

АНОТАЦІЯ

Шульга М.В. Метод та програмні засоби мультимодального аналізу медичних даних на основі глибокого навчання. – Кваліфікаційна наукова праця на правах рукопису.

Дисертація на здобуття наукового ступеня доктора філософії за спеціальністю 121 – Інженерія програмного забезпечення з галузі знань 12 – Інформаційні технології. – Національний Технічний Університет України «Київський Політехнічний Інститут імені Ігоря Сікорського», Київ, 2023.

Дисертаційна робота присвячена розробці методу та програмних засобів мультимодального аналізу медичних даних на основі глибокого навчання, що дозволяє підвищити точність багатокласової класифікації.

Останнім часом сфера медицини активно використовує можливості штучного інтелекту. Зокрема, глибокі нейронні мережі довели свою ефективність і придатність для автоматизованого виявлення та класифікації захворювань. Інтеграція методів штучного інтелекту має потенціал для оптимізації та покращення точності програм скринінгу шляхом автоматизації аналізу медичних даних, усуваючи необхідність безпосередньої участі медичного персоналу на етапі скринінгу. Нагальною проблемою в цьому контексті є діагностика діабетичної ретинопатії, поширеного ускладнення діабету, що призводить до погіршення зору серед дорослого населення світу. Отже, в даний час існує актуальна потреба в застосуванні підходу з використанням комп'ютерного зору та глибоких нейронних мереж для надання передових медичних послуг за допомогою штучного інтелекту, з особливим акцентом на класифікації діабетичної ретинопатії.

Тема дисертаційної роботи входить в план наукової роботи затвердженому на кафедрі обчислювальної техніки КПІ ім. Ігоря Сікорського, що враховує розпорядження Кабінету Міністрів України від 2 грудня 2020 р. № 1556-р про схвалення Концепції розвитку штучного інтелекту в Україні.

Метою дисертаційної роботи є підвищення точності багатокласової класифікації захворювань глибокими нейронними мережами, шляхом розробки

методу мультимодального аналізу медичних даних. Об'єктом дослідження є процеси аналізу вимог, розробки, впровадження і супроводження програмного забезпечення для мультимодального аналізу медичних даних на основі глибокого навчання, які сприяють підвищенню точності визначення окремих класів для завдання багатокласової класифікації медичних даних; а предметом дослідження – методи і моделі розробки і супроводу програмного забезпечення для мультимодального аналізу медичних даних на основі глибоких нейронних мереж різної архітектури і способів організації додаткових модальностей.

Методичною основою дослідження є системне опрацювання та аналіз теоретичного матеріалу, присвяченого підвищенню точності вирішення завдання багатокласової класифікації, шляхом аналізу медичних даних за допомогою моделей глибоких нейронних мереж.

Для досягнення поставленої мети були вирішені такі завдання:

- було досліджено сучасні методи та системи виявлення та багатокласової класифікації захворювань на прикладі діабетичної ретинопатії;
- було виявлено шляхи покращення існуючих методів та систем виявлення та багатокласової класифікації захворювань на прикладі діабетичної ретинопатії;
- було розроблено метод мультимодального аналізу медичних даних на основі доповнення метаданих;
- було розвинуто спосіб підвищення точності багатокласової класифікації завдяки використанню методу мультимодального аналізу для різних архітектур згорткової компоненти нейронної мережі;
- було розроблено метод недетермінованого штучного доповнення метаданих;
- було розвинуто спосіб підвищення точності багатокласової класифікації завдяки використанню методу недетермінованого штучного доповнення метаданих для різних стандартних і спеціалізованих медичних наборів даних;
- було розроблено математичний опис для оцінки підвищення точності багатокласової класифікації на різних стандартних і спеціалізованих медичних

наборах даних.

За результатами проведеного дослідження та згідно поставленого завдання було запропоновано комплексний метод мультимодального аналізу медичних даних на основі глибокого навчання, при розробці якого були проведені наступні дослідження:

- дослідження доповнення метаданих для вирішення завдання багатокласової класифікації;
- дослідження впливу складності мультимодальної моделі на вирішення завдання багатокласової класифікації;
- дослідження недетермінованого доповнення метаданих для вирішення завдання багатокласової класифікації;
- дослідження впливу мультимодального доповнення метаданих на точність багатокласової класифікації.

Розглянуто проблему багатокласової класифікації для одноmodalьної (з введенням зображення) моделі та мультимодальної (з введенням зображення та тексту) моделі та створено кілька варіантів вхідних значень і відповідних моделей на основі аналізу способів та методів використання глибокого навчання для виявлення захворювань на прикладі діабетичної ретинопатії і аналізу існуючих методологій виявлення захворювань: одноmodalьна модель (SM) лише з вхідним зображенням і мультимодальні моделі з вхідними зображеннями та текстом, як-от мультимодальна модель із думкою пацієнта (MP), мультимодальна модель із думкою експерта (ME), мультимодальна модель із думкою пацієнта та експерта (MPE) та мультимодальна модель з недетермінованою думкою експерта (MMFE). Вплив додаткових даних, таких як суб'єктивна думка “пацієнта” про свій стан здоров'я та думка “експерта” (що забезпечує “витік даних”), може бути корисним у деяких практичних ситуаціях. Думки пацієнтів та експертів були імітовані додатковими (доповненими) даними, отриманими з змодельованих анкет.

Проведено дослідження доповнення метаданих для вирішення завдання багатокласової класифікації, яке показало, що всі створені мультимодальні моделі (MP, ME, MPE) у порівнянні з одноmodalьною моделлю (SM) дозволили досягти

різних статистично значущих покращень точності багатокласової класифікації за значенням площі під кривою похибок (AUC) для всіх класів у діапазоні від 4% до 27%, що виходять за межі стандартного відхилення 2-3% виміряного перехресною перевіркою.

Проведено дослідження впливу складності мультимодальної моделі на вирішення завдання багатокласової класифікації, яке показало, що мультимодальна модель (MP) у порівнянні з одномодальною моделлю (SM) дозволила досягти різних статистично значущих покращень точності багатокласової класифікації за значенням AUC для деяких класів у діапазоні від 15% до 26% (в залежності від складності архітектури згорткової компоненти нейронної мережі), що виходять за межі стандартного відхилення 3-8% виміряного перехресною перевіркою.

Проведено дослідження недетермінованого доповнення метаданих для вирішення завдання багатокласової класифікації, яке показало, що мультимодальна модель (MMFE) у порівнянні з одномодальною моделлю (SM) дозволила досягти різних статистично значущих покращень точності багатокласової класифікації за значенням AUC у діапазоні від 12% до 26%.

Наведено теоретичні оцінки можливих (найкращих – коли точність за класами підвищується на сумарну похибку точності за окремими класами для одномодальної моделі, де похибку вдається зменшити завдяки застосуванню маркування у додатковій модальності; проміжних – коли точність за класами підвищується на сумарну похибку точності, помножену на функцію розподілу ймовірностей, де продемонстровано можливу залежність покращення від середньої точності для одномодальної моделі; і найгірших – коли точність за класами не підвищується взагалі) рівнів покращення точності багатокласової класифікації за допомогою простого математичного опису з акцентом на деяких практичних випадках.

На основі теоретичних оцінок проведено дослідження впливу мультимодального доповнення метаданих на точність багатокласової класифікації з використанням різних стандартних (CIFAR10) і спеціалізованих медичних

(PathMNIST, RetinaMNIST) наборів даних, яке показало, що точність мультимодальних моделей значно змінюється залежно від складності набору даних, розміру вибірки та мінливості даних.

Запропоновано комплексний метод мультимодального аналізу медичних даних на основі глибокого навчання, який полягає у використанні “витоку даних” на крайніх і подібних класах, що дозволяє підвищити точність визначення окремих класів для вирішення завдань багатокласової класифікації.

Проведено аналіз результатів використання запропонованого комплексного методу мультимодального аналізу медичних даних, який показав, що запропонований метод на відміну від існуючих рішень за рахунок використання додаткових модальностей на основі доповнення метаданих дозволив отримати підвищення точності визначення окремих класів для завдань багатокласової класифікації захворювань на 4-27% у порівнянні зі стандартним одномодальним підходом для розглянутих ідентичних наборів даних і архітектур нейронних мереж.

Ключові слова: багатокласова класифікація, нейронні мережі, глибоке навчання, доповнення метаданих, недетерміноване маркування, мультимодальна модель, сітківка, діабетична ретинопатія.

ABSTRACT

Shulha M.V. Method and software tools of multimodal analysis of medical data based on deep learning. – Qualified scientific work on the rights of the manuscript.

Dissertation for the degree of Doctor of Philosophy in the specialty 121 – Software Engineering and 12 – Information Technologies. – National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute», Kyiv, 2023.

Recently, the field of medicine has been actively using the capabilities of artificial intelligence. In particular, deep neural networks have proven their effectiveness and suitability for automated disease detection and classification. The integration of artificial intelligence techniques has the potential to optimize and improve the accuracy of screening programs by automating the analysis of medical data, eliminating the need for direct involvement of medical personnel at the screening stage. An urgent problem in this context is the diagnosis of diabetic retinopathy, a common complication of diabetes that leads to visual impairment among the adult population of the world. Hence, there is currently an urgent need to apply a computer vision and deep neural network approach to provide advanced healthcare services with the help of artificial intelligence, with special emphasis on the classification of diabetic retinopathy.

The topic of the dissertation is included in the plan of scientific work approved by Computer engineering department of National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, which takes into account the order of the Cabinet of Ministers of Ukraine of December 2, 2020 # 1556-r on the approval of the Concept of the Development of Artificial Intelligence in Ukraine.

The goal of the dissertation is to increase the accuracy of multiclass classification of diseases by deep neural networks, by developing a method of multimodal analysis of medical data. The object of research is the processes of analysis of requirements, development, implementation and maintenance of software for multimodal analysis of medical data based on deep learning, which contribute to increasing the accuracy of determining individual classes for the task of multi-class classification of medical data; and the subject of research is methods and models of software development and support

for multimodal analysis of medical data based on deep neural networks of various architectures and ways of organizing additional modalities.

The methodological basis of the study is the systematic processing and analysis of theoretical material devoted to increasing the accuracy of solving the problem of multiclass classification by analyzing medical data using deep neural network models.

To achieve the goal, the following tasks were solved:

- modern methods and systems of detection and multiclass classification of diseases were studied using the example of diabetic retinopathy;
- ways of improving the existing methods and systems of detection and multiclass classification of diseases using the example of diabetic retinopathy were identified;
- a method of multimodal analysis of medical data based on the metadata augmentation was developed;
- a method of increasing the accuracy of multiclass classification was developed due to the use of the multimodal analysis method for different architectures of the convolutional component of the neural network;
- a method of fuzzy metadata augmentation was developed;
- a method of increasing the accuracy of multiclass classification was developed through the use of the method of fuzzy metadata augmentation for various standard and specialized medical data sets;
- a mathematical model was developed to evaluate the accuracy improvement of multiclass classification on various standard and specialized medical datasets.

According to the results of the research and according to the task, a complex method of multimodal analysis of medical data based on deep learning was proposed, during the development of which the following research were conducted:

- research of the metadata augmentation for the multiclass classification;
- research of the impact of multimodal model complexity on multiclass classification;
- research of fuzzy metadata augmentation for the multiclass classification;
- research of the effect of multimodal metadata augmentation on multiclass

classification.

The problem of multiclass classification for a single modal (with image input) model and a multimodal (with image and text input) model was considered, and several options for input values and corresponding models were created based on the analysis of ways and methods of using deep learning to detect diseases using the example of diabetic retinopathy and analysis of existing disease detection methodologies: single modal model (SM) with input image only and multimodal models with input images and text, such as multimodal model with patient opinion (MP), multimodal model with expert opinion (ME), multimodal model with patient and expert opinion (MPE) and multimodal model with fuzzy expert opinion (MMFE). The influence of additional data, such as the subjective opinion of the “patient” about his state of health and the opinion of the “expert” (which provides “data leakage”), can be useful in some practical situations. The opinions of patients and experts were simulated with additional (augmented) data obtained from simulated questionnaires.

Research was conducted on the metadata augmentation to solve the problem of multiclass classification, which showed that all the created multimodal models (MP, ME, MPE) in comparison with the single modal model (SM) made it possible to achieve various statistically significant improvements in the accuracy of multiclass classification by the value of the area under the curve (AUC) for all classes ranged from 4% to 27%, exceeding the 2-3% standard deviation measured by cross-validation.

Research of the influence of the complexity of the multimodal model on the solution of the multiclass classification task was conducted, which showed that the multimodal model (MP) compared to the single modal model (SM) allowed to achieve various statistically significant improvements in the accuracy of the multiclass classification according to the AUC value for some classes in the range of 15% up to 26% (depending on the complexity of the architecture of the convolutional component of the neural network), which are beyond the standard deviation of 3-8% measured by cross-validation.

Research of fuzzy metadata augmentation for solving the multiclass classification task was conducted, which showed that the multimodal model (MMFE) compared to the

single modal model (SM) achieved various statistically significant improvements in multiclass classification accuracy in terms of AUC ranging from 12% to 26%.

Theoretical estimates of the possible ones are given (the best – when the accuracy by classes increases by the total error of accuracy by individual classes for a single modal model, where the error can be reduced thanks to the application of labeling in an additional modality; intermediate – when the accuracy by classes increases by the total error of accuracy multiplied by the distribution function probabilities, where the possible dependence of the improvement on the average accuracy for a single modal model is demonstrated; and the worst – when the accuracy by class does not improve at all) levels of improvement in the accuracy of multiclass classification using a simple mathematical model with an emphasis on some practical cases.

Based on theoretical evaluations, research of the impact of multimodal metadata addition on the accuracy of multiclass classification using different standard (CIFAR10) and specialized medical (PathMNIST, RetinaMNIST) data sets was conducted, which showed that the accuracy of multimodal models varies significantly depending on the complexity of the data set, sample size and data variability.

A complex method of multimodal analysis of medical data based on deep learning is proposed, which consists in the use of "data leakage" on extreme and similar classes, which allows to increase the accuracy of determining individual classes for solving multiclass classification tasks.

An analysis of the results of using the proposed complex method of multimodal analysis of medical data was conducted, which showed that the proposed method, unlike existing solutions, due to the use of additional modalities based on the metadata augmentation, made it possible to increase the accuracy of determining individual classes for the tasks of multiclass classification of diseases by 4-27% in comparison with the standard single modal approach for the considered identical datasets and neural network architectures.

Keywords: multiclass classification, neural networks, deep learning, metadata augmentation, fuzzy labeling, multimodal model, retina, diabetic retinopathy.