УДК 004.89
DOI https://doi.org/10.32782/2663-5941/2026.2.1/09

Денисенко І.В.
https://orcid.org/0009-0004-1096-3860
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

Терейковський І.А.
https://orcid.org/0000-0003-4621-9668
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

ПАРАМЕТРИ ОЦІНКИ ЕФЕКТИВНОСТІ ЗАСОБІВ РОЗПІЗНАВАННЯ ЕМОЦІЙНОГО СТАНУ В СИСТЕМАХ ДИСТАНЦІЙНОГО НАВЧАННЯ

У статті проведено комплексний та системний аналіз сучасних методів і програмних засобів розпізнавання емоційного стану користувачів у специфічному контексті систем дистанційного навчання. Актуальність даного дослідження зумовлена масовим та незворотним переходом до онлайн-освіти, де відсутність безпосереднього невербального контакту між викладачем та аудиторією критично ускладнює об'єктивну оцінку рівня залученості здобувачів освіти та своєчасне виявлення їхньої когнітивної втоми або втрати інтересу. З метою ефективного подолання цих викликів у роботі теоретично обґрунтовується та практично формується уніфікована багатовимірна система критеріїв для оцінювання існуючих алгоритмів комп'ютерного зору (Computer Vision). Запропонований розширений набір параметрів глибоко охоплює як експлуатаційні, так і функціональні вимоги до систем: мінімізацію навантаження на апаратне забезпечення кінцевих клієнтських пристроїв, адаптивність та стійкість до реальних умов проведення відеоконференцій (зокрема, нестабільне або слабке освітлення, часткові оклюзії обличчя, повороти голови), гарантування абсолютної приватності та автономності обробки чутливих біометричних даних, високу семантичну релевантність отриманих метрик для педагогіки (детектування глибокої залученості, розгубленості та втоми, а не лише дискретних базових емоцій), простоту інженерної інтеграції у сучасне веб-середовище та здатність моделей до пролонгованого аналізу складних часових рядів. На основі розроблених критеріїв у статті детально та критично розглянуто архітектурні особливості легковагових згорткових нейронних мереж (CNN), новітніх візуальних трансформерів (Vision Transformers), а також провідних комерційних хмарних інструментів (у вигляді SDK та API). У результаті проведеного порівняльного аналізу виявлено суттєву науково-практичну проблему – виражену поляризацію існуючих технологічних рішень: високоточні хмарні та важкі трансформерні моделі абсолютно не відповідають сучасним вимогам конфіденційності та жорстким апаратним обмеженням студентських лептопів, тоді як швидкі локальні алгоритми часто залишаються надто примітивними та обмеженими у контекстному аналізі тривалої поведінки. Для успішного вирішення цієї ключової суперечності у статті формалізовано концепцію збалансованого гібридного підходу. Він органічно базується на парадигмі граничних обчислень (Edge AI), де первинна оптимізована обробка просторових ознак обличчя виконується суто локально за допомогою стиснутих CNN, а подальший аналіз динаміки емоцій та складних мімічних патернів у часі здійснюється через рекурентні модулі. Такий інноваційний підхід дозволяє кардинально перенести фокус із ресурсоємної статичної покадрової детекції на цілісне та неперервне розуміння когнітивних станів студента, забезпечуючи при цьому максимально можливу точність та абсолютну конфіденційність без жодної потреби у передачі відеопотоку на зовнішні сторонні сервери.

Ключові слова: *розпізнавання емоцій, дистанційне навчання, Computer Vision, залученість студентів, Edge AI, візуальні трансформери, приватність даних.*

Постановка проблеми. У сучасному освітньому просторі дистанційне навчання стало домінуючим, проте відсутність невербального контакту ускладнює педагогам оцінку залученості аудиторії та своєчасне виявлення втоми. Це актуалізує потребу в автоматизованому аналізі емоційного стану через відеопотік для відновлення зворотного зв'язку.

Однак технічні обмеження користувацьких пристроїв, специфіка відеоконференцій та суворі вимоги до приватності роблять існуючі алгоритми малоефективними для масового впровадження. На цьому ґрунтується актуальність розробки збалансованих засобів розпізнавання емоцій, що поєднують точність, швидкодію та конфіденційність для забезпечення адаптивності навчального процесу.

Аналіз останніх досліджень і публікацій. Розпізнавання емоційного стану за відеопотоком є одним із ключових та найбільш динамічних напрямків у галузі комп'ютерного зору (Computer Vision) та афективних обчислень. У контексті систем дистанційного навчання ця задача трансформується з класичної класифікації базових емоцій (радість, гнів, смуток тощо) у більш комплексну проблему визначення специфічних когнітивних станів, таких як рівень залученості, уваги, втоми або нудьги. Сучасні наукові рішення фокусуються на розробці алгоритмів, здатних детектувати ці стани в реальному часі, однак проведений аналіз виявляє суттєву поляризацію: високоточні моделі вимагають значних обчислювальних ресурсів, тоді як легковагові рішення часто мають обмеження щодо глибини аналізу часових рядів.

Для вирішення проблеми жорстких апаратних обмежень клієнтських пристроїв розробляються спеціалізовані оптимізовані архітектури. Подібні підходи до побудови ефективних обчислювальних моделей для систем, що функціонують в умовах дефіциту ресурсів та відсутності стабільного зв'язку з хмарою, ґрунтовно досліджено у спільній праці I. A. Tereikovskyi та A. V. Didus [21, с. 343] на прикладі обробки голосових сигналів. Своєю чергою, вагомий внесок у мінімізацію мережевих затримок (latency) у системах комп'ютерного зору зробив колектив дослідників на чолі з N. E. Rashidy [1, с. 6]. Запропонована ними архітектура RERS-FOG поєднує оптимізований детектор YOLOv8 для миттєвої локалізації обличчя та метод головних компонент (PCA), що дозволило досягти точності 96–98% на датасеті CK+ при суттєвому скороченні часу відгуку системи.Проблематику створення легковагових згорткових мереж (CNN) для автоматизованого визначення залученості безпосередньо на мобільних пристроях глибоко розкрито у працях J. A. Ramirez-Quintana зі співавторами [2, с. 5] та науковців X. Chen і L. Huang [4, с. 4]. Так, у розробленій командою J. A. Ramirez-Quintana моделі LiExNet [2, с. 10] застосування спеціалізованих легковагових згорток та механізму ефективної канальної уваги дозволило досягти точності понад 90% при значному зниженні навантаження на процесори вбудованих пристроїв. Зі свого боку, запропонована X. Chen та L. Huang архітектура LFNSB [4, с. 7] завдяки інтеграції модуля просторового зміщення (Spatial Bias) успішно розв'язує проблему розрізнення візуально схожих емоційних станів зі збереженням конкурентної точності.

З метою підвищення розрізнювальної здатності моделей в умовах візуального шуму та оклюзій, дослідницька група на чолі з J. Yu [3, с. 2] пропонує інтеграцію механізму глобальної канально-просторової уваги (Global Channel-Spatial Attention), що динамічно перезважує важливість ознак та фокусує мережу на ключових регіонах обличчя. Подібний підхід раніше був успішно реалізований командою S. Minaee в архітектурі Deep-Emotion [15, с. 3], де використання просторових трансформерів для попереднього вирівнювання зображень дозволило досягти точності 98.0% на лабораторному наборі CK+ та 70.02% на складному датасеті FER-2013.

Проблема моніторингу пролонгованих станів ефективно вирішується за допомогою аналізу часових рядів та мультимодальності. У спільному дослідженні M. Aly та N. S. Alotaibi [6, с. 9] запропоновано комплексну архітектуру, що комбінує згорткові мережі ResNet-50 та 3D-CNN, гарантуючи точність класифікації на рівні 94%. Науковий колектив під керівництвом S. Gupta [9, с. 28598] розширює цей підхід мультимодальною інтеграцією мікро-виразів обличчя та векторів уваги (Gaze Estimation), забезпечуючи загальну точність 91–93%. Водночас адаптація розпізнавання до 16 унікальних емоційних категорій успішно валідована в алгоритмах моніторингу через смартфони, які розробила M. Keinert та її колеги [10, с. 8], де досягнуто середньої точності 83–86%.

Важливим вектором досліджень є практична інтеграція афективних систем у педагогічні сценарії. Дослідники J. Kim та G. Choi [5, с. 6] розробили гібридну архітектуру ArecaNet, що оптимізована під розпізнавання станів у системах людино-машинної взаємодії та показує безпрецедентну точність 97.8%. Водночас науковець N. Rahmeisi [7, с. 129] обґрунтовує доцільність використання веб-орієнтованих no-code платформ, які дозволяють досягати точності 90–95% без залучення значних потужностей. Динамічне

ж використання отриманих метрик досліджено групою авторів на чолі з R. Gutierrez [8, с. 5], де розпізнані емоції слугують тригером для автоматичної адаптації складності контенту з точністю класифікації у межах 88–92%.

Значний масив сучасних досліджень зосереджений на впровадженні архітектур візуальних трансформерів (Vision Transformer, ViT) [11–14] для подолання обмежень традиційних CNN. Зокрема, гібридна інтеграція алгоритму YOLO з ViT, запропонована V. Sareen та K. R. Seeja [11, с. 5], забезпечує стійкість до оклюзій і швидкодію понад 35 FPS. Оскільки стандартні трансформери часто ігнорують дрібні локальні зміни міміки, колектив Y. Tian [12, с. 4] розробив модифікацію з механізмами гібридної локальної уваги. Для боротьби з перенавчанням S. Min та співавтори [13, с. 3] успішно застосовують стратегію випадкового маскування візуальних токенів. Крім того, можливості трансформерів адаптуються для детекції мікро-виразів у дослідженні J. Hong та його колег [14, с. 4] шляхом застосування механізмів пізнього злиття у Video Transformer, що дозволяє незалежно обробляти просторові та часові компоненти відеопотоку [14, с. 4]. У сукупності ці підходи демонструють значну перевагу ViT-архітектур над класичними CNN у захопленні глобальних контекстних залежностей, проте їх використання суворо обмежене надмірними апаратними вимогами.

Окремий клас інструментів представляють комерційні хмарні API та серверні фреймворки. Рішення, що працюють на стороні клієнта (Client-side), такі як Affectiva [16] та JavaScript-бібліотека MorphCast [17], аналізують розширений набір метрик (Action Units, Arousal, Valence) локально, гарантуючи відповідність стандартам приватності GDPR. Натомість мультимодальна платформа Hume AI [18], хмарні когнітивні сервіси Face++ [20] та Python-фреймворк DeepFace [19] забезпечують глибокий семантичний аналіз понад 28 складних станів на основі комбінації відео, аудіо та лінгвістичного контексту з винятковою стійкістю до складного освітлення. Проте орієнтація цих систем на серверні обчислення (Backend processing) створює критичні перешкоди для дистанційного навчання: постійна трансляція відеопотоку генерує мережеві затримки, комерційна тарифікація робить їх масштабне впровадження економічно неефективним, а необхідність передачі біометричних даних третім сторонам несе суттєві ризики порушення конфіденційності.

Незважаючи на стрімкий прогрес в архітектурах комп'ютерного зору, більшість проаналізованих рішень проектувалися для лабораторних умов або високопродуктивних серверів. Питання їхньої комплексної оцінки та адаптації саме для слабких клієнтських пристроїв здобувачів освіти (в умовах нестабільного зв'язку та обмежених обчислювальних ресурсів) залишається недостатньо розкритим. Це зумовлює необхідність формування спеціалізованої системи критеріїв оцінки, що враховуватиме специфіку систем дистанційного навчання.

Постановка завдання. Основною метою публікації є формування уніфікованої системи критеріїв для комплексної оцінки ефективності засобів автоматичного розпізнавання емоційного стану користувачів за відеопотоком, а також проведення на їх основі порівняльного аналізу сучасних методів та інструментів комп'ютерного зору в контексті систем дистанційного навчання. Окрім цього, завданням дослідження є обґрунтування концепції збалансованого гібридного підходу, що дозволить подолати виявлені недоліки існуючих локальних та хмарних систем і забезпечить високу точність детекції складних когнітивних станів здобувачів освіти за умови збереження повної приватності даних.

Виклад основного матеріалу. Для проведення об'єктивного порівняльного аналізу ефективності існуючих рішень та вибору оптимального підходу необхідно формалізувати систему оцінювальних критеріїв. На основі аналізу специфіки функціонування платформ дистанційного навчання та виявлених обмежень комп'ютерного зору, було сформовано набір параметрів оцінки методів розпізнавання емоцій N1-N7. Для визначення цього переліку критеріїв використано методологічний підхід до автоматизованого розпізнавання емоційних станів, обґрунтований у дисертаційній роботі Л. О. Терейковської [22, с. 5] та доповнений результатами досліджень I. A. Tereikovskyi і A. V. Didus [21, с. 343]. Дані параметри доцільно розділити на дві групи.

Дані параметри доцільно розділити на дві групи: експлуатаційні (технічні) та функціональні (педагогічні) параметри. До першої групи віднесено характеристики, що визначають технічну можливість розгортання системи: мінімізацію вимог до апаратного забезпечення клієнта (N1), гарантії приватності біометричних даних (N3), ступінь мережевої автономності (N4) та простоту інженерної інтеграції (N6). До другої групи увійшли критерії, що визначають якість розпізнавання: здатність алгоритмів працювати в реальних умовах оклюзій (N2), семантичну релевантність отримуваних метрик для цілей педагогіки (N5) та здатність моделей до пролонгованого аналізу часових рядів (N7).

Оцінювання алгоритмів за наведеними критеріями здійснювалося з використанням бінарної шкали (0 або 1), де 1 означає повну або часткову, але задовільну відповідність методу вимогам критерію, а 0 – незадовільну відповідність або повну невідповідність цим вимогам. Зокрема: для N1 одиниця означає достатню здатність працювати на клієнтських процесорах (Edge), а нуль – критичну потребу в GPU; для N3 та N4 одиниця свідчить про локальну обробку (Client-side) або прийнятний рівень автономності, а нуль ставиться за обов'язкову трансляцію біометричних даних на сервери; для N5 одиниця виставляється за здатність визначати комплексні стани (залученість, увага), а нуль – лише базові емоції; для N6 одиниця означає наявність готового SDK, API або загальну простоту розгортання, а нуль – потребу складної самостійної імплементації; для N7 одиниця присвоюється алгоритмам із механізмами аналізу часової динаміки. Результати оцінювання наведено в Таблиці 2.

Виходячи з аналізу таблиці, проведеного за сформованим переліком бінарних критеріїв N1-N7, можна зробити висновок, що більшість розглянутих методів та засобів мають суттєві обмеження в контексті їх застосування у системах дистанційного навчання.

Таблиця 1

Параметри оцінки методів розпізнавання емоцій

Скорочене позначення	Опис характеристики	Інтерпретація шкали
N1	Вимоги до апаратного забезпечення	1 – працює на слабкому пристрої, 0 – вимагає потужного пристрою
N2	Здатність працювати в реальних умовах	1 – стійкість до перешкод, адаптованість до реальних умов, 0 – працює лише в лабораторних умовах
N3	Приватність	1 – дані не покидають пристрій, 0 – передача даних третій стороні
N4	Автономність	1 – повна незалежність від інтернету, 0 – залежність від хмарних рішень
N5	Релевантність для систем дистанційного навчання	1 – розпізнає не лише базові емоції, а й складніші, такі як увага та залученість, 0 – розпізнає лише базові емоції
N6	Простота інтеграції	1 – готове рішення, 0 – складна архітектура; потреба реалізації “з нуля”
N7	Пролонгованість аналізу	1 – аналіз динаміки емоційного стану в часі, 0 – статичний покадровий аналіз

Таблиця 2

Оцінки характеристик методів розпізнавання емоцій

Назва роботи	Характеристики						
	N1	N2	N3	N4	N5	N6	N7
Rashidy (Fog Computing) [1]	1	1	0	0	0	0	0
Ramirez-Quintana (LiExNet) [2]	1	1	1	1	1	0	0
Yu (Global Attention) [3]	0	1	1	1	0	0	0
Chen (Spatial Bias + Loss) [4]	1	1	1	1	0	0	0
Kim (ArecaNet) [5]	1	1	1	1	0	0	0
Aly (Hybrid ResNet+3D-CNN) [6]	0	1	1	1	1	0	1
Rahmeisi (Google Teachable Machine) [7]	1	1	0	0	1	1	0
Gutierrez (Adaptive Learning System) [8]	1	1	0	0	1	0	1
Gupta (Student Engagement) [9]	1	1	1	1	1	0	1
Keinert (Smartphone 16 Emotions) [10]	1	1	1	1	1	0	0
Sareen (YOLO + ViT)	0	1	1	1	0	0	1
Tian (ViT Hybrid Attention) [12]	0	1	1	1	0	0	0
Min (Transformers Masking) [13]	0	1	1	1	0	0	0
Hong (Late Fusion ViT) [14]	0	1	1	1	0	0	1
Minaee (Deep-Emotion CNN) [15]	0	1	1	1	0	0	0
Affectiva [16]	1	1	1	1	1	1	1
MorphCast [17]	1	1	1	1	1	1	1
Hume AI [18]	1	1	0	0	1	1	1
DeepFace [19]	0	1	1	1	0	0	0
Face++ [20]	1	1	0	0	1	1	0

Зокрема, спостерігається чітка поляризація рішень. З одного боку, існують високоточні архітектури на базі глибокого навчання (наприклад, трансформери ViT у роботах [11], [12], [14] та згорткові мережі з механізмами просторової уваги [15]), які повністю задовольняють вимоги щодо стійкості до реальних умов (N2 = 1) та приватності (N3 = 1), проте отримують критичну оцінку (0) за критеріями N1 (вимоги до апаратного забезпечення) та N6 (простота інтеграції). Це робить їх пряме розгортання на слабких клієнтських пристроях студентів технічно неможливим.

З іншого боку, існують веб-орієнтовані та хмарні рішення (наприклад, Google Teachable Machine [7], Hume AI [18], Face++ [20]), які вирішують проблему апаратних обмежень та семантичної релевантності метрик (N5 = 1). Однак вони повністю залежать від зовнішньої інфраструктури (N4 = 0), що автоматично призводить до порушення вимог приватності біометричних даних (N3 = 0) через необхідність трансляції відеопотоку на сервери третьої сторони.

Повну відповідність більшості висунутих критеріїв демонструють лише спеціалізовані клієнтські рішення (наприклад, Affectiva [16], MorphCast [17] та архітектура з роботи [9, с. 28598]), однак їх закритий комерційний код або надмірна оптимізація під конкретні датасети обмежують гнучкість їхнього використання у незалежних освітніх платформах.

Формування даного переліку характеристик та проведений порівняльний аналіз дозволяють окреслити напрямок наступного етапу досліджень – розробку методу та програмних засобів, які б стабільно отримували найвищі оцінки (1) за всіма критичними показниками: забезпечували баланс між точністю розпізнавання складних когнітивних станів (N5=1), ефективністю роботи на пристроях з обмеженими ресурсами (N1=1) та абсолютною приватністю даних користувачів (N3=1). Доцільним вбачається створення гібридного підходу, що поєднує локальні легковагові Edge-архітектури з механізмами рекурентного аналізу часових рядів (N7=1) для підвищення інформативності результатів.

Висновки. В ході проведених досліджень було обґрунтовано базовий перелік характеристик засобів для автоматичного розпізнавання емоційного стану за відеопотоком. Цей перелік, що включає функціональні та експлуатаційні параметри (N1-N7), дозволяє комплексно оцінювати, наскільки існуючі методи відповідають специфічним вимогам систем дистанційного навчання, зокрема щодо приватності даних та автономності роботи. Визначено перспективність подальших досліджень в напрямку розробки методу та програмних засобів, які базуються на гібридному підході, для забезпечення балансу між точністю аналізу складних когнітивних станів та ефективністю використання ресурсів клієнтських пристроїв.

Список літератури:

1. Rashidy N. E., Allogmani E., Hassan E., Alnowaiser K., Elmannai H., Ali Z. H. Toward real-time emotion recognition in fog computing-based systems: leveraging interpretable PCA_CNN, YOLO with self-attention mechanism. Frontiers in Computer Science. 2026. Vol. 7. P. 1714394. DOI: 10.3389/fcomp.2025.1714394.

2. Ramirez-Quintana J. A., Muñoz-Pacheco J. J., Ramirez-Alonso G., Medrano-Hermosillo J. A., Corral-Saenz A. D. Lightweight Convolutional Neural Network with Efficient Channel Attention Mechanism for Real-Time Facial Emotion Recognition in Embedded Systems. Sensors. 2025. Vol. 25, No 23. P. 7264. DOI: 10.3390/s25237264.

3. Yu J., Zheng Y., Wang L., Wang Y., Xu S. Design of an Expression Recognition Solution Based on the Global Channel-Spatial Attention Mechanism and Proportional Criterion Fusion. arXiv preprint arXiv:2503.11935. 2025. DOI: 10.48550/arXiv.2503.11935.

4. Chen X., Huang L. A Lightweight Model Enhancing Facial Expression Recognition with Spatial Bias and Cosine-Harmony Loss. Computation. 2024. Vol. 12, No 10. P. 201. DOI: 10.3390/computation12100201.

5. Kim J., Choi G. ArecaNet: Robust Facial Emotion Recognition via Assembled Residual Enhanced Cross-Attention Networks for Emotion-Aware Human-Computer Interaction. Sensors. 2025. Vol. 25, No 23. P. 7375. DOI: 10.3390/s25237375.

6. Aly M., Alotaibi N. S. A comprehensive deep learning framework for real time emotion detection in online learning using hybrid models. Scientific Reports. 2025. Vol. 15. P. 42012. DOI: 10.1038/s41598-025-26381-7.

7. Rahmeisi N. Real-Time Emotion Recognition in Online Learning Using Google Teachable Machine. Indonesian Journal of Informatics Education. 2025. Vol. 9, No 2. P. 125. DOI: 10.20961/ijie.v9i2.110565.

8. Gutierrez R., Villegas-Ch W., Govea J. Development of adaptive and emotionally intelligent educational assistants based on conversational AI. Frontiers in Computer Science. 2025. Vol. 7. P. 1628104. DOI: 10.3389/fcomp.2025.1628104.

9. Gupta S., Kumar P., Tekchandani R. A multimodal facial cues based engagement detection system in e-learning context using deep learning approach. Multimedia Tools and Applications. 2023. Vol. 82. P. 28589–28615. DOI: 10.1007/s11042-023-14392-3.

10. Keinert M., Pistrosch S., Mallol-Ragolta A., Schuller B. W., Berking M. Facial Emotion Recognition of 16 Distinct Emotions From Smartphone Videos: Comparative Study of Machine Learning and Human Performance. Journal of Medical Internet Research. 2025. Vol. 27. e68942. DOI: 10.2196/68942.

11. Sareen V., Seeja K. R. Video-Based Facial Emotion Recognition using YOLO and Vision Transformer. EPJ Web of Conferences. 2025. Vol. 328. P. 01040. DOI: 10.1051/epjconf/202532801040.
12. Tian Y., Zhu J., Yao H., Chen D. Facial Expression Recognition Based on Vision Transformer with Hybrid Local Attention. Applied Sciences. 2024. Vol. 14. P. 6471. DOI: 10.3390/app14156471.
13. Min S., Yang J., Lim S. Emotion Recognition Using Transformers with Random Masking. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops. 2024. P. 4734–4743.
14. Hong J., Lee C., Jung H. Late Fusion-Based Video Transformer for Facial Micro-Expression Recognition. Applied Sciences. 2022. Vol. 12. P. 1169. DOI: 10.3390/app12031169.
15. Minaee S., Minaei M., Abdolrashidi A. Deep-Emotion: Facial Expression Recognition Using Attentional Convolutional Network. Sensors. 2021. Vol. 21. P. 3046. DOI: 10.3390/s21093046.
16. Emotion AI SDK: Humanizing Technology. Affectiva. 2024. URL: https://www.affectiva.com (дата звернення: 24.02.2026).
17. Emotion AI HTML5 SDK: Client-side Facial Emotion Recognition. MorphCast. 2024. URL: https://www.morphcast.com (дата звернення: 24.02.2026).
18. The Empathic Voice Interface (EVI) & Expression Measurement API. Hume AI. 2025. URL: https://hume.ai (дата звернення: 24.02.2026).
19. Serengil S. I., Ozpinar A. DeepFace: A Lightweight Face Recognition and Facial Attribute Analysis Framework for Python. 2021 International Conference on Computer Science and Engineering (UBMK). 2021. P. 1–6. DOI: 10.1109/UBMK52708.2021.9558915.
20. Face++ Cognitive Services API: Emotion Recognition. Megvii Technology. 2023. URL: https://www.faceplusplus.com (дата звернення: 24.02.2026).
21. Tereikovskyi I. A., Didus A. V. A model for keyword spotting in voice signal for specialized computer systems. Herald of Advanced Information Technology. 2025. Vol. 8, No 3. P. 341–351. DOI: 10.15276/hait.08.2025.22.
22. Терейковська Л. О. Методологія автоматизованого розпізнавання емоційного стану слухачів системи дистанційного навчання : автореф. дис. ... д-ра техн. наук : 05.13.06. Київ, 2022. 40 с. URL: http://diser.ntu.edu.ua/Tereykovska_aref.pdf (дата звернення: 03.03.2026).

Denysenko I.V., Tereikovskyi I.A. PARAMETERS FOR ASSESSING THE EFFICIENCY OF EMOTION RECOGNITION TOOLS IN DISTANCE LEARNING SYSTEMS

The article provides a comprehensive and systematic analysis of modern methods and software tools for recognizing the emotional state of users within the specific context of distance learning systems. The relevance of this study is driven by the massive and irreversible transition to online education, where the lack of direct non-verbal contact between the teacher and the audience critically complicates the objective assessment of students' engagement levels and the timely detection of cognitive fatigue or loss of interest. To effectively overcome these challenges, the paper theoretically substantiates and practically develops a unified multidimensional system of criteria for evaluating existing computer vision algorithms. The proposed expanded set of parameters deeply covers both the operational and functional requirements for such systems: minimizing the computational load on the hardware of end-user client devices, adaptability and robustness to real-world video conferencing conditions (including unstable or poor lighting, partial facial occlusions, head rotations), ensuring absolute privacy and autonomy in processing sensitive biometric data, high semantic relevance of the obtained metrics for pedagogy (detecting deep engagement, confusion, and fatigue, rather than just discrete basic emotions), the simplicity of engineering integration into modern web environments, and the models' capability for prolonged analysis of complex time series. Based on the developed criteria, the article provides a detailed and critical review of the architectural features of lightweight convolutional neural networks (CNNs), advanced Vision Transformers (ViTs), and leading commercial cloud-based tools (in the form of SDKs and APIs). As a result of the comparative analysis, a significant scientific and practical problem is identified–the pronounced polarization of existing technological solutions: highly accurate cloud-based and heavy transformer models completely fail to meet modern privacy requirements and the strict hardware constraints of students' laptops, whereas fast local algorithms often remain too primitive and limited in the contextual analysis of prolonged behavior. To successfully resolve this key contradiction, the article formalizes the concept of a balanced hybrid approach. It is organically based on the Edge AI paradigm, where the initial optimized processing of spatial facial features is performed strictly locally using compressed CNNs, and the subsequent analysis of emotional dynamics and complex facial patterns over time is carried out through recurrent modules. Such an innovative approach makes it possible to fundamentally shift the focus from resource-intensive static frame-by-frame detection to a holistic and continuous understanding of the student's cognitive states, thereby ensuring the highest possible accuracy and absolute confidentiality without any need to transmit the video stream to external third-party servers.

Keywords: *emotion recognition, distance learning, Computer Vision, student engagement, Edge AI, vision transformers, data privacy.*

Дата першого надходження статті до видання: 05.03.2026
Дата прийняття статті до друку після рецензування: 30.03.2026
Дата публікації (оприлюднення) статті 11.05.2026