

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/357780576>

Medical Image Recognition by Genetic-inductive Forest Algorithm [in Ukrainian]

Conference Paper · November 2021

CITATIONS

0

READS

9

4 authors:



[Vitalii Babenko](#)

National Technical University of Ukraine Kyiv Polytechnic Institute

25 PUBLICATIONS 14 CITATIONS

[SEE PROFILE](#)



[Ie. Nastenko](#)

National Technical University of Ukraine Kyiv Polytechnic Institute

18 PUBLICATIONS 11 CITATIONS

[SEE PROFILE](#)



[Vladimir Anatolievich Pavlov](#)

National Technical University of Ukraine Kyiv Polytechnic Institute

9 PUBLICATIONS 5 CITATIONS

[SEE PROFILE](#)



[Olena Nosovets](#)

National Technical University of Ukraine Kyiv Polytechnic Institute

28 PUBLICATIONS 24 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Optimal Treatment Strategy Searching [View project](#)



Coronary Heart Disease Classifier [View project](#)

здійсненні лікувального процесу дотримуватися правил санітарії, як це робиться при лікуванні хворих з різними видами інфекційних хвороб.

Висновки. Запропоновано гіпотезу про можливу структуру нової коронавірусної інфекції Covid 2019 у вигляді складових лега-вірусів. Розглянуто особливості нового типу інфекції і витікаючі з цього вимоги до санітарно епідемічних заходів щодо запобігання її поширення.

Перелік посилань:

1. Михаил Пинкус Вчені-ядерники створили модель поширення COVID-19. – URL: <http://surl.li/aejup>.
2. Ereemeeva N. I. Building a modification of the SEIRD model of epidemic spread that takes into account the features of COVID-19. *Vestnik TvGU. Seriya: Prikladnaya Matematika [Herald of Tver State University. Series: Applied Mathematics]*. 2020. № 4. 14–27 (in Russian). <https://doi.org/10.26456/vtpmk602>.
3. Ольга Дорохова Сибірські вчені представили прототип моделі поширення COVID-19.– URL: <http://surl.li/aejul>.
4. Bar-On SARS-CoV-2 (COVID-19) by the numbers / Bar-On, Yinon & Flamholz, Avi & Phillips, Rob & Milo, Ron. // eLife. 2020. 9. 10.7554.

УДК 004.048 + 616-079.4

РОЗПІЗНАВАННЯ МЕДИЧНИХ ЗОБРАЖЕНЬ АЛГОРИТМОМ ГЕНЕТИЧНО-ІНДУКТИВНОГО ЛІСУ

В. О. Бабенко¹, Є. А. Настенко², В. А. Павлов³, О. К. Носовець⁴

Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського»

E-mail: ¹vbabenko2191@gmail.com, ²nastenko.e@gmail.com, ³pavlov.vladimir264@gmail.com, ⁴o.nosovets@gmail.com

The given work is devoted to the modern developments in the field of medicine, namely medical image recognition by innovative algorithm called "Genetic-inductive Forest". The main purpose of creating such an algorithm is to improve already existing Machine Learning approaches in order to obtain more efficient results in recognition tasks. To test the performance of the algorithm, liver ultrasound images and speckle-tracking echocardiography frames were used.

Вступ. Зображення ультразвукового дослідження (УЗД) несуть у собі цінну діагностичну інформацію, яка використовується лікарями та медичними експертами задля оцінки стану певного органу пацієнта. Однією із найголовніших властивостей подібних зображень є «ехогенність», яка необхідна для охарактеризування тканини досліджуваного органу. Кожен орган володіє власним рівнем ехогенності, і тому зміна рівню у більшості випадків свідчить про появу патології у пацієнта, що призводить до розвинення захворювання. Спеціалісти УЗД володіють знаннями про те, який орган має володіти тою чи іншою ехогенністю, тим самим визначаючи його стан за кольором зображення. Хоча спеціалісти з великим стажем у більшості випадків здатні абсолютно точно визначати зміну ехогенності, не виключені випадки, коли ручна діагностика стає надскладною задачею (особливо на ранніх етапах появи патології), і уникнути помилок стає неможливо. Складності також додає об'єм тих даних, які надходять до подібних спеціалістів. Таким чином набуває актуальності задача розпізнавання медичних ультразвукових зображень та за допомогою підходів штучного інтелекту і машинного навчання, які останнім часом набули широкої популярності у сфері медицини, а їхня властивість швидко обробляти й аналізувати надвеликі масиви даних є дуже ефективною для досягнення головної мети дослідження. У даному дослідженні пропонується застосування інноваційного алгоритму машинного навчання під назвою «Генетично-індуктивний ліс», який поєднує у собі властивості таких методів, як: генетичний алгоритм, метод групового урахування аргументів (МГУА) та випадковий ліс.

Основна частина. Основна ідея генетично-індуктивного лісу полягає у використанні принципу ансамблевого навчання (підвиду машинного навчання) під назвою «бегінг». Головною властивістю даного підходу є «навчання» не одної моделі штучного інтелекту, а відразу «ансамблю» (множини)

подібних моделей, які різняться між собою тим, що навчаються на різних підмножинах вхідних даних. Дана характеристика описується «множина середніх моделей краще ніж одна сильна модель». Оскільки найважливіша ціль будь-якого моделювання не апроксимація вхідних даних дослідження, а їх узагальнення, то недостатньо отримати сильну модель, яка показує ідеальні результати на тренувальній вибірці, оскільки в більшості випадків це свідчить про «перенавчання», і на нових даних (які модель не бачила до цього) будуть отримані значно гірші результати. Відповідно, під вище вказаною характеристикою мається на увазі те, що множина середніх моделей дають змогу досягти цілі моделювання.

Поняття бегінгу появилось після запровадження методу ансамблевого навчання під назвою випадковий ліс, який ідейно є множиною різних дерев прийняття рішень (базовий алгоритм машинного навчання), що навчаються на різних підмножинах даних. Хоч даний метод і вважається одним із найбільш точних в сучасному світі науки про дані, в залежності від задачі точність на екзаменаційній вибірці може не досягати навіть і 90%. Тому було запропоновано удосконалити випадковий ліс за допомогою МГУА і генетичного алгоритму.

МГУА є алгоритмом структурно-параметричного синтезу, тобто, ставиться задача отримання моделі оптимальної складності (з якомога простішою структурою моделі та високою точністю на тестових даних). Для цього зазвичай використовують дві вибірки, де на першій навчаються різні моделі, а на другій робиться підсумкова оцінка кожної моделі, на основі якої і виконується вибір. Було прийнято рішення використати даний підхід для побудови дерев прийняття рішень генетико-індуктивного лісу, таким чином отримуючи оптимальну складність для кожного із них.

Генетичний алгоритм відомий як метод оптимізації, який використовує евристичні техніки для пошуку необхідного рішення задачі. Оскільки дерева у випадковому лісі, окрім того, що навчаються на різних підмножинах даних, вони містять у собі підмножини вхідних ознак, які підбираються випадковим чином – генетичний алгоритм було вирішено застосувати для знаходження оптимальної підмножини ознак для кожного із дерев лісу.

Для перевірки працездатності отриманого алгоритму були використані дані зображень, що надані:

а) спеціалістами ДУ «Інституту ядерної медицини та променевої діагностики НАМН України» [4], які забезпечили базою знімків ультразвукового дослідження (УЗД) печінки;

б) спеціалістами ДУ «Національного інституту серцевої-судинної хірургії імені Амосова НАМН України» [5], які в свою чергу надали доступ до бази відеоданих спекл-трекінг ехокардіографії (СТЕ) серця, що для проведення дослідження були розбиті на кадри (фрейми).

Генетико-індуктивний ліс порівнювався з такими відомими аналогами, як: логістична регресія, випадковий ліс і адаптивний бустинг (AdaBoost). Вибірки даних були розбиті на: навчальну (80%), тестову (10%) і екзаменаційну (10%). Як результат, у задачі класифікації УЗД печінки генетико-індуктивний ліс досягнув 93.2% точності на екзаменаційній вибірці (серед інших алгоритмів найкращий результат показав метод логістичної регресії у 64.3% на екзамені), у задачі класифікації СТЕ серця – 72.6% точності на екзаменаційній вибірці (серед інших – 65.5% точності на екзамені найкращий результат у випадкового лісу). Отриманий результат підтверджує, що розроблений алгоритм виявився потужним, і має перспективи у використанні в інших подібних задачах.

Висновки. В результаті даного дослідження було запропоновано інноваційний алгоритм машинного навчання для вирішення задач розпізнавання медичних зображень. Порівнявши отримані результати розробленого алгоритму з більш відомими алгоритмами, можна зробити висновок, що генетико-індуктивний ліс є абсолютно ефективним інструментом для розв'язання задач не лише медицини, а будь-якої іншої предметної області.

Перелік посилань:

1. Ghaheri A., Shoar S., Naderan M., Hoseini S. S. The applications of genetic algorithms in medicine. *Oman Medical Journal*. 2015. Vol. 30, No. 6. pp. 406–416. DOI: 10.5001/omj.2015.82.
- 2 Anastasakis L., Mort N. The development of self-organization techniques in modelling: a review of the group method of data handling (GMDH). *Research Report-University of Sheffield Department of Automatic Control and Systems Engineering*. 2001. No. 813. pp. 39–39.

3. Fawagreh K., Gaber M. M., Elyan E. Random forests: From early developments to recent advancements. *Systems Science and Control Engineering*. 2014. Vol. 2, No. 1. pp. 602–609. DOI: 10.1080/21642583.2014.956265.

4. Nastenko I., Maksymenko V., Dykan I., Nosovets O., Tarasiuk B., Pavlov V., Babenko V., Kruhlyi V., Soloduschenko V., Dyba M., Umanets V. *Liver Pathological States Identification in Diffuse Diseases with Self-Organization Models Based on Ultrasound Images Texture Features*. Zbarazh, Ukraine:2020. DOI: 10.1109/CSIT49958.2020.9321999.

5. Petrunina O., Shevaga D., Babenko V., Pavlov V., Rysin S., Nastenko I. Comparative Analysis of Classification Algorithms in the Analysis of Medical Images from Speckle Tracking Echocardiography Video Data. *Innovative Biosystems and Bioengineering*. 2021. Vol. 5, No. 3. DOI: 10.20535/ibb.2021.5.3.234990.

УДК 004.9

ВИКОРИСТАННЯ ПЕРЕТВОРЕННЯ ГІЛЬБЕРТА-ХУАНГА ДЛЯ АНАЛІЗУ БІОМЕДИЧНИХ СИГНАЛІВ

К. О. Безвершнюк, М. Ю. Буриченко, О. В. Мельников, Х. С. Мусійчук
Національний авіаційний університет, м. Київ
E-mail: bikam@i.ua

The Hilbert-Huang transformation (HHT) consists of empirical mode decomposition (EMD) and Hilbert spectral analysis. EMD uses signal characteristics to adaptively decompose them into several internal mode (IMF) functions. Then the Hilbert transformations (HTs) are used to convert the IMF to instantaneous frequencies (IMs) for obtain the signal distribution by frequency and energy frequency. The article considers HHT application to different types of biomedical signals.

Вступ. Біомедичні сигнали у своїй більшості є багатокомпонентними нестационарними сигналами, властивості яких змінюються в часі. Часово-частотні характеристики компонент біосигналів мають високу діагностичну цінність і тому актуальним є застосування часово-частотних методів аналізу. Одним із таких методів є перетворення Гільберта-Хуанга (ПГХ) [1], яке дає змогу розкласти сигнал на компоненти (моди) з амплітудною і частотною модуляцією, з подальшим обчисленням миттєвої амплітуди і частоти цих компонент. Зазначені компоненти мають діагностичний сенс, хоча що на даний час не отримали широкого застосування для аналізу біомедичних сигналів. Метою даної роботи є застосування ПГХ для аналізу нестационарних біомедичних сигналів з метою для отримання діагностичної інформації. Завданням роботи було проведення аналізу біомедичних сигналів різного виду з використанням розробленого для даної мети програмного забезпечення.

Основна частина. Відповідно до ПГХ обробку цифрових сигналів (часових рядів) виконують в два етапи. На першому етапі вхідні сигнали за допомогою емпіричної модової декомпозиції розкладають на декілька компонент, що мають назву функція внутрішньої моди (ФВМ) або емпірична мода (ЕМ). Таким чином, дані розкладають за емпіричним базисом, який обумовлений властивостями саме даних, а не вибирають з заздалегідь встановленого набору. Другий етап ПГХ – застосування перетворення Гільберта до кожної ФВМ і побудові спектра Гільберта, який відображає локальні часові і частотні особливості аналізованих часових рядів і дозволяє визначити миттєві амплітуди і частоти для ФВМ. В роботі проведено аналіз декількох видів біосигналів: електрокардіографічних, електроенцефалографічних, кефалографічних та ін. [2]. Оброблення біосигналів у такий спосіб дає змогу виявити нові діагностичні ознаки, що не є видимими при інших методах аналізу. Для аналізу біосигналів в системі MATLAB розроблено програмний код, в якому використані деякі з функцій, що опубліковано в [3].

Висновки. Розроблений програмний код дозволив виявити діагностично важливі дані, що приховані в біосигналах. За допомогою створеного програмного забезпечення на основі ПГХ можна досліджувати різні види біомедичних сигналів і отримувати діагностичні дані, що не могли бути виявлені іншими методами. Одним з недоліків методу є значна похибка оцінювання миттєвої частоти ФВМ, що спричинена її шумами.

Перелік посилань:

1. Huang, N. The Hilbert-Huang transform and its applications / N. Huang, S. S. P. Shen. Singapore: World scientific publishing, 2005. 526 p.