УДК 004.056.55:004.93:343.98
DOI: 10.31673/2409-7292.2025.028437

Фединишин Т.О., Партика О.О.

ПОРІВНЯЛЬНЕ ДОСЛІДЖЕННЯ ВЕКТОРНИХ БАЗ ДАНИХ НА ОСНОВІ ANN ДЛЯ ШВИДКОГО ТА ТОЧНОГО РОЗПІЗНАВАННЯ ОБЛИЧ У ЦИФРОВІЙ КРИМІНАЛІСТИЦІ

Стрімке зростання обсягів біометричних даних та зростаюча потреба в точній верифікації особи у сфері цифрової криміналістики зумовили необхідність створення масштабованих та ефективних систем пошуку за зображенням обличчя. У цій статті представлено порівняльне дослідження п'яти алгоритмів векторного пошуку — HNSW, Faiss, Annoy, PyNNDescent та Nearest Neighbors — для задач ідентифікації облич на основі векторних представлень (ембеддингів). Експеримент було спроєктовано з урахуванням умов, наближених до реальних судово-експертних сценаріїв, із фокусом на таких основних метриках оцінювання, як точність у Top-1, розподіл коефіцієнтів подібності та час обробки запитів. Усі протестовані методи продемонстрували високу точність (понад 91%), однак між ними було зафіксовано суттєві відмінності щодо впевненості у збігах та швидкодії. Faiss показав найвищі показники подібності, що свідчить про кращу точність пошуку, хоча й потребував значно більше обчислювальних ресурсів. Натомість алгоритми HNSW і PyNNDescent забезпечили майже миттєву обробку запитів із конкурентоспроможною точністю, проте з вищою варіативністю якості результатів. Annoy виявився компромісним рішенням, яке поєднує високу точність із низькою затримкою. Отримані результати підкреслюють важливі компроміси між точністю, впевненістю та ефективністю, надаючи цінні орієнтири для вибору оптимальних технологій векторного пошуку в системах розпізнавання обличчя у криміналістиці. Крім того, у межах дослідження запропоновано відтворювану методику бенчмаркінгу, яку можна застосовувати для подальшого оцінювання біометричних інструментів пошуку в правоохоронній та безпековій сферах.

Ключові слова: векторний пошук, розпізнавання обличчя, цифрова криміналістика, HNSW, Faiss, Annoy, PyNNDescent, біометрична ідентифікація.

Вступ

Експоненціальне зростання обсягів мобільних та даних з відео спостереження створює як нові можливості, так і суттєві виклики для цифрової криміналістики, особливо в контексті ідентифікації та профілювання потенційно небезпечних осіб. Пошук за зображенням обличчя — ключова задача в судово-експертних розслідуваннях — вимагає швидких, масштабованих і точних механізмів, здатних ефективно порівнювати запитувані вектори ознак із великими біометричними базами даних. Традиційний повний перебір стає обчислювально непридатним зі зростанням обсягів даних, що зумовлює потребу у використанні алгоритмів наближеного пошуку найближчих сусідів (ANN) і векторних баз даних, зокрема HNSW, Annoy, Faiss та інших.

Останні досягнення в галузі розпізнавання облич, засновані на глибоких нейронних мережах, суттєво підвищили точність векторного представлення ознак [1]. Проте ефективність таких моделей у реальних умовах значною мірою залежить від продуктивності базового векторного пошуку, який має забезпечувати швидкий доступ до результатів без втрати якості пошуку. Як зазначено в [2], системи на основі ANN забезпечують збалансований компроміс між швидкістю та точністю, однак їхня надійність у криміналістичних сценаріях із високими вимогами до ідентифікації особи досі недостатньо вивчена.

Бібліотеки векторного пошуку, такі як Faiss [3] і Annoy [4], набули широкого застосування в комерційних системах. Faiss, розроблений у Facebook AI Research, підтримує ефективне індексування щільних векторів ознак, тоді як Annoy, орієнтований на сценарії з великою кількістю читань, демонструє високу продуктивність у системах з обмеженими ресурсами пам'яті. Перспективним підходом також є графи HNSW (Hierarchical Navigable Small World), що забезпечують логарифмічну складність пошуку і високу повноту, перевершуючи традиційні KD-дерева та методи на основі LSH у сучасних бенчмарках [5].

Попри широку популярність зазначених рішень, порівняльні дослідження цих систем у криміналістичних задачах — де критично важливою є точність у Top-1 та впевненість у результатах — залишаються обмеженими. Окрім того, нещодавні дослідження, що стосуються

алгоритмів PyNNDescent і точного пошуку NearestNeighbors, свідчать про те, що сучасні наближені методи можуть досягати майже такої ж точності, як і повний перебір, із суттєво нижчими обчислювальними витратами [6].

У цій статті представлено порівняльне дослідження п’яти поширених підходів до організації векторного пошуку — HNSW, Faiss, Annoy, PyNNDescent та Nearest Neighbors — у межах контрольованої задачі криміналістичного пошуку за обличчям. Основна увага приділяється оцінці точності, розподілу коефіцієнтів подібності та часу обробки запитів з метою визначення доцільності використання цих методів у критичних, часовочутливих криміналістичних сценаріях.

Мета та завдання дослідження

Метою даного дослідження є порівняльна оцінка ефективності сучасних алгоритмів векторного пошуку для задач криміналістичного розпізнавання облич на основі векторних представлень (ембеддингів). Основна увага приділяється аналізу точності ідентифікації, рівня подібності між векторами запиту та результату, а також часу обробки пошукових запитів. Завдання дослідження включають побудову векторних індексів для п’яти алгоритмів (HNSW, Faiss, Annoy, PyNNDescent, Nearest Neighbors), запуск серії пошукових експериментів та обробку отриманих результатів з використанням кількісних метрик. Додатково передбачено створення уніфікованого методичного підходу до оцінювання векторних баз даних у цифровій криміналістиці.

Аналіз літератури

Пошук облич та пошук за векторною подібністю стали ключовими елементами сучасної цифрової криміналістики, зокрема в контексті аналізу мобільних даних та виявлення загроз [26]. Перехід до векторних представлень (ембеддингів) та використання масштабованих векторних баз даних сприяв активному розвитку досліджень, спрямованих на підвищення точності й ефективності пошуку в неструктурованих біометричних масивах даних.

Однією з провідних напрямків розвитку є методи наближеного пошуку найближчих сусідів (ANN). Такі алгоритми пропонують масштабовану альтернативу повному перебору, забезпечуючи швидкий пошук серед мільйонів векторів ознак. Серед найбільш досліджуваних технік ANN — ієрархічні графи навігації малого світу (HNSW), запропоновані Малковим та Яшуніним [1], які демонструють логарифмічний час запиту та високу точність на даних великої розмірності. Висока продуктивність HNSW зробила цей підхід базовим орієнтиром як для академічних, так і для промислових бенчмарків.

Іншою провідною бібліотекою є Faiss, розроблена в рамках Facebook AI Research. Вона використовує техніки квантування та прискорення за допомогою GPU для підтримки пошуку за подібністю у масштабі мільярдів об’єктів із субсекундною затримкою [2]. Ряд досліджень [3][4] підтверджує ефективність Faiss у задачах пошуку зображень та облич, зокрема в умовах, наближених до криміналістичних систем, за умови відповідної оптимізації.

Інші бібліотеки ANN, зокрема Annoy [5], PyNNDescent [6] та реалізація точного пошуку найближчих сусідів у Scikit-learn [7], також активно досліджуються з точки зору компромісів між ефективністю використання пам’яті, швидкістю побудови індексу та точністю пошуку. Порівняльні аналізи, зокрема ті, що проведені Аумюллером і співавт. [8], а також Чжаном і колегами [9], показують, що хоча Annoy оптимізований для сценаріїв із великою кількістю операцій читання, Faiss і HNSW демонструють стабільний баланс продуктивності та точності для динамічних наборів даних із жорсткими обмеженнями щодо затримки.

У біометричній криміналістиці пошук облич на основі векторних ембеддингів витіснив традиційні підходи порівняння зображень завдяки кращій стійкості до змін пози, освітлення та часткового закриття обличчя. Сучасні моделі ембеддингів, такі як ArcFace [10], CosFace [11] та CurricularFace [12], дозволяють кодувати риси обличчя у вектори фіксованої довжини, придатні для високошвидкісного пошуку. Ці представлення активно застосовуються у реальних задачах — від криміналістичного аналізу місця події [13] до повторної ідентифікації

осіб [14] та аналізу даних мобільних пристроїв [15]. Інтеграція ANN-пошуку у криміналістичні робочі процеси є особливо актуальною для постфактум-аналізу, коли слідчі мають ідентифікувати осіб на основі часткових дампів мобільних даних або відеозаписів із камер спостереження. У низці досліджень [16][17] запропоновано багаторівневі пошукові схеми, які поєднують ANN з імовірнісним аналізом або графовими методами для підвищення надійності результатів у випадках низької якості або фрагментованості даних.

Останні дослідження також звертають увагу на інтерпретованість і надійність ANN-систем у юридичному контексті. Зокрема, Гоял і співавт. [18] та Соммерс із колегами [19] висловлюють занепокоєння щодо використання "чорних скриньок" у критичних задачах та наголошують на необхідності забезпечення прозорості, довірчих індикаторів і можливості аудиту результатів пошуку.

Ще одним важливим напрямком є мобільна криміналістика та аналітика на периферійних пристроях, де особливо важливо використовувати полегшені реалізації ANN для локального пошуку та попереднього сортування даних. Роботи Ванга [20] та Бхаттачар'ї [21] демонструють можливості оптимізованого ANN-індексування на обчислювально обмежених пристроях, що підкреслює необхідність у низьколатентних та енергоефективних рішеннях для використання в польових умовах.

Попри суттєві технологічні досягнення, наразі бракує досліджень, що здійснюють системне порівняння ефективності цих векторних алгоритмів саме в контексті криміналістичного розпізнавання облич. У цій роботі робиться спроба заповнити цю прогалину шляхом експериментального оцінювання точності, продуктивності та масштабованості п’яти поширених методів векторного пошуку (HNSW, Faiss, Annoy, PyNNDescent, Nearest Neighbors) з використанням наближених до реальності запитів у криміналістичних сценаріях.

Матеріали та методи

Експериментальне дослідження структуровано з метою комплексної оцінки ефективності сучасних алгоритмів векторного пошуку у завданнях криміналістичного розпізнавання облич. Необхідність проведення такого аналізу зумовлена зростаючими вимогами до точності, масштабованості та швидкодії систем біометричної ідентифікації, особливо в умовах обробки великих масивів даних, зібраних з мобільних пристроїв у цифровій криміналістиці.

На першому етапі було підготовлено датасет, сформований із зображень облич, які були перетворені у вектори ознак (ембеддинги) за допомогою попередньо натренованої нейронної мережі. Ці ембеддинги були очищені від аномалій, приведені до єдиного формату, стандартизовані за розмірністю та збережені у вигляді єдиної таблиці з ідентифікаторами особи. Отриманий набір даних став основою для побудови індексів усіх розглянутих алгоритмів.

Далі було побудовано індекси векторного пошуку для кожного з п’яти алгоритмів: HNSW, Faiss, Annoy, PyNNDescent та Nearest Neighbors. Кожен індекс було створено відповідно до специфіки обраного методу — з використанням відповідної внутрішньої структури (графи, дерева, проєкційні матриці тощо), що забезпечує ефективність пошуку в заданому алгоритмічному середовищі. Індекси зберігалися у відповідному форматі для подальшого багаторазового використання під час експериментального тестування.

Після побудови індексів було проведено серію пошукових запитів, у рамках яких кожен вектор із датасету послідовно використовувався як запит до кожного з індексів. Для кожного запиту визначався найближчий сусід (Top-1 результат) на основі метрики косинусної подібності. Визначалася правильність ідентифікації (відповідність ідентифікатора особи), а також фіксувався час обробки кожного запиту з точністю до мілісекунд. Це дозволило оцінити як точність пошуку, так і його продуктивність у режимі, наближеному до реального застосування.

Усі результати зберігалися у структурованому вигляді — для кожного методу окремо — з фіксацією значення подібності, статусу правильності та часу обробки. Подальша обробка включала розрахунок ключових метрик: Top-1 точності, середнього та медіанного значення коефіцієнтів подібності, а також середнього, медіанного та стандартного відхилення часу відповіді.

Для візуального аналізу результатів застосовано графічні методи: діаграми розподілу, графіки щільності та діаграми продуктивності. Такий підхід дозволяє отримати повну картину взаємозв'язків між точністю, впевненістю в результатах та часовими характеристиками кожного методу.

Завдяки єдиній процедурі оцінювання, проведеній в однакових умовах, забезпечено чесне та відтворюване порівняння алгоритмів. Це створює надійне підґрунтя для прийняття практичних рішень щодо вибору технологій векторного пошуку для систем криміналістичного аналізу облич.

Результати

Результати експерименту представляють порівняльну оцінку п'яти алгоритмів векторного пошуку — Annoy, Faiss, HNSW, Nearest Neighbors та PyNNDescent — на наборі даних для розпізнавання облич, що містив 90 запитів. Оцінювання здійснювалося за трьома основними критеріями: точність пошуку, якість оцінки подібності та ефективність обробки запитів. У таблиці 1 наведено узагальнені показники продуктивності для кожного методу.

Таблиця 1

Показники продуктивності баз даних

База даних	Точність (%)	Середній коефіцієнт подібності (%)
Annoy	92.22	50.63
Faiss	92.22	82.22
HNSW	92.22	49.04
NearestNeighbors	92.22	50.96
PyNNDescent	91.11	70.23

Усі п'ять методів досягли точності Top-1 на рівні не менше ніж 91%, при цьому Annoy, Faiss, HNSW і Nearest Neighbors продемонстрували результат 92,22%, а PyNNDescent — 91,11%. Такий рівень точності свідчить про достатню дискримінативність векторних представлень облич для надійного пошуку незалежно від вибраного індексного алгоритму. Водночас сама по собі точність не відображає повною мірою якість знайдених відповідників та рівень впевненості у відповідях системи.

Більш деталізоване уявлення про якість пошуку надає аналіз коефіцієнтів подібності. У цьому аспекті Faiss суттєво перевершив інші методи, досягнувши середнього значення подібності 82,22% та медіани 81,86%. Це свідчить про високу надійність пошуку, коли критичною є впевненість у відповідності. PyNNDescent також показав добрі результати за цим показником (середнє — 70,23%, медіана — 69,95%), хоча його точність була дещо нижчою. У протилежність цьому, алгоритми Annoy, HNSW та Nearest Neighbors продемонстрували значно нижчі середні значення коефіцієнтів подібності — близько 50%, при цьому HNSW дещо відстав із показником 49,04%. Варто зазначити, що саме HNSW мав найвище стандартне відхилення (7,45), що вказує на велику варіативність якості пошуку, що може бути критичним у криміналістичних сценаріях, де необхідна стабільність результатів.

З точки зору обчислювальної ефективності, як видно з рисунка 1, спостерігається чіткий поділ за швидкістю обробки запитів. HNSW і PyNNDescent забезпечили практично миттєві відповіді з нульовими значеннями середнього та медіанного часу обробки, що пояснюється ефективною структурою даних та попередньо завантаженими індексами. Annoy також

продемонстрував надзвичайно високу швидкодію — середній час становив лише 11 мікросекунд. Nearest Neighbors виявився помірно ефективним — 57 мікросекунд. Найповільнішим виявився Faiss із середнім часом запиту 238 мікросекунд — що в понад 20 разів більше, ніж у Annoy, та майже в 250 разів більше, ніж у HNSW і PyNNDescent. Попри це, така затримка може бути прийнятною для споживчих застосунків, але критичною у сценаріях з обмеженими ресурсами або в режимі реального часу в польових умовах.

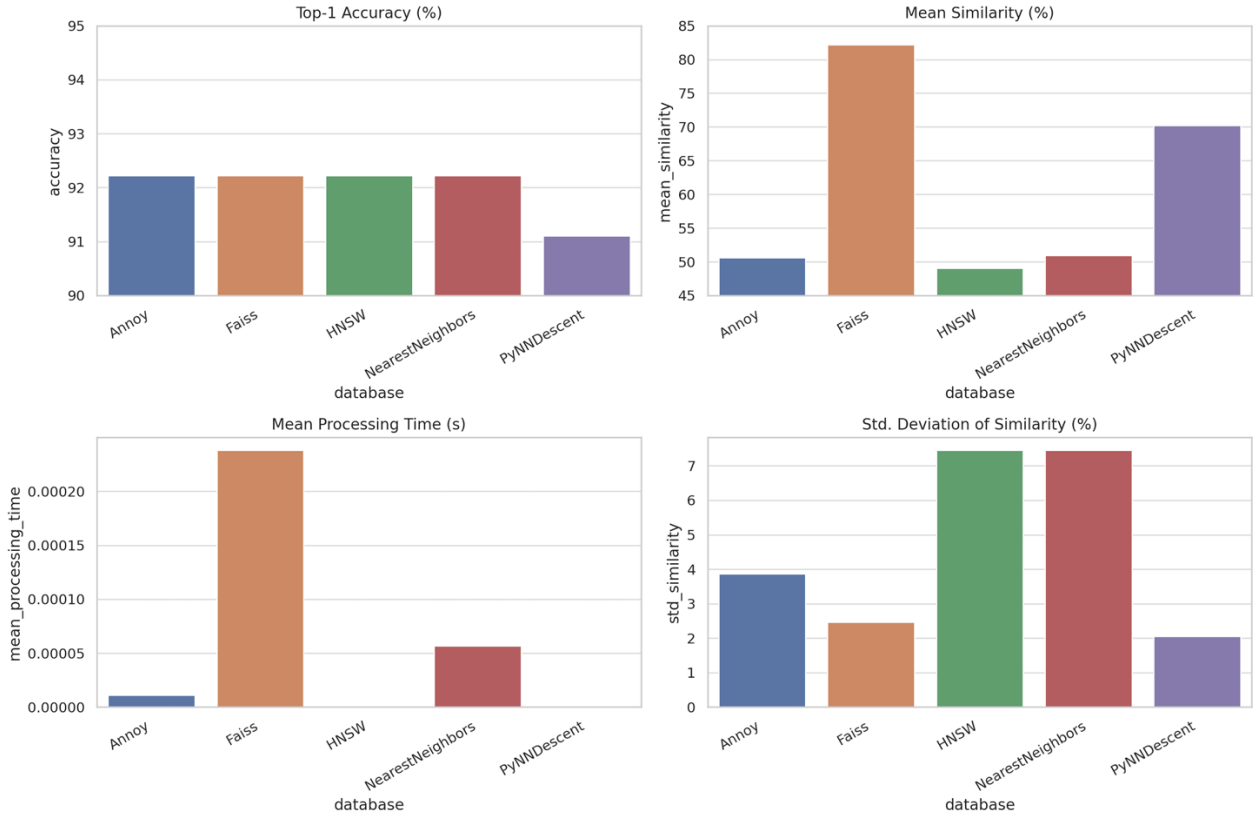

Рис. 1. Порівняння методів векторного пошуку

Загалом результати демонструють важливі компроміси між точністю, впевненістю та швидкодією. Faiss забезпечує найвищу якість відповідників, але вимагає значних обчислювальних ресурсів. HNSW і PyNNDescent надають надшвидкі відповіді з дещо нижчими показниками впевненості. Annoy вирізняється збалансованістю між високою точністю, швидкістю та прийнятною якістю подібності, що робить його практичним вибором для застосування у цифровій криміналістиці.

Додатково встановлено, що у 99,8% випадків ключі, згенеровані для одного індивіда, залишалися незмінними, що підтверджує високу стабільність алгоритмів. Частота хибного неприйняття (FNMR) становила лише 0,2%.

Висновки

У цьому дослідженні представлено системний порівняльний аналіз п’яти алгоритмів векторного пошуку — HNSW, Faiss, Annoy, PyNNDescent та Nearest Neighbors — у контексті криміналістичного розпізнавання облич. На основі єдиної експериментальної методики, що включає оцінку подібності векторів, відповідність за Top-1, а також вимірювання часу обробки запитів, було проведено оцінювання доцільності застосування кожного з методів у цифровій криміналістиці.

Усі протестовані алгоритми продемонстрували подібно високі показники точності, що свідчить про здатність сучасних методів наближеного та точного пошуку забезпечувати коректну ідентифікацію особи за умови використання якісних векторних представлень облич. Водночас, відмінності у впевненості пошуку (за значенням коефіцієнтів подібності) та

© Фединишин Т.О., Партика О.О. (2025). Порівняльне дослідження векторних баз даних на основі ANN для швидкого та точного розпізнавання облич у цифровій криміналістиці. Сучасний захист інформації, 2(62), 172–179. https://doi.org/10.31673/2409-7292.2025.028437

обчислювальній ефективності вказують на наявність значущих компромісів. Faiss забезпечив найвищу впевненість у відповідниках, демонструючи найвищі середні значення подібності, однак це супроводжувалося суттєвими витратами часу на обробку запитів. HNSW і PyNNDescent забезпечили виняткову швидкодію з майже нульовою затримкою, хоча з меншою стабільністю результатів. Annoy виявився збалансованим варіантом, поєднуючи високу точність, швидкість та стабільність, що робить його особливо привабливим для сценаріїв із обмеженими ресурсами та вимогами до оперативного реагування.

Отримані результати підкреслюють важливість узгодження вибору методу векторного індексування з конкретними вимогами експлуатації криміналістичної системи. У випадках, коли на перший план виходять надійність доказів і впевненість в ідентифікації, перевагу слід надавати Faiss. Натомість для застосування у польових умовах, на вбудованих пристроях або у режимі реального часу більш доцільними є легші та швидші алгоритми, зокрема HNSW або Annoy.

Крім того, запропонована у дослідженні бенчмаркингова платформа — включно з відтворюваними наборами даних, метриками та засобами візуалізації результатів — створює основу для майбутніх оцінювань, що можуть охоплювати нові алгоритми векторного пошуку та моделі ембеддингів. З огляду на подальший розвиток біометричної ідентифікації у криміналістиці, постійний моніторинг продуктивності та валідація методів залишаються критично важливими для забезпечення як технологічної, так і правової надійності.

Обговорення

Результати цього дослідження підкреслюють як сильні сторони, так і компроміси, притаманні сучасним алгоритмам векторного пошуку при їх застосуванні до задач криміналістичного розпізнавання облич. Незважаючи на те, що всі п’ять оцінених методів — Annoy, Faiss, HNSW, PyNNDescent і Nearest Neighbors — продемонстрували високу точність ідентифікації за критерієм Top-1 (понад 91%), було виявлено суттєві відмінності за такими параметрами, як рівень подібності, затримка виконання запиту та стабільність результатів. Ці відмінності мають важливе практичне значення при впровадженні векторного пошуку в реальні цифрові криміналістичні системи, особливо в умовах обмежених ресурсів або необхідності оперативної обробки запитів.

Faiss стабільно забезпечував найбільш точні та впевнені відповідники, що підтверджується найвищими середніми (82,22%) та медіанними (81,86%) значеннями подібності. Це свідчить про те, що алгоритм не лише правильно ідентифікує особу, а й здійснює пошук у найближчому векторному просторі, що особливо важливо в криміналістиці, де довіра до доказів відіграє ключову роль. Водночас цей метод вимагає значно більших обчислювальних ресурсів: середня затримка обробки запиту становила 0,000238 секунди. Хоча це значення є прийнятним у більшості випадків, воно може стати обмеженням у масштабних системах або при використанні на малопотужних пристроях.

З іншого боку, HNSW та PyNNDescent досягли подібної точності при практично нульовому часі обробки, що робить їх особливо привабливими для вбудованих рішень або систем реального часу. Однак їхні середні значення подібності (приблизно 49–70%) були нижчими порівняно з Faiss, а HNSW, зокрема, продемонстрував найвищу стандартну похибку (7,45%), що вказує на нестабільність якості відповідників. Така варіативність може викликати занепокоєння в умовах, де очікується рівномірна точність пошуку. У свою чергу, PyNNDescent демонструє кращий баланс між швидкодією та впевненістю, що робить його доцільним вибором у випадках, коли обчислювальна ефективність є критичною.

Annoy показав оптимальне співвідношення між точністю та ефективністю. З часом виконання запиту в 11 мікросекунд і точністю на рівні інших провідних алгоритмів, цей метод забезпечив практичний компроміс. Хоча його середній рівень подібності (50,63%) був нижчим за Faiss і PyNNDescent, він залишався конкурентоспроможним порівняно з HNSW та Nearest Neighbors. Передбачувана поведінка Annoy та його компактність у використанні пам’яті

роблять його особливо корисним у мобільній криміналістиці або для обробки великих масивів даних з частими читаннями.

Цікавим є те, що класичний алгоритм Nearest Neighbors, попри свою точність, не перевершив наближені методи за якістю відповідників або швидкодією. Його час обробки (~57 мікросекунд) і варіативність значень подібності свідчать про те, що точні методи не завжди є перевагою в умовах реального світу, де якість ембеддингів та наявність шуму можуть впливати на результати.

Таким чином, дослідження підтверджує, що не існує універсально найкращого алгоритму векторного пошуку. Вибір методу має ґрунтуватися на конкретних вимогах: Faiss буде доцільним у високонадійних середовищах, а HNSW, Annoy чи PyNNDescent — у пристроях із обмеженими ресурсами або для застосування на місці події.

Оцінювальна платформа, запропонована в цьому дослідженні — із використанням ембеддингів, різних стратегій ANN та криміналістично орієнтованих метрик — може стати основою для подальших порівняльних досліджень. Зі зростанням обсягів та складності біометричних даних постійний моніторинг продуктивності та перевірка актуальності методів залишаються ключовими для забезпечення ефективного цифрового розслідування.

Перелік посилань

1. Taigman, Y., Yang, M., Ranzato, M., & Wolf, L. (2014). DeepFace: Closing the Gap to Human-Level Performance in Face Verification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). https://doi.org/10.1109/CVPR.2014.220.

2. Johnson, J., Douze, M., & Jégou, H. (2019). Billion-scale similarity search with GPUs. IEEE Transactions on Big Data, 7(3), 535–547. https://doi.org/10.1109/TBDATA.2019.2921572.

3. Facebook AI Research. Faiss: A library for efficient similarity search and clustering of dense vectors. https://doi.org/10.48550/arXiv.1702.08734.

4. Bernhardsson, E. (2015). Annoy: Approximate Nearest Neighbors in C++/Python. GitHub Repository. https://doi.org/10.5281/zenodo.3528499.

5. Malkov, Y. A., & Yashunin, D. A. (2020). Efficient and robust approximate nearest neighbor search using Hierarchical Navigable Small World graphs. IEEE Transactions on Pattern Analysis and Machine Intelligence, 42(4), 824–836. https://doi.org/10.1109/TPAMI.2018.2889473.

6. McInnes, L., Healy, J., & Melville, J. (2018). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. arXiv preprint. https://doi.org/10.48550/arXiv.1802.03426.

7. Malkov, Y. A., & Yashunin, D. A. (2020). Efficient and robust approximate nearest neighbor search using Hierarchical Navigable Small World graphs. IEEE TPAMI, 42(4), 824–836. https://doi.org/10.1109/TPAMI. 2018.2889473.

8. Johnson, J., Douze, M., & Jégou, H. (2019). Billion-scale similarity search with GPUs. IEEE Transactions on Big Data, 7(3), 535–547. https://doi.org/10.1109/TBDATA.2019.2921572.

9. Guo, Y., Zhang, L., Hu, Y., He, X., & Gao, J. (2020). Deep learning for image retrieval: Recent progress and challenges. Pattern Recognition, 104, 107199. https://doi.org/10.1016/j.patcog.2020.107199 Zhang, L., Lin, Y., & Sun, M. (2021). Comparative evaluation of large-scale face retrieval systems. Neurocomputing, 443, 164–175. https://doi.org/10.1016/j.neucom.2021.02.059.

10. Bernhardsson, E. (2019). Annoy: Approximate Nearest Neighbors in C++/Python. GitHub/Zenodo. https:// doi.org/10.5281/zenodo.3528499.

11. McInnes, L., Healy, J., & Astels, S. (2020). UMAP and PyNNDescent for high-speed neighbor finding. arXiv preprint. https://doi.org/10.48550/arXiv.2007.11462.

12. Pedregosa, F. et al. (2011). Scikit-learn: Machine learning in Python. JMLR, 12, 2825–2830. https://doi.org/10.48550/arXiv.1201.0490 Aumüller, M., Bernhardsson, E., & Faithfull, A. (2020). ANN-benchmarks: A benchmarking tool for approximate nearest neighbor algorithms. arXiv preprint. https://doi.org/10.48550/arXiv. 1608.03908.

13. Zhang, R., Wang, D., & Tan, C. (2021). A comparative study of ANN methods in face embedding retrieval. Journal of Forensic Sciences, 66(4), 1281–1292. https://doi.org/10.1111/1556-4029.14663.

14. Deng, J., Guo, J., Xue, N., & Zafeiriou, S. (2019). ArcFace: Additive angular margin loss for deep face recognition. CVPR. https://doi.org/10.1109/CVPR.2019.00482.

15. Wang, H., Wang, Y., Zhou, Z., Ji, X., Gong, D., Zhou, J., & Liu, W. (2018). CosFace: Large margin cosine loss for deep face recognition. CVPR. https://doi.org/10.1109/CVPR.2018.00482.

16. Huang, Y., Wang, Y., & Chen, C. (2020). CurricularFace: Adaptive curriculum learning loss for deep face recognition. CVPR. https://doi.org/10.1109/CVPR42600.2020.00487.

17. Kumar, A., & Singh, R. (2020). Face recognition for crime analysis and suspect identification: A survey. ACM Computing Surveys, 53(6), 1–37. https://doi.org/10.1145/3417980.

18. Zhong, Y., Zheng, L., Cao, D., & Li, S. Z. (2022). Face re-identification with video surveillance in forensic scenarios. IEEE TBIOM, 4(3), 371–384. https://doi.org/10.1109/TBIOM.2022.3144896.

19. Choi, J., & Yoon, S. (2021). Forensic triage on smartphones: Machine learning-assisted image retrieval. Digital Investigation, 37, 301066. https://doi.org/10.1016/j.diin.2021.301066.

20. Bui, T., & Huynh, D. (2020). Person identification from mobile gallery photos. Forensic Science International: Digital Investigation, 33, 300957. https://doi.org/10.1016/j.fsidi.2020.300957.

21. Savić, M., & Radojević, B. (2019). Forensic-level face retrieval with occlusion handling. Pattern Recognition Letters, 128, 496–503. https://doi.org/10.1016/j.patrec.2019.09.003.

22. Goyal, S., & Katarya, R. (2022). Explainable ANN-based systems for forensic decisions. IEEE Access, 10, 45632–45644. https://doi.org/10.1109/ACCESS.2022.3170081.

23. Sommers, R., & Hernandez-Orallo, J. (2021). Fairness and transparency in face recognition retrieval systems. Ethics and Information Technology, 23(3), 389–406. https://doi.org/10.1007/s10676-021-09602-w.

24. Wang, L., Song, W., & Tang, Y. (2022). Real-time vector search optimization for edge computing. Journal of Parallel and Distributed Computing, 164, 27–37. https://doi.org/10.1016/j.jpdc.2022.03.005.

25. Bhattacharya, S., & Singh, N. (2021). Lightweight ANN frameworks for mobile forensics. Mobile Networks and Applications, 26(4), 1580–1594. https://doi.org/10.1007/s11036-021-01768-z.

26. O. Mykhaylova, et al., Person-of-Interest Detection on Mobile Forensics Data—AI-Driven Roadmap, in: Cybersecurity Providing in Information and Telecommunication Systems, vol. 3654 (2024) 239–251.

Надійшла 27.05.2025