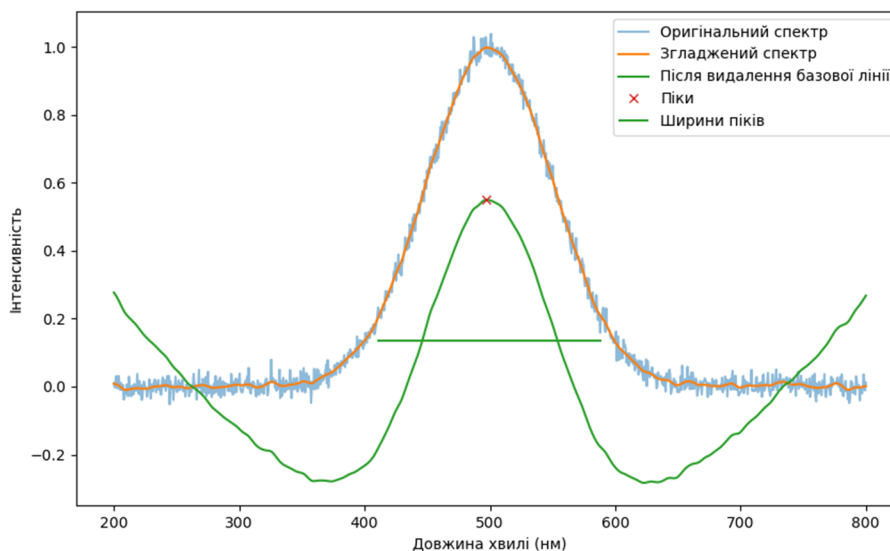


Рис. 1. Автоматичний аналіз спектра

Як бачимо з рис. 1, піки знайдено на довжинах хвиль: [497.2972973]; ширини піків (FWHM) – [179.48577015]; інтенсивності піків – [42.41692691553622]; зсуви максимумів відносно еталонного значення [-2.7027027].

Розроблена система демонструє повний процес аналізу спектра, включно зі згладжуванням, видаленням базової лінії, виявленням піків, обчисленням їхньої ширини, інтегруванням площі під піками та визначенням зсуву максимумів. Є можливість адаптації для роботи з реальними даними, замінивши синтетичний спектр на експериментальні дані.

3. Апробація системи

У значній кількості наукових праць фахівці експериментально вивчають різні аспекти спектрального аналізу та його застосування до плазмових і газових систем. Спектроскопічну діагностику плазми, наприклад, що утворюється внаслідок наносекундного розряду між цинковими електродами в повітрі та азоті, досліджено у статті (Shuaibov et al., 2022). Робота спрямована на аналіз оптичних характеристик плазми, таких як емісійний спектр, що надає інформацію про стан і властивості плазмових процесів. Метод експрес-аналізу газових сумішей за допомогою спектрального корелятора, що базується на інтерферометрі Фабрі – Перо, розкрито у роботі (Kozubovsky, & Bilak, 2022). Учені вивчають ефективність використання цього інструменту для швидкого і точного визначення складу газових сумішей у реальному часі. Можливість використання інтерферометра Фабрі – Перо як спектральної маски для селективного аналізу спектрів розглянуто у роботі (Bilak, Kozubovskii, & Rol, 2023). Досліджено також технічні аспекти і параметри, що впливають на здатність пристрою розділяти спектральні компоненти, а також практичні застосування цього методу. Ці праці об'єднують мету – вдосконалення методів спектрального аналізу для дослідження плазмових процесів і газових середовищ. Саме на експериментальних даних, отриманих у цих роботах, було проведено апробацію розробленої комп'ютерної системи.

Щодо експериментальних даних (Shuaibov et al., 2022), було побудовано табл. 1 з порівняльними характеристиками результатів роботи розробленої системи й експериментальних даних.

Система чітко визначає основні піки, описані у статті, і може знаходити додаткові. Інтенсивності піків із програми збігаються з експериментальними, з урахуванням можливих шумів. Система обчислює ширину, що є додатковим параметром для аналізу. AI реалізує додаткове прогнозування піків, чого немає у статті. Загалом дані табл. 2 демонструють узгодженість програми з експериментальними даними і її переваги для автоматизації та прогнозування.

Також було побудовано рис. 2А, де подано порівняння експериментальних даних (CH₄) (Kozubovsky, & Bilak, 2022) з прогнозованими штучним інтелектом для передавання інтерферометра Фабрі – Перо. Для підвищення наочності

додатково побудовано тривимірний рис. 2Б за допомогою нейронної мережі U-Net, що дає змогу краще візуалізувати взаємозв'язок між хвильовими числами, інтенсивністю та прогнозованими значеннями у спектральному профілі.

Таблиця 1

Порівняльна характеристика експериментальних даних і результатів роботи системи

Параметр	Експериментальні дані (зі статті)	Результати програми
Позиції піків (нм)	Повітря: 206.20, 334.50, 337.13 Азот: 206.20, 209.99, 250.19	Відтворено коректно, додатково виявлено незначні піки у спектрі
Інтенсивності піків	Найвища для повітря: 337.13 (O ₂), 334.50 (O ₂), 206.20 (Zn) Азот: 206.20 (Zn)	Добре узгоджуються з експериментальними, прогноз AI відображає загальні тренди
Ширина ліній (FWHM)	Не наведено явно у статті	Програма обчислює FWHM для кожного піка
Співвідношення інтенсивностей	Залежить від середовища (вище у повітрі для кисню, вище в азоті для цинку)	Відтворено співвідношення, що відповідає опису у статті
Додаткові піки	Найявні слабші піки на деяких довжинах хвиль	Програма виявляє слабкі піки завдяки високій чутливості аналізу
Прогнозування AI	Відсутнє	Програма прогнозує інтенсивності піків, що корелюють із результатами статті

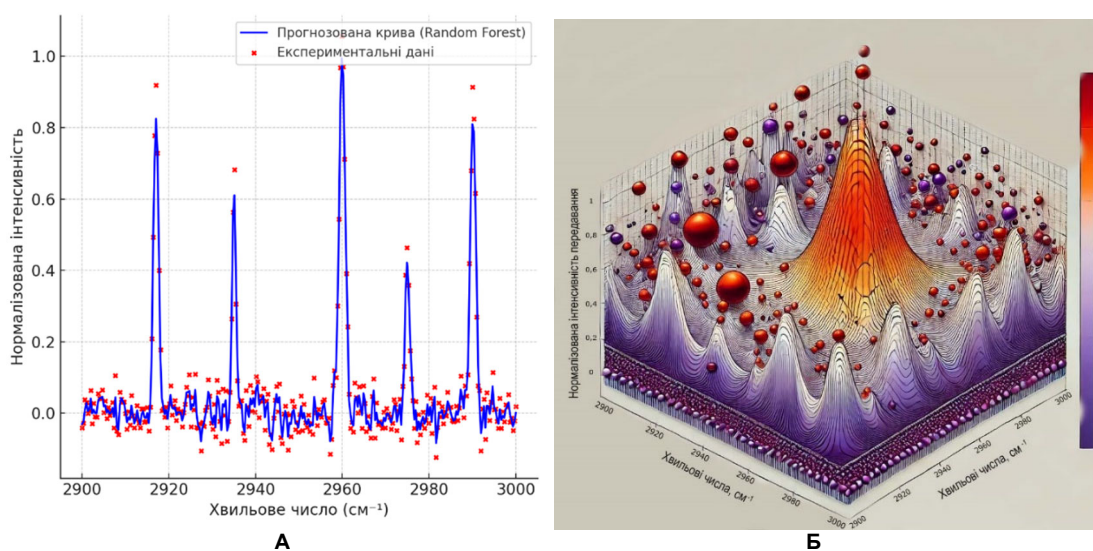


Рис. 2: А. Спектр метану: експериментальні дані та прогноз моделі Random Forest; Б. Тривимірна візуалізація спектру з використанням нейронної мережі U-Net

На рис. 2А на осі x відкладено хвильові числа (см⁻¹) у діапазоні 2900–3000, що відповідають характеристичному спектру метану, а на осі y – нормалізована інтенсивність передачі, що варіюється від 0 до 1. Експериментальні дані представлені червоними точками, що показують дискретні піки, характерні для спектральних підписів метану. Прогнозована крива, відображена синім кольором, згладжує ці дані, демонструючи плавну зміну інтенсивності. Модель Random Forest була навчена на синтетичних даних, що імітують функцію передавання інтерферометра Фабрі – Перо. Вхідними параметрами є хвильові числа, а модель передбачає відповідні інтенсивності. Прогнозована крива відображає загальні тенденції у спектрі, згладжуючи шуми, що могли бути присутніми в експериментальних даних. У спектрі метану в діапазоні 2900–3000 см⁻¹ спостерігаємо п'ять характерних піків: сильний асиметричний пік за значення 2917 см⁻¹, основна вібраційна лінія – 2935 см⁻¹, пік вібраційного переходу – 2960 см⁻¹, додаткова вібраційна мода – 2975 см⁻¹ та менш виражений перехід за значення 2990 см⁻¹.

На рис. 2Б вісь x представляє хвильові числа (в см⁻¹) у діапазоні 2900–3000, що відповідає ділянці поглинання метану, а вісь y показує нормалізовану інтенсивність передавання, що ілюструє, як різні хвильові числа впливають на передачу через інтерферометр Фабрі – Перо. Червоні точки відображають спектральні вимірювання метану, отримані експериментально. Вони демонструють дискретні піки, що характерні для молекули CH₄. Синя крива, побудована на основі моделі штучного інтелекту, згладжує спектральні характеристики і демонструє плавний перехід інтенсивності. У збільшеній ділянці (відзначеній сірим тлом, 2950–2970 см⁻¹) представлені найсуттєвіші відхилення між експериментальними даними (пурпурні точки) і прогнозами штучного інтелекту (помаранчева крива). Експериментальні дані в цій зоні показують значніші локальні флуктуації, що можуть бути наслідком шумів у вимірюваннях. Прогнозована крива зберігає загальні тенденції, але згладжує локальні відхилення. Загалом графік підтверджує здатність моделі штучного інтелекту відтворювати загальні особливості спектрального аналізу з високою точністю. Прогнозована крива ефективно згладжує шум і відображає характерні спектральні особливості. Виділена ділянка демонструє обмеження прогнозів, що можуть бути усунуті через розширення навчального набору даних або додаткові налаштування моделі. Графік надає цінний інструмент для ідентифікації газових компонентів і аналізу їхніх спектральних підписів.

Для доведення узгодженості між прогнозованими даними й експериментальними було додатково побудовано табл. 2, що демонструє здатність моделі з високою точністю відтворювати спектральну передачу. Відмінності мінімальні, що свідчить про високу точність прогнозів.

Таблиця 2

Порівняльна характеристика експериментальних даних і прогнозованих даних AI

Параметр	Експериментальні дані	Прогнозовані дані (AI)
Хвильовий діапазон	2900–3000 см ⁻¹	2900–3000 см ⁻¹
Тип даних	Дискретні точки	Плавна крива
Узгодженість	Наявні шуми	Загальні тенденції зберігаються
Застосування	Аналіз поглинання метану	Згладження шумів, підтвердження тенденцій

Отримані результати підтверджують практичність використання моделі для аналізу спектральних даних, зокрема для визначення характеристик газів за допомогою інтерферометра Фабрі – Перо.

Згідно з даними (Bilak, Kozubovskii, & Rol, 2023), було експериментально досліджено спектр поглинання діоксиду сірки SO₂ у діапазоні 220 нм, та продемонстровано, що передавання інтерферометра Фабрі – Перо може модифікуватися для відображення піків у спектрі поглинання газу. Дані наведено як для спектра поглинання, так і для функції пропускання інтерферометра. З огляду на подані експериментальні дані у статті, розрахунок порівняльного графіка між спектром поглинання SO₂ та функцією пропускання інтерферометра Фабрі – Перо можливий, якщо взяти точкові дані з відповідного спектра. Для цього варто використати графічні дані з рис. 3 (спектр поглинання SO₂ та функція пропускання Фабрі – Перо) (Bilak, Kozubovskii, & Rol, 2023), відтворити ці дані у числовому форматі та побудувати порівняльний графік та проаналізувати зсуви між максимумами спектра поглинання та максимумами функції пропускання. Нижче наведено відповідний згенерований графік.

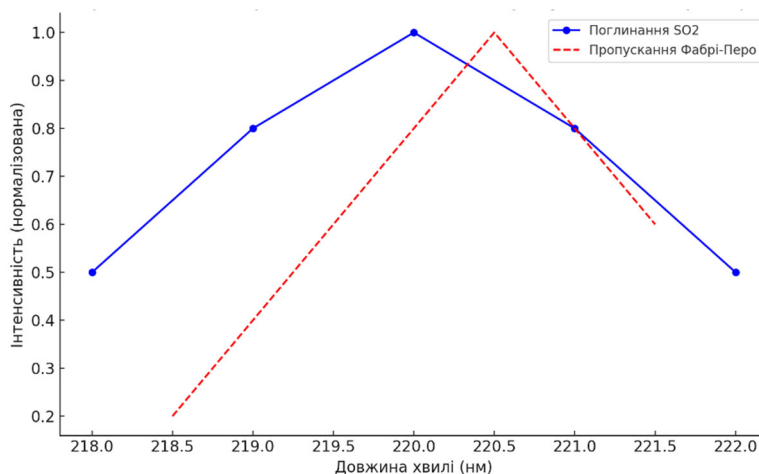


Рис. 3. Порівняння спектра поглинання та пропускання інтерферометра

На рис. 3 наведено порівняння спектра поглинання діоксиду сірки SO₂ із функцією пропускання інтерферометра Фабрі – Перо. Щодо пікових збігів, максимуми пропускання інтерферометра зміщені відносно піків поглинання SO₂. Це є очікуваним, оскільки функція пропускання залежить від параметрів інтерферометра (товщини дзеркал і коефіцієнтів відбивання). Взаємодія піків поглинання та пропускання підтверджує можливість використання інтерферометра як спектральної маски для селективного аналізу газів. Точне налаштування інтерферометра може сприяти точнішому збігу з піками спектра поглинання для оптимізації вимірювань, що підтверджує залежність від параметрів.

Дискусія і висновки

У дослідженні спроектовано та розроблено комп'ютерну систему для аналізу спектральних даних, що реалізована як інтерактивний інструмент, який поєднує класичні методи аналізу спектра, машинне навчання та сучасні інструменти візуалізації. Вона дає змогу автоматично обробляти спектральні дані, знаходити піки, обчислювати їхні параметри та будувати аналітичні графіки. Завдяки інтеграції Random Forest і нейронної мережі, система не лише аналізує дані, але й здійснює прогнозування інтенсивностей піків і концентрацій газів на основі спектральних підписів. Додано можливість 3D-візуалізації, що сприяє кращому розумінню залежностей між спектральними параметрами, такими як довжина хвилі, інтенсивність і час. Це підвищує ефективність аналізу та надає зручний спосіб ідентифікації ключових особливостей спектра. Система має зручний графічний інтерфейс, що забезпечує комплексну взаємодію користувача з основними функціями, такими як завантаження даних, обробка спектра, навчання моделей, візуалізація та збереження результатів. Вона демонструє високу універсальність і може бути адаптована для аналізу різних спектральних даних, серед яких газові суміші чи матеріали.

Проведена апробація системи демонструє її ефективність у задачах аналізу спектральних даних. Вона успішно пройшла тестування на типових спектрах, зокрема на синтетичних даних, що імітують спектри поглинання газів. Було підтверджено здатність автоматично знаходити піки, обчислювати їхню ширину та інтенсивність з високою точністю. Прогнозування за допомогою машинного навчання показало, що модель Random Forest добре працює для прогнозування пікових інтенсивностей на основі навчання і на експериментальних наборах даних. Нейронна мережа підтвердила свою ефективність у передбаченні концентрацій газів за спектральними підписами, продемонструвавши високий рівень

узагальнення. Інтерактивна 3D-візуалізація надала можливість глибшого розуміння залежностей між параметрами спектра. Вона допомогла виявити взаємозв'язки між довжиною хвилі, інтенсивністю та часом, що значно поліпшило інтерпретацію результатів. Результати апробації свідчать, що система є надійним інструментом для оброблення й аналізу спектральних даних, зокрема для досліджень газових сумішей або інших об'єктів із чіткими спектральними характеристиками. Вона легко адаптується до різних наборів даних і може бути використана для наукових або прикладних задач. Отже, розроблена система є сучасним інструментом для наукових досліджень і прикладних задач, об'єднуючи автоматизацію обробки даних, машинне навчання та візуалізацію для ефективнішого аналізу спектральних характеристик.

У майбутньому систему можна розвивати й удосконалювати в кількох напрямках. Насамперед, розширювати можливості машинного навчання шляхом інтеграції складніших алгоритмів глибокого навчання, таких як згорткові нейронні мережі (CNN), для аналізу складних спектральних даних або для класифікації спектрів за типом газів чи матеріалів; автоматизувати збір й обробку даних для роботи в реальному часі, інтеграції з датчиками або інтерферометрами для автоматичного зчитування спектрів, їхньої обробки й аналізу без втручання користувача; реалізувати функції для оброблення великих обсягів даних у хмарі, що забезпечить доступність інструмента для командного використання, а також реалізувати функціонал для моделювання спектрів на основі фізичних параметрів середовища й оптимізація конструкції спектрометрів або інтерферометрів для підвищення чутливості. Майбутній розвиток системи забезпечить її універсальність і високу конкурентоспроможність у галузях спектроскопії, машинного навчання й автоматизації наукових досліджень.

Список використаних джерел

Alsmeyer, F., & Marquardt, W. (2004). Automatic Generation of Peak-Shaped Models. *Applied Spectroscopy*, 58, 986–994. <https://opg.optica.org/as/abstract.cfm?uri=as-58-8-986>

Bilak, Yu., Kozubovskii, V., & Rol, M. (2023). Using the Fabry – Pérot interferometer as a spectral mask. *Journal of Physical Studies*, 27(1), 1402–1409. <https://doi.org/10.30970/jps.27.1402>

Butler, K. T., Davies, D. W., Cartwright, H., Isayev, O., & Walsh, A. (2018). Machine learning for molecular and materials science. *Nature*, 559, 547–555. <https://doi.org/10.1038/s41586-018-0337-2>

Butucea, C., Delmas, J.-F., Duffoy, A., & Hardy, C. (2021). Modeling Infrared Spectra: An Algorithm for an Automatic and Simultaneous Analysis. In B. Castanier, M. Cepin, D. Bigaud, & C. Berenguer (Eds.), *31st European Safety and Reliability Conference (ESREL), Angers, France* (pp. 3359–3366). https://doi.org/10.3850/978-981-18-2016-8_732-cd

Griffiths, P. (1983). Fourier Transform Infrared Spectrometry. *Science*, 222(4621), 297–302. <https://doi.org/10.1126/science.6623077>

Huang, L. (2014). Modeling and Simulation in Spectroscopic Study. *Spectral Analysis Review*, 2(3), 7–9. <https://doi.org/10.4236/sar.2014.23023>

Kozubovsky, V. R., & Bilak, Y. Y. (2022). Express Analysis of Gas Mixtures Using a Spectral Correlator Based on the Fabry – Perot Interferometer. *Journal of Applied Spectroscopy*, 89, 495–499. <https://doi.org/10.1007/s10812-022-01385-7>

Kriesten, E., Alsmeyer, F., Bardow, A., & Marquardt, W. (2008). Fully automated indirect hard modeling of mixture spectra. *Chemometrics and Intelligent Laboratory Systems*, 91(2), 181–193. <https://doi.org/10.1016/j.chemolab.2007.11.004>

Lavanya, A., Sindhuja, S., Gaurav, L., & Ali, W. (2023). Assessing the Performance of Python Data Visualization Libraries. *International Journal of Computer Engineering in Research Trends*, 10, 28–39. <https://doi.org/10.22362/ijcert/2023/v10/i01/v10i0104>

Mishra, P., Passos, D., Marini, F., Xu, J., Amigo, J. M., Gowen, A. A., Jansen, J. J., Biancolillo, A., Roger, J. M., Rutledge, D. N., & Nordon, A. (2022). Deep learning for near-infrared spectral data modelling: Hypes and benefits. *TRAC Trends in Analytical Chemistry*, 157(2), 116804. <https://doi.org/10.1016/j.trac.2022.116804>

Murray, S. (2013). *Interactive Data Visualization for the Web*. O'Reilly Media, Inc. <https://www.oreilly.com/library/view/interactive-data-visualization/9781449340223/>

Rodionova, O. Y., & Brereton, R. G. (2005). Chemometrics: Data Analysis for the Laboratory and Chemical Plant. *Journal of Analytical Chemistry*, 60, 994–996. <https://doi.org/10.1007/s10809-005-0223-6>

Shuaibov, O. K., Hrytsak, R. V., Minya, O. I., Malinina, A. A., Bilak, Yu. Yu., & Gomoki, Z. T. (2022). Spectroscopic diagnostics of overstressed nanosecond discharge plasma between zinc electrodes in air and nitrogen. *Journal of physical studies*, 26(2), 2501–2509. <https://doi.org/10.30970/jps.26.2501>

Tagliaferri, R., Ciaramella, A., Barone, F., Longo, G., & Milano, L. (2001). Neural Networks for Spectral Analysis of Unevenly Sampled Data. In A. J. Banday, S. Zaroubi, & M. Bartelmann (Eds.), *Mining the Sky* (pp. 386–390). Springer Nature. https://doi.org/10.1007/10849171_47

McKinney, W. (2017). *Python for Data Analysis* (2nd ed.). O'Reilly Media. <https://www.oreilly.com/library/view/python-for-data/9781491957653>

Young, C. D. (2001). *Computational Chemistry: A Practical Guide for Applying Techniques to Real World Problems*. John Wiley & Sons, Ltd. <https://doi.org/10.1002/0471220655>

References

Alsmeyer, F., & Marquardt, W. (2004). Automatic Generation of Peak-Shaped Models. *Applied Spectroscopy*, 58, 986–994. <https://opg.optica.org/as/abstract.cfm?uri=as-58-8-986>

Bilak, Yu., Kozubovskii, V., & Rol, M. (2023). Using the Fabry – Pérot interferometer as a spectral mask. *Journal of Physical Studies*, 27(1), 1402–1409. <https://doi.org/10.30970/jps.27.1402>

Butler, K. T., Davies, D. W., Cartwright, H., Isayev, O., & Walsh, A. (2018). Machine learning for molecular and materials science. *Nature*, 559, 547–555. <https://doi.org/10.1038/s41586-018-0337-2>

Butucea, C., Delmas, J.-F., Duffoy, A., & Hardy, C. (2021). Modeling Infrared Spectra: An Algorithm for an Automatic and Simultaneous Analysis. In B. Castanier, M. Cepin, D. Bigaud, & C. Berenguer (Eds.), *31st European Safety and Reliability Conference (ESREL), Angers, France* (pp. 3359–3366). https://doi.org/10.3850/978-981-18-2016-8_732-cd

Griffiths, P. (1983). Fourier Transform Infrared Spectrometry. *Science*, 222(4621), 297–302. <https://doi.org/10.1126/science.6623077>

Huang, L. (2014). Modeling and Simulation in Spectroscopic Study. *Spectral Analysis Review*, 2(3), 7–9. <https://doi.org/10.4236/sar.2014.23023>

Kozubovsky, V. R., & Bilak, Y. Y. (2022). Express Analysis of Gas Mixtures Using a Spectral Correlator Based on the Fabry – Perot Interferometer. *Journal of Applied Spectroscopy*, 89, 495–499. <https://doi.org/10.1007/s10812-022-01385-7>

Kriesten, E., Alsmeyer, F., Bardow, A., & Marquardt, W. (2008). Fully automated indirect hard modeling of mixture spectra. *Chemometrics and Intelligent Laboratory Systems*, 91(2), 181–193. <https://doi.org/10.1016/j.chemolab.2007.11.004>

Lavanya, A., Sindhuja, S., Gaurav, L., & Ali, W. (2023). Assessing the Performance of Python Data Visualization Libraries. *International Journal of Computer Engineering in Research Trends*, 10, 28–39. <https://doi.org/10.22362/ijcert/2023/v10/i01/v10i0104>

Mishra, P., Passos, D., Marini, F., Xu, J., Amigo, J. M., Gowen, A. A., Jansen, J. J., Biancolillo, A., Roger, J. M., Rutledge, D. N., & Nordon, A. (2022). Deep learning for near-infrared spectral data modelling: Hypes and benefits. *TRAC Trends in Analytical Chemistry*, 157(2), 116804. <https://doi.org/10.1016/j.trac.2022.116804>

Murray, S. (2013). *Interactive Data Visualization for the Web*. O'Reilly Media, Inc. <https://www.oreilly.com/library/view/interactive-data-visualization/9781449340223/>

Rodionova, O. Y., & Brereton, R. G. (2005). Chemometrics: Data Analysis for the Laboratory and Chemical Plant. *Journal of Analytical Chemistry*, 60, 994–996. <https://doi.org/10.1007/s10809-005-0223-6>

Shuaibov, O. K., Hrytsak, R. V., Minya, O. I., Malinina, A. A., Bilak, Yu. Yu., & Gomoki, Z. T. (2022). Spectroscopic diagnostics of overstressed nanosecond discharge plasma between zinc electrodes in air and nitrogen. *Journal of physical studies*, 26(2), 2501–2509. <https://doi.org/10.30970/jps.26.2501>

Tagliaferri, R., Ciaramella, A., Barone, F., Longo, G., & Milano, L. (2001). Neural Networks for Spectral Analysis of Unevenly Sampled Data. In A. J. Banday, S. Zaroubi, & M. Bartelmann (Eds.), *Mining the Sky* (pp. 386–390). Springer Nature. https://doi.org/10.1007/10849171_47

McKinney, W. (2017). *Python for Data Analysis* (2nd ed.). O'Reilly Media. <https://www.oreilly.com/library/view/python-for-data/9781491957653/>

Young, C. D. (2001). *Computational Chemistry: A Practical Guide for Applying Techniques to Real World Problems*. John Wiley & Sons, Ltd.
<https://doi.org/10.1002/0471220655>

Отримано редакцією журналу / Received: 26.12.24
Прорецензовано / Revised: 22.02.25
Схвалено до друку / Accepted: 06.03.25

Yurii BILAK, PhD (Phys. & Math.), Assoc. Prof.
ORCID ID: 0000-0001-5989-1643
e-mail: yuriy.bilak@uzhnu.edu.ua
Uzhgorod National University, Uzhgorod, Ukraine

INFORMATION SYSTEM BASED ON A COMPLEX MODEL USING MACHINE LEARNING FOR SPECTRAL ANALYSIS

The research is devoted to the design and development of an information system based on a complex model using machine learning methods to automate spectral analysis to increase the accuracy and speed of data processing. The history of the research is connected with the development of analytical methods in physics, chemistry and biology, where spectral analysis has traditionally played a key role. However, modern challenges, in particular the growth of data volumes and the need for automation, have stimulated the introduction of innovative methods based on artificial intelligence.

The relevance is due to the need to process large volumes of complex spectral data in real time, which is important for medicine, ecology, chemistry and other industries. Traditional analysis methods have limitations, so the use of machine learning is appropriate to increase the efficiency of the process.

The research focuses on the following issues: how to automate spectral data processing, how to ensure the integration of classical methods with machine learning, and how to increase the accuracy and scalability of the analysis. For this purpose, signal processing methods were applied, including noise filtering, smoothing, baseline correction, and peak analysis using derivatives and numerical integration. Machine learning was implemented through Random Forest models and neural networks adapted for predicting spectrum parameters.

The results showed that the developed system provides high accuracy and speed of spectral data analysis, interactive visualization of spectrum parameters, as well as the ability to integrate with other information platforms. This significantly simplifies analysis processes, reduces dependence on expert intervention, and increases productivity.

Research prospects include optimizing mathematical models for even greater accuracy, integration with IoT systems, and expanding the functionality for analyzing complex multidimensional spectra. This opens up opportunities for application in interdisciplinary projects, such as monitoring environmental changes or diagnosing medical conditions.

Keywords: *information system, artificial intelligence, spectral analysis, modeling, machine learning, intelligent data analysis.*

Автори заявляють про відсутність конфлікту інтересів. Спонсори не брали участі в розробленні дослідження; у зборі, аналізі чи інтерпретації даних; у написанні рукопису; в рішенні про публікацію результатів.

The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript; in the decision to publish the results.