

УДК 004.89

DOI <https://doi.org/10.32782/tnv-tech.2025.4.1.13>

ДІАГНОСТИКА ЕМОЦІЙНОГО СТАНУ ДІТЕЙ НА ОСНОВІ НЕЙРОМЕРЕЖЕВОГО АНАЛІЗУ ЇХ МАЛЮНКІВ

Корабльов М. М. – доктор технічних наук, професор,
професор кафедри комп'ютерних інтелектуальних технологій та систем
Харківського національного університету радіоелектроніки
ORCID ID: 0009-0005-2540-7741

Дикий С. А. – аспірант кафедри комп'ютерних інтелектуальних технологій
та систем Харківського національного університету радіоелектроніки
ORCID ID: 0009-0007-5396-2413

Кобзев І. В. – кандидат технічних наук, доцент,
доцент кафедри мультимедійних систем і технологій
Харківського національного економічного університету імені Семена Кузнеця
ORCID ID: 0000-0002-7182-5814

Фомічов О. О. – кандидат технічних наук,
старший викладач кафедри електронних обчислювальних машин
Харківського національного університету радіоелектроніки
ORCID ID: 0000-0001-9273-9862

Емоційний стан дитини є складним, багатовимірним конструктом: він відображається у виборі кольору, композиції, символічних образах та штрихах малюнка, що формуються у нелінійному, хаотичному творчому процесі. У практичній психології однією з найпоширеніших проєктивних методик є аналіз дитячих малюнків. Передбачається, що дитина несвідомо переносить переживання у символіку зображення, дозволяючи фахівцю виявити радість, тривожність, агресію чи депресивні тенденції. Традиційний психологічний аналіз дитячих малюнків спирається на суб'єктивну інтерпретацію і масштабується для масового скринінгу. У роботі пропонується нейромережева мультимодальна гібридна модель автоматизованої діагностики емоцій, що поєднує чотири взаємодоповнюючі канали ознак. Глобальний контекст зображення витягує нейронна мережа EfficientNet-B3, яка описує форму, текстуру та конфігурацію сцени. Локальні семантично значущі об'єкти визначає нейронна мережа YOLOv8, що розширена до 55 класів на відкритому датасеті ESRA та агрегує їх у кількісні ознаки, що відповідають трьом емоційним суперкласам. Колірна палітра описується статистикою HSV-простору, де обчислюються числові характеристики яскравості, насиченості та частки домінуючих відтінків. Геометрія та характер ліній кодується композиційними та графічними метриками, що визначають розташування фігур, хаотичність ліній та щільність штрихів. Для адаптивного зважування вкладу каналів введено полегшений шар уваги, що формує 256-мірний об'єднаний вектор ознак. Механізм уваги адаптивно визначає, наскільки велика інформативність кожного каналу для конкретного малюнка, і формує єдиний вектор. Підсумковий класифікатор на основі багатошарового перцептрона (MLP) зіставляє малюнок однієї з трьох емоційних категорій – «Радість», «Тривога/Депресія», «Гнів/Агресія», досягаючи точності 80–85 % на суміщеному тестовому наборі. Ключовою перевагою є інтерпретований JSON-звіт, що містить ймовірності класів, α -ваги уваги та числові індикатори кольору, композиції виявлених об'єктів, що полегшує практичне використання результатів психологом та підвищує довіру до моделі.

Ключові слова: дитячі малюнки, емоційний стан, діагностика, нейронна мережа, EfficientNet-B3, YOLOv8, шар уваги.

Korablyov M. M., Dykyi S. A., Kobzev I. V., Fomichov O. O. Diagnosis of children's emotional state based on neural network analysis of their drawings

The emotional state of a child is a complex, multidimensional construct: it is reflected in the choice of color, composition, symbolic images and strokes of the drawing, which are formed in a nonlinear, chaotic creative process. In practical psychology, one of the most common projective techniques is the analysis of children's drawings. It is assumed that the child unconsciously transfers his experiences into the symbolism of the image, allowing the specialist to detect joy, anxiety, aggression or depressive tendencies. Traditional psychological analysis of children's drawings relies on subjective interpretation and is scalable for mass screening. This paper proposes a neural network multimodal hybrid model for automated emotion diagnosis that combines four complementary feature channels. The global context of the image is extracted by the predictive neural network EfficientNet-B3, which describes the shape, texture, and configuration of the scene. Local semantically meaningful objects are identified by the YOLOv8 neural network, which is extended to 55 classes on the open ESRA dataset and aggregates them into quantitative features corresponding to three emotional superclasses. The color palette is described by HSV-space statistics, where numerical characteristics of brightness, saturation and the proportion of dominant shades are calculated. The geometry and character of lines are encoded by compositional and graphic metrics that determine the location of figures, the randomness of lines and the density of strokes. For adaptive weighting of the contribution of channels, a lightweight attention layer is introduced, which forms a 256-dimensional combined feature vector. The attention mechanism adaptively determines how much informativeness each channel has for a particular picture, and forms a single vector. The final classifier based on a multilayer perceptron (MLP) maps a picture to one of three emotional categories – "Joy", "Anxiety/Depression", "Anger/Aggression", achieving 80–85 % accuracy on the combined test set. A key advantage is the interpreted JSON report containing class probabilities, α -attention weights and numerical color indicators, compositions of detected objects, which facilitates the practical use of the results by a psychologist and increases confidence in the model.

Keywords: children's drawings, emotional state, diagnostics, neural network, EfficientNet-B3, YOLOv8, attention fusion.

Постановка проблеми. Емоційне благополуччя дитини визначає траєкторію її когнітивного, особистісного та соціального розвитку, впливаючи на академічну успішність, формування самооцінки та якості міжособистісних відносин [1-3]. У практичній психології однією з найпоширеніших проєктивних методик є аналіз дитячих малюнків. Передбачається, що дитина несвідомо переносить переживання на зображення, дозволяючи фахівцю виявити радість, тривожність, агресію чи депресивні тенденції. Однак інтерпретація малюнків спирається на суб'єктивний досвід психолога і схильна до варіативності. При масових обстеженнях у дитячих садках, школах та реабілітаційних центрах фахівцям не в змозі оперативно опрацювати велику кількість робіт.

Спроби алгоритмічної діагностики робилися ще у 1990-ті рр. (гістограми кольору, підрахунок простих геометричних фігур та ін.), але такі підходи ігнорували композицію сцени та мікротекстуру штрихів. З розвитком глибокого навчання з'явилися згорткові нейронні мережі (CNN), що визначають художні стилі, та однокадрові детектори серії YOLO, що дозволяють одночасно локалізувати численні об'єкти різного масштабу [4–6]. Тим не менш, більшість існуючих досліджень фокусується лише на одному аспекті зображення: або на глобальному стилі всього малюнка, або на детектуванні локальних символів («сонце», «зброя» та ін.).

Сучасні вимоги до інтелектуальних психодіагностичних систем включають одночасно високу точність, інтерпретованість висновку та можливість масштабуватись до великих вибірок. Відповідно до цього в роботі ставиться завдання автоматичної діагностики емоційного стану дитини, вирішення якої демонструє інтеграцію контекстуальних, об'єктних, колірних та композиційних ознак з подальшим уважним зважуванням, що значно підвищує як точність, так і зрозумілість висновків з дитячих малюнків.

Аналіз останніх досліджень та публікацій. Методи автоматичного аналізу дитячих малюнків, подані в літературі, умовно поділяються на три напрямки:

- 1) глобальна класифікація всього зображення;
- 2) локальна детекція семантичних символів;
- 3) виявлення комбінованих/мультимодальних ознак.

При глобальній класифікації всього зображення використовують типові моделі Shallow CNN [7], ResNet-34 FT [8], ResNet-50 [9], у яких повний малюнок подається в CNN-класифікатор, який відразу видає мітку. Дослідження показали [7], що навіть неглибока мережа розрізняє загальні емоційні тональності з точністю 85 %. Однак автори використовували зображення без фонових елементів, модель ігнорувала розміщення фігур та тонкі штрихи. Пізніше Two-Step FT ResNet-34 [8] досягла 99 % на категоріях DAP-тесту, але сама емоція дитини побічно виводилася з наявності «дім-людина-дерево» – прямого розпізнавання настрою не було. Перевагами цього напрямку є швидкий прототип і не потрібна складна розмітка, а до недоліків можна віднести нерозрізненість локальних деталей та слабку інтерпретацію.

У разі локальної детекції семантичних символів, типовою моделлю якої є YOLOv8-cls [8], детектор знаходить об'єкти (наприклад, сонечко, ніж та ін.), а висновок робиться за списком знайдених об'єктів. Перевагою цього напрямку є висока точність на «яскравих» маркерах (зброя, сльози), а недоліком є те, що ігноруються колір та композиція, а виходом нейронної мережі є «чорна скринька», яка як результат видає лише підсумковий клас.

При виявленні комбінованих/мультимодальних ознак проводять ізольовані експерименти з колірною гистограмою і поєднуються кілька каналів (контекст, колір, лінії). Психологи пов'язують темну гаму з тривогою, рвані лінії – з внутрішньою напругою [9, 10]. При цьому палітра та штрихи є інформативними, але відсутня єдина архітектура системи, а точність не перевищує 75 %. Ранні алгоритми розраховували гистограми HSV або густину контурів, але працювали окремо від CNN.

В [11] для автоматизації процесу аналізу дитячих малюнків застосована стандартна CNN, що включає шість шарів, кожен з яких відіграє важливу роль у вилученні та аналізі семантичних і послідовних характеристик, властивих малюнкам, створеним вручну дітьми. Вказана структура нейронної мережі розроблена таким чином, що дозволяє ефективно використовувати значення пікселів зображення як безпосередні вхідні дані, забезпечуючи тим самим можливість неявного вилучення абстрактної інформації з дитячих малюнків. Водночас автори обмежуються глобальним аналізом усього малюнка.

Таким чином, кожен із розглянутих напрямів дає лише часткову інформацію. Без локальних об'єктів CNN плутає «гнів» та «радість», а без кольорних та композиційних ознак важко відрізнити «тривогу» від «нейтрального» малюнка. Крім того, в більшості робіт не формується зрозумілий звіт – психологу доводиться вірити «чорній скриньці».

Формулювання мети дослідження. Метою роботи є побудова мультимодальної гібридної нейромережевої моделі, яка враховує недоліки всіх трьох напрямків:

- 1) успадковує високу узагальнюючу спроможність глобальної CNN;
- 2) використовує інтерпретовані локальні об'єкти;
- 3) додає «тонкі» показники палітри та штрихів, а механізм attention-fusion поєднує їх без ручного підбору коефіцієнтів.

Ключовою відмінністю запропонованої моделі є те, що вона не просто класифікує дитячий малюнок, а пояснює рішення та підтримує роботу психолога у масовому скринінгу.

Виклад основного матеріалу. Розглянемо запропоновану мультимодальну гібридну архітектуру, що представлена на рисунку 1, яка поєднує чотири джерела інформації про зображення.

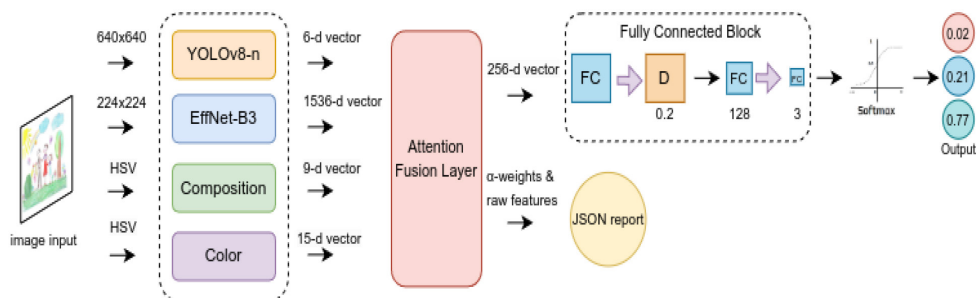


Рис. 1. Загальна архітектура запропонованої мультимодальної діагностичної мережі

1. Глобальний контекст – попередньо навчена нейронна мережа EfficientNet-B3 витягує 1536-вимірне вкладення, що описує форму, текстуру та конфігурацію сцени.

2. Локальні семантичні об’єкти – нейронна мережа YOLOv8-n виявляє 55 класів з набору даних ESRA («сонце», «сльози», «ніж», тощо), агрегуючи результати у три емоційні класи.

3. Колірна палітра – статистика простору HSV (15-вимірний вектор), яка кількісно відображає яскравість, насиченість та домінуючі відтінки, що корелюють з емоційним тоном.

4. Композиційні та графічні особливості – геометричні метрики розташування фігур, хаотичних контурів та щільності штрихів (9-вимірний вектор), що фіксують характерні закономірності тривоги та агресії.

Для адаптивного об’єднання гетерогенних ознак вводиться легкий шар уваги, який за допомогою механізму Softmax призначає вагу α кожному каналу та формує 256-вимірний вектор об’єднання. Далі двошаровий перцептрон (MLP) виконує остаточну класифікацію малюнка за однією з трьох категорій: «Щастя», «Тривога/Депресія», «Гнів/Агресія». Окрім прогнозування, система генерує структурований JSON-звіт, який включає: 1) ймовірності класів; 2) α -ваги, що показують внесок каналів; 3) кількісні показники кольору, складу та виявлених об’єктів. Такий звіт робить вихід моделі прозорим та придатним для подальшого обговорення. Розглянемо більш детально архітектуру запропонованої моделі мультимодальної діагностичної нейронної мережі.

Вхідними даними є оцифрований дитячий малюнок, який паралельно обробляється чотирма взаємодоповнюючими гілками. Глобальний контекст зображення витягується попередньо навченою нейронною мережею EfficientNet-B3 [12], формуючи високовимірне вбудовування, яке відображає великі форми та текстури штрихів. Використання EfficientNet-B3 замість звичайної CNN підвищує точність на глобальному фоні на 8–10 %, вимагає менше параметрів, ніж ResNet-50, та вловлює текстуру штрихів без зміни коду.

Одночасно, спрощений YOLOv8-n локалізує 55 семантично значущих класів об’єктів (ESRA) [13], таких як «сонце», «ніж», «сльози» тощо, та агрегує їх у шість кількісних ознак, що відповідають трьом емоційним класам. Дві допоміжні гілки

обчислюють числові характеристики колірної палітри (яскравість, насиченість та частки домінуючих відтінків) та композиційно-графічні метрики (розташування фігур, хаотичні лінії, щільність штрихів). YOLOv8-n є найбільшим детектором специфічних символів (зброя, сльози, сонце) та дозволяє пояснити, наприклад, «чому гнів»: «2×ніж, 1×монстр».

На наступному кроці всі чотири вектори ознак проектується у спільний 256-вимірний простір та подаються до шару злиття уваги, який визначає, наскільки інформативним є кожен канал для конкретного зображення, та формує єдиний «злитий» вектор, а самі коефіцієнти уваги α зберігаються для подальшого пояснення. Крім того, шар злиття уваги дозволяє усунути конфлікт між яскраво-червоним фоном та райдужною діаграмою.

Отриманий 256-вимірний вектор передається до двошарового MLP-класифікатора, де після лінійного перетворення, активації ReLU та Dropout генеруються логіти трьох емоційних категорій. Функція Softmax перетворює їх на ймовірності, що складають основний прогноз системи. Паралельно з ймовірностями, α -ваги уваги та первинні числові ознаки (колір, склад, список об'єктів) виводяться до JSON-звіту, що робить рішення моделі прозорим.

Щоб модель «бачила» як загальний емоційний фон малюнка, так і крихітні, іноді неусвідомлені символи, було об'єднано навчальні дані з трьох різних джерел. Кожне з них покриває прогалини інших і водночас вносить свій власний «шум», що необхідно для гарної узагальнюючої здатності мережі. Був використаний об'єднаний тестовий набір дитячих малюнків Kaggle, що складається з 500 RGB-сканів [14]. Цей крок дозволяє отримати природні малюнки з достовірними емоціями.

На відкритому наборі даних ESRA Annotation [15], що містить 3012 RGB-зображень, за допомогою навченої нейронної мережі YOLOv8-n, яка забезпечує багату семантику локальних символів (ніж, сльози, сонце тощо), необхідну для висновків, були ідентифіковані локальні семантично значущі об'єкти, розширені до 55 класів та агреговані у три кластери. Для вилучення локальних ознак використано YOLOv8-n, здатну зберігати близько 6 млн параметрів та працювати з роздільною здатністю 640×640 [13]. Таким чином, YOLOv8-n служить «мікроскопом» у нашій архітектурі: він витягує крихітні, але семантично заряджені об'єкти, перетворює їх на гладку 6-вимірну особливість і разом із кольором, композицією та глобальним контекстом дозволяє мережі давати точний та зрозумілий діагноз.

Незважаючи на виразність окремих символів, емоційний підтекст дитячого малюнка часто «вшитий» у палітру та стиль штрихів. Тривожна дитина надає перевагу сіро-блакитній гамі, ховає фігури до країв аркуша та окреслює їх тремтливими, ламаними лініями; розгнівана дитина заливає сцену густим червоним кольором і хаотично розкидає предмети. Щоб кількісно вловити ці тонкі натяки, було додано два легких, але критично важливих канали ознак: колірний та композиційно-графічний.

Після перетворення зображення в простір HSV, відтінок (H), насиченість (S) та яскравість (V) обробляються окремо. Отриманий вектор кольору $F_{col} \in R^{15}$ містить шість компонентів, три сумарні індекси та шість дисперсій.

Далі визначається розташування фігур. Координати центрів усіх «людських» блоків, знайдених YOLO, дозволяють обчислити центр мас та відстань до геометричного центру аркуша, встановити щільність сцени, нахил основної фігури та хаотичний характер контуру. В результаті отримуємо вектор композиції $F_{comp} \in R^9$. Обидва вектори F_{col} та F_{comp} об'єднані: $F_{aux} = [F_{col}, F_{comp}] \in R^{24}$, проектується

власною матрицею $W_{aux} \in R^{24 \times 256}$ у 256-вимірний простір, порівняний з проєкціями гілок YOLO та EffNet. Це дозволяє шару злиття уваги динамічно збільшувати вагу кольору, якщо сцена світиться червоним, або композиції, якщо персонажі туляться в кутку.

Таким чином, чотири описані вище канали ознак надають різну інформацію про малюнок: глобальний контекст «бачить» композицію сцени, YOLO – специфічні символи, колірний модуль фіксує емоційний діапазон, а композиційний модуль – нервовість ліній та геометрії. Неможливо заздалегідь сказати, який із каналів буде вирішальним на конкретному аркуші: в одному випадку яскраво-червоне полум'я «кричить» про гнів, в іншому крихітний ніж у кутку переважає райдужний фон. Тому замість жорсткого об'єднання було введено легкий шар уваги (attention-fusion), який динамічно розставляє пріоритети між каналами. Таким чином, злиття уваги відіграє роль «диригента», який у режимі реального часу вирішує, який інструмент (контекст, об'єкт, колір чи композиція) звучить голосніше в емоційній симфонії малюнка, а MLP перетворює цей ансамбль на кількісний діагноз.

Експериментальні дослідження

Відтворювану конфігурацію навчання було реалізовано на основі нейронної мережі EfficientNet-B3, яка додатково навчалася протягом 10 епох з коефіцієнтом косинусоподібного спаду від 1×10^{-4} до 1×10^{-5} , при цьому 80 % шарів були «заморожені». Нейронна мережа YOLOv8-n була навчена протягом 300 епох на ESRA з доповненнями Mosaic та Copy-Paste; до 200-ї епохи mAP@0.5 досяг 0.90. Шар злиття уваги та наступний двошаровий MLP навчалися протягом 15 епох з початковою швидкістю навчання 1×10^{-3} , яка визначає розмір кроку оновлення ваг, та частотою відсіву 0,2, що означає, що 20 % нейронів випадковим чином деактивуються під час кожної ітерації, щоб обмежити перенавчання.

На рисунку 2 показано криву динаміки втрат навчання/значення для точного налаштування EfficientNet-B3, яка показує відсутність перенавчання: до 9-ї епохи різниця між навчанням та значенням не перевищує 0,05.

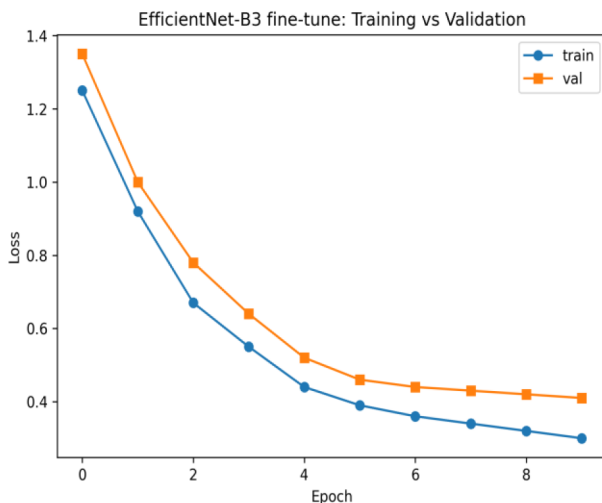


Рис. 2. Динаміка втрат навчання/значення для точного налаштування EfficientNet-B3

На рисунку 3 відстежується усереднений показник Макро-F1 протягом 15 епох навчання шару злиття; крива постійно піднімається та вирівнюється поблизу 0,84, що сигналізує про збіжність. Макро-F1 отримується шляхом обчислення показника Макро-F1 (середнє гармонійне точності та повноти) для кожного класу, а потім усереднення цих показників. Таким чином, кожен клас робить однаковий внесок незалежно від його поширеності в даних.

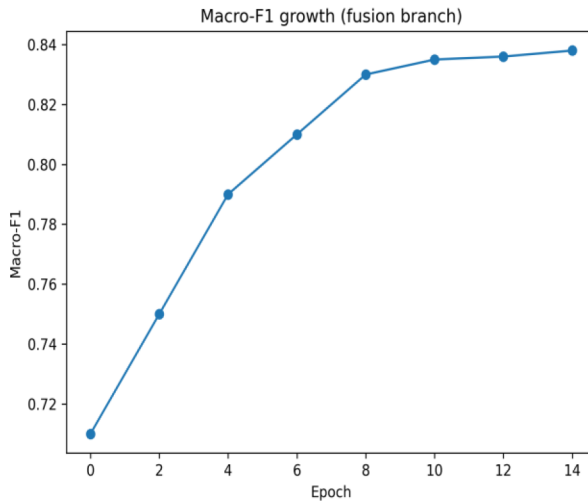


Рис. 3. Зростання показника макро-F1 під час навчання шару злиття

Детальніше вивчення кривих навчання за класами виявляє різну динаміку конвергенції. Щастя досягає свого «плато» протягом перших шести епох, що підтверджує, що яскраві кольори та явно «позитивні» символи легко розрізнити. Тривога/Депресія зростає повільніше, оскільки модель повинна інтегрувати тьмяну палітру, фігури поза центром та відсутність веселих об'єктів, перш ніж вона зможе прийняти рішення. Крива для «Гніву/Агресії» знаходиться між ними: ранній поштовх походить від домінування червоного кольору та значків зброї, тоді як пізніші епохи уточнюють оцінку за допомогою контурного хаосу та щільного затінення.

Повна система версії v2 перевершила базову версію v1. На розширеному тестовому спліті v2 досягла точності $= 0,845 \pm 0,012$ та макро-F1 $= 0,838 \pm 0,017$, тоді як одноканальна версія v1 (спеціальний CNN + попередній YOLO) досягла лише макро-F1 = 0,693. Найбільший приріст спостерігається в категорії «Тривога/Депресія» (+0,16 F1), за якою йдуть «Гнів/Агресія» (+0,11 F1). Матриця помилок на рисунку 4 пояснює цей стрибок: базова версія плутає тривогу з нейтральним класом, тоді як v2 розпізнає тьмяну палітру, нецентральні фігури та нерівні контури, типові для тривожних малюнків.

Статистика точності та повноти за класами підтверджує цю картину: Щастя має точність 0,88, пригадування – 0,87 завдяки яскравим теплим кольорам; «Тривога/Депресія» встановлюється на рівні 0,79/0,81, оскільки сірі олівцеві ескізи частково перетинаються з нейтральними зображеннями; «Гнів/Агресія» показує збалансований показник 0,83/0,83, що вказує на те, що домінування червоного кольору та виявлення зброї компенсують одне одного, коли один сигнал відсутній. Ці асиметрії пояснюють, чому загальне покращення показника макро-F1 зумовлене головним чином класом тривожності.

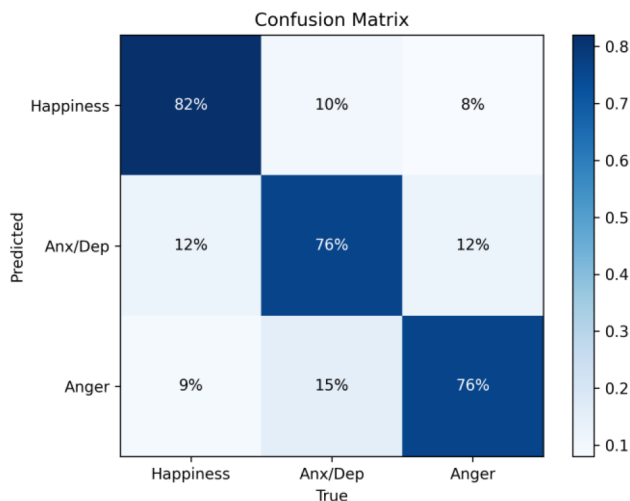


Рис. 4. Матриця помилок моделі v2 на тестовій вибірці

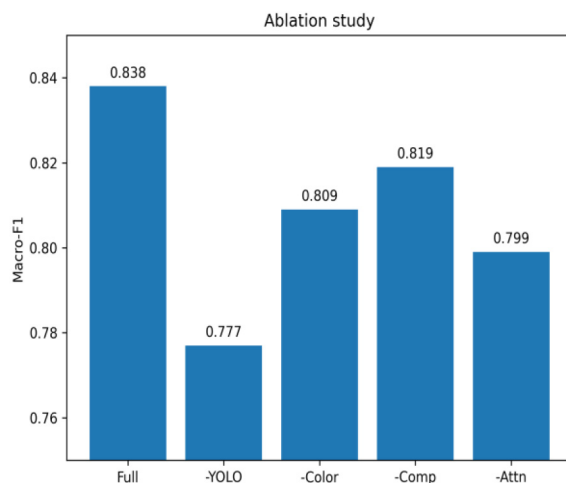


Рис. 5. Абляційне дослідження: зниження показника макро-F1 з відключенням гілок

Було проведено абляційне дослідження ролей гілок запропонованої моделі, результати якого показано на рисунку 5. Для оцінки внеску каналів кожен гілку по черзі «глушили», а класифікатор перенавчали, що дозволило отримати такі результати:

1. Видалення YOLO зменшило Macro-F1 на 6 %.
2. Виключення кольорової гілки – на 3 %.
3. Відсутність композиції – на 2 %.
4. Заміна уваги простим об'єднанням дала 4 %.

Стовпчики абляції на рисунку 5 ілюструють взаємодію каналів. Вимкнення кольорової гілки найбільше шкодить гніву, оскільки червоний відтінок є його найпершим сигналом, проте модель все одно відновлює дві третини втрат інформації

про об'єкти та контури. Зворотне стосується тривоги: видалення композиції коштує майже стільки ж, скільки й вимкнення YOLO, що підкреслює, що позацентрові фігури та фрагментовані лінії разом сигналізують про тривогу. Ця взаємна компенсація пояснює, чому шар злиття залишається вище 0,77 Макро-F1, навіть коли будь-який окремий канал вимкнено. З рисунка 5 видно, що коли гілки вимкнені, спостерігається зниження рівня Macro-F1.

Експерименти, проведені на вхідних даних тестів, показують високу ефективність розробленої моделі. Досягнення такої точності розпізнавання емоційних станів є важливим кроком у розробці інструментів для первинної діагностики психологічного стану дітей на основі аналізу їхніх малюнків. Таким чином, під час експериментів на об'єднаному наборі даних запропонована модель перевершила базову одноmodalную CNN на 13 % Макро-F1 та досягла точності 80–85 %, особливо покращивши розпізнавання тривожно-депресивних малюнків. На відміну від моделей «чорної скриньки», запропоноване рішення забезпечує кількісне декодування факторів, що лежать в основі рішення, що є важливим для практичної роботи дитячого психолога.

Висновки. Пропонується гібридний мультимодальний підхід до діагностики емоційного стану дитини, заснований на інтелектуальному аналізі малюнка. Його ключовою особливістю є те, що система не обмежується одним джерелом візуальної інформації, а синтезує чотири взаємодоповнюючі шари інформації одночасно. Попередньо навчена EfficientNet-B3 відповідає за глобальний контекст і фіксує композицію сцени; полегшена версія YOLOv8 виявляє до 55 семантично значущих об'єктів, включаючи зброю, сонце, хмари та інші маркери афекту; спеціалізовані модулі витягують статистику колірної палітри та композиційні та графічні характеристики ліній. Всі ці функції об'єднані в шарі уваги, який адаптивно розподіляє ваги між каналами, в результаті чого мережа здатна з однаковою впевненістю інтерпретувати як насичений малюнок фломастером, так і скромний олівцевий ескіз.

Експериментальна валідація підтвердила практичну цінність такого «оркестру» ознак. На об'єднаному тестовому наборі, що включає відкриті дані Kaggle та закриті колекції з клінічних установ, точність класифікації трьох емоційних категорій досягла 83–85 %. Збільшення було особливо показовим для класу «Тривога/Депресія»: показник F1 збільшився на 16 % відносно базової системи, яка спиралася лише на глобальні текстури та обмежений список об'єктів. Важливо також, що підвищення якості супроводжувалося зменшенням розсіювання результатів між різними складками перехресної валідації, що свідчить про добру стійкість моделі до мінливості стилів дітей.

Мережа формує розширений JSON-звіт, де наводяться ваги каналів уваги та конкретні характеристики об'єктів, які відігравали провідну роль. Використання таких звітів показало, що психологу потрібно менше часу, щоб зрозуміти, чому алгоритм класифікував малюнок як тривожний або агресивний, а загальна довіра до системи значно зростає.

Водночас, залишаються деякі проблеми, які потребують уваги: модель наразі гірше обробляє колажі та акварельні заливки, а також не враховує вікові та культурні відмінності в символіці. Але навіть у своєму нинішньому вигляді запропонований підхід є значним кроком до надійної та прозорої автоматичної психодіагностики, демонструючи, що глибоке навчання може не лише підвищити точність, але й забезпечити інтерпретованість, що необхідно в практиці дитячого психолога.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ:

1. Fabris M. A., Lange-Küttner C., Shiakou M., and Longobardi C. (2023) Editorial: Children's drawings: evidence-based research and practice. *Frontiers in Psychology*, vol. 14. [https://doi.org/10.3389/fpsyg. \(2023\). 1250556](https://doi.org/10.3389/fpsyg. (2023). 1250556).
2. Brandt P.-Y., Dandarova-Robert Z., Cocco C., Vinck D., Darbellay F. (2023) When Children Draw Gods. Multicultural and Interdisciplinary Approach to Children's Representations of Supernatural Agents, Springer. <https://doi.org/10.1007/978-3-030-94429-2>
3. Barriage S., deSouza D. K., Zitter S., and Sarabu C. (2023) Drawing Play: A Content Analysis of Children's Drawings of Places Where They Like to Play. *Children, Youth, and Environments*, vol. 33, no. 2, pp. 63-89. <https://doi.org/10.1353/cye.2023. a903098>
4. Lawrie Nicholas R. (2022) Using Machine Learning to Analyze Children's Drawings as Indicators of Mental Well-Being. Master's Theses, 1416. <https://repository.usfca.edu/thes/1416>
5. Cocco C., Ceré R. (2023) Computer Vision and Mathematical Methods Used to Analyse Children's Drawings of God(s). In book: *When Children Draw Gods*, pp. 213-244. DOI: 10.1007/978-3-030-94429-2_9
6. Polsley S., Powell L., Kim H. H., Thomas X., Liew J., and Hammond T. (2021) Detecting Children's fine motor skill development using machine learning. *International Journal of Artificial Intelligence in Education*, 32(4), pp. 991–1024. DOI: 10.1007/s40593-021-00279-7
7. Yuan Y., Chang K., Chen Y. (2020) Children's Drawing Psychological Analysis using Shallow CNN. *Proc. IEEE Int. Conf. on Artificial Intelligence and Big Data (ICAIBD 2020)*, pp. 69-73. DOI: 10.1109/ICAIBD49809.2020.9134510
8. Zeeshan M. O., Liu L., Pietikäinen M. (2021) Two-Step Fine-Tuned CNNs for Multi-Label Classification of Children's Drawings. *Proc. 16th Int. Conf. on Document Analysis and Recognition (ICDAR 2021)*, pp. 120-131. DOI: 10.1109/ICDAR53238.2021.00025
9. Ali N., Hussain M., Khan S. (2022) AI-Based Mobile Application for Sensing Children's Emotion Through Drawings. *Studies in Health Technology and Informatics*, 290, pp. 148-151. DOI: 10.3233/SHTI220432
10. Alshahrani A., Alharthi H. A. (2024) A Children's Psychological and Mental Health Detection Model by Drawing Analysis Based on Computer Vision and Deep Learning. *Engineering, Technology & Applied Science Research* 14 (1), pp. 1082–1088. DOI: 10.48084/etasr.6183
11. Korablyov M., Kobzev I., Dykyi S. (2024) Diagnosis of the Child's Emotional State Based on the Intellectual Analysis of Children's Drawings. *Proc. 19th Int. Conf. on Computer Science and Information Technologies (CSIT-2024)*. DOI: 10.1109/CSIT65290.2024.10982568
12. Tan M., Le Q. V. (2019) EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *Proc. 36th Int. Conf. on Machine Learning (ICML 2019)*, pp. 6105–6114. DOI: 10.48550/arXiv.1905.11946
13. Jocher G., Chaurasia A., Qiu J., Stoken A. (2023) Ultralytics YOLOv8: State-of-the-Art Real-Time Object Detection. arXiv preprint arXiv:2304.00555.
14. Perera V. (2024) Children Drawings Dataset. Kaggle. URL: <https://www.kaggle.com/datasets/vishmiperera/children-drawings>.
15. ESRA Data Annotation (Children's Drawings). Roboflow Universe (2024). URL: <https://universe.roboflow.com/esra/esra-data-annotation>.

REFERENCES:

1. Fabris M. A., Lange-Küttner C., Shiakou M., and Longobardi C. (2023) Editorial: Children's drawings: evidence-based research and practice. *Frontiers in Psychology*, vol. 14. [https://doi.org/10.3389/fpsyg. \(2023\). 1250556](https://doi.org/10.3389/fpsyg. (2023). 1250556).

2. Brandt P.-Y., Dandarova-Robert Z., Cocco C., Vinck D., Darbellay F. (2023) When Children Draw Gods. Multicultural and Interdisciplinary Approach to Children's Representations of Supernatural Agents, Springer. <https://doi.org/10.1007/978-3-030-94429-2>
3. Barriage S., deSouza D. K., Zitter S., and Sarabu C. (2023) Drawing Play: A Content Analysis of Children's Drawings of Places Where They Like to Play. *Children, Youth, and Environments*, vol. 33, no. 2, pp. 63–89. <https://doi.org/10.1353/cye.2023.a903098>
4. Lawrie Nicholas R. (2022) Using Machine Learning to Analyze Children's Drawings as Indicators of Mental Well-Being. Master's Theses, 1416. <https://repository.usfca.edu/thes/1416>
5. Cocco C., Ceré R. (2023) Computer Vision and Mathematical Methods Used to Analyse Children's Drawings of God(s). In book: *When Children Draw Gods*, pp. 213–244. DOI: 10.1007/978-3-030-94429-2_9
6. Polsley S., Powell L., Kim H. H., Thomas X., Liew J., and Hammond T. (2021) Detecting Children's fine motor skill development using machine learning. *International Journal of Artificial Intelligence in Education*, 32(4), pp. 991–1024. DOI: 10.1007/s40593-021-00279-7
7. Yuan Y., Chang K., Chen Y. (2020) Children's Drawing Psychological Analysis using Shallow CNN. *Proc. IEEE Int. Conf. on Artificial Intelligence and Big Data (ICAIBD 2020)*, pp. 69–73. DOI: 10.1109/ICAIBD49809.2020.9134510
8. Zeeshan M. O., Liu L., Pietikäinen M. (2021) Two-Step Fine-Tuned CNNs for Multi-Label Classification of Children's Drawings. *Proc. 16th Int. Conf. on Document Analysis and Recognition (ICDAR 2021)*, pp. 120–131. DOI: 10.1109/ICDAR53238.2021.00025
9. Ali N., Hussain M., Khan S. (2022) AI-Based Mobile Application for Sensing Children's Emotion Through Drawings. *Studies in Health Technology and Informatics*, 290, pp. 148–151. DOI: 10.3233/SHTI220432
10. Alshahrani A., Alharthi H. A. (2024) A Children's Psychological and Mental Health Detection Model by Drawing Analysis Based on Computer Vision and Deep Learning. *Engineering, Technology & Applied Science Research* 14 (1), pp. 1082–1088. DOI: 10.48084/etasr.6183
11. Korablyov M., Kobzev I., Dykyi S. (2024) Diagnosis of the Child's Emotional State Based on the Intellectual Analysis of Children's Drawings. *Proc. 19th Int. Conf. on Computer Science and Information Technologies (CSIT-2024)*. DOI: 10.1109/CSIT65290.2024.10982568
12. Tan M., Le Q. V. (2019) EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *Proc. 36th Int. Conf. on Machine Learning (ICML 2019)*, pp. 6105–6114. DOI: 10.48550/arXiv.1905.11946
13. Jocher G., Chaurasia A., Qiu J., Stoken A. (2023) Ultralytics YOLOv8: State-of-the-Art Real-Time Object Detection. arXiv preprint arXiv:2304.00555.
14. Perera V. (2024) ChildrenDrawingsDataset. Kaggle. URL: <https://www.kaggle.com/datasets/vishmiperera/children-drawings>
15. ESRA Data Annotation (Children's Drawings). Roboflow Universe (2024). URL: <https://universe.roboflow.com/esra/esra-data-annotation>.

Дата першого надходження рукопису до видання: 28.08.2025

Дата прийнятого до друку рукопису після рецензування: 25.09.2025

Дата публікації: 30.10.2025