

КЛАС-ОРІЄНТОВАНИЙ МЕТОД АУГМЕНТАЦІЇ ФУНДУС-ЗОБРАЖЕНЬ

¹Харківський національний університет радіоелектроніки

Запропоновано інноваційний клас-орієнтований метод аугментації фондус-зображень. Описано переваги цього методу у порівнянні з відомими. Обґрунтовано та підібрано стратегію аугментації для зображень з наявними ознаками глаукоми, катаракти, діабетичної ретинопатії та здорового ока відповідно до специфіки кожного класу. Аугментації наближені до реальних клінічних варіацій. Покращено стійкість саме до тих варіацій, які характерні для конкретної хвороби. Підвищено чутливість і специфічність саме до патологій на медичних зображеннях. Розроблено нейромережеву модель з використанням клас-орієнтованого методу. До базової мережі EfficientNetB3 методами трансферного навчання додано шари GlobalAveragePooling, Dropout з rate 0,5, Dense з 1024 нейронами, l2-регуляризацією зі значенням 0,001 та функцією активації relu, класифікаційний шар Dense з 4 нейронами та функцією активації softmax. Половину шарів базової моделі заморожено. Модель компілювалася за допомогою компілятора Adam з початковим learning rate 0,0001 та функцією втрач категоричної кроссентропії. Під час попередньої обробки розміри зображень зменшувалися до 224×224. Нормалізація зображень автоматично відбувалася під час генерації даних для навчання. Для коригування навчання моделі використовувалися такі функції зворотного виклику: ModelCheckpoint — для збереження найкращої моделі, EarlyStopping — для зупинки за відсутності покращення метрики val accuracy протягом 15 епох, ReduceLROnPlateau — для зменшення learning rate у 3 рази за стагнації. В результаті навчання отримано високі показники метрик. Навчену модель стиснено шляхом квантифікації для подальшого використання на мобільних пристроях та на пристроях з обмеженими можливостями. Запропонований підхід дозволяє підвищити загальну точність, робастність нейронної мережі, подолати обмеження традиційного методу.

Ключові слова: машинне навчання, аугментація, нейронні мережі, медичні зображення.

Вступ

Однією з проблем у сучасному машинному навчанні є незбалансованість даних. У разі незбалансованості окремі класи представлені недостатньо. Це призводить до зниження якості класифікації у випадках з рідкісними або малими класами. Традиційні методи аугментації, що застосовуються однаково до всіх класів, не завжди ефективні для таких задач. Необхідно застосовувати підхід, за якого долаються складнощі з балансуванням. В оглядах зазначається, що методи аугментації поділяються на кілька груп: прості геометричні та фотометричні перетворення, генеративні моделі, методи на рівні ознак [1], [2]. Просте збільшення кількості зображень через стандартну аугментацію дозволяє підвищити recall, але може спричинити перенавчання [3], [4]. Науковці застосовують об'єктно-рівневі методи для сегментації та виявлення об'єктів [5], [6]. Генеративні методи, зокрема умовні GAN (cGAN) та дифузійні моделі, дозволяють створювати нові приклади для заданих класів. Такі підходи особливо поширені в медичній діагностиці, де кількість прикладів обмежена [7], [8]. За такого підходу недоліком є значне спотворення вхідного зображення. Через це дослідники наголошують на необхідності додаткової валідації синтетичних даних за допомогою кількісних та якісних критеріїв [9]. Ще одним напрямом розвитку є аугментація на рівні ознак, яка передбачає модифікацію векторів ознак залежно від класової належності. Наведений механізм забезпечує підвищення різноманітності репрезентацій без створення нових зображень [10], [11]. Поєднання різних стратегій відкриває нові можливості для побудови універсальних систем обробки даних [12]—[14].

Метою статті є розроблення клас-орієнтованого методу аугментації фондус-зображень для здійснення оптимізації, формування оптимізаційного потенціалу, підвищення робастності та точності нейронної мережі.

Результати дослідження

Аугментація даних є важливою технікою для підвищення узагальнювальної здатності моделей глибокого навчання, зокрема для задач, де доступна обмежена кількість тренувальних даних. Ефективна аугментація не тільки збільшує обсяг даних, а й має значний позитивний вплив на робастність та оптимізацію моделі [13], [14]. Без аугментації модель намагається ідеально підлаштуватися під кожний тренувальний приклад. Її навчання стає не робастним, з безліччю локальних мінімумів. Стандартні підходи до аугментації даних у задачах комп'ютерного зору базуються на застосуванні універсальних стратегій перетворень. Такі стратегії, які зазвичай включають геометричні (обертання, зсуви) та фотометричні (зміна яскравості, контрасту) перетворення, не враховують фізичну природу та специфічні візуальні характеристики зображень, що аналізуються. У галузі медичної діагностики такий підхід є неоптимальним та інженерно невинуватим. Кожна модальність візуалізації — оптична когерентна томографія, фундус-фотографія — має унікальний набір артефактів, шумових характеристик та значущих ознак. Застосування універсальних аугментацій може призвести до негативних наслідків. Перший негативний наслідок — це генерація нереалістичних прикладів. Система генерує дані, які не можуть існувати в реальному світі. Наприклад, застосування адитивного гаусівського шуму до зображень оптичної когерентної томографії не моделює реальний спекл-шум, що має мультиплікативну природу і є когерентним артефактом лазерного сканування. Такий вид аугментації вводить в оману нейронну мережу, змушуючи її вивчати нерелевантні ознаки. Другий негативний наслідок — руйнування ключових ознак хвороби. Агресивна зміна контрасту корисна для загальних фотографій. Проте ця аугментація може повністю приховати аневризми з низьким контрастом на фундус-зображенні або згладити ледь помітні межі між шарами сітківки на зображеннях оптичної когерентної томографії, що є ключовою ознакою глаукоми. Третя проблема — збільшення семантичного розриву. Універсальні аугментації можуть створювати зображення, що є синтаксично коректними, але семантично хибними для лікаря-експерта, що ускладнює інтерпретацію поведінки моделі. Таким чином, виникає актуальна задача комп'ютерної інженерії — перейти від універсального підходу до розробки науково обґрунтованого методу клас-орієнтованої — модально-специфічної аугментації. Мета цього методу — не просто збільшити варіативність даних, а симулювати реалістичні варіації, що зустрічаються у клінічній практиці, для кожного типу зображень індивідуально. Переваги клас-орієнтованої аугментації у порівнянні зі звичайною подано у табл.

Порівняльний аналіз звичайної та клас-орієнтованої аугментації

Критерій	Звичайна аугментація	Клас-орієнтована аугментація
Підхід	однакові трансформації застосовуються до всіх класів	для кожного класу підбираються аугментації відповідно до його специфіки
Збереження ознак	може спотворити або замаскувати важливі патологічні патерни	максимально зберігає ознаки, оскільки підбирається з урахуванням патології
Узагальнення	підвищує стійкість моделі, але іноді до «неправильних» варіацій	підвищує стійкість до тих варіацій, які характерні для конкретної хвороби
Дисбаланс класів	не вирішує проблему — дані всіх класів множаться однаково	дозволяє спеціально підсилювати «слабкі» класи
Ризик шуму даних	є ризик створення штучних, неприродних прикладів	аугментації наближені до реальних клінічних варіацій
Інтерпретація результатів	складніше пояснити, чому вибрані саме такі аугментації	легше пояснити, чому саме такі аугментації
Використання у medical AI	більш загальний підхід, зручний для простих задач класифікації	підвищує чутливість і специфічність до патологій на медичних зображеннях

В дослідженні використано набір даних Kaggle Eye Diseases, що знаходиться у публічному доступі. Цей набір даних містить зображення з наявними ознаками глаукоми, діабетичної ретинопатії, катаракти та зображення без патологій. Він використаний в дослідженні для аугментації та подальшого навчання нейронної мережі. Наведемо аугментації для кожного типу зображень. Для катаракти доцільно використовувати обертання. Обертання безпечне, якщо кут не надто великий, щоб уникнути обрізання важливих ділянок, тому що ця аугментація змінює координати пікселів. Нова позиція (x', y') кожного пікселя розраховується зі старої позиції (x, y) шляхом обертання на певний кут навколо центру зображення за формулою

$$P'(x', y') = ((x - x_c) \cos \theta - (y - y_c) \sin \theta + x_c; (x - x_c) \sin \theta + (y - y_c) \cos \theta + y_c). \quad (1)$$

Горизонтальне віддзеркалення також доцільно застосувати для зображень з наявністю ознак катаракти. Якщо ширина зображення дорівнює W , то нова позиція кожного пікселя для зазначеної аугментації визначається за формулою

$$(x', y') = (W - x, y). \quad (2)$$

Також до зображення з наявністю ознак катаракти доцільно застосовувати аугментацію зміни яскравості та контрасту через те, що вона моделює реальні умови зйомки. У пацієнтів із катарактою зображення сітківки або переднього сегмента ока можуть бути контрастними. Ця аугментація допомагає навчити нейромережу бути стійкою до таких варіацій, коли одне зображення буде чітким, а інше — тьмяним. Зображення з наявністю ознак катаракти можуть мати неоднорідну освітленість. Зміна яскравості і контрасту під час тренування моделі дозволяє алгоритму не «збиватися» через різну експозицію. У разі легкої форми катаракти зображення контрастніше, у разі важкої — темніше. Така аугментація дозволяє моделі бачити спектр можливих варіантів, що збільшує здатність до узагальнення. Таким чином, ця аугментація допомагає нейромережевій моделі працювати з нечіткими, затемненими зображеннями, які типові для катаракти. На рис. 1 показано результат застосування аугментацій для зображення з наявними ознаками катаракти.

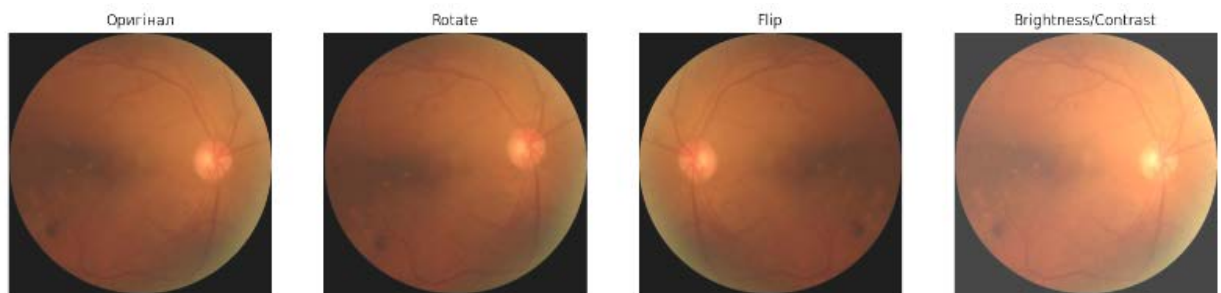


Рис. 1. Результат застосування аугментацій для зображення з наявністю катаракти

Глаукома — найважче для виявлення захворювання. Тому для нього застосовується найбільша кількість аугментацій. Доцільно застосувати аугментацію повороту в межах $\pm 10^\circ$, що задається формулою (1) через те, що вона не спотворює диск зорового нерва та навчає модель бути стійкою до орієнтації знімка. Аугментацію віддзеркалення, що задається формулою (2), також доцільно застосувати у разі глаукоми через те, що віддзеркалення дозволяє моделі розпізнавати глаукому незалежно від сторони. Фотографії можуть бути занадто світлими чи темними через різне освітлення та використання різних медичних апаратів. Глаукома проявляється у структурі диска зорового нерва. Зміна яскравості тренує модель бути робастною до різного освітлення. Формула для цієї аугментації має вигляд

$$I'(x, y) = \alpha \cdot I(x, y) + \beta, \quad (3)$$

де α — коефіцієнт контрасту, β — коефіцієнт зміни яскравості.

Ще одна доцільна аугментація у разі глаукоми — Color Jitter, за допомогою якої можна отримати зображення зі зміненою кольоровою гамою. Вищезгадані аугментації у разі глаукоми імітують клінічні умови, знижують ризик перенавчання. Для зображень з наявною ознакою діабетичної ретинопатії застосовано такі аугментації як зсув, масштабування та поворот, горизонтальне віддзеркалення, зміна яскравості та контрасту, а також гаусівський шум, що задається формулою

$$I'(x, y) = I(x, y) + \eta(x, y), \quad (4)$$

де $\eta(x, y) \sim \mathcal{N}(0, \sigma^2)$, $\sigma^2 \in [\sigma_{\min}, \sigma_{\max}]$.

Для зображень без наявності патологій доцільні аугментації — горизонтальне віддзеркалення та невеликий поворот. Для подальшого розв'язання задачі класифікації захворювань очей вибрано нейромережну архітектуру EfficientNetB3. EfficientNetB3 — це сучасна згортова нейронна мережа, яка відзначається високою ефективністю та точністю класифікації зображень. Ця модель застосовує метод координатного масштабування, який дозволяє отримати високу точність. Модель масштабується рівномірно за шириною, глибиною та роздільною здатністю. Через зазначену особливість модель має кращу продуктивність у порівнянні з іншими моделями. Зміна кількості параметрів та їхнє точне налаштування дозволяють використовувати EfficientNetB3 для задач кла-

сифікації захворювань очей завдяки своїй здатності вловлювати дрібні деталі. Ця мережа дозволяє отримати вищу точність, ніж EfficientNetB0. EfficientNetB3 має набагато менше параметрів, ніж EfficientNetB7. Вказана особливість дозволяє використовувати EfficientNetB3 після стиснення на мобільних телефонах та інших пристроях з обмеженою можливістю. Таким чином, ця архітектура забезпечує оптимальний баланс між точністю та ефективністю.

До базової мережі методами трансферного навчання додано шари GlobalAveragePooling, Dropout з $lr = 0,5$, Dense з 1024 нейронами, l2-регуляризацією зі значенням 0,001 та функцією активації relu, класифікаційний шар Dense з 4 нейронами та функцією активації softmax. Половину шарів базової моделі заморожено. Модель компілювалася за допомогою компілятора Adam з початковим learning rate 0,0001 та функцією втрат categorical crossentropy. Під час попередньої обробки розміри зображень зменшувалися до 224×224 . Нормалізація зображень автоматично відбувалася під час генерації даних для навчання. Для коригування навчання моделі використовувалися такі функції зворотного виклику: ModelCheckpoint — для збереження найкращої моделі, EarlyStopping — для зупинки за відсутності поліпшення метрики val accuracy протягом 15 epoch, ReduceLROnPlateau — для зменшення learning rate у 3 рази за стагнації. В результаті навчання отримано високі показники метрик.

На рис. 2 показано графік точності та функції витрат на тренувальній та валідаційній множині.

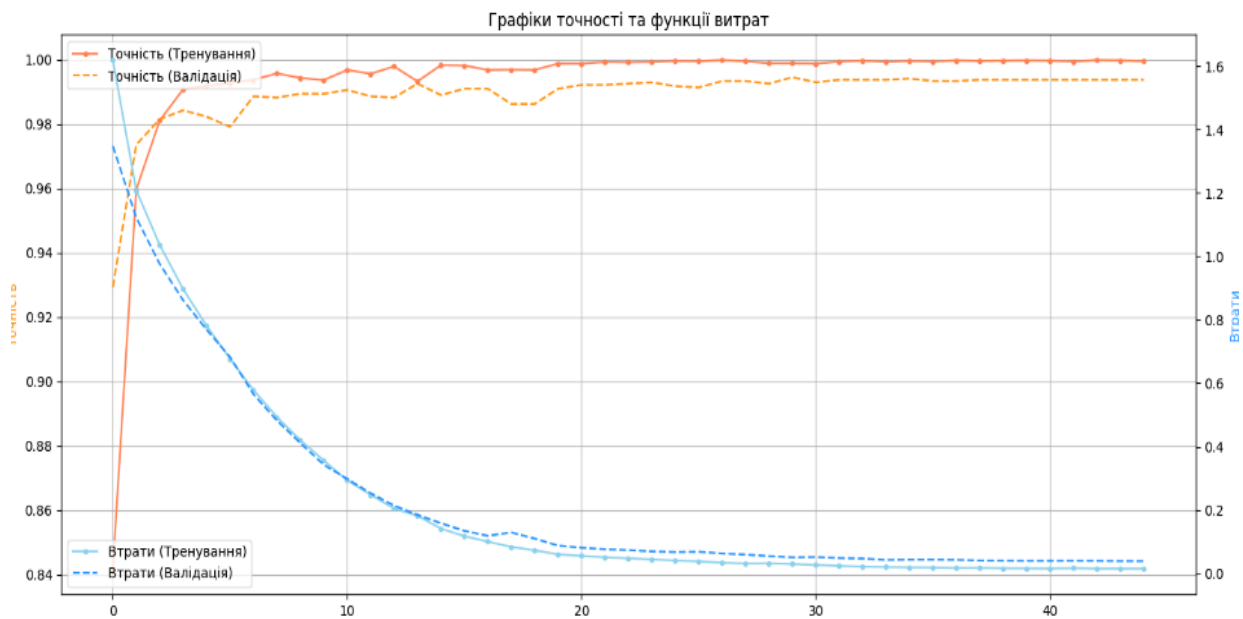


Рис. 2. Показники точності та функції втрат на тренувальній та валідаційній множині

На рис. 3 показана матриця помилок моделі.

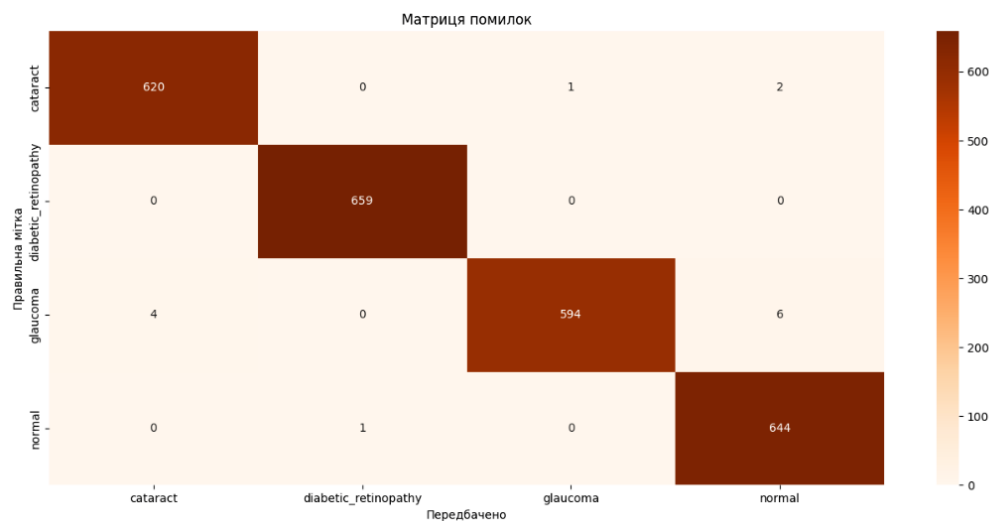


Рис. 3. Матриця помилок

На рис. 4 показано показник F1-Score за класами.

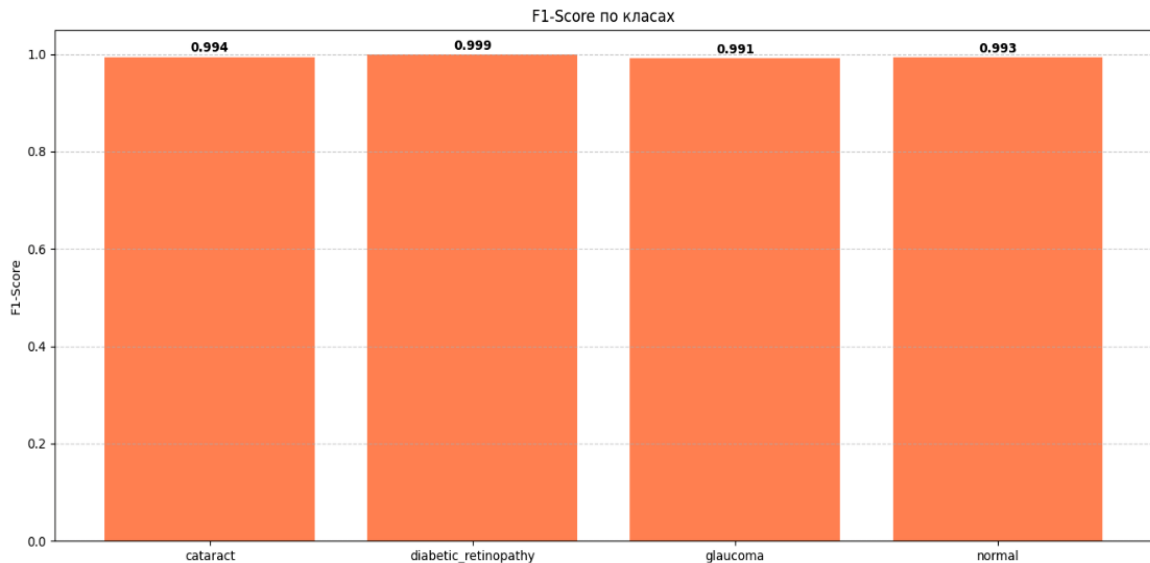


Рис. 4. Показник F1-Score за класами

Навчену модель стиснено шляхом квантифікації для подальшого використання на мобільних пристроях та на пристроях з обмеженими можливостями. Квантифікація дозволила суттєво зменшити розмір моделі (в 6 разів) і підвищити швидкість inference на мобільних пристроях без значної втрати точності.

Висновки

Запропонований підхід дозволяє підвищити загальну точність, робастність нейронної мережі, подолати обмеження традиційного методу. Також клас-орієнтований метод дозволяє отримати оптимізовану модель, що може використовуватися як на потужних серверах, так і на пристроях з обмеженими можливостями після стиснення. Перспективою подальших досліджень є використання запропонованого методу для інших задач виявлення захворювань очей.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

- [1] Z. Wang, "A comprehensive survey on data augmentation," *arXiv preprint*, arXiv:2401.12345, 2024.
- [2] M. Xu, "A comprehensive survey of image augmentation," *Information Fusion*, vol. 97, pp. 1-23, 2023.
- [3] M. Buda, A. Maki, and M. A. Mazurowski, "A systematic study of the class imbalance problem in convolutional neural networks," *Neural Networks*, vol. 106, pp. 249-259, 2018.
- [4] C. Shorten, and T. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data*, vol. 6, no. 60, 2019.
- [5] G. Ghiasi, Y. Cui, and A. Srinivas, "Simple copy-paste is a strong data augmentation method for instance segmentation," in *Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2021, pp. 2918-2928.
- [6] Z. Gong, L. Duan, F. Xiao, and Y. Wang, "MSAug: Multi-Strategy Augmentation for rare classes in semantic segmentation of remote sensing images," *Displays*, vol. 84, Art.102779, 2024.
- [7] M. Frid-Adar, et al., "GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification," *Neurocomputing*, vol. 321, pp. 321-331, 2018.
- [8] Y. Chen, et al., "Conditional GAN-based data augmentation for medical imaging," *IEEE Transactions on Medical Imaging*, vol. 42, no. 5, pp. 1123-1134, 2023.
- [9] L. Zhang, et al., "Evaluation metrics for synthetic data in computer vision," *Pattern Analysis and Applications*, vol. 27, no. 3, pp. 455-470, 2024.
- [10] G. Hu, et al., "Semantics-preserved graph siamese networks with class-oriented feature vector generation," *Neurocomputing*, vol. 527, pp. 123-135, 2023.
- [11] T. Li, et al., "Balanced contrastive learning with class-aware augmentation," *Pattern Recognition*, vol. 142, Art. 109702, 2025.
- [12] L. Perez, and J. Wang, "The effectiveness of data augmentation in image classification using deep learning," *ArXiv preprint*, ArXiv:1712.04621, 2017.
- [13] D. V. Prochukhan, "Features of the modification of the inceptionresnetv2 architecture and the creation of a diagnostic system for determining the degree of damage to retinal vessels," *Computer systems and information technologies*, no. 1, pp. 27-32, 2024, <https://doi.org/10.31891/csit-2024-1-3>.
- [14] Д. В. Прочухан. «Особливості конкатенації згорткових нейронних мереж для скринінгу діабетичної ретинопатії», *Системи обробки інформації*, № 1 (176), с. 89-94, 2024, <https://doi.org/10.30748/soi.2024.176.11>.

Рекомендована кафедрою комп'ютерних наук ВНТУ

Прочухан Дмитро Володимирович — аспірант кафедри комп'ютерних інтелектуальних технологій та систем, e-mail: dmytro.prochukhan@nure.ua .

Харківський національний університет радіоелектроніки, Харків

D. V. Prochukhan¹

Class-oriented Method of Fundus Images Augmentation

¹Kharkiv National University of Radio Electronics

Innovative class-oriented method of fundus images augmentation is proposed. The advantages of this method compared to the standard method are given. The augmentation strategy for images with the existing signs of glaucoma, cataracts, diabetic retinopathy and a healthy eye is substantiated and selected according to the specifics of each class. Augmentation approximates real clinical variation. Resistance to precisely those variations that are characteristic of a specific disease has been improved. Sensitivity and specificity specifically to pathologies in medical images is increased. A neural network model was developed using the class-oriented method. GlobalAveragePooling layers, Dropout with rate 0.5, Dense with 1024 neurons, l2-regularization with a value of 0.001 and categorical crossentropy loss function, Dense classification layer with 4 neurons and softmax activation function were added to the EfficientNetB3 core network by learning methods transfer. Half of the layers of the base model were frozen. The model was compiled using an Adam compiler with an initial learning rate of 0.0001 and a cost feature of categorical crossentropy. In pre-processing, image sizes resized to 224 by 224. The normalization of images automatically occurred during the generation of data for training. The following callback functions were used to adjust the model learning: ModelCheckpoint — to save the best model, EarlyStopping — with to stop in the absence of improvement of the val accuracy metric during 15 eras, ReduceLROnPlateau — to reduce the learning rate by 3 times when stagnating. As a result of the training, high indicators of metrics were obtained. The trained model was compressed by quantification for later use on mobile and disabled devices. The proposed approach makes it possible to increase the overall accuracy and robustness of the neural network, to overcome the limitations of the traditional method.

Keywords: machine learning, augmentation, neural networks, medical images.

Prochukhan Dmytro V. — Post-Graduate Student of Chair of Computer Science Intelligent Technologies and Systems, e-mail: dmytro.prochukhan@nure.ua