

УДК 004.82:[004.932.2:159.942]

ДІАГНОСТИКА ЕМОЦІЙНОГО СТАНУ ДІТЕЙ НА ОСНОВІ МУЛЬТИМОДАЛЬНОГО ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ МАЛЮНКІВ

Дикий С. Д.

e-mail: stanislav.dykyi@nure.ua

Науковий керівник – д.т.н., проф. Корабльов М. М.

Харківський національний університет радіоелектроніки, каф. ІСТ
м. Харків, Україна

This paper proposes a hybrid multimodal model for automated assessment of children's emotional state from drawings. The approach combines EfficientNet-B3 global features, YOLOv8 symbol detection, HSV palette statistics and composition/graphical metrics. A lightweight attention-fusion layer weights channels and forms a 256-D embedding, which a two-layer MLP classifies into Happiness, Anxiety/Depression and Anger/Aggression. The attention weights quantify the contribution of each feature channel and support explainable decision-making. Experiments on a mixed test set (including Kaggle) reach 80–85% accuracy and produce an interpretable JSON report, which is useful for practical screening in educational and rehabilitation settings.

Для автоматизованої діагностики емоційного стану дітей за їхніми малюнками запропоновано гібридну мультимодальну нейромережеву модель, що поєднує глобальний контекст зображення (EfficientNet-B3), виявлення локальних символів (YOLOv8), статистичний опис палітри у просторі HSV та композиційно-графічні метрики зі зваженим злиттям ознак на основі механізму уваги.

Емоційний стан проявляється у виборі кольорів, композиції, символах і штрихах; традиційна експертна оцінка суб'єктивна й погано масштабується для шкільного чи реабілітаційного скринінгу [1].

Модель об'єднує чотири канали ознак: попередньо навчена EfficientNet-B3 [2] створює 1536-вимірний дескриптор сцени й штрихів; YOLOv8-n [3] виявляє 55 символічних класів, стискаючи їх у 6-вимірний вектор; додаткові модулі додають 15 HSV-статистик кольору та 9 композиційних метрик (щільність штрихування, хаотичність контурів, симетрія, геометрія розміщення). Таке узгоджене злиття композиції, кольору, символів і характеру штрихів відтворює експертний підхід, але робить його об'єктивним і придатним до масштабування.

Для адаптивного злиття різнорідних ознак застосовано lightweight attention-fusion шар: кожен канал проєктується у спільний латентний простір, після чого механізм уваги обчислює ваги внеску каналів (softmax-нормалізація) та формує узгоджений 256-вимірний fused-вектор. На виході двошаровий MLP-класифікатор (256→128→3, ReLU та dropout) генерує логіти трьох категорій: «Щастя», «Тривожність/депресивність», «Гнів/агресія», а

Softmax перетворює їх на ймовірності класів. Перед подачею у гілки зображення автоматично обрізається за областю непорожніх пікселів із невеликим полем та масштабується до 224×224 (EfficientNet) і 640×640 (YOLO), що стабілізує якість ознак на різних сканах. Важливою особливістю є формування інтерпретованого JSON-звіту з імовірностями, вагами уваги та переліком виявлених об'єктів/індикаторів — це спрощує використання результатів практичним психологом.

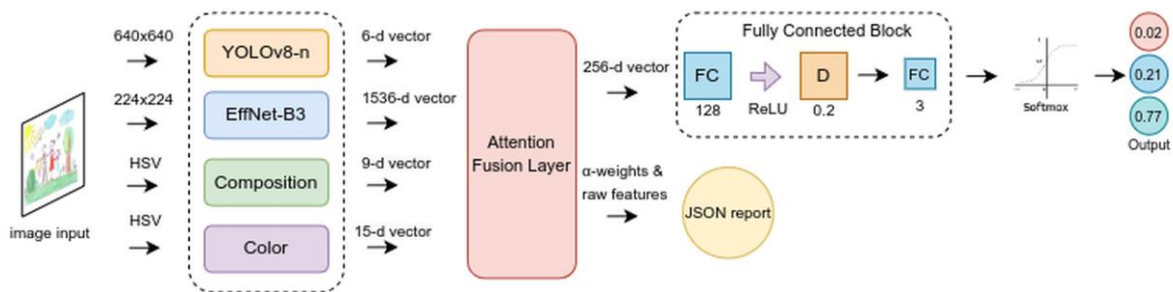


Рисунок 1 – Загальна архітектура запропонованої мультимодальної нейромережевої моделі діагностики

Експериментальна перевірка на комбінованій тестовій вибірці (зокрема, дані Kaggle) показала точність класифікації 80–85% та покращення розпізнавання класу «Тривожність/депресивність» завдяки поєднанню палітри, композиції та локальних символів. Модель формує інтерпретований JSON-звіт (ймовірності, ваги каналів, ключові ознаки), що підвищує придатність для практичного застосування. До обмежень відносяться складнощі з колажами та акварельними заливками, а також відсутність метаданих про вік і культурний контекст дитини.

Запропоновано мультимодальну гібридну модель для діагностики емоційного стану дітей за малюнками, яка об'єднує глобальні ознаки (EfficientNet-B3), локальні символи (YOLOv8), палітру HSV та композиційно-графічні метрики із attention-fusion злиттям. Поєднання різнорідних інформаційних каналів забезпечує конкурентну точність класифікації, підвищує пояснюваність отриманих результатів і є особливо важливим для задач масового психологічного скринінгу.

Список використаних джерел:

1. U. Podobnik, J. Jerman, J. Selan. Understanding analytical drawings of preschool children: the importance of a dialogue with a child. *International Journal of Early Years Education*, 32(1), 2024, 189–203.
2. M. Tan, Q. V. Le. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *Proc. ICML*, 2019.
3. G. Jocher et al. Ultralytics YOLOv8: State-of-the-Art Real-Time Object Detection. *arXiv:2304.00555*, 2023.