

Молчанова М.О.

<https://orcid.org/0000-0001-9810-936X>

Хмельницький національний університет

Тимофієв А.А.

<https://orcid.org/0009-0008-8878-3383>

Хмельницький національний університет

Залуцька О.О.

<https://orcid.org/0000-0003-1242-3548>

Хмельницький національний університет

Мазурець О.В.

<https://orcid.org/0000-0002-8900-0650>

Хмельницький національний університет

ПІДХІД ДО ФОРМУВАННЯ ВИБІРКИ МІКРОЗОБРАЖЕНЬ ТЕКСТИЛЬНИХ МАТЕРІАЛІВ ДЛЯ НЕЙРОМЕРЕЖЕВОГО АНАЛІЗУ СИРОВИННОГО СКЛАДУ

У роботі сформовано та експериментально верифіковано відтворювану методологію профілювання вибірки мікроскопічних зображень текстильних матеріалів для нейромережевого аналізу сировинного походження. Дослідження виконано на контрольованому авторському датасеті, що містить 3107 мікрозображень, розподілених на матеріали натурального походження (1547) та матеріали синтетичного походження (1560), отриманих за стандартизованих умов зйомки.

Побудова ознакових представлень текстури та їх проєкційний аналіз у зниженому просторі підтвердили наявність вираженої геометричної структури даних і переважне просторове відокремлення класів матеріалів натурального та синтетичного походження, що свідчить про наявність дискримінативної інформації у структурно-текстурних властивостях мікрозображень. Структурний аналіз із виділенням підгруп показав, що вибірка має кілька стійких кластерів, узгоджених із класовими мітками; кількісно узгодженість кластерної структури з мітками для невіднесених до шуму спостережень становить $ARI = 0.4429$ та $NMI = 0.5606$, при високій чистоті кластерів ($Purity = 0.9975$), а неоднозначні спостереження концентруються в локальних зонах перекриття.

Кількісна перевірка відтворюваності висновків підтвердила стабільність роздільності класів у просторі сформованих представлень: за 5-кратною кросвалідацією отримано $ROC-AUC = 1.0000 \pm 0.0000$ та $F_1 = 0.9995 \pm 0.0010$. Контроль інформаційного витоку через близькі повтори шляхом групування зображень за $rHash$ та застосування $GroupKFold$ показав збереження метрик на тому самому рівні, що підтверджує, що отримані результати не є наслідком потрапляння дублів кадрів до різних підвбірок. Додатковий аналіз граничних прикладів дозволив локалізувати структурно схожі випадки, які формують зони перекриття, не змінюючи загального висновку про високу роздільність класів.

Отримані результати підтверджують, що нейромережеві та ознакові підходи здатні здійснювати коректне розмежування матеріалів натурального і синтетичного походження за мікроскопічними зображеннями, а запропонована схема профілювання та контролю відтворюваності є придатною як методична основа для подальшого розширення дослідження на багатокласову класифікацію текстильних відходів у задачах циркулярної економіки.

Ключові слова: текстильні матеріали, нейромережева класифікація, структурний аналіз, кластеризація, UMAP, HDBSCAN.



Постановка проблеми. Нейромережеве визначення сировинного складу текстильних матеріалів за мікроскопічними зображеннями ускладнюється тим, що показники якості класифікації істотно залежать від принципів формування вибірки та коректності протоколу її оцінювання [1, с. 15197]. У багатьох практичних наборах даних не регламентуються правила відбору кадрів, недостатньо контролюються репрезентативність текстур і внутрішньокласова варіативність, а серійні зображення одного й того самого зразка можуть потрапляти до різних підвбірок під час валідації [2, с. 6360]. Це створює ризик інформаційного витоку та завищення метрик за рахунок близьких повторів, коли фактично оцінюється здатність моделі відтворювати вже спостережені зразки, а не узагальнювати на нові матеріали [3, с. 100603].

Окремою складовою проблеми є вплив фотометричних факторів (яскравість, контраст, різкість), які навіть за контрольованих умов зйомки можуть систематично відрізнятися між класами як через фізичні властивості волокон і переплетення, так і через параметри оптики [4, с. 417]. Без профілювання цих характеристик і кількісної оцінки їхнього внеску неможливо обґрунтовано встановити, чи роздільність класів забезпечується макроструктурою текстури, чи зумовлюється спрощеними фотометричними сигналами, що знижує інтерпретованість і відтворюваність результатів [5, с. 63, 6, с. 495].

Отже, актуальною є формалізація підходу до формування вибірки мікрозображень текстильних матеріалів для нейромережевого аналізу сировинного складу, який включає контроль якості та узгодженості даних, виявлення близьких повторів, групові схеми розбиття під час валідації та перевірку стійкості висновків. Такий підхід є необхідною передумовою коректної інтерпретації експериментальних результатів і подальшого перенесення розроблених рішень на практичні задачі сортування текстильних відходів у контексті циркулярної економіки.

Аналіз останніх досліджень і публікацій. В українських публікаціях останніх років проблематика текстильних відходів здебільшого розглядається через призму формування потоків відходів, їхнього повторного використання та прогнозування перероблення як передумов для впровадження циркулярних практик. Зокрема, у роботі з аналізу даних для прогнозування перероблення текстильних відходів в Україні акцентовано, що керованість рішень у ланцюгу переро-

блення потребує кількісних моделей і даних про склад та структуру відходів, а не лише описових підходів [7]. Водночас дослідження щодо стану та перспектив перероблення й використання текстильних виробів в Україні підкреслюють обмеженість інфраструктури сортування і дефіцит інструментів, що забезпечують стабільну ідентифікацію матеріалу у змішаних потоках [8, с. 38]. Ці роботи формують предметну мотивацію: без надійної ідентифікації сировинного складу неможливо забезпечити технологічно коректні маршрути повторного використання та перероблення.

У сегменті цифрових методів контролю матеріалів українські публікації частіше фокусуються на задачах візуального контролю якості (дефекти, неоднорідності, технологічні артефакти), що методологічно близьке до вашої постановки як аналіз макроструктури та текстури зображення. Прикладом є робота про автоматизований контроль дефектів матеріалів легкої промисловості [9, с. 353], у якій продемонстровано придатність нейромережевого аналізу зображень для виробничих сценаріїв контролю поверхні. Водночас саме задачі визначення сировинного складу (натуральні/синтетичні, суміші) у відкритих українських публікаціях трапляються рідше, що підсилює актуальність робіт, орієнтованих на формалізацію вибірки, профілювання даних і коректні протоколи оцінювання.

Для промислового сортування текстильних відходів значний масив робіт базується на спектральних вимірюваннях, насамперед NIR, як на найбільш практичному не руйнівному методі ідентифікації волокон і сумішей. Показовою є стаття про NIR-орієнтоване сортування поліестер-вмісних текстильних відходів, де оцінювання робить спробу переходу від лабораторної реєстрації спектрів до сценарію промислової лінії сортування [10]. У близькому напрямі робота з поєднання NIR і глибинного навчання для класифікації постспоживчих текстильних відходів [11, с. 2475]. Загальна методологічна риса цих підходів – використання спектральних ознак, але такі дослідження залежать від спеціалізованих сенсорів і калібрування, що ускладнює масштабування у сценарії сортування на рівні пунктів збору або первинної інспекції.

Постановка завдання. Метою статті є наукове обґрунтування та експериментальна верифікація підходу до формування вибірки мікроскопічних зображень текстильних матеріалів, придатної для нейромережевого визначення сировинного

складу, з доведенням того, що класифікаційна роздільність забезпечується структурно-текстурними ознаками матеріалу за коректного протоколу оцінювання та контролю повторів.

У межах цієї мети розглядається бінарна постановка задачі класифікації з двома класами: текстильні матеріали натурального походження та текстильні матеріали синтетичного походження.

Виклад основного матеріалу. Для доведення коректності класифікації сировинного походження текстильних матеріалів за мікроскопічними зображеннями виконано послідовне профілювання датасету та верифікацію стійкості отриманих висновків. Методологія організована як ланцюжок з п'яти кроків, що забезпечують: формальну підготовку вибірки, контроль якості та узгодженості даних, кількісний опис фотометричних і різкісних характеристик, побудову числових представлень текстури, аналіз структури вибірки з виділенням підгруп, а також перевірку відтворюваності висновків із контролем повторів. На рисунку 1 подано узагальнену схему експерименту, де вхідними даними є датасет мікрозображень із класовими мітками, а вихідними даними є профіль датасету, зведені статистики та кількісні метрики роздільності класів.

Крок 1. Підготовка та верифікація даних. На цьому кроці виконується індексація всіх зображень датасету з формуванням єдиної таблиці спостережень, у якій кожному файлу однозначно відповідають шлях до зображення та класова мітка [12, с. 321]. Далі встановлюється технічна придатність датасету до обчислювального аналізу: кожне зображення перевіряється на читабельність і коректність формату, фіксуються базові атрибути кадру (розмір, співвідношення сторін), а також визначається співвідношення кількості зображень між класами [13]. Очікуваним результатом кроку є верифікована індексована вибірка без пошкоджених файлів і з однозначно заданими мітками, яка є вихідною основою для подальшого профілювання.

Крок 2. Обчислення базових профільних показників. На цьому кроці виконується кількісний опис фотометричних і різкісних властивостей зображень у розрізі класів. Для кожного зображення визначається середня яскравість як середня інтенсивність пікселів після приведення до єдиного масштабу, контраст як стандартне відхилення інтенсивності, а різкість як варіація другого похідного компонента яскравості, що відображає насиченість кадру високочастотними деталями. Для кожного класу обчислюються узагальнювальні статистики цих показників у вигляді середнього значення, стандартного відхилення та медіани. Очікуваним результатом кроку є профіль датасету за трьома показниками (яскравість, контраст, різкість), який дозволяє встановити наявність або відсутності систематичних зсувів між класами та надалі коректно інтерпретувати результати класифікаційної роздільності.

Крок 3. Побудова ознакових представлень. На цьому кроці кожне зображення переводиться у числове представлення, яке кодує структурно-текстурні властивості поверхні матеріалу та придатне для подальшого аналізу в просторі ознак. Побудовані ознаки формують матрицю представлень, де кожному зображенню відповідає вектор фіксованої розмірності. На цьому ж кроці забезпечується узгодженість масштабу ознак та технічна коректність матриці (відсутність пропусків, нечислових значень і вироджених векторів). Очікуваним результатом кроку є матриця ознак, яка є основою для аналізу структури вибірки та кількісної оцінки роздільності класів.

Крок 4. Виявлення структури та підгруп даних. На цьому кроці виконується структурний аналіз вибірки в просторі ознакових представлень. Результат подається у вигляді карти розміщення зображень у зниженому просторі та у вигляді розбиття вибірки на підгрупи, що відображають природну організацію даних [14, с. 315]. Узгодженість виділених підгруп із класовими мітками оцінюється кількісно через метрики відповідності



Рис. 1. Кроки аналізу вибірки мікрозображень текстильних матеріалів для нейромережевого аналізу сировинного складу

розбиття класам, що характеризують ступінь узгодження структури даних із заданими класами. Очікуваним результатом кроку є встановлення характеру структури вибірки та рівня узгодженості підгруп із класами, що на пряму відображає наявність або відсутність структурної роздільності за сировинним походженням.

Крок 5. Перевірка відтворюваності висновків. На цьому кроці підтверджується стійкість отриманих висновків і виключається можливість завищення метрик через повтори [15, с. 42]. Виконується контроль близьких повторів зображень шляхом групування майже ідентичних кадрів та застосування протоколу оцінювання, у якому елементи однієї групи не можуть одночасно потрапляти до навчальної і тестової частин. Після цього обчислюються метрики роздільності класів і порівнюються з результатами оцінювання без урахування груп повторів. Очікуваним результатом кроку є підтвердження того, що роздільність класів зберігається за контрольованого протоколу перевірки та не пояснюється витоком інформації між підвибірками.

Експеримент виконано на авторському датасеті мікроскопічних зображень текстильних матеріалів «Natural and synthetic fabrics dataset» [16], сформованому з контролем умов зйомки. Зйомка здійснювалася в одному приміщенні за штучного освітлення, а також за стандартизованої геометрії зйомки: зображення отримано з однакової відстані до мікроскопа, що зменшує варіативність, не пов'язану зі структурою матеріалу, і підвищує коректність порівнянь між класами. Датасет містить 3107 мікрозображень, розподілених на два класи за сировинним походженням: 1547 зображень належать до класу матеріалів натурального походження, 1560 – до класу матеріалів синтетичного походження.

У результаті підготовки та верифікації (виконання Крок 1) підтверджено індексовану вибірку з 3107 мікроскопічних зображень текстильних матеріалів, для яких однозначно задано класову належність до матеріалів натурального або синтетичного походження. Усі зображення є читабельними, мають коректні формати та узгоджені базові параметри, що підтверджує технічну придатність датасету до подальшого аналізу. Співвідношення кількості зображень між класами є близьким до рівномірного, що виключає домінування одного класу як джерело потенційного зсуву результатів [17, с. 85].

У межах Кроку 2 (обчислення базових профільних показників) виконано кількісне профілювання мікроскопічних зображень за трьома узгодженими характеристиками: яскравість (brightness) як середнє значення інтенсивності, контраст (contrast) як стандартне відхилення інтенсивності та різкість (sharpness) як дисперсія оператора Лапласа [18, с. 5], що відображає насиченість кадру високочастотними деталями (рисунок 2). Метою цього етапу було зафіксувати класові відмінності у базових фотометричних і різкісних властивостях зображень та оцінити їх потенційні впливи на подальші висновки щодо структурної роздільності класів.

На рисунку 2а подано розподіли яскравості: для матеріалів синтетичного походження характерні вищі значення яскравості, тоді як для матеріалів натурального походження розподіл зміщений у бік нижчих значень; при цьому наявна зона перекриття розподілів. На рисунку 2б наведено розподіли контрасту: для синтетичного класу спостерігається зсув у бік більших значень контрасту, тоді як для натурального класу розподіл є компактнішим; перекриття розподілів зберігається. На рисунку 2в показано розподіли різкості: обидва

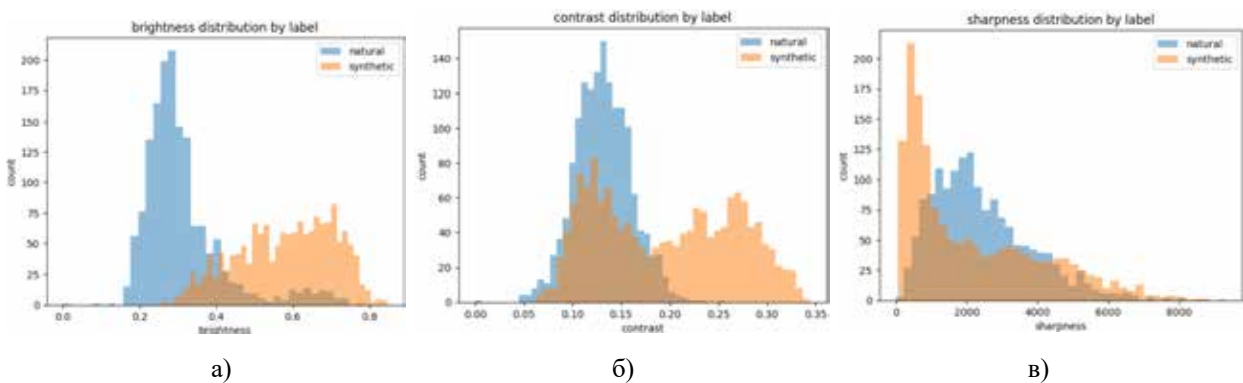


Рис. 2. Розподіли базових профільних показників зображень за класами: а) за яскравістю, б) за контрастом, в) за різкістю

класи характеризуються значною варіативністю та довгими хвостами розподілу; водночас медіанні значення різкості відрізняються між класами, що показує відмінності в рівні деталізації мікроструктурних елементів.

Кількісні зведені статистики для кожного показника (mean, STD, median) наведено у таблиці 1.

Узагальнено встановлено, що матеріали синтетичного походження демонструють вищі середні значення яскравості та контрасту, тоді як для показника різкості спостерігаються відмінності як у медіані, так і в дисперсії. Наявність перекриття розподілів за всіма трьома показниками підтверджує, що базові фотометричні та різкісні характеристики не забезпечують повної класової роздільності та розглядаються в роботі як компонент профілю датасету, необхідний для коректної інтерпретації подальших результатів аналізу структурно-текстурних представлень [19, с. 941].

Результати Кроку 3. Побудова ознакових представлень. У результаті побудови ознакових представлень кожне мікроскопічне зображення було відображене у векторний простір ознак сталої розмірності, придатний для порівняння та аналізу структури вибірки. Для візуальної перевірки інформативності отриманих представлень виконано проєкцію цього простору у двовимірний вимір [20]. На рисунку 3а показано, що у сформованому ознаковому просторі спостерігається виражена геометрична організація даних: зразки утворюють компактні області, а класи матеріалів натурального та синтетичного походження демонструють переважне просторове відокремлення. Наявність обмежених зон перекриття відповідає граничним спостереженням, що є очікуваним для реальних даних із внутрішньокласовою варіативністю та структурно близькими текстурами. Отриманий результат підтверджує, що побудовані ознакові представлення несуть дискримінативну інформацію, релевантну задачі розмежування

матеріалів натурального і синтетичного походження за мікроскопічними зображеннями.

Результати Кроку 4. Виявлення структури та підгруп даних. Для встановлення внутрішньої організації вибірки у просторі ознакових представлень виконано кластеризацію зразків та оцінено узгодженість отриманих підгруп із класовими мітками матеріалів натурального та синтетичного походження. На рисунку 3б наведено результат розбиття у проєктованому просторі, який демонструє наявність стійких компактних підгруп, що відповідають природній неоднорідності текстур у межах кожного класу. Кількісно узгодженість кластерної структури з класовими мітками для невіднесених до шуму спостережень становить $ARI = 0.4429$ та $NMI = 0.5606$, що підтверджує наявність статистично вираженої відповідності між природними підгрупами даних та класовою належністю. При цьому чистота кластерів є високою: переважна більшість виділених підгруп містить зразки одного класу, а класифікаційно неоднозначні спостереження концентруються у локальних зонах перекриття, які відповідають граничним випадкам структурної схожості. Отриманий результат підтверджує, що класи матеріалів розрізняються не лише на рівні глобального відокремлення у просторі ознак, а й через формування підгруп, що відображають внутрішньокласову варіативність текстур.

Результати Кроку 5. Перевірка відтворюваності висновків. Для підтвердження стійкості висновків щодо роздільності класів виконано кількісне оцінювання за 5-кратною кросвалідацією. Отримано $ROC-AUC = 1.0000 \pm 0.0000$ та $F_1 = 0.9995 \pm 0.0010$, що доводить практично повну роздільність класів у просторі сформованих представлень. Для контролю можливого інформаційного витоку через близькі повтори зображення згруповано за pHash та застосовано GroupKFold із розбиттям за групами повторів; за цієї схеми метрики

Таблиця 1

Зведені статистики базових профільних показників мікрозображень

Показник	Матеріали натурального походження	Матеріали синтетичного походження
Яскравість: середнє	0.313048	0.581429
Яскравість: σ	0.107862	0.125281
Яскравість: медіана	0.285904	0.594957
Контраст: середнє	0.131512	0.195734
Контраст: σ	0.029109	0.070952
Контраст: медіана	0.131107	0.195707
Різкість: середнє	2440.362685	2171.336693
Різкість: σ	1403.128190	1907.737990
Різкість: медіана	2166.240694	1433.420147

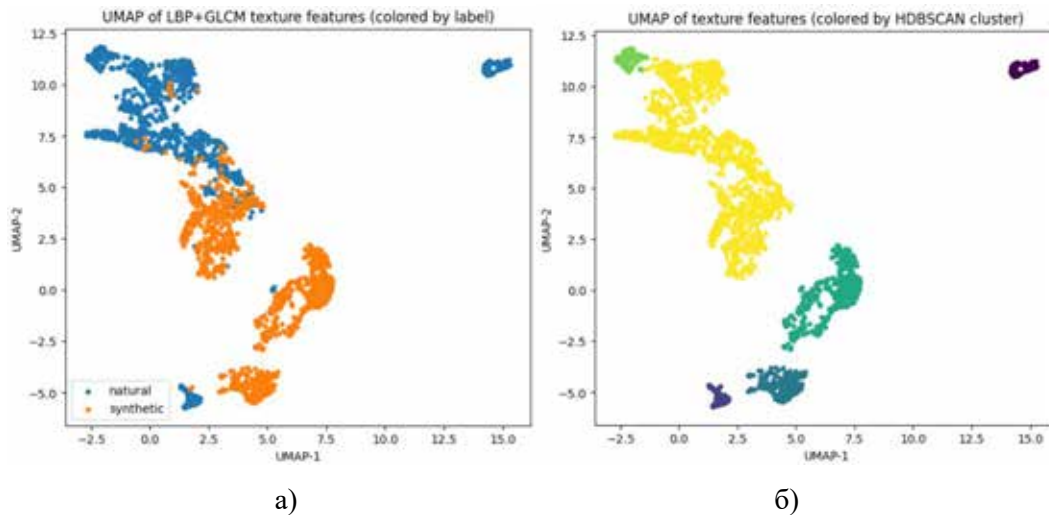


Рис. 3. Проекція ознакових представлень текстури у двовимірний простір та виявлення підгруп вибірки: а) UMAP текстурних особливостей LBP+GLCM (забарвлених за міткою), б) UMAP текстурних ознак (забарвлених кластером HDBSCAN)



Рис. 4. Граничні спостереження вибірки за критерієм близькості до центроїда протилежного класу

зберігаються на тому самому рівні ($ROC-AUC = 1.0000 \pm 0.0000$, $F_1 = 0.9995 \pm 0.0010$), що підтверджує відсутність завищення якості за рахунок дублів кадрів. Додатково виконано аналіз граничних спостережень як зразків, що мають мінімальний відступ до центроїда протилежного класу, та сформовано їх візуальну добірку для інтерпретації причин неоднозначності (рисунок 4).

На рисунку 4 візуально наведено підмножину зображень, які за сформованими поданнями розташовані найближче до центроїда протилежного класу (ambiguous samples). Такі кадри формують інтерпретовану групу «складних» випадків, що відповідають локальним зонам перекриття у просторі ознак і відображають структурну схожість мікротекстур між класами. Використання цієї добірки дозволяє якісно підтвердити, що похибки

та потенційні змішані області зосереджені в обмеженій кількості граничних прикладів і не визначають загальну роздільність класів.

Висновки. У роботі сформовано відтворювану схему профілювання датасету мікрозображень текстильних матеріалів для задачі розмежування матеріалів натурального та синтетичного походження, яка охоплює підготовку вибірки, контроль якості та узгодженості даних, побудову числових представлень, аналіз структури вибірки з виділенням підгруп і перевірку відтворюваності отриманих висновків. Профілювання яскравості, контрасту та різкості виявило міжкласові зсуви, однак перекриття розподілів підтвердило, що ці показники не забезпечують повної роздільності класів і мають розглядатися як контрольні характеристики датасету при інтерпретації результа-

тів. Побудовані ознакові представлення забезпечили виражену структурну організацію вибірки, а узгодженість виділених підгруп із класовими мітками підтверджено значеннями $ARI = 0.4429$, $NMI = 0.5606$ та $Purity = 0.9975$. Кількісна перевірка продемонструвала стабільно високі показ-

ники роздільності класів (5-кратна кросперевірка: $ROC-AUC = 1.0000 \pm 0.0000$, $F_1 = 0.9995 \pm 0.0010$), а додатковий контроль повторів із груповим розбиттям за перцептивним хешем не змінив значень метрик, що виключає завищення результатів через близькі дублікати зображень.

Список літератури:

1. Seçkin M., Seçkin A.Ç., Demircioglu P., Bogrekcı I. FabricNET: a microscopic image dataset of woven fabrics for predicting texture and weaving parameters through machine learning. *Sustainability*. 2023. Vol. 15, № 21. P. 15197. DOI: <https://doi.org/10.3390/su152115197>.
2. Zhang Y., Han T., Wei B., Hao K., Gao L. A spatial-spectral adaptive learning model for textile defect images recognition with few labeled data. *Complex & Intelligent Systems*. 2023. Vol. 9, № 6. P. 6359–6371. DOI: <https://doi.org/10.1007/s40747-023-01070-y>.
3. Kumar K.S., Bai M.R. LSTM based texture classification and defect detection in a fabric. *Measurement: Sensors*. 2023. Vol. 26. P. 100603. DOI: <https://doi.org/10.1016/j.measen.2022.100603>.
4. Мазурець О.В., Жарновський О.В., Гладун О.В., Собко О.В. Нейромережеве виявлення фейкових зображень людей. *Науковий журнал «Вісник Хмельницького національного університету» серія: Технічні науки*. Хмельницький. 2025. № 5, Т. 1. С. 416–422. DOI: <https://doi.org/10.31891/2307-5732-2025-357>.
5. Hanbay K. Classification of Circular Knitting Fabric Defects Using MobileNetV2 Model. *Türk Doğa ve Fen Dergisi*. 2023. Vol. 12, № 4. P. 63–68. DOI: <https://doi.org/10.46810/tdfd.1327971>
6. Похитун А.В., Мазурець О.В., Дидо Р.А., Молчанова М.О. Програмна архітектура для нейромережевого виявлення модифікованих фотографій обличчя людей. *Науковий журнал «Вісник Хмельницького національного університету» серія: Технічні науки*. Хмельницький. 2025. № 3, Т. 2. С. 493–500. DOI: <https://doi.org/10.31891/2307-5732-2025-353-68>.
7. Bukhantsova L., Zacharkevich O., Lushevskaya O., Krasniuk L., Koshevko Y., Ditkovska O., Shvets G. Data analysis for the prediction of textilewaste recycling in Ukraine. *Technická univerzita v Liberci*. 2024. DOI: <https://doi.org/10.15240/tul/008/2024-2-009>.
8. Romaniuk I., Kurushkina A. Current state and prospects of the development of recycling and use of textile products in Ukraine. *Fashion Industry*. 2025. № 2. P. 38–46. DOI: <https://doi.org/10.30857/2706-5898.2024.2.2>
9. Горященко С., Солом'яний Є. Особливості автоматизованого контролю дефектів матеріалів легкої промисловості. *Науковий журнал «Вісник Хмельницького національного університету» серія: Технічні науки*. Хмельницький. 2025. №. 349, № 2. С. 353–357. DOI: <https://doi.org/10.31891/2307-5732-2025-349-51>.
10. Becker A., Datko A., Kroell N., Küppers B., Greiff K., Gries T. Near-infrared-based sortability of polyester-containing textile waste. *Resources, Conservation and Recycling*. 2024. Vol. 206. P. 107577. DOI: <https://doi.org/10.1016/j.resconrec.2024.107577>.
11. Riba J.R., Cantero R., Riba-Mosoll P., Puig R. Post-consumer textile waste classification through near-infrared spectroscopy, using an advanced deep learning approach. *Polymers*. 2022. Vol. 14, № 12. P. 2475. DOI: <https://doi.org/10.3390/polym14122475>.
12. Derzhak V.V., Mazurets O.V. Approach to image preprocessing for household waste classification in circular economy. Resource-Saving Technologies of Apparel, Textile & Food Industry. *Proceedings of International Scientific and Practical Conference*. November 20, 2025. Khmelnytskyi, Ukraine. P. 319–323. URL: <https://elar.khmnu.edu.ua/items/ff44524d-5231-451f-a573-879d96baa479>.
13. Ma C., Shi Z., Lu Z., Xie S., Chao F., Sui Y. A survey on image quality assessment: Insights, analysis, and future outlook. *arXiv preprint arXiv:2502.08540*. 2025. DOI: <https://doi.org/10.48550/arXiv.2502.08540>.
14. Kazmirchuk Y.M., Mykytiuk M.O., Molchanova M.O., Mazurets O.V. Algorithmic scaling of textile prints for serial multicolor printing with palette reproduction using neural networks. Resource-Saving Technologies of Apparel, Textile & Food Industry. *Proceedings of International Scientific and Practical Conference*. November 20, 2025. Khmelnytskyi, Ukraine. P. 314–318. URL: <https://elar.khmnu.edu.ua/items/d24e0d7e-ed27-48d1-9d5c-98d226ef6cf6>.
15. Barz B., Denzler J. Do we train on test data? Purging CIFAR of near-duplicates. *Journal of Imaging*. 2020. Vol. 6, № 6. P. 41. DOI: <https://doi.org/10.3390/jimaging6060041>.
16. Natural and Synthetic Fabrics Dataset. *Kaggle*. URL: <https://www.kaggle.com/datasets/olyarnncnn/natural-and-synthetic-fabrics-dataset>.
17. Sobko O., Mazurets O., Molchanova M., Krak I., Barmak O. Method for analysis and formation of representative text datasets. *CEUR Workshop Proceedings*. 2025. Vol. 3899. P. 84–98. URL: <https://ceur-ws.org/Vol-3899/paper9.pdf>.

18. Jamil S. Review of image quality assessment methods for compressed images. *Journal of Imaging*. 2024. Vol. 10, № 5. P. 113. DOI: <https://doi.org/10.3390/jimaging10050113>.

19. Залуцька О.О., Мазурець О.В. Нейромережева технологія аналізу макроструктури тканин для інтелектуальних виробничих систем легкої промисловості. Інформаційні технології і автоматизація. *Матеріали XVIII міжнародної науково-практичної конференції*. 30–31 жовтня 2025 р. Одеса, ОНТУ. 2025. С. 941–943. URL: <https://elar.khmnmu.edu.ua/items/82814de5-ab5f-4a6f-9616-7936a6d7f5f4>.

20. Zhou Y., Tang C., Xiao X., Zhan X., Wang T., Xiao G., Xu L. Dimensionality reduction for visualizing spatially resolved profiling data using SpaSNE. *GigaScience*. 2025. Vol. 14. P. g1af002. DOI: <https://doi.org/10.1093/gigascience/g1af002>.

Molchanova M.O., Timofiev A.A., Zalutska O.O., Mazurets O.V. APPROACH TO FORMING A MICROIMAGES SET OF TEXTILE MATERIALS FOR NEURAL NETWORK ANALYSIS OF RAW MATERIAL COMPOSITION

The reproducible methodology for profiling a dataset of microscopic images of textile materials for neural-network-based analysis of material origin has been developed and experimentally validated. The study was conducted on a controlled author-created dataset containing 3,107 micro-images, divided into materials of natural origin (1,547) and materials of synthetic origin (1,560), acquired under standardized imaging conditions.

The construction of texture feature representations and their projection analysis in a reduced-dimensional space confirmed a pronounced geometric structure of the data and predominant spatial separation between the classes of materials of natural and synthetic origin, indicating the presence of discriminative information in the structural-textural properties of the micro-images. The structural analysis with subgroup identification showed that the dataset contains several stable clusters consistent with the class labels; quantitatively, the agreement between the cluster structure and the labels for non-noise observations is $ARI = 0.4429$ and $NMI = 0.5606$, with high cluster purity ($Purity = 0.9975$), while ambiguous observations are concentrated in local overlap regions.

The quantitative reproducibility assessment confirmed the stability of class separability in the space of the constructed representations: under 5-fold cross-validation, $ROC-AUC = 1.0000 \pm 0.0000$ and $F1 = 0.9995 \pm 0.0010$ were obtained. Leakage control for near-duplicate images by grouping samples using pHash and applying GroupKfold preserved the metrics at the same level, confirming that the reported results are not driven by duplicates being split across different subsets. An additional analysis of borderline samples localized structurally similar cases forming overlap zones without changing the overall conclusion regarding high class separability.

The obtained results confirm that neural-network-based and feature-based approaches can provide correct discrimination between materials of natural and synthetic origin from microscopic images, and that the proposed profiling and reproducibility control scheme is suitable as a methodological basis for extending the study to multi-class classification of textile waste within circular-economy applications.

Keywords: *textile materials, neural network classification, structural analysis, clustering, UMAP, HDBSCAN.*

Дата першого надходження статті до видання: 16.01.2026

Дата прийняття статті до друку після рецензування: 12.02.2026

Дата публікації (оприлюднення) статті: 08.04.2026