

УДК 004.93

ЕФЕКТИВНІСТЬ ЗАСТОСУВАННЯ КОМП'ЮТЕРНОЇ ТОМОГРАФІЇ ТА СУЧАСНИХ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ У ВИЯВЛЕННІ СТРУКТУРНИХ ЗМІН СЕРЦЯ ПРИ ГОСТРІЙ ФОРМІ COVID-19

Гончарук Максим Олександрович¹

maksymhoncharuk42@gmail.com

Настенко Євген Арнольдович^{1,2}

nastenko.e@gmail.com

Линник Микола Іванович³

nicklinnik1957@gmail.com

¹ Національний технічний університет України
«КПІ імені Ігоря Сікорського»
м. Київ, Україна

² ДУ «Національний інститут серцево-судинної хірургії
ім. М.Амосова НАМН України»
м. Київ, Україна

³ ДУ «Національний інститут фтизіатрії і пульмонології
ім. Ф.Г. Яновського НАМН України»
м. Київ, Україна

Анотація – За допомогою комп'ютерної томографії та сучасних методів інформаційних технологій можливо оцінити не тільки стан ураження легень, а й визначити зміни в серці, які непомітні неозброєним оком безпосередньо під час КТ-діагностики.

Основною метою даного дослідження є аналіз структурних змін в серці, пов'язаних із гострим перебігом COVID-19. Отримані результати можуть бути приводом для призначення специфічних кардіологічних обстежень та лікувальних заходів.

Використано відкритий датасет COVID-CT-MD та дані КТ обстежень, виконаних в Національному інституті фтизіатрії і пульмонології імені Ф.Г. Яновського НАМН України і Національного інституту серцево-судинної хірургії імені М.М. Амосова НАМН України.

Спостереження були розділені на два класи: 1- особи, які не хворіли на COVID-19 та 2 – особи із гострою формою COVID-19.

В результаті розмітки КТ-серій і виключення обстежень із явними візуальними артефактами залишено 167 і 68 випадків (пацієнтів) COVID-19 і норми відповідно. Для балансування класів застосовано метод андерсемплінгу. Даний метод скоротив кількість випадків до 68 для кожного з класів.

Для оцінки структурних змін в серці застосовано текстурну матрицю суміжності градацій сірого (GLCM) і ансамблеві (Random Forest, LightGBM, XGBoost) та класичні (SVM) класифікатори. Загальну вибірку розділено на тренувальну та валідаційну з частками 70% та 30% відповідно.

Розроблено оригінальний алгоритм відбору інформативних ознак з GLCM, який включає відбір найбільш повторюваних відтінків сірого та їх комбінацій у сусідніх точках зображення у двох класах.

Якість класифікації оцінено за наступними критеріями: точність, чутливість, специфічність та F-міра. Найкращий результат продемонстрував метод SVM з точністю 63% та F-мірою у 63% на тестовій вибірці.

Встановлено, що специфічні структурні зміни в серці при COVID-19 можуть бути виявлені за даними КТ із застосуванням сучасних інформаційних технологій.

Ключові слова: COVID-19, комп'ютерна томографія, комп'ютерний зір, машинне навчання, радіометрія, серце.

I. ВСТУП

Створення швидких, безпечних та високоточних засобів діагностики найбільш поширених та тяжких хвороб є вельми актуальним. Одним з найбільш перспективних інструментів для цього є радіометрія, яка базується на даних комп'ютерної томографії [1,2,3]. Цей метод надає можливість отримати точні дані та кількісний аналіз особливостей рентгенівських зрізів, що важливо для діагностики та дослідження патології органів і тканин.

Радіометрія дозволяє отримувати кількісні дані щодо структурних змін серця при різних захворюваннях. Ці дані мають не тільки діагностичне, але й наукове значення. Завдяки застосуванню сучасних інформаційних технологій стало можливим оцінити специфічність цих змін, зокрема при гострому перебігу COVID-19.

Важливим завданням для підтримки прийняття діагностичних рішень є можливість отримання якомога об'ємної та змістовної діагностичної інформації. Цьому сприяє сучасний розвиток інформаційних технологій та засобів штучного інтелекту. Ці інструменти дозволяють зібрати та обробити великі обсяги даних та отримувати більш точні та повні дані для діагностики хвороб.

Використання комп'ютерної томографії у поєднанні із застосуванням сучасних інформаційних технологій є вельми актуальним у виявленні структурних змін серця при гострій формі COVID-19. Ці методи дозволяють покращити діагностичні можливості, скоротити час діагностичних обстежень та допомагають у вивченні хвороби для розробки ефективних методів лікування.

II. МЕТА ДОСЛІДЖЕННЯ

Метою даного дослідження є пошук структурних змін в серці при Covid-19 на медичних зображеннях комп'ютерної томографії за допомогою методів текстурного аналізу.

III. МАТЕРІАЛИ ТА МЕТОДИ

3.1. Клінічні дані

Для проведення дослідження в мережі інтернет було знайдено набір даних із КТ знімками [4] який складається з 169 підтверджених позитивних випадків COVID-19 (середній вік складає $50,8 \pm 16,8$, відношення чоловічої до жіночої статі приблизно 2:1), 76 випадків норми (середній вік складає $43,3 \pm 14,3$, відношення чоловічої до жіночої статі приблизно 1:1) та 60 випадків негоспітальної пневмонії (середній вік складає $57,8 \pm 21,9$, відношення чоловічої до жіночої статі приблизно 1:1).

Для перевірки результатів дослідження використано дані КТ обстежень, виконаних в Національному інституті фтизіатрії і пульмонології імені Ф.Г. Яновського НАМН України та Національного інституту серцево-судинної хірургії імені М.М. Амосова.

В дослідженні використано зрізи КТ із серцем у нормі та при гострій формі COVID-19. У кожному зрізі із серцем виділялися області інтересу. Область інтересу являє собою прямокутник в якому знаходиться серце. Після анотації і аналізу отриманих даних із загального набору було виключено знімки із артефактами. В результаті відбору залишилося 167 і 68 випадків COVID-19 і норми відповідно.

3.2. Методи дослідження

Вхідними даними для аналізу були області інтересу в яких знаходиться серце (Рисунок 1).

Для перетворення знімків КТ дослідження у монохромне зображення з відтінками сірого від 0 до 255 використано стандартний серцевий режим $level = 200$ HU (одиниць Хаунсфілда) і $width = 1000$ HU [5].

Для вирішення поставленої задачі використано методи текстурного аналізу, а саме матрицю суміжності градацій сірого (GLCM - Gray Level Co-occurrence Matrix) [6].



Рисунок 1. Приклад анотованих даних у вигляді області інтересу

Отримані зазначеним далі методом текстурні ознаки з областей інтересу досліджувалися на інформативність та використовувались для навчання класифікаційних алгоритмів.

Для вирішення задачі класифікації були використані наступні алгоритми: Random Forest, LightGBM, XGBoost та SVM. Ці алгоритми обрані через їх високу ефективність у різних класифікаційних завданнях.

Random Forest базується на ансамблі моделей дерев рішень [7], а LightGBM є оптимізованим алгоритмом градієнтного підсилення [8]. XGBoost є алгоритмом градієнтного підсилення, який використовує моделі на основі дерев рішень [9]. SVM (Support Vector Machines) є методом, що використовується для задач бінарної та багатокласової класифікації [10]. Він працює шляхом розділення даних за допомогою гіперплощини та знаходження оптимальної границі, що розділяє класи.

Кожним із зазначених методів були побудовані класифікаційні моделі та проведено порівняння отриманих результатів на тестових вибірках. Таке порівняння дозволяє визначити ефективність кожного класифікатора в даному контексті для ансамблевих (XGBoost, Random Forest, LightGBM) та класичних (SVM) методів класифікації. Результати цих оцінок допомагають вибрати найкращу модель для подальшого застосування в реальних задачах класифікації.

3.3. Балансування вибірок

Так як кількість об'єктів в класах

відрізняється майже в два рази, було прийнято рішення збалансувати тренувальну вибірку методом випадкового видалення об'єктів переважаючого класу до співвідношення приблизно 1:1.

Це призводить до зменшення кількості об'єктів у тренувальній вибірці, проте забезпечує більшу рівновагу між класами.

Використання цього підходу дає можливість зменшити похибку в результатах класифікації, що може виникати через нерівномірний розподіл об'єктів у класах. Таким чином забезпечується краща точність і надійність моделі, яка навчається на збалансованій тренувальній вибірці.

В результаті загальна кількість досліджень КТ для кожного з класів сягає 68 та близько 1800 областей інтересу.

Тренувальна і тестова вибірки складають 70% і 30% відповідно від загальної кількості КТ-досліджень.

IV. РЕЗУЛЬТАТИ ТА ОБГОВОРЕННЯ

Було сформульоване досить очевидне припущення, що шукані патерни КТ зображень, які мають діагностичне значення, представлені найбільш повторюваними відтінками сірого.

Таке ж припущення та схожий підхід використані у роботах [11, 12].

Отже на початковому етапі пошуку інформативних ознак здійснено наступні кроки.

1. Для всіх об'єктів обох класів норми і COVID-19 розраховано середню гістограму розподілу відтінків сірого на областях інтересу
2. На основі отриманої гістограми відібрано ті відтінки сірого, що найчастіше зустрічаються. В даному випадку обрано 25% відтінків сірого від загальної кількості в монохромному зображенні (256). На рисунку 2 червоним помічені шукані відтінки сірого (далі інформативні відтінки сірого) на середній гістограмі розподілу частот відтінків сірого при нормі і Covid-19.

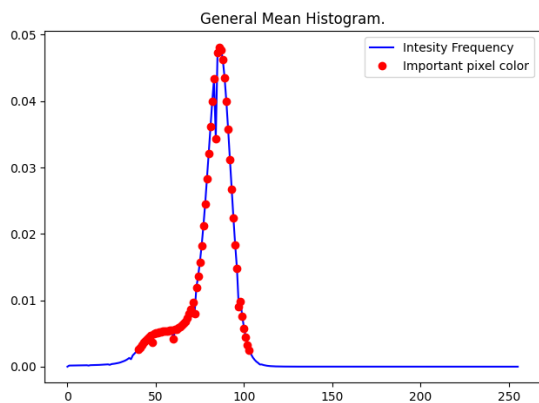


Рисунок 2. Середня гістограми розподілу частот відтінків сірого із інформативними відтінками сірого, що позначені червоним.

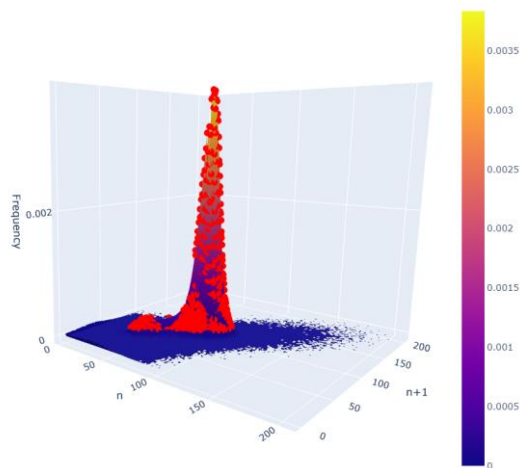


Рисунок 3. Середня гістограми розподілу частот пар відтінків сірого із відібраними патернами на основі відношень

3. Далі для кожної області інтересу сформовано нормалізовану, симетричну матрицю GLCM з параметрами $\theta: 0^\circ; d: 1$. На основі отриманих матриць розраховано загальну середню матрицю по суміші об'єктів обох класів
4. З отриманої матриці відібрано тільки ті пари відтінків сірого (патерни), відтінки яких входять в множину інформативних відтінків сірого, тобто отримані на кроці 2.
5. Так як середня матриця є симетричної можливо скоротити кількість ознак у двічі, розділивши її по головній діагоналі. З отриманих патернів з кроку 4 розраховано середнє значення частоти і залишено ті, значення яких більше половини середнього.

7. В результаті для кожної маски з області інтересу знімка сформовано вектор ознак, що подається на вхід класифікатору. Тобто, в якості ознак використовується значення (частота появи певної пари відтінків сірого) з матриці GLCM.

Результати класифікації на тестовій вибірці за обраними методами XGBoost, Random Forest, LightGBM та SVM наведені у таблиці 1. Результати отримані на даних досліджень, виконаних в Національному інститут фтизіатрії і пульмонології імені Ф.Г. Яновського НАМН України та Національного інституту серцево-судинної хірургії імені М.М. Амосова, не впливали на кінцевий результат

Таблиця 1 Результати класифікації на тестовій вибірці

	RandomForest	XGBoost	LightGBM	SVM
Точність	0.585515	0.564572	0.589878	0.63438
Precision	0.587754	0.564897	0.59142	0.63208
Recall	0.58797	0.565243	0.591778	0.63157
F1-міра	0.585477	0.564118	0.589753	0.631742

6. Фінальним етапом відбору патернів є розрахунок їх відношень між класами. З двох відношень (клас1/клас2 і клас2/клас1) відібрано по 30% найбільших та унікальних (Рисунок 3).

З результатів видно, що найкраще себе показав класичний метод класифікації SVM. За всіма метриками він переважає інші методи. Сама якість класифікації середня (F1-score = 0,63). На це впливає ряд факторів серед яких:

- відсутня стратифікація за станом здоров'я пацієнтів, не хворих на COVID-19: із вибірки було виключено лише ті випадки в яких була інформація пов'язана із патологіями серця.

- відсутня стратифікація за ступенем важкості або симптомів пацієнтів хворих на COVID-19: із вибірки було виключено лише ті випадки в яких була інформація пов'язана із патологіями серця. Імовірність розвитку патологій серця, або інших систем організму, значно вища, якщо форма хвороби переходить у так званий Long-COVID [13].

Враховуючи отримані результати та фактори впливу на них, можна вважати, що специфічні зміни серця хворих на COVID-19 можна розпізнати за допомогою радіометрії, текстурного аналізу та сучасних методів машинного навчання.

V. ВИСНОВКИ

Відповідно до отриманих результатів, встановлено, що на зображеннях КТ серця текстурні ознаки в осіб, які не хворіли COVID-19, і у хворих з гострою формою COVID-19 мають суттєві відмінності.

Ідентифікація структурних змін стає можливою із застосуванням радіометрії, текстурного аналізу та методів машинного навчання.

Найкращий результат показав класичний метод класифікації SVM зі значенням F1-міри=0,63.

Отримана додаткова інформація під час обстеження, дозволяє терміново призначити необхідні діагностичні та лікувальні заходи.

Таким чином, комп'ютерна томографія у поєднанні з сучасними інформаційними технологіями дозволяє отримати додаткову цінну інформацію для прийняття медичних рішень щодо діагностики та лікування хворих.

Даний підхід може мати наукове значення для в сенсі отримання нових знань щодо стану пацієнта під час перебігу COVID-19 або інших патологічний процесів.

Напрямки подальших досліджень

Для покращення результатів класифікації та ідентифікації патернів структурних змін міокарду планується покращення поточного підходу шляхом:

- стратифікації хворих пацієнтів на основі ступені ураження легень або важкості перебігу хвороби та здорових пацієнтів на основі їх клінічних характеристик;

- пошук оптимальних параметрів алгоритму гістограмного перетворення «вікно-рівень» для візуалізації структур міокарду;

- перехід від областей інтересу у вигляді прямокутників до масок з чіткими межами.

Так як результати показали, що структурні зміни в серці при COVID-19 наявні, варто спробувати проаналізувати чи наявні ці зміни при інших патологіях, таких як негоспітальна пневмонія.

Фінансування. Дане дослідження не отримувало зовнішнього фінансування.

Конфлікт інтересів. Автори заявляють про відсутність конфлікту інтересів.

ORCID ID та внесок авторів.

0000-0003-1537-4198 (B, C, D)

Honcharuk Maksym.

0000-0002-1076-9337 (A, E, F) Nastenko Ievgen.

0000-0002-0011-7482 (G) Linnik Mykola

A - Концепція роботи та дизайн, B-аналіз літератури, C - Відповідальність за статистичний аналіз, D - Написання статті, E – Критичний огляд, F - Остаточне схвалення статті, G – Експертна думка.

VI. ПЕРЕЛІК ПОСИЛАНЬ

1. Land DV. Medical microwave radiometry and its clinical applications. In: IEE Colloquium on The Application of Microwaves in Medicine. IEE; 1995. DOI: <https://doi.org/10.1049/ic:19950253>
2. Drakopoulou M, Moldovan C, Toutouzas K, Tousoulis D. The role of microwave radiometry in carotid artery disease. Diagnostic and clinical prospective. Curr Opin Pharmacol. 2018; 39:99–104. DOI: <http://dx.doi.org/10.1016/j.coph.2018.02.008>

3. Laskari K, Siores E, Tektonidou MM, Sfikakis PP. Microwave radiometry for the diagnosis and monitoring of inflammatory arthritis. *Diagnostics (Basel)*. 2023;13(4). DOI: <http://dx.doi.org/10.3390/diagnostics13040609>
4. Afshar P, Heidarian S, Enshaei N, Naderkhani F, Rafiee MJ, Oikonomou A, et al. COVID-CT-MD, COVID-19 computed tomography scan dataset applicable in machine learning and deep learning. *Sci Data*. 2021;8(1):121. DOI: <http://dx.doi.org/10.1038/s41597-021-00900-3>
5. Gheorghe AG, Fuchs A, Jacobsen C, Kofoed KF, Møgelvang R, Lynnerup N, et al. Cardiac left ventricular myocardial tissue density, evaluated by computed tomography and autopsy. *BMC Med Imaging*. 2019;19(1):29. DOI: <http://dx.doi.org/10.1186/s12880-019-0326-4>
6. Haralick RM, Shanmugam K, Dinstein I. Textural Features for Image Classification. *IEEE Trans Syst Man Cybern*. 1973; SMC-3(6):610–21. DOI: <http://dx.doi.org/10.1109/tsmc.1973.4309314>
7. Parmar A, Katariya R, Patel V. A review on random forest: An ensemble classifier. In: *International Conference on Intelligent Data Communication Technologies and Internet of Things (ICICI) 2018*. Cham: Springer International Publishing; 2019. p. 758–762. DOI: https://dx.doi.org/10.1007/978-3-030-03146-6_86
8. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. 2017.
9. Budholiya K, Shrivastava SK, Sharma V. An optimized XGBoost based diagnostic system for effective prediction of heart disease. *J King Saud Univ - Comput Inf Sci*. 2022;34(7):4514–23. DOI: <http://dx.doi.org/10.1016/j.jksuci.2020.10.013>
10. Suykens JAK, Vandewalle J. Least Squares Support Vector Machine Classifiers. *Neural Process Lett*. 1999;9(3):293–300. DOI: <http://dx.doi.org/10.1023/a:1018628609742>
11. Nastenka IA, Pavlov VA, Honcharuk MO, Hrishko DY. Forming the stack of texture features for liver ultrasound images classification. *Herald of Advanced Information Technology*. 2021;3(4):240–51. DOI: <http://dx.doi.org/10.15276/hait.04.2020.3>
12. Nastenka E, Dykan I, Pavlov OV, Tarasyuk B, Babenko V, Kruglyi V, et al. Analysis of ultrasound images of chronic diffuse liver diseases using texture parameters of gray gradations. *Radiation Diagnostics, Radiation Therapy*. 2022;13(3):7–29. DOI: <http://dx.doi.org/10.37336/2707-0700-2022-3-1>
13. Світайло ВС, Чемиш МД, Саєнко ОС. LONG-COVID ТА АСОЦІЙОВАНІ УРАЖЕННЯ СЕРЦЕВО-СУДИННОЇ ТА НЕРВОВОЇ СИСТЕМ. *Інфекційні хвороби*. 2023;(4):49–54. DOI: <http://dx.doi.org/10.11603/1681-2727.2022.4.13701>

UDC 004.93

THE EFFECTIVENESS OF THE USE OF COMPUTED TOMOGRAPHY AND MODERN INFORMATION TECHNOLOGIES IN DETECTING STRUCTURAL CHANGES OF THE HEART IN THE ACUTE FORM OF COVID-19

Maksym Honcharuk¹
maksymhoncharuk42@gmail.com

Ievgen Nastenka^{1,2}
nastenka.e@gmail.com

Mykola Linnik³
nicklinnik1957@gmail.com

¹National Technical University of Ukraine
«Igor Sikorsky Kyiv Polytechnic Institute»,
Kyiv, Ukraine

²Amosov National Institute of Cardiovascular Surgery
Kyiv, Ukraine

³Yanovsky National Institute of Phthiology and Pulmonology
Kyiv, Ukraine

Abstract - Using computed tomography (CT) and advanced information technology methods, it is possible to assess not only the condition of lung damage but also to detect changes in the heart that are invisible to the naked eye during CT diagnostics.

The main goal of this study is to analyze structural changes in the heart associated with the acute course of COVID-19. The results obtained may provide a basis for prescribing specific cardiological examinations and therapeutic measures.

The open dataset COVID-CT-MD and data from CT examinations collected at the National Institute of Phthiatry and Pulmonology named after F.G. Yanovsky and the National Institute of Cardiovascular Surgery named after M.M. Amosov National Academy of Sciences of Ukraine were used during this study.

Observations were divided into two classes: 1 - persons who did not have COVID-19 and 2 - persons with an acute form of COVID-19.

As a result of marking the CT series and excluding examinations with obvious visual artifacts, there were remained 167 and 68 cases (patients) of COVID-19 and the norm, respectively. The undersampling method was used to balance the classes. This method reduced the number of cases to 68 for each of the classes.

To assess the structural changes in the heart, the gray level co-occurrence matrix (GLCM), ensemble (Random Forest, LightGBM, XGBoost) and classical (SVM) classifiers were used. The total sample is divided into training and validation with ratio of 70% and 30%, respectively.

In the course of the work, a proprietary algorithm for selecting informative features from the GLCM matrix was developed. It includes the selection of the most repeated shades of gray and their combinations in neighboring points of the image.

The quality of the classification was evaluated according to the following criteria: accuracy, sensitivity, specificity and F-measure. The best result was demonstrated by the SVM method with an accuracy of 63% and an F-measure of 63% on the test sample.

It has been established that specific structural changes in the heart with COVID-19 can be detected using CT data using modern information technologies.

Keywords: computed tomography, computer vision, COVID-19, heart, machine learning, radiometry.