

**ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
РАДІОЕЛЕКТРОНІКИ**

РЕФЕРАТ РОБОТИ

**ГІБРИДНІ НЕЙРО-ФАЗЗИ СИСТЕМИ ДЛЯ
ЗАВДАНЬ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ
ПОТОКІВ МЕДИЧНИХ ДАНИХ**



Перова І.Г.
д.т.н., доцент, професор кафедри
біомедичної інженерії

Харків
2021 р

Актуальність теми. В даний час методи інтелектуального аналізу даних (Data Mining) отримали широке розповсюдження для розв'язання широкого класу завдань у промисловості, економіці, фінансах, банківській сфері, сільському господарстві. Не оминули ці методи і медичні програми, де в силу специфіки таких завдань сформувався напрям, відомий як Medical Data Mining. Основними в рамках цього напрямку є завдання діагностування, які вирішуються засобами розпізнавання образів на основі парадигм як навчання, так і самонавчання, і з математичної точки зору зводяться до розв'язання завдань класифікації або кластеризації.

В Україні у 2018 році введено систему eHealth, кінцевою метою якої буде створення бази медичних карт усіх українців. На цьому етапі актуальною буде оброблення медичної інформації в послідовному режимі з використанням підходів Medical Data Mining. Але особливістю цієї сфери є неможливість використання традиційних методів Data Mining в чистому вигляді, що пов'язане з цілою низкою обставин: обмеженість вибірок, які підлягають класифікації, істотне перекриття класів, які відносяться до різних захворювань, нелінійний характер гіперповерхонь, що поділяють класи, наявність аномальних спостережень – тобто викидів, що спотворюють первинну інформацію, значна роль суб'єктивного людського фактора, що не забезпечує отримання точних даних, необхідність оброблення медичних даних в послідовному онлайн-режимі, можливість роботи з потоками даних.

Всі перелічені обставини призводять до формування неопуклих і нечітких класів, водночас адекватним математичним апаратом для роботи з такими даними є методи обчислювального інтелекту, перш за все, штучні нейронні мережі, системи нечіткого висновування і гібридні нейро-фаззі-системи, побудовані на їхній основі. Тут слід зазначити, що все одно в цьому класі задач перелічені системи не є панацеєю, оскільки для свого навчання вимагають великих обсягів інформації, які найчастіше лікарі не мають у своєму розпорядженні та не пристосовані до необхідності опрацьовувати дані в послідовному онлайн-режимі. Також варто наголосити, що при

проведенні медичного діагностування обов'язковим етапом є етап підготовки даних, який охоплює завдання заповнення пропущених значень (якщо такі мають місце), нормування та кодування даних і зменшення кількості ознак шляхом компресії даних або відбору найінформативніших ознак. Наявні підходи не забезпечують дотримання зазначених вимог щодо роботи в режимі онлайн для завдань медичного діагностування.

Таким чином, актуальною є науково-технологічна проблема розроблення методів для аналізу потоків медичної інформації в послідовному режимі з використанням підходів Medical Data Mining для підтримки впровадження системи eHealth в Україні і підвищення ефективності медичного діагностування в цілому. Для вирішення цієї актуальної проблеми в роботі сформульовано завдання розроблення методів аналізу різних за обсягом медичних даних для медичного діагностування в онлайн-режимі на основі створення адаптивних гібридних нейро-фаззі систем і удосконалення методів їхнього навчання.

Мета і завдання дослідження. Метою роботи є підвищення ефективності медичного діагностування в онлайн-режимі на основі інтелектуального аналізу масивів медичних даних з використанням гібридних нейро-фаззі систем.

Згідно з поставленою метою в роботі розв'язано такі завдання:

1. Розробити адаптивний метод попереднього оброблення медичних даних в послідовному режимі з урахуванням типів даних, включаючи заповнення пропущених значень, гібридний метод оцінювання інформативності медичних показників в онлайн-режимі.

2. Розробити методи аналізу медичних даних під час медичного онлайн-діагностування в режимі контрольованого навчання на основі гібридних адаптивних нейро-фаззі систем для ситуацій репрезентативної навчальної вибірки.

3. Розробити методи аналізу медичних даних під час медичного онлайн-діагностування в режимі активного навчання та асоціації на основі

гібридних нейро-фаззі систем для ситуацій нерепрезентативної навчальної вибірки.

4. Розробити методи аналізу медичних даних під час медичного онлайн-діагностування в режимі самонавчання, що базуються на кластеризації та асоціативній кластеризації за допомогою нейро-фаззі мереж.

5. Провести апробацію усіх розроблених методів на клінічних даних.

Об'єктом дослідження є процес аналізу потоків медичних даних в онлайн-режимі.

Предметом дослідження є моделі й методи послідовного аналізу масивів медичних даних в умовах обмеженої інформації в онлайн-режимі.

Методи дослідження: отримані в роботі результати ґрунтуються на застосуванні методів: теорії факторного аналізу, що дають змогу оцінювати інформативну цінність медичних ознак; теорії штучних нейронних мереж, що надають можливість синтезувати нові типи гібридних архітектур нейро-фаззі систем; теорії нечіткої логіки, що дають змогу у комбінації зі штучними нейронними мережами створювати нові типи нейро-фаззі систем та враховувати медичні ознаки, виміряні в різних шкалах, зі збереженням їхнього змістовного інтерпретування; теорії оптимізації, лінійної алгебри, математичної статистики, що уможливають синтез процедур навчання гібридних нейро-фаззі систем; інтелектуальний аналіз даних та їх потоків (Data Stream Mining), що дає змогу знаходити додаткову корисну інформацію в первинних медичних даних.

Наукова новизна отриманих результатів:

1. Вперше запропоновано метод оцінювання інформативності медичних показників, який на основі гібридизації підходів визначення першої головної компоненти за допомогою нейрону Ойя та методів виділення інформативних ознак з множини наявних забезпечує попереднє оброблення медичних даних у послідовному онлайн-режимі та уможливорює виділення з потоку медичних даних найінформативніших медичних ознак зі збереженням їхнього змістовного інтерпретування.

2. Удосконалено метод просторової екстраполяції діагностичних ознак пацієнта шляхом використання нечіткого рівня належності, що надає можливість виконання обґрунтованого заповнення пропусків у потоках медичних даних.

3. Удосконалено метод формування багатовимірного нео-фаззі нейрона за рахунок введення в його архітектуру нелінійних активаційних функцій, що надає можливість використовувати його у завданнях нечіткого онлайн-діагностування в умовах наявності класів-діагнозів з перетином у просторі ознак та одночасно підвищити як точність діагностування, так і швидкодію роботи системи.

4. Вперше запропоновано метод формування еволюційної нейро-фаззі мережі в умовах змінної кількості медичних ознак і діагнозів, який відрізняється здатністю адаптувати свою архітектуру в разі зміни розмірності простору вхідних ознак та діагнозів, що дає змогу підвищити якість діагностування за рахунок використання усіх вимірних даних.

5. Вперше запропоновано метод активного навчання нейро-фаззі мережі для медичного діагностування, який здійснює автоматичне перемикання між режимами контрольованого навчання та самонавчання і забезпечує процес налаштування синаптичних ваг нейро-фаззі мережі за ознаками пацієнта як з відомим діагнозом (контрольоване навчання), так і з невідомим (самонавчання), що дає змогу зменшити необхідні обсяги даних у навчальній вибірці та час налаштування мережі.

6. Вперше запропоновано метод адаптивної робастної нечіткої кластеризації багатовимірного простору ознак пацієнта, який використовує манхеттенську метрику як основу нечіткої належності, що надає можливість розв'язувати завдання кластеризації в умовах спотвореності медичних даних аномальними викидами.

7. Вперше розроблено метод асоціативної кластеризації на основі нейро-фаззі автоасоціативної пам'яті, який відрізняється тим, що послідовно уточнює свої параметри під час надходження нових медичних даних та

автоматично реалізує асоціативне нечітке висновування на основі аналізу прецедентів, завдяки чому до аналізу додаються медичні випадки з нетиповим поєднанням симптомів.

Практичне значення отриманих результатів. Розроблене в роботі математичне забезпечення надає можливість заповнювати пропущені значення в медичних показниках пацієнта в послідовному режимі для максимально ефективного використання усіх вимірних ознак, проводити оцінювання інформативної цінності кожної з ознак для обґрунтування необхідності їхнього подальшого використання або вилучення, виконувати діагностування у послідовному онлайн-режимі з розрахунком нечіткої функції належності кожного пацієнта до кожного діагнозу. Завдяки цьому істотно знижується ризик помилкового встановлення діагнозу пацієнта під час лікувально-діагностичного процесу, що підтверджено результатами проведених клінічних випробувань, відображеними в отриманих актах впровадження. Результати роботи апробовано і впроваджено в КЗОЗ «Харківська міська клінічна лікарня №13», м. Харків (акт впровадження від 16 жовтня 2017р.); у відділі популяційних досліджень, відділ науково-організаційної роботи та медичної інформації з бібліотекою та відділ атеросклерозу та ішемічної хвороби серця ДУ «Національний інститут терапії ім. Л.Т. Малої Національної академії наук України», м. Харків (акти впровадження від 27 листопада 2017р., акт впровадження від 24 листопада 2017р.); в ТзОВ «Клініка сучасних методів діагностики та інноваційних медичних технологій «Водолій», м. Хуст, Закарпатська область (акт впровадження від 01 вересня 2017р.); в Діагностично-лікувальний медичний центр «Ужмед», м. Ужгород (акт впровадження від 04 жовтня 2017р.); в Інституті електроніки та інформаційних технологій Люблінського технологічного університету, м. Люблін, Польща (акт впровадження від 22 травня 2018р.); на кафедрі гігієни та екології №2 Харківського національного медичного університету (акт впровадження від 15 січня 2019р.).

Особистий внесок. Наукова робота виконана на кафедрі біомедичної інженерії Харківського національного університету радіоелектроніки. Основні результати роботи належать особисто автору і повністю опубліковані в фаховій літературі [1–35].

Публікації: Основні результати роботи з 2009 року опубліковано в 34 наукових працях, із яких 2 виконані без співавторства. До розгляду подаються наукові праці в період з 2009 по 2019 рік, за який було опубліковано 34 працях, серед яких 19 статей (4 з них входить до наукометричної бази SCOPUS, 8 – в англomовних журналах із імпаkт-фактором), 14 доповідей (з них 4 представлені на конференціях, що проходили під егідою IEEE, вони викладені в базах IEEEExplore, SCOPUS, Web of Science та Google Shcolar) та 1 патент України на винахід. Таким чином, всі матеріали роботи опубліковані на науково-практичних конференціях, форумах і симпозіумах галузевого, державного та міжнародного рівнів, 4 публікації викладено в науково-метричній базі даних Web of Science (загальна кількість цитувань публікацій, опублікованих за темою досліджень складає 16, h-індекс – 3), 8 публікацій викладені в науково-метричній базі даних Scopus (загальна кількість цитувань публікацій, опублікованих за темою досліджень складає 62, h-індекс – 6), 18 публікації викладено в науково-метричній базі даних Google Shcolar (загальна кількість цитувань публікацій, опублікованих за темою досліджень складає 139, h-індекс – 8). Отримано патент України на винахід. В рамках роботи у 2019 році було захищено дисертаційну роботу на здобуття наукового ступеня доктора технічних наук «Інформаційна технологія аналізу медичних даних на основі гібридних нейро-фаззі систем» за спеціальністю 05.13.09 – медична та біологічна інформатика та кібернетика.

Висновки. У роботі вирішено актуальну науково-технологічну проблему розроблення методів для аналізу потоків медичної інформації в послідовному режимі з використанням підходів Medical Data Mining для підтримки впровадження системи eHealth в Україні і підвищення

ефективності медичного діагностування в цілому. Ефективність розроблених методів і засобів підтверджено під час їхнього застосування для розв'язання завдань медичного діагностування за клінічними даними.

Основними науковими й практичними результатами роботи є:

1. На основі аналізу підходів, методів та систем діагностування і препроцесингу медичних даних виявлено низьку ефективність їхньої роботи з потоками таких даних, встановлено необхідність об'єднання (гібридизації) кількох типів систем аналізу даних в одну та розроблення методів навчання таких систем з урахуванням особливостей первинних медичних даних, що дало змогу визначити необхідні вимоги до програмно-технічних складників систем діагностування, спрямовані на підвищення ефективності медичного діагностування.

2. Удосконалений метод просторової екстраполяції діагностичних ознак пацієнта за рахунок використання нечіткого рівня належності в режимі заповнення пропущених значень надав можливість виконання обґрунтованого заповнення пропусків у початкових медичних даних з метою їхнього подальшого використання для розв'язання завдань діагностування.

3. Розроблення гібридного методу оцінювання інформативності медичних показників за рахунок доповнення нейрону Ойя контуром адаптивного налаштування дає змогу зменшити час опрацювання інформації та підвищити швидкодію процесу самонавчання нейронної мережі для виділення інформативних ознак, що надало можливість виділити з потоку медичних даних найінформативніших медичних ознак зі збереженням їхнього змістовного інтерпретування, що сприяє кращому розумінню процесу подальшого діагностування лікарями.

4. Розроблені методи аналізу медичних даних під час медичного онлайн-діагностування в режимі контрольованого навчання на основі гібридних адаптивних нейро-фаззі систем для ситуацій репрезентативної навчальної вибірки уможливили проведення нечіткого онлайн-діагностування з перетином класів-діагнозів у просторі ознак з підвищенням

точності діагностування і швидкодії роботи системи. Під час використання модифікованого багатовимірного нео-фаззі нейрону значення критерію ефективності для пацієнтів із сечокам'яною хворобою з повною кількістю ознак складає $PerCr = 0.085$ (точність діагностування 88,89%), з залученням тільки найінформативніших ознак $PerCr = 0.054$ (точність діагностування 93,6%). Точність діагностування у модифікованого багатовимірного нео-фаззі нейрона для пацієнтів із фармакорезистентною формою епілепсії склала 94.18%, значення критерію ефективності – $PerCr = 0.0501$.

5. Розроблений метод аналізу медичних даних під час медичного онлайн-діагностування в режимі контрольованого навчання з можливістю архітектурної адаптації нейро-фаззі мережі для розширення простору медичних ознак та діагнозів дав змогу підвищити якість діагностування за рахунок використання всіх вимірних даних.

6. Запропонований метод аналізу медичних даних під час медичного онлайн-діагностування в режимі активного навчання на основі гібридної нейро-фаззі мережі здійснює автоматичне перемикавання між режимами контрольованого навчання та самонавчання і забезпечує відповідні зміни процесу налаштування ваг нейро-фаззі мережі за індивідуальними ознаками пацієнта, допомагає зменшити необхідні обсяги даних у навчальній вибірці та час налаштування мережі. Значення критерію ефективності на даних пульмонологічних досліджень склало $PerCr = 0.0961$ (точність 84,38%).

7. Розроблений метод аналізу медичних даних під час медичного онлайн-діагностування в режимі самонавчання, який базується на адаптивній робастній нечіткій кластеризації багатовимірного простору ознак пацієнта з використанням манхеттенської метрики як основи нечіткої належності, надав можливість розв'язувати завдання кластеризації в умовах спотворення медичних даних аномальними викидами.

8. Розроблений метод аналізу медичних даних під час медичного онлайн-діагностування в режимі самонавчання, який базується на асоціативній кластеризації на основі нейро-фаззі автоасоціативної пам'яті,

яка послідовно уточнює свої параметри під час надходження нових медичних даних та автоматично реалізує асоціативне нечітке висновування на основі аналізу прецедентів, дав змогу долучати до аналізу медичні випадки з нетиповим поєднанням симптомів. Значення критерію ефективності у діагностуванні гіпертонічної хвороби та ішемічної хвороби серця склали $PerCr = 0,468$, значення точності діагностування 82,11%.



І.Г. Перова

Перелік наукових публікацій претендента, які увійшли до роботи

1. Є. Бодянський, М. Мустецов та І. Чурюмова, «Спосіб оцінки біологічних станів, заснований на нечіткій кластеризації даних множини вимірюваних показників». Україна Патент UA 91767 C2 МПК (2009), G06F 19/00, G06F 17/00, G06F 7/00, G01N 33/48, 25 серпня 2010.
2. Е. Бодянский, Е. Винокурова, П. Мулеса и И. Перова, «Диагностирующая нейро-фаззи-система и ее адаптивное обучения в задачах интеллектуальной обработки данных медико-биологических исследований» Системні технології. Регіональний міжвузівський збірник наукових праць, т.91, № 2, pp. 125-135, 2014.
3. И. Перова, «Адаптивная обработка данных медико-биологических исследований методами вычислительного интеллекта» Восточно-европейский журнал передовых технологий, т. 67, № 1, pp. 24-28, 2014.
4. Е. Бодянский и И. Перова, «Нечеткая классификация данных медико-биологических исследований в условиях дефицита информации» Системи обробки інформації, т. 136, № 11, pp. 161-163, 2015.
5. И. Перова и Е. Бодянский, «Восстановление пропусков в таблицах данных на основе метода нечеткой пространственной экстраполяции с использованием манхэттенской метрики,» в Интеллектуальные системы принятия решения и проблемы вычислительного интеллекта ISDMCI'2015, Железный Порт, 2015.
6. P. Mulesa and I. Perova, "Fuzzy Spacial Extrapolation Method Using Manhattan Metrics for Tasks of Medical Data Mining," in Computer Science and Information Technologies CSIT'2015, Lviv, Ukraine, 2015. (SCOPUS)
7. O. Turuta, I. Perova and A. Deineko, "Evolving flexible neuro-fuzzy system for medical diagnostic tasks" International Journal of Computer Science and Mobile Computing IJCSMC, vol. 4, no. 8, pp. 475-480, 2015.
8. O. Turuta, A. Deineko, I. Perova, Y. Kutsenko and M. Shalamov, "Evolving Neural Network for Kernel Principal Component Analysis" International Journal of Computer Science and Mobile Computing IJCSMC, vol. 9, no. 4, pp. 356-363, 2015.
9. I. Perova and Y. Bodyanskiy, "Adaptive fuzzy clustering based on Manhattan metrics in medical and biological applications" Вісник національного університету "Львівська політехніка", vol. 826, pp. 8-12, 2015.
10. Е. Бодянский и И. Перова, «Нейро-фаззи система для задач обработки медицинских данных в ситуациях множества диагнозов» Бионика интеллекта, т. 85, № 2, pp. 86-89, 2015.
11. I. Perova and Y. Bodyanskiy, "Fast evolving diagnostic neuro-fuzzy system and its learning in medical data mining tasks" System technologies, vol. 101, no. 6, pp. 109-116, 2015.
12. И. Перова, Е. Бражникова и И. Плисс, «Нео-фаззи подход в задачах online медико-биологической диагностики» Прикладная радиоэлектроника, т. 15, № 1, pp. 51-57, 2016.
13. I. Perova, I. Pliss, G. Churyumov, F. Eze and S. Mahmoud, "Neo-Fuzzy Approach for Medical Diagnostics Tasks in Online-Mode," in 2016 IEEE First International Conference on Data Stream Mining & Processing (DSMP), Lviv, Ukraine, 2016. (SCOPUS)
14. І. Перова, Є. Бражникова та О. Бойко, «Діагностуюча нейро-фаззі система зі змінною кількістю входів та виходів,» в VIII Міжнародна школа-семінар «Теорія прийняття рішень», Ужгород, Україна, 2016.
15. И. Перова и Е. Бодянский, «Об одной метрике в задачах интеллектуального анализа медицинских данных,» в 19-th International Conference on System Analysis and Information Technology "SAIT 2017", Киев, 2017.
16. Е. Бодянский, И. Перова и Н. Мирошниченко, «Адаптация центров функций принадлежности эволюционирующей диагностирующей нейро-фаззи системы,» в

Интеллектуальные системы принятия решений и проблемы вычислительного интеллекта ISDMCI'2017, Железный порт, Украина, 2017.

17. Е. Бодянский и И. Перова, «Оценка информативности медицинских показателей на основе нейро-фаззи подхода,» в Материалы VI Международной конференции "Information Management System and Technologies", Одесса, Украина, 20-22 сентября 2017.

18. И. Перова и Т. Генъ, «Нечеткая пространственная экстраполяция при заполнении пропусков в медицинских выборках данных в последовательном режиме,» в IV Міжнародна науково-практична конференція IT&I'2017, Киев, Украина, 2017.

19. I. Perova and Y. Bodyanskiy, "Fast medical diagnostics using autoassociative neuro-fuzzy memory" International Journal of Computing, vol. 16, no. 1, pp. 34-40, 2017. (SCOPUS)

20. I. Perova and I. Pliss, "Deep hybrid System of Computational Intelligence with Architecture Adaptation for Medical Fuzzy Diagnostics" I.J. Intelligent System and Applications, vol. 7, pp. 12-21, 2017. (SCOPUS)

21. I. Perova and Y. Bodyanskiy, "Medical online neuro-fuzzy diagnostics system with active learning" International Journal of Advances in Computer and Electronics Engineering, vol. 2, no. 7, p. 1-10, July 2017.

22. S. Mahmoud, I. Perova and I. Pliss, "Multidimensional neo-fuzzy-neuron for solving medical diagnostics tasks in online-mode" Journal of Applied Computer Science, vol. 25, no. 1, pp. 39-48, 2017.

23. I. Pliss and I. Perova, "Diagnostic Neuro-Fuzzy System and Its Learning in Medical Data Mining Tasks in Conditions of Uncertainty about Numbers of Attributes and Diagnoses" Automatic Control and Computer Sciences, vol. 51, no. 6, pp. 391-398, 2017. (SCOPUS)

24. Н. Мірошніченко та І. Перова, «Медичне діагностування на основі нейромережних технологій» Прикладная радиоэлектроника, т. 16, № 1, pp. 46-50, 2017.

25. Є. Бодяньський, І. Перова та Г. Стойка, «Оптимізація оцінювання інформативності медичних показників на основі гібридного підходу,» Розвиток транспорту, т. 1, № 1, pp. 108-115, 2017.

26. I. Perova and Y. Bodyanskiy, "Adaptive Human Machine Interaction Approach for Feature Selection-Extraction Task in Medical Data Mining," International Journal of Computing, vol. 17, no. 2, pp. 113-119, 2018. (SCOPUS)

27. І. Перова, Є. Бражнікова та Н. Мірошніченко, «Про один з підходів до попереднього аналізу медичних даних,» в XXVI Міжнародної науково-практичної конференції MicroCAD-2018, Харків, Украина, 2018.

28. І. Перова, Є. Бражнікова та О. Літовченко, «Нечітка кластеризація багатовимірних часових рядів показників сироватки крові для онлайн діагностування,» в Интеллектуальные системы принятия решения и проблемы вычислительного интеллекта ISDMCI'2018, Железный порт, Украина, 2018.

29. I. Perova, O. Litovchenko, Y. Bodyanskiy, Y. Brazhnykova, I. Zavgorodnii and P. Mulesa, "Medical Data-Stream Mining in the Area of Electromagnetic Radiation and Low Temperature Influence on Biological Objects," in Proc. 2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP), Lviv, Ukraine, August 21-25, 2018. (SCOPUS)

30. I. Perova, Y. Bodyanskiy, Y. Brazhnykova and P. Mulesa, "Neural Network for Online Principal Component Analysis in Medical Data Mining Tasks," in IEEE First International Conference on System Analysis & Intelligent Computing (SAIC), Kyiv, Ukraine, 8-12 October 2018. (SCOPUS)

31. I. Perova, "Hybrid systems of computational intelligence for online medical diagnostics tasks," in The Fourth China-Ukraine Forum On Science And Technology, Harbin, China, 2018.

32. И. Перова и Н. Мирощниченко, «Медицинское диагностирование сердечно-сосудистых заболеваний при помощи нейро-фаззи-системы,» в Збірник наукових праць І

Міжнародної науково-практичної конференції «Інформаційні системи та технології в медицині» (ISM-2018), Харків, Україна, 28-30 листопада 2018.

33. Y. Bodyanskiy, I. Perova and P. Zhernova, "Online fuzzy clustering of high-dimensional data based on ensembles in data stream mining tasks," Сучасний стан наукових досліджень та технологій в промисловості, vol. 1, no. 7, pp. 16-24, 2019.

34. Perova I., Bodyanskiy Ye., Online Medical Data Stream Mining Based on Adaptive Neuro-Fuzzy Approaches, Cybernetics and computer engineering, 2019. – №4 (198). – p. 3-25
DOI: 10.15407/kvt198.04.003