



УДК 004.8+004.9

О. Д. Яремченко, П. Я. Пукач

Національний університет "Львівська політехніка", м. Львів, Україна

**ВДОСКОНАЛЕННЯ МЕТОДУ РОЗПІЗНАВАННЯ ЕМОЦІЙ НА БАЗІ НЕЙРОННИХ МЕРЕЖ**

У статті проаналізовано використання мікровиразів – тонких рухів обличчя, які важко помітити людському оку, а ще важче одразу проаналізувати, щоб оцінити психологічний стан за допомогою методів штучного інтелекту. Навіть спеціалістам у галузі не завжди ідеально це вдається, адже зміна цих мікровиразів відбувається за 1/5 до 1/3 секунди. Дослідження спрямоване на вдосконалення аналізу мікроміміки для точної ідентифікації емоцій і психологічного стану. Наведено огляд реалізованих технологічних рішень на базі CNN та запропоновано метод для їх покращення. Експериментальна перевірка, здійснена на відеозаписах людей, які переживають різні емоції, підтвердила високу точність розробленого методу в розпізнаванні емоцій і психологічного стану. Незважаючи на проблеми, пов'язані з дефіцитом наборів даних мікровиразів і тонкістю рухів обличчя, розроблено модель CapsuleNet для розпізнавання мікровиразів, сформовано архітектуру системи та здійснено тестування. Завдяки об'єднанню трьох основних наборів даних (SMIC, CASME II і SAMM) в уніфіковану кросбазу даних запропонований у роботі метод перевіряє можливість узагальнення моделі за різними предметними характеристиками. Продуктивність CapsuleNet, оцінена за допомогою кросбазового порівняльного аналізу та перевірки Leave-One-Object-Out, істотно перевершує базову (LBP-TOP) та інші вдосконалені моделі CNN. У роботі показано, що продуктивність розробленої моделі, визначена незваженим середнім показником запам'ятовування та результатами F1, перевершує як базову лінію LBP-TOP, так і інші передові моделі CNN. У комплексній системі розпізнавання мікровиразів спершу обробляємо дані, щоб ідентифікувати верхівкові кадри в послідовностях і виокремити область обличчя на цих кадрах. Ці оброблені зображення обличчя потім подаються в CapsuleNet для класифікації. Результати роботи розвивають та доповнюють методи емоційного штучного інтелекту, покращуючи оцінку за допомогою мікроміміки психологічного стану, який впливає на психічне здоров'я, взаємодію людини з комп'ютером і соціальну робототехніку. Ця технологія має потенціал для розвитку та розширення у різних галузях, зокрема, надає додаткову можливість для фахівців, котрі працюють з людьми, стежити за їхньою продуктивністю, психологічним станом, та оцінювати стрес.

**Ключові слова:** CapsuleNet, мікроміміка, емоційний штучний інтелект, розпізнавання лицьових емоцій, взаємодія людина-комп'ютер.

**Вступ / Introduction**

Мікровирази (МВ) – це швидкі й мимовільні вирази обличчя, які виникають, коли людина намагається приховати свої справжні емоції. На відміну від звичайної міміки або макровиразу, який триває від пів секунди до чотирьох секунд і передбачає значні рухи обличчя, мікровираз набагато швидкоплинніший, з'являється лише на частку секунди (звичай між 1/5 і 1/25, як припускають Мацумото та Хванг [1]) і на локалізованих ділянках обличчя. Їх короткочасність і тонка природа ускладнюють виявлення мікровиразів у реальному часі навіть кваліфікованим фахівцям. Мікровирази важко свідомо контролювати або підробити, тому вони є цінним інструментом для виявлення справжніх емоцій. Це має вагомий наслідок для таких сфер, як кримінальні допити [2], національна безпека [3] та виявлення брехні [4], [5], у яких важливе розуміння справжніх емоційних станів.

Micro-Expression Grand Challenge (MEGC) об'єднав три ключові спонтанні набори даних мікровиразу, а саме SAMM[6], CASME II[7] і SMIC[8] в єдину перехресну базу (кросбазу) даних. MEGC спростив класифікацію емоцій за трьома широкими категоріями

[9] – негативні, позитивні та здивовані, щоб мінімізувати розбіжності в інтерпретації в різних наборах даних. Завдання полягає в тому, щоб точно класифікувати кожну послідовність мікровиразів у один із цих трьох класів за допомогою кросбазу даних.

Згорткові нейронні мережі (CNN), які досягли надзвичайного успіху в різних задачах візуального розпізнавання, також ефективно використовували для розпізнавання мікровиразів, або як екстрактори ознак, або як класифікатори. CapsuleNet використовує механізм маршрутизації за домовленістю, щоб розпізнати зв'язки "частина – ціле". Продемонструвавши вищу продуктивність порівняно із CNN, CapsuleNet досяг успіху в таких завданнях, як розпізнавання цифр і об'єктів [10].

На основі розроблених капсульних моделей у 2017 р. Sabour та ін. [16] запропонували CapsuleNet, яка призначена спеціально для розпізнавання мікровиразів, зосереджуючись на верхинних кадрах.

*Об'єкт дослідження* – системи штучного інтелекту для аналізу мікровиразів – тонких рухів обличчя, які важко помітити людському оку.

*Предмет дослідження* – методи та засоби аналізу мікроміміки людей на основі штучного інтелекту.

*Мета роботи* – оцінити здатність CapsuleNet розуміти зв'язки “частина – ціле” та її ефективність на порівняно невеликих наборах даних, характерних для завдань розпізнавання мікроекспресій, а також покращити метод виявлення мікроекспресій стосовно відомих рішень для якіснішого аналізу мікроекспресій та її потенційного застосування для оцінювання психологічного стану людей, порівняти її продуктивність з наявними системами розпізнавання мікроекспресій.

Для досягнення вказаної мети визначено такі основні завдання дослідження:

- огляд літератури щодо аналізу мікроекспресій та емоційного штучного інтелекту;
- розроблення архітектури системи машинного навчання для аналізу мікроекспресій в реальному часі;
- аналізування отриманих результатів.

**Аналіз останніх досліджень та публікацій.** Після створення наборів даних для спонтанних мікроекспресій у численних дослідженнях використовували ручні методи розроблення функцій, зокрема, локальний бінарний шаблон із трьома ортогональними площинами (LBP-TOP) [11]. Ян та ін. [7] застосували LBP-TOP для вилучення ознак, інтегрувавши його з класифікатором SVM для завдань розпізнавання. Wang та ін. [12] використали розширене уточнення функцій за допомогою LBP-SixIntersectionPoints (LBP-SIP), тоді як Li та ін. [13] об'єднали LBP-TOP з гістограмою орієнтованих градієнтів і гістограмою орієнтації градієнта зображення для створення повних векторів ознак для розпізнавання.

**Дослідження капсульних мереж.** Після появи CapsuleNetworks (CapsuleNet) опубліковано численні дослідницькі праці для оцінювання ефективності CapsuleNet у різних галузях досліджень. Афшар та ін. у [14] продемонстрували, що CapsuleNet, розроблений для ідентифікації пухлин головного мозку, ефективніший від звичайних моделей CNN.

**Матеріали і методи дослідження. Запропоноване удосконалення методу.** В роботі побудовано структуру для розпізнавання мікроекспресій за допомогою CapsuleNet та здійснено аналіз результатів розпіз-

навання завдяки цій технології. Спочатку структура ідентифікує та обробляє вершинні кадри з послідовностей МВ, якщо це ще не зазначено, а потім продовжує генерувати прогнози для МВ. Застосовано підхід навчання передачі для використання попередньо навчених ваг ImageNet для початкових згорткових шарів під час навчання мережі. Структуру, із модулями попереднього оброблення та класифікації, подано на рис. 1.

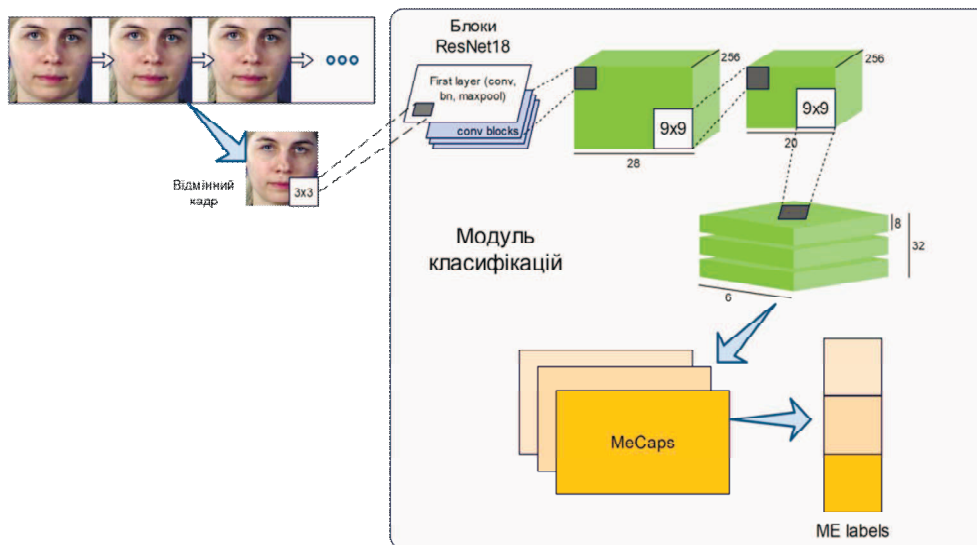
**Попереднє оброблення.** Використовуємо набір інструментів для обличчя з відкритої бази, щоб виявити 68 орієнтирів обличчя в кожному кадрі МВ. Дотримуючись [15], визначаємо десять ділянок обличчя на основі цих орієнтирів, що відповідають областям, які часто беруть участь у рухах м'язів. Розміри кожної ділянки евристично визначають як половину ширини рота. Ці десять ділянок показано на рис. 2.

Щоб ідентифікувати вершинний кадр, обчислюємо абсолютну різницю пікселів у цих десяти регіонах між поточним кадром і кадрами початку та зсуву. Щоб зменшити шум навколишнього середовища, нормалізуємо суму цих відмінностей різницею між поточним кадром і наступним. Це дає середнє значення на піксель для кожного кадру в послідовності. Передбачено, що верхівковий кадр є кадром із найвищою різницею середньої інтенсивності порівняно з кадрами початку та зміщення, тому зазвичай вибираємо кадр із піксельним значенням МВ на піксель. Рис. 3 демонструє зміну середнього значення МВ у послідовності.

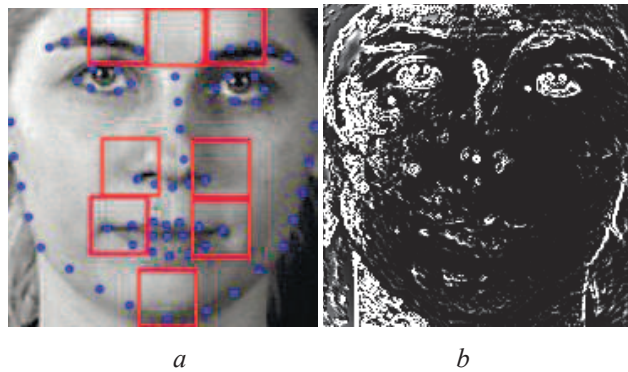
$$f(\text{frame}_i, \text{frame}_j) = \frac{|\text{frame}_i - \text{frame}_j| + 1}{|\text{frame}_i - \text{frame}_{i-3}| + 1} \quad (1)$$

$$M_i = f(\text{frame}_i, \text{frame}_{\text{onset}}) + f(\text{frame}_i, \text{frame}_{\text{offset}}) \quad (2)$$

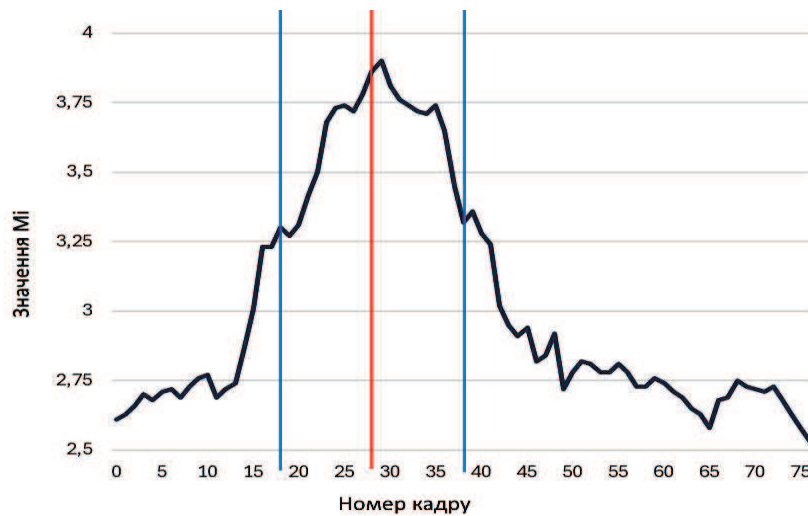
Після ідентифікації вершинного кадру в наборі даних SMIC переходимо до виділення ділянки обличчя в цих вершинних кадрах послідовностей мікроекспресій (МВ). Ця ізоляція ґрунтується на орієнтирах обличчя. Ці обрізані ділянки обличчя потім вводимо у CapsuleNet для класифікації. Етапи роботи модуля попереднього оброблення подано на рис. 4.



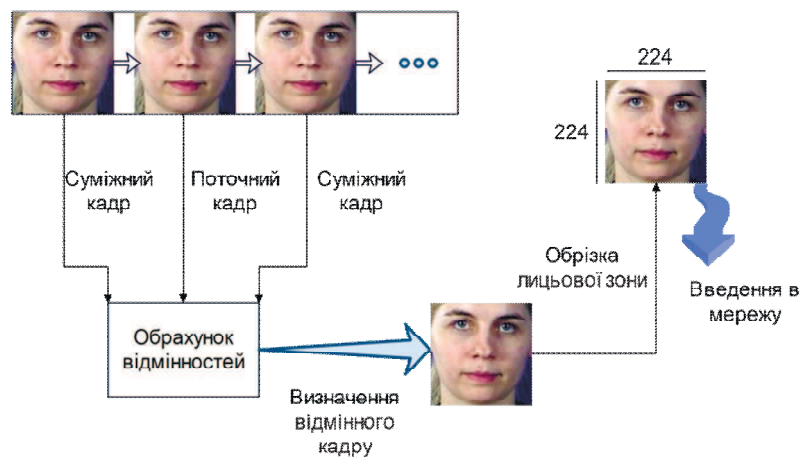
**Рис. 1.** Повна структура для розпізнавання мікроекспресій за допомогою архітектури CapsuleNet / A complete framework for microexpression recognition using the CapsuleNet architecture



**Рис. 2.** Демонстрація визначення апекс-кадру (приклад із CASME II): а – 10 комірок – регіони інтересу на основі 68 орієнтирів обличчя; б – відмінності кадрів для аналізу МВ / Demonstration of determining the apex of the frame (example from CASME II) а – 10 cells – regions of interest based on 68 facial landmarks; б – Differences in frames for MV analysis



**Рис. 3.** Середнє значення варіації  $M_i$  в послідовності МВ. Основна істина, зображена в послідовності МВ, позначена вертикальною червоною лінією / The average value of the variation  $M_i$  in the sequence MV. The ground truth mark depicted in the MV sequence is indicated by a vertical red line



**Рис. 4.** Модуль попереднього оброблення в нашій структурі / The preprocessing module in our framework

**Реалізація вибраного підходу із CapsuleNet.** У пропонованому підході використано архітектуру CapsuleNet, розроблену для розпізнавання мікровиразів з кадру апексу в заданій послідовності МВ. Як показано на рис. 2, ця архітектура вводить вершинні кадри. Ці кадри, зокрема області обличчя, видобуті з верхівкових кадрів, перетворюються на кольорові зображення та змінюються до [224, 224, 3]. Цей вхідний розмір істотно більший, ніж цифровий вхідний розмір, вказаний у [16].

Необхідна адаптація, оскільки мікровирази спричиняють тонкі рухи обличчя, а зменшення розміру може призвести до істотних втрат інформації. Щоб перетворити інтенсивність пікселів на значущі локальні характеристики форми [28, 28, 256], зображення обробляють через перший згортковий шар, а відтак три згорткові блоки моделі ResNet18 [17]. Вибір ResNet18 замість більших варіантів, таких як ResNet50 або ResNet101, мінімізує ризик переповнення, врахо-



вуючи малий масштаб наборів даних мікрОВИРАЗІВ. Модель використовує попередньо підготовані вагові коефіцієнти з шарів ResNet18 на ImageNet, ефективно зменшуючи кількість параметрів і полегшуючи передавання знань про розпізнавання об'єктів із великого набору даних ImageNet.

Ці особливості, своєю чергою, передаються у первинний шар капсули для отримання первинних капсул (позначених як PrimaryCaps) у результаті множення на згортки. Кожна первинна капсула інкапсулює інформацію та характеристики сутності у векторі нейронів замість скалярного значення в CNN. Активність нейронів у капсулі описує параметри екземпляра цієї сутності. Позначимо  $i$  як капсулу на первинному шарі капсули, а  $j$  як капсулу на вихідному шарі капсули. Активація капсули  $j$  визначається на основі активації всіх капсул у первинному шарі. Коефіцієнт зв'язку  $c_{ij}$  вимірює узгодженість між капсулою  $i$  та капсулою  $j$ , яку обчислюють за допомогою маршрутизації *softmax* так:

$$c_{ij} = \frac{\exp(b_{ij})}{\sum_k \exp(b_{ik})}, \quad (3)$$

де  $b_{ij}$  позначено як логарифм імовірності того, чи повинна капсула  $i$  на первинному шарі капсули з'єднуватися з материнською капсулою  $j$ . Значення  $b_{ij}$  спочатку встановлено на 0, а  $c_{ij}$  ітеративно уточнюється за допомогою процесу динамічної маршрутизації. Детально вказаний процес описано в [21], де також міститься вся необхідна додаткова інформація. Потім вхідні дані для капсули  $j$  (позначеної як  $x_j$ ) обчислюють так:

$$x_i = \sum_i c_{ij} W_{ij} u_i. \quad (4)$$

У вихідній капсулі її довжина відображає ймовірність існування певної сутності, тоді як її орієнтація примусово представляє властивості сутності. Щоб забезпечити довжину вихідних векторів між інтервалом  $[0,1]$  за незмінної його орієнтації, застосовано функцію стиснення нелінійності. Остаточний результат капсули  $j$ , тобто  $v_j$ , обчислюємо за допомогою функції *своиш*, як показано нижче:

$$v_j = \frac{\|x_j\|^2 x_j}{1 + \|x_j\|^2 \|x_j\|}. \quad (5)$$

У розробленій CapsuleNet вихідні капсули (позначають як MeCaps) використовують лише з метою розпізнавання мікрОВИРАЗІВ, але не для реконструкції вхідних зображень. Додаткову підмережу реконструкції в [21] введено як регуляризацію для загальної мережі з вхідними цифровими зображеннями розміру [28, 28]. Однак варто видалити цю частину реконструкції з мережі, оскільки розмір вхідного зображення набагато більший [224, 224], а кількість навчальних наборів даних доволі невелика.

## Результати дослідження та їх обговорення / Research results and their discussion

**Оптимізація нейронної мережі.** Використаємо втрату маржі як цільову функцію для навчання мережі:

$$L^{net} = \sum_k L_k^{margin}, \quad (6)$$

де  $L_k^{margin}$  – втрата маржі для відповідного МВ  $k$ . Якщо зображення містить МВ  $k$ , капсулу відповідної довжини робимо довгою, оскільки очікуємо, що вона буде короткою, якщо на зображенні немає МВ. Щоб задовольнити такі умови, використовуємо такі втрати маржі для кожного МВ  $k$ :

$$L_k^{margin} = T_k \max(0, m^+ - \|v_k\|)^2 + \lambda_k (1 - T_k) \max(0, \|v_k\| - m^-)^2, \quad (7)$$

де  $T_k = 1$ , якщо МВ  $k$  існує на зображенні, і  $T_k = 0$ , якщо  $k$  не існує на зображенні. Коефіцієнт  $\lambda_k$  визначає вплив втрат, які виникають, коли МЕ наявний або відсутній на зображенні. Нарешті, втрата маржі дорівнює нулю, якщо  $\|v_k\| > m^+$ , коли  $T_k = 1$ , і  $\|v_k\| < m^-$ , якщо  $T_k = 0$ .

**Експериментальна частина.** У пропонованій реалізації введення розмірів [224, 224, 3] через початковий згортковий блок забезпечує розмір тензора [20, 20, 256]. На рівні PrimCaps є 32 канали, що містять 8D-капсули, кожна капсула містить вісім згорткових нейронів і використовує розмір ядра [9×9]. Розкладка сітки [6×6] використовується для капсул, що приводить до спільної ваги. Отримуємо [6×32] виходів із 8D-капсул. З'єднання від кожної капсули в шарі PrimCaps до капсул у шарі MeCaps полегшуються за допомогою вагової матриці  $W_{ij}$  [16×8]. Шар MeCaps завершується трьома вихідними капсулами. Параметри маржинальних втрат ( $L^{margin}$ ) встановлено з  $m^+ = 0,9$ ,  $m^- = 0,1$  і  $\lambda_k = 0,5$ . Оптимізатор Адама використовуємо для оптимізації, починаючи зі швидкості навчання 0,0001 і впроваджуючи стратегію зниження швидкості навчання 0,9 протягом 20 епох. Динамічна маршрутизація здійснюється через три ітерації. Для покращення навчального набору даних застосовують такі методи збільшення даних, як зміна розміру, довільне кадрування, дзеркальне відображення, обертання та коливання кольорів.

**Основний підхід.** Щоб викликати MEGC 2019, ми вирішили повторно застосувати метод локального бінарного шаблону на трьох ортогональних площинах (LBP-TOP) як базову лінію на основі дослідження. Застосовуючи техніку LBP-TOP, кадри послідовності МЕ сегментуємо на [5×5] блоків, що не перекриваються, із встановленими параметрами LBP-TOP: радіусами RXY, RXT, RYT = {1,1,4}, кількість сусідніх точок  $P = 4$  на всіх площинах, а часовий інтервал (ТІМ) дорівнює 10.

Крім того, продуктивність пропонованої моделі порівнювали з двома провідними архітектурами розпізнавання об'єктів, ResNet (зокрема, версією ResNet18) [14] і VGG (версією VGG11) [18]. В обидві базові моделі внесено зміни із заміною їхніх оригінальних кінцевих повністю зв'язаних шарів, які мали 1000 категорій виведення, новими повністю зв'язаними шарами, призначеними для трьох категорій виведення. Окрім коригувань останнього шару, початкові ваги моделі були призначені за допомогою попередньо навчених ваг ImageNet. Функцію крос-ентропії з декількома мітками використано як критерій втрат. Під

час навчання для цих базових моделей дотримувалися ідентичної швидкості навчання, кількості епох і налаштувань оптимізатора, які використовують для навчання нашої запропонованої моделі CapsuleNet.

**Набори даних.** Перехресна база даних Micro-Expression Grand Challenge (MEGC 2019) об'єднує три відомі набори даних спонтанного мікроекспресу.

База даних SMIC [8]: цей набір даних містить 164 мікроекспресії від 16 суб'єктів, із кожним МВ, зафіксованим зі швидкістю 100 кадрів за секунду та класифікованим за трьома широкими емоційними ярликами: позитивним, негативним та здивування. Нова ітерація бази даних, SMIC-E, містить додаткові фрейми без виразів, що оточують мічені мікроекспресії.

CASME II [7]: детальний спонтанний набір даних мікроекспресії із 247 зразками відео від 26 азійських учасників, середній вік яких 22,03 року. Він охоплює п'ять категорій мікроекспресії: щастя, відраза, страх, здивування та сум.

SAMM [6]: база даних спонтанних дій і мікроекспресії складається із 159 зразків мікроекспресії від 32 демографічно різноманітних учасників із середнім віком 33,24 року та однаковою гендерною розподілом. Зосереджений на семи основних емоціях, SAMM був розроблений для вивчення мікроекспресії обличчя.

І CASME II, і SAMM записані зі швидкістю 200 кадрів за секунду, тоді як SMIC із меншою швидкістю – 100 кадрів за секунду. Щоб спростити класифікацію у цих наборах даних, у виклику прийнято уніфіковану систему позначок, подібну до SMIC. Після об'єднання консолідована база даних містить 68 записів (16 із SMIC, 24 із CASME II та 28 із SAMM) і використовує спрощені загальні класи.

Враховуючи обмежений розмір перехресної бази даних, автори розширили набір даних, вибравши верхній кадр і чотири його сусідні кадри із кожної послідовності МВ для навчання. Крім того, для пом'якшення впливу дисбалансу набору даних використано методи повторної вибірки.

**Табл. 1.** Підсумок міток для кожного набору даних і кросбази даних (три бази – комбіновані) / Label summary for each data set and cross-database (3 databases combined)

Набори даних	Клас			
	Негатив	Позитив	Здивування	Сума
SMIC	70	51	43	164
CASME II	88	32	25	145
SAMM	92	26	15	132
Крос-база даних	250	109	83	442

**Метрики оцінювання.** Ефективність методів розпізнавання в попередньому виклику MEGC оцінювали за допомогою методів Holdout-DatabaseEvaluation (HDE) і CompositeDatabaseEvaluation (CDE). Однак підхід HDE є менш сприятливим через можливість надмірної кількості комбінацій у численних комбінаційних даних для тренувального тестування з кросбази даних. Використовуємо перехресну перевірку Leave-One-Subject-Out (LOSO) як техніку CDE для звітування про ефективність розпізнавання ME. Цей метод враховує різноманітність походження учасників, зокрема етнічну належність, стать та емоційну чутливість, зафіксовану в різних середовищах, полегшуючи оцінку, незалежну від суб'єкта. Враховуючи притаманний кросбазам даних дисбаланс, ефективність розпізнавання оцінюють за допомогою двох збалансованих показників:

Незважений показник F1 (UF1), який також називають макроусередненим показником F1, передбачає обчислення істинно позитивних ( $TP_c$ ), хибно позитивних ( $FP_c$ ) і хибно негативних результатів ( $FN_c$ ) для всіх  $k$  згортків LOSO для кожного класу  $c$  (з усієї кількості  $C$  класів). Після обчислення їх відповідних балів F1 загальний збалансований бал  $F1$  отримуємо, усереднюючи ці бали F1 для кожного класу:

$$F1_c = \frac{2TP_c}{2TP_c + FP_c + FN_c}, \quad (8)$$

$$UF1 = \frac{1}{C} \sum_c F1_c. \quad (9)$$

Середній відгук (UAR), також відомий як збалансована точність системи. Спочатку розраховуємо оцінку точності  $UAR_c$  для кожного класу, усереднюємо всі  $Acc_c$  за кількістю класів, щоб отримати остаточну оцінку UAR, як показано нижче:

$$Acc_c = \frac{TP_c}{N_c}, \quad (10)$$

$$UAR = \frac{1}{C} \sum_c Acc_c. \quad (11)$$

І  $UF1$ , і  $UAR$  дають об'єктивну оцінку того, наскільки добре модель працює на всіх класах, а не різницю лише на кількох певних класах.

**Обговорення результатів дослідження.** Подано незважену оцінку  $F1$  ( $UF1$ ) і показники відгуку для моделі, їх порівняно із базовим рівнем, наданим викликом, і додатковими базовими показниками, які встановлено у табл. 2. Ця табл. 2 містить  $UF1$  і незважене середнє запам'ятовування (UAR) для всієї кросбази даних і її окремих компонентів – SMIC, CASME II і SAMM – як для базових моделей, так і для запропонованого рішення. Пропонована модель досягла оцінки  $UF1$  0,6517 й оцінки  $UAR$  0,6504, тому її можна вважати найефективнішою серед базових порівнянь. Зокрема, запропонований метод перевершує просунуті мережі CNN, такі як ResNet18 і VGG11, приблизно на 10%, що підкреслює ефективність архітектури CapsuleNet. На відміну від методу LBP-TOP, який покладається на всю послідовність МВ для виділення ознак, запропонований у роботі підхід використовує лише кадр апексу з кожної послідовності МВ як вхідні дані. Незважаючи на це, він усе ще досягає 6,5% покращення показників UAR та  $UF1$  порівняно з методом LBP-TOP.

На рис. 5 подано матрицю помилок “Виключити одного суб'єкта” (LOSO) для презентованого методу. Підкреслено помітно високий рівень запам'ятовування для негативного класу (0,780) через його поширеність у кросбазі даних. Ефективність коефіцієнтів відгуку для

двох інших класів також на рівні 0,596 і 0,575 відповідно, демонструючи стійкість нашої CapsuleNet до проблеми незбалансованих даних завдяки ефективним методам повторної вибірки.

**Вимірювання впливу компонентів на продуктивність системи.** Інтеграція додаткових рівнів у оригінальну архітектуру CapsuleNet вирішує обчислювальні проблеми. Вибираємо додаткові рівні із найсучасніших мереж, таких як ResNet і VGG, яким надаємо перевагу. Це зумовлено їх відомою продуктивністю на великих наборах даних, таких як ImageNet, з технікою передавання знань. Зокрема, запропоновано ввести

шари з ResNet18, як описано раніше, і для порівняння, перші три згорткових і відповідні шари максимального об'єднання з VGG11. Обидві адаптації приводять до ідентичних розмірів вихідних характеристик [28, 28, 256]. Однак, як показано в табл. 3, отримані з ResNet18 шари виявилися ефективнішими для вилучення ознак у CapsuleNet, ніж шари з VGG11. Можливим поясненням є те, що часте максимальне об'єднання VGG11 може призвести до більшої втрати просторової інформації, яка є критичною для навчання CapsuleNet, порівняно з використанням згорткових шарів ResNet18 із кроком 2 для зменшення розмірності.

**Табл. 2.** Ефективність перехресної перевірки LOSO за запропонованими методами та базовими лініями / LOSO cross-validation performance of the proposed methods and baselines.

Метод	кросбаза даних							
	UF1 (повна)	UAR (повна)	UF1(SMIC)	UAR (SMIC)	UF1 (CASME II)	UAR (CASME II)	UF1 (SAMM)	UAR (SAMM)
LBP-TOP	0.5885	0.5791	0.2000	0.5280	0.7026	0.7429	0.3954	0.4102
VGG11	0.5264	0.5392	0.3461	0.3558	0.5315	0.5381	0.2871	0.4056
ResNet18	0.5392	0.5459	0.3576	0.3602	0.5367	0.5441	0.4821	0.4322
<b>CapsuleNet</b>	<b>0.6517</b>	<b>0.6504</b>	<b>0.5820</b>	<b>0.5877</b>	<b>0.7068</b>	<b>0.7018</b>	<b>0.6209</b>	<b>0.5989</b>

**Табл. 3.** Результати перехресної перевірки LOSO для абляційного дослідження / LOSO cross-validation results for ablation study.

Метод	Значення	
	Незважений F1	Незважений середній відклик
CapsuleNetra VGG11	0.6126	0.6257
CapsuleNetra ResNet18	0.6513	0.6496



**Рис. 5.** Матриця помилок у перехресній базі даних із методом перехресної оцінки LOSO. Число в кожній комірці вказує на кількість передбачень / Cross-database error matrix with the LOSO cross-estimation method. The number in each cell indicates the number of predictions

*Наукова новизна отриманих результатів дослідження* – удосконалено наявну систему для розпізнавання мікроекспресій та побудовано ефективну архітектуру для аналізу мікроекспресій, котра ефективніша за відомі CNN системи і забезпечує приріст за показниками UF1 та UAR до 10 %.

*Практична значущість результатів дослідження* – полягає у тому, що запропонована система дає змогу покращити відомі методи для розпізнавання мікроекспресій, а також зменшити ресурсне навантаження для реалізації системи.

## Висновки / Conclusions

У цьому дослідженні запропоновано комплексну структуру, яка використовує CapsuleNet для розпізнавання мікроекспресій, зосереджуючись виключно на вершинних кадрах. Нова архітектура істотно (до 10 %) зменшує обчислювальні вимоги CapsuleNet і підвищує її продуктивність на обмежених наборах даних мікроекспресій. Використовуючи лише вершинні кадри, CapsuleNet уникає великих і складних обчислень, необхідних для аналізу усєї послідовності мікроекспресій. Експериментальні результати демонструють перевагу розробленого у статті підходу, який перевершує базовий рівень LBP-TOP і різні вдосконалені архітектури CNN у розпізнаванні мікроекспресій, а отже, має перспективи для розвитку і використання у сучасних умовах.

## References

- Matsumoto, D., Hwang, H. S. (2011). Evidence for training the ability to read microexpressions of emotion. *Motivation Emotion*, 35, 181–191. <https://doi.org/10.1007/s11031-011-9212-2>
- Russell, T. A., Chu, E., & Phillips, M. L. (2006). A pilot study to investigate the effectiveness of emotion recognition remediation in schizophrenia using the micro-expression training tool. *British Journal of Clinical Psychology*, 45(4), 579–583. <https://doi.org/10.1348/014466505X90866>
- Weinberger, S. (2010). Airport security: intent deceive? *Nature*, 465, 412–415. <https://doi.org/10.1038/465412a>
- Ekman, Paul (2009). Lie catching and microexpressions. In *The Philosophy of Deception*, ed. C.W. Martin (pp. 118–133). Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195327939.003.0008>
- Ekman, P. “Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage”, revised ed.; WW Norton Company: New York, NY, USA, 2009.



6. Davison, A. K., Lansley, C., Costen, N., Tan, K. & Yap, M. H. (2018). Sann: A spontaneous micro-facial movement dataset. *IEEE Transactions on Affective Computing*, 9(1), 116–129. <https://doi.org/10.1109/TAFFC.2016.2573832>
7. Yan, W.-J., Li, X., Wang, S.-J., Zhao, G., Liu, Y.-J., Chen, Y.-H., & Fu, X. (2014). Casme ii: An improved spontaneous micro-expression database and the baseline evaluation. *PLoS one*, 9(1). <https://doi.org/10.1371/journal.pone.0086041>
8. Li X, Pfister T, Huang X, Zhao G, Pietika inen M (2013). A Spontaneous Micro-expression Database: Inducement, Collection and Baseline, 10th Proc Int Conf Autom Face Gesture Recognit (FG2013). Shanghai, China. <https://doi.org/10.1109/FG.2013.6553717>
9. Davison, A., Merghani, W., & Yap, M. H. (2018). Objective classes for microfacial expression recognition. *Journal of Imaging*, 4(10), 119. <https://doi.org/10.3390/jimaging4100119>
10. Xi, E., Bing, S., & Jin, Y. (2017). Capsule network performance on complex data. arXiv preprint arXiv:1712.03480. <https://doi.org/10.48550/arXiv.1712.03480>
11. Zhao, G., & Pietikainen, M. (2007). Dynamic texture recognition using local binary patterns with an application to facial expressions. *IEEE Trans. on PAMI*, 29(6), 915–928. <https://doi.org/10.1109/TPAMI.2007.1110>
12. Wang, Y., See, J., Phan, C.W. et al. (2014). LBP with Six Intersection Points: Reducing Redundant Information in LBP-TOP for Microexpression Recognition, 12th Asian Conference on Computer Vision 2014, Singapore, Singapore. [https://doi.org/10.1007/978-3-319-16865-4\\_34](https://doi.org/10.1007/978-3-319-16865-4_34)
13. Li, X., Hong, X., Moilanen, A., Huang, X., Pfister, T., Zhao, G., & Pietikainen, M. (2017). Towards Reading Hidden Emotions: A Comparative Study of Spontaneous Micro-expression Spotting and Recognition Methods, *IEEE Transactions on Affective Computing*, 9(4), 563–577. <https://doi.org/10.1109/TAFFC.2017.2667642>
14. Afshar, P., Mohammadi, A., & Plataniotis, K. N. (2018). Brain tumor type classification via capsule networks. 2018 25th IEEE International Conference on Image Processing (ICIP), Athens, Greece, pp. 3129–3133. <https://doi.org/10.1109/ICIP.2018.8451379>
15. Borza, Diana, Danescu, Radu, Itu, Razvan, & Darabant, Adrian Sergiu (2017). High-Speed Video System for Micro-Expression Detection and Recognition. *Sensors* 2017, 17(12), 2913. <https://doi.org/10.3390/s17122913>
16. Sabour, S., Frosst, N., & Hinton, G. E. (2017). Dynamic Routing Between Capsules. ArXiv, abs/1710.09829. <https://doi.org/10.48550/arXiv.1710.09829>
17. Simonyan, K., & Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. ArXiv preprint arXiv:1512.03385. <https://doi.org/10.48550/arXiv.1409.1556>
18. Wang, Y., Sun, A., Han, J., Liu, Y., & Zhu, X. (2018). Sentiment analysis by capsules. WWW'18: Proceedings of the 2018 World Wide Web Conference, Lyon, France, 1165–1174. <https://doi.org/10.1145/3178876.3186015>

**O. D. Yaremchenko, P. Ya. Pukach**

*Lviv Polytechnic National University, Lviv, Ukraine*

## IMPROVEMENT OF EMOTION RECOGNITION METHODS BASED ON NEURAL NETWORKS

This article analyzes the use of microexpressions – subtle facial movements that are difficult for the human eye to notice, and even more difficult to immediately analyze, even specialists in the field do not always succeed in this perfectly, because their speed is only 1/5 to 1/3 of a second, for assessment of psychological state using artificial intelligence methods. The research is aimed at improving the analysis of micro-mimicry for accurate identification of emotions and psychological state. An overview of implemented technological solutions based on CNN was conducted, and a method for their improvement was found. An experimental test conducted on video recordings of people experiencing various emotions showed the high accuracy of the developed method in recognizing emotions and psychological state. Despite the challenges of the scarcity of microexpression datasets and the subtlety of facial movements, the paper presents a CapsuleNet model for microexpression recognition, builds a system architecture, and conducts testing. By combining three main data sets (SMIC, CASME II and SAMM) into a unified cross-database, the method developed in the work tests the possibility of generalization of the model by different subject characteristics. The performance of CapsuleNet, evaluated by cross-baseline benchmarking and Leave-One-Object-Out validation, significantly outperforms the baseline (LBP-TOP) and other improved of an CNN models. The paper shows that the performance of the developed model, determined by unweighted average recall and F1 scores, outperforms both the LBP-TOP baseline and other state-of-the-art CNN models. In a comprehensive microexpression recognition system. First, we process the data to identify the peak frames in the sequences and isolate the face region in these frames. These processed face images are then moved to CapsuleNet for the classification. The results of the work is to develop and complement methods of emotional artificial intelligence, offering new insights into micromimic assessment of psychological states that affect mental health, human-computer interaction, and social robotics. This technology has potential for development and expansion. This is an additional opportunity for companies that work with people and it is important for them to monitor their productivity, as it is directly related to the psychological state.

**Keywords:** CapsuleNet, micro-mimics, emotional AI, facial expression recognition, human-computer interaction.

### Інформація про авторів:

**Яремченко Олександр Дмитрович**, аспірант, кафедра систем штучного інтелекту.

Email: Oleksandr.D. Yaremchenko@lpnu.ua; <https://orcid.org/0009-0001-2002-2704>

**Пукач Петро Ярославович**, д-р техн. наук, професор, директор інституту ІМФН.

Email: petro.y.pukach@lpnu.ua; <https://orcid.org/0000-0002-0359-5025>

**Цитування за ДСТУ:** Яремченко О. Д., Пукач П. Я. Вдосконалення методів розпізнавання емоцій на базі нейронних мереж.

*Український журнал інформаційних технологій*. 2024, т. 6, № 1. С. 58-64.

**Citation APA:** Yaremchenko, O. D., & Pukach, P. Ya. (2024). Improvement of emotion recognition methods based on neural networks.

*Ukrainian Journal of Information Technology*, 6(1), 58-64. <https://doi.org/10.23939/ujit2024.01.058>