

DOI: 10.26565/2313-6693-2023-47-04

УДК: 616.1:616.8-09.832]-053.2

Ковальчук Т. А.^{А, В, С, D, F}, **Боярчук О. Р.**^{А, В, Е, F}, **Богай С. Є.**^{А, С, F}

tetianakovalchuk@gmail.com

АЛГОРИТМ МАШИННОГО НАВЧАННЯ XGBOOST ДЛЯ ДИФЕРЕНЦІЙНОЇ ДІАГНОСТИКИ СИНКОПЕ У ДИТЯЧОМУ ВІЦІ

А – концепція та дизайн дослідження; В – збір даних; С – аналіз та інтерпретація даних; D – написання статті; Е – редагування статті; F – остаточне затвердження статті

Анотація. Пошук нових методів диференційної діагностики видів синкопе дозволить покращити діагностику вазовагальних синкопе, синкопе внаслідок ортостатичної гіпотензії та кардіогенних синкопе у дитячому віці з метою своєчасного прийняття адекватних діагностичних та терапевтичних рішень.

Метою дослідження було розробити ефективну модель машинного навчання для диференційної діагностики вазовагальних синкопе, синкопе внаслідок ортостатичної гіпотензії та кардіогенних синкопе у дітей.

Матеріали і методи. Було обстежено 140 пацієнтів із синкопе віком 8–17 років: 92 дитини з діагнозом вазовагальне синкопе, 28 – з синкопе внаслідок ортостатичної гіпотензії та 20 – з кардіогенним синкопе. Модель машинного навчання було побудовано на основі алгоритму XGBoost для мультикласової класифікації, беручи за основу вхідні клінічні, лабораторні та інструментальні дані пацієнтів.

Результати. Розроблена модель машинного навчання на основі алгоритму XGBoost є ефективною у диференційній діагностиці синкопе, про що свідчать метрики точності (0,93), влучності (0,93 у групі вазовагальних синкопе; 1,00 у групі синкопе внаслідок ортостатичної гіпотензії; 0,80 у групі кардіогенних синкопе), чутливості (0,96 для вазовагальних синкопе; 1,00 для синкопе внаслідок ортостатичної гіпотензії; 0,67 у групі кардіогенних синкопе), f1-міри (0,95 у групі вазовагальних синкопе; 1,00 у групі синкопе внаслідок ортостатичної гіпотензії; 0,73 у групі кардіогенних синкопе), ROC AUC (0,95 у дітей із вазовагальними синкопе; 1,00 у дітей із синкопе внаслідок ортостатичної гіпотензії; 0,89 у дітей із кардіогенними синкопе), PR AUC (0,96 для вазовагальних синкопе; 1,00 для синкопе внаслідок ортостатичної гіпотензії; 0,79 у групі кардіогенних синкопе), коефіцієнт Каппа (0,85) та коефіцієнт кореляції Метьюса (0,85). В алгоритмі диференційної діагностики видів синкопе найінформативнішими показниками є ортостатична гіпотензія, пароксизмальна суправентрикулярна тахікардія, коефіцієнт Хільдебранта, шкала Calgary Syncope Seizure Score, вітамін В6, середня тривалість інтервалу P-Q за добу, частота тахікардії за добу, ударний індекс, гомоцестин, об'єм серця, систолічний об'єм крові.

Висновки. Дана модель машинного навчання володіє достатньою ефективністю та може використовуватися педіатрами та спеціалістами з дитячої кардіоревматології для диференційної діагностики вазовагальних синкопе, синкопе внаслідок ортостатичної гіпотензії та кардіогенних синкопе у дитячому віці.

КЛЮЧОВІ СЛОВА вазовагальне синкопе, синкопе внаслідок ортостатичної гіпотензії, кардіогенне синкопе, диференційна діагностика, модель XGBoost, діти

ІНФОРМАЦІЯ ПРО АВТОРІВ

Тетяна Анатоліївна Ковальчук, к. мед. н., доцентка кафедри педіатрії №2 Тернопільського національного медичного університету імені І. Я. Горбачевського; Майдан Волі, 1, Тернопіль, Україна, 46001; e-mail: tetianakovalchuk@gmail.com; ORCID ID: <http://orcid.org/0000-0003-2455-3278>

Оксана Романівна Боярчук, д. мед. н., професорка, завідувачка кафедри дитячих хвороб з дитячою хірургією Тернопільського національного медичного університету імені І. Я. Горбачевського; Майдан Волі, 1, Тернопіль, Україна, 46001; e-mail: boyarchuk@tdmu.edu.ua; ORCID ID: <http://orcid.org/0000-0002-1234-0040>

Святослав Євгенович Богай, IT-фахівець; вул. Назарія Яремчука 41/2, Тернопіль, Україна, 46009; e-mail: slavabogaj@gmail.com

Для цитування:

Ковальчук ТА, Боярчук ОР, Богай СС. АЛГОРИТМ МАШИННОГО НАВЧАННЯ XGBOOST ДЛЯ ДИФЕРЕНЦІЙНОЇ ДІАГНОСТИКИ СИНКОПЕ У ДИТЯЧОМУ ВІЩІ. Вісник Харківського національного університету імені В. Н. Каразіна. Серія «Медицина». 2023;47: 33–46. DOI: 10.26565/2313-6693-2023-47-04

ВСТУП

Диференційна діагностика транзиторних втрат свідомості у будь-якому віці залишається складною та досі чітко не стандартизованою. Основною компетенцією лікарів, які оцінюють пароксизмальні події, є знання в першу чергу клінічних особливостей втрати свідомості транзиторного характеру, які допомагають визначити різну етіологію непритомності та встановити найбільш імовірний діагноз [1, 2]. Хоча вазовагальні синкопе (ВС) є найпоширенішою причиною розвитку синкопе у дітей, часто-густо діагноз є необґрунтованим, і особливо тоді, коли деякі клінічні симптоми насторожують з приводу ймовірного аритмогенного, а не нейрогенного синкопе [3, 4]. З іншого боку, диференційна діагностика непритомності ускладнюється тим фактом, що у дітей із захворюваннями серцево-судинної системи можуть виникати як кардіогенні синкопе (КС), так і ВС, при чому останні з такою ж частотою, як і в загальній популяції [5]. Крім того, жодного окремо взятого симптому у зборі анамнезу не достатньо для диференціальної діагностики синкопе, а тригери або продромальні симптоми можуть бути супутніми як при ВС, так і синкопе внаслідок ортостатичної гіпотензії (СВОГ) та КС [6–8]. Варто зазначити і те, що деякі неспецифічні симптоми, які узагальнено називають запамороченням, можуть бути викликані рядом захворювань та співіснувати в однієї людини, що ще більше ускладнює диференційну діагностику синкопе [9, 10]. Усе це є свідченням актуальності проблеми та необхідності пошуку нових, значно ефективніших методів диференційної діагностики синкопе у пацієнтів дитячого віку.

МЕТА

Метою дослідження було розробити модель машинного навчання, яка дозволить покращити диференційну діагностику ВС, СВОГ та КС у дітей. Дана модель має мати високу ефективність за низкою показників

оцінки, що забезпечить лікарям педіатрам та дитячим кардіоревматологам точність діагностики видів синкопе для прийняття своєчасних і обґрунтованих діагностично-лікувальних рішень.

МАТЕРІАЛИ ТА МЕТОДИ

Було обстежено 140 пацієнтів із синкопе віком 8–17 років: 92 дитини з діагнозом ВС, 28 – з СВОГ та 20 – з КС. Для діагностики ВС, СВОГ, КС застосовували критерії Європейської асоціації кардіологів (2018) [11]. Пацієнти з проявами транзиторної втрати свідомості внаслідок травми голови були вилучені з дослідження.

У роботі дотримані етичні принципи щодо людей, які виступають суб'єктами дослідження, з урахуванням основних положень Гельсінської декларації Всесвітньої медичної асоціації з біомедичних досліджень, у яких людина виступає їх об'єктом (World Medical Association Declaration of Helsinki, 1964, 2000, 2008), Конвенції Ради Європи про права людини й біомедицину (2007). Усі діти та їх батьки надали добровільну інформовану згоду на участь у дослідженні.

Усім пацієнтам були виконані наступні дослідження:

- ретельний збір скарг та анамнезу, включаючи оцінку за шкалами Calgary Syncope Seizure Score (CSSS) та Modified Calgary Syncope Seizure Score (MCSSS) [12];
- фізикальне обстеження, у т. ч. індекс маси тіла (ІМТ) та тест активного ортостазу. Тест активного ортостазу виконували покроково усім залученим у дослідження дітям відповідно до протоколу, розробленого Н. Tanaka et al. [13]. У положенні пацієнта лежачи на спині дослідник накладав манжету тонометра на плече, фіксував стетоскоп на руці над плечовою артерією. Після десятихвилинного відпочинку у положенні лежачи вимірювали частоту пульсу (ЧП) та артеріального тиску (АТ). Далі дослідник надував манжету тонометра до рівня середнього значення систолічного артеріального тиску (САТ) і перекидав гумову трубку затискачем,

аускультуючи при цьому приглушені тони Короткова. В момент коли дитина активно підіймалася із горизонтального у вертикальне положення, дослідник вмикав секундомір і намагався вислухати тони Короткова. При ортостазі тони Короткова зникають внаслідок фізіологічного падіння АТ, але в нормі відновлюються в середньому через 17 секунд. Час відновлення вихідного рівня АТ фіксували у всіх випадках. В подальшому вимірювання АТ і ЧП провадились загальноприйнятими методами на 1, 3, 5, 7 та 10-й хвилинах. Нормальною реакцією на ортостаз вважали САТ ± 10 мм рт. ст., ДАТ ± 5 мм рт. ст., ЧП $+(5-15)$ уд/хв. Ортостатичну гіпотензію (ОГ) під час тесту активного ортостазу діагностували за умови падіння САТ ≥ 20 мм рт. ст. або ДАТ ≥ 10 мм рт. ст. від вихідних значень, або зниження САТ до < 90 мм рт. ст.;

- загальний аналіз крові, глюкоза крові натще;

- визначення рівнів піридоксину (вітамін В6), фолатів (вітамін В9), кобаламіну (вітамін В12), 25-гідроксिवітаміну D (25-ОН-D) та гомоцистеїну в сироватці крові колориметричним методом імуноферментного аналізу;

- інтегральні показники функціонального стану серцево-судинної системи (сistolічний об'єм крові, величина хвилиного об'єму кровотоку, коефіцієнт економічності системи кровообігу, потужність роботи лівого шлуночка серця, об'єм серця, індекс Робінсона, систолічний показник, коефіцієнт функціонального стану системи кровообігу, коефіцієнт функціонального стану кардіореспіраторної системи, коефіцієнт Хільдебранта, величина адаптаційного потенціалу серцево-судинної системи, серцевий індекс, ударний індекс, загальний периферичний опір, питомий периферичний опір) [14];

- електрокардіографія у 12 відведеннях (ритм, електрична вісь серця, тривалість інтервалів R-R, P-Q, Q-T, T-P, S-T, QRS-комплексу, а також вольтаж зубців P, Q, R, S, T);

- ехокардіографія (кінцевий діастолічний та систолічний розміри лівого шлуночка, діаметр правого шлуночка, лівого передсердя, аорти та легеневої артерії, товщина міжшлуночкової

перетинки, товщина задньої стінки лівого шлуночка, фракція укорочення лівого шлуночка, кінцевий діастолічний та систолічний об'єми крові, ударний об'єм, хвилиний об'єм крові, фракція викиду лівого шлуночка, відносна товщина лівого шлуночка, маса міокарда лівого шлуночка, індекс маси міокарда лівого шлуночка);

- холтеровське моніторування електричної активності серця впродовж 24 годин (середня ЧСС за день, ніч та добу, циркадний індекс, тривалість епізодів тахікардії, брадикардії, аритмії, екстрасистолії, аналіз атріовентрикулярного проведення, а також тривалості електричної систоли та варіабельності серцевого ритму за часовими (SDANN, RMSSD, рNN50) та частотними (TP, VLF, LF, HF, співвідношення LF/HF) показниками);

- добове моніторування артеріального тиску (середні значення систолічного артеріального тиску (САТ), діастолічного артеріального тиску (ДАТ), середнього артеріального тиску (СрАТ) та пульсового артеріального тиску (ПАТ) за добу, день і ніч, варіабельність артеріального тиску (АТ), нічне зниження АТ, ранкове підвищення АТ (величина і швидкість), показники «навантаження тиском» (індекс часу гіпертензії, індекс площі під графіком) за добу, день та ніч, індекс резистентності судин, частоту серцевих скорочень за добу, день та ніч; ступені нічного зниження АТ (dippers, non-dippers, over-dippers, night-peakers);

- електроенцефалографія;

- консультації спеціалістів – невролога, кардіолога, окуліста.

Процес розробки моделі машинного навчання для диференційної діагностики синкопе у дітей складався з п'яти послідовних етапів: (1) вибору типу моделі машинного навчання, (2) підготовки даних, (3) розподілу даних на навчальний і тестовий набори, (4) навчання моделі на навчальному наборі даних, (5) оцінки ефективності запропонованої моделі на тестовому наборі даних.

Вибір типу моделі машинного навчання

Відповідно до характеру набору даних та специфіки завдань у цьому дослідженні, було обрано модель XGBoost [15, 16]. Для мультикласової класифікації використову-

вали функцію втрат «multi:softmax», яка дозволяла прогнозувати найбільш ймовірний діагноз для кожного окремого прикладу набору даних.

Підготовка даних

Спершу усі клінічні та інструментальні показники 140 пацієнтів із синкопе було завантажено в програмне середовище JupyterLab за допомогою бібліотеки Pandas, перевірено їх на наявність пропущених значень та інші невідповідності, визначено типи ознак. Для кодування категоріальних ознак було вибрано метод One-Hot Encoding.

Розподіл даних на навчальний і тестовий набори.

Розподіл даних було виконано за допомогою функції `train_test_split` бібліотеки `sklearn` зі співвідношенням 70 % для навчальних і 30 % для тестових даних. Так як розподіл даних по мітках «BC», «СВОГ», «КС» представлений нерівномірно, було використано стратифікований розподіл даних, який дозволив адекватно представити кожен групу захворювань у загальному наборі даних

Навчання моделі на навчальному наборі даних.

Для створення моделі використовували пакет XGBoost для Python. Для побудови продуктивної моделі було виконано оптимізацію гіперпараметрів за допомогою методу «GridSearchCV» пакету `scikit-learn`. Для запобігання перенавчання моделі застосовували гіперпараметр «`early_stopping_rounds`», який дозволив визначити кількість раундів підсилення для визначення покращення показників у моделі. Оскільки дана модель належить до мульти класової класифікації, додатково була використана метрика логістичних втрат (Mlogloss) та обраховано коефіцієнт Каппа та коефіцієнт кореляції Метьюса на кожному раунді підсилення.

Оцінка ефективності запропонованої моделі на тестовому наборі даних.

Для оцінки ефективності моделі використовували наступні метрики: точність (Accuracy), влучність (Precision), чутливість (Recall), F1-міра, коефіцієнт Каппа,

коефіцієнт кореляції Метьюса, ROC AUC (площа робочої характеристики приймача під кривою), PR AUC (площа точності запам'ятовування під кривою). Для кожної метрики обраховували довірчий інтервал за допомогою інверсії функції нормального розподілу із стандартним значенням рівня довіри 0.95. Окремо було визначено показник важливості для кожної ознаки моделі машинного навчання.

За результатами побудови дерев рішень обчислювали суму листкових значень за допомогою функції `softmax` за формулами:

$$\sigma(x_1) = e^{x_1} / (e^{x_1} + e^{x_2} + e^{x_3}), \quad (1)$$

$$\sigma(x_2) = e^{x_2} / (e^{x_1} + e^{x_2} + e^{x_3}), \quad (2)$$

$$\sigma(x_3) = e^{x_3} / (e^{x_1} + e^{x_2} + e^{x_3}), \quad (3)$$

де $\sigma(x_1)$ – ймовірність приналежності до класу BC;

$\sigma(x_2)$ – ймовірність приналежності до класу СВОГ;

$\sigma(x_3)$ – ймовірність приналежності до класу КС;

x_1 – сума листкових значень 1 і 4 дерева рішень;

x_2 – сума листкових значень 2 і 5 дерева рішень;

x_3 – сума листкових значень 3 і 6 дерева рішень.

РЕЗУЛЬТАТИ

Запропонована модель машинного навчання була розроблена на основі алгоритму XGBoost для мультикласової класифікації. Використання гіперпараметру «`early_stopping_rounds`» дозволило встановити, що після другого раунду підсилення деякі метрики досягають точності з числовим значенням 1, а інші змінюються в незначній мірі, отже було прийнято рішення зупинитися на другому раунді підсилення моделі.

Матриця плутанини для нашої моделі представлена на рисунку 1. Тестовий набір даних містив 42 пацієнти, у якому 28 пацієнтів відносилися до класу «BC», 8 – до «СВОГ», 6 – до «КС». Дана модель правильно знаходила 27 (96,4 %) із 28 (100,0 %) випадків «BC», 8 (100,0 %) із 8 (100,0 %) випадків «СВОГ» та 4 (66,7 %) із 6 (100,0 %) випадків «КС».

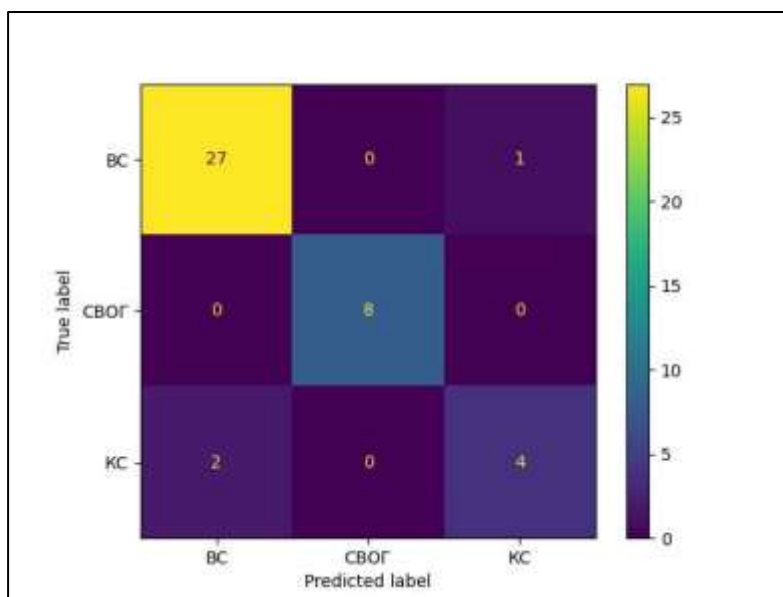


Рис. 1. Матриця плутанини для моделі диференційної діагностики ВС, СВОГ і КС на тестовому наборі даних

Fig. 1. Confusion matrix for the model of differential diagnosis vasovagal syncope, syncope due to orthostatic hypotension and cardiac syncope on the test data set

Також, на тестовому наборі даних цієї моделі були отримані метрики точності, влучності, чутливості, f1-міри, ROC AUC,

PR AUC, коефіцієнт Каппа та коефіцієнт кореляції Метьюса (рис. 2).

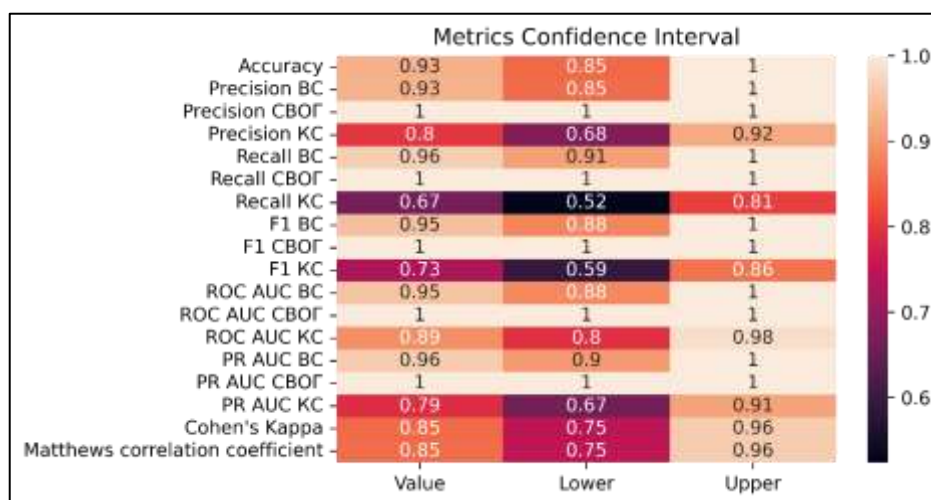


Рис. 2. Метрики точності, влучності, чутливості, F1-міри, ROC AUC, PR AUC, коефіцієнт Каппа та коефіцієнт кореляції Метьюса для тестового набору даних

Fig. 2. Metrics of precision, accuracy, sensitivity, F1-measure, ROC AUC, PR AUC, Kappa coefficients, and Matthews correlation coefficients for the test set data

Модель на основі XGBoost дала змогу визначити показник важливості для кожної ознаки, яка використовується для оцінки впливу показника на прогнозування. Дана оцінка вказувала, наскільки корисною була кожна ознака при побудові дерева рішень –

чим частіше ознака використовувалася для прийняття ключових рішень у дереві моделі, тим вищою є її важливість. Важливість ознак у запропонованій нами моделі диференційної діагностики ВС, СВОГ та КС відображена на рисунку 3.

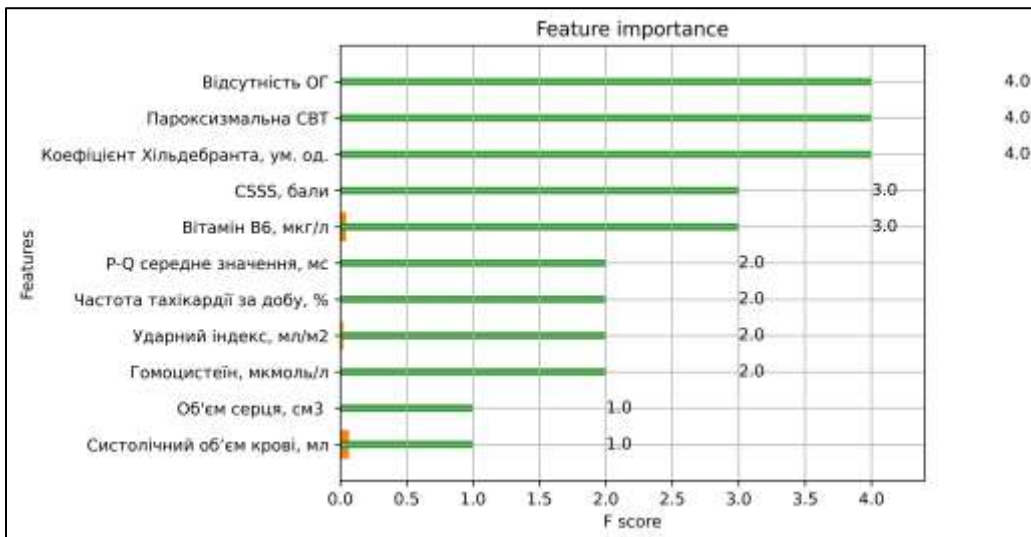


Рис. 3. Важливість ознак в моделі диференційної діагностики видів синкопе
 Fig. 3. Feature importance for the model of differential diagnosis of syncope types

XGBoost – це алгоритм посилення градієнта, який створює ансамбль дерев рішень для прогнозування цільової змінної. Алгоритм працює шляхом повторного додавання дерев рішень до ансамбля, при цьому кожне нове дерево намагається виправити помилки попередніх дерев. Під час навчання алгоритм оптимізує

функцію втрат, яка визначала різницю між прогнозованими та фактичними значеннями цільової змінної. Застосування методу градієнтного бустінгу XGBoost для створення моделі машинного навчання дозволило побудувати шість дерев, які в сумі дають змогу диференціювати ВС, СВОГ і КС у дитячому віці (рис. 4–9).

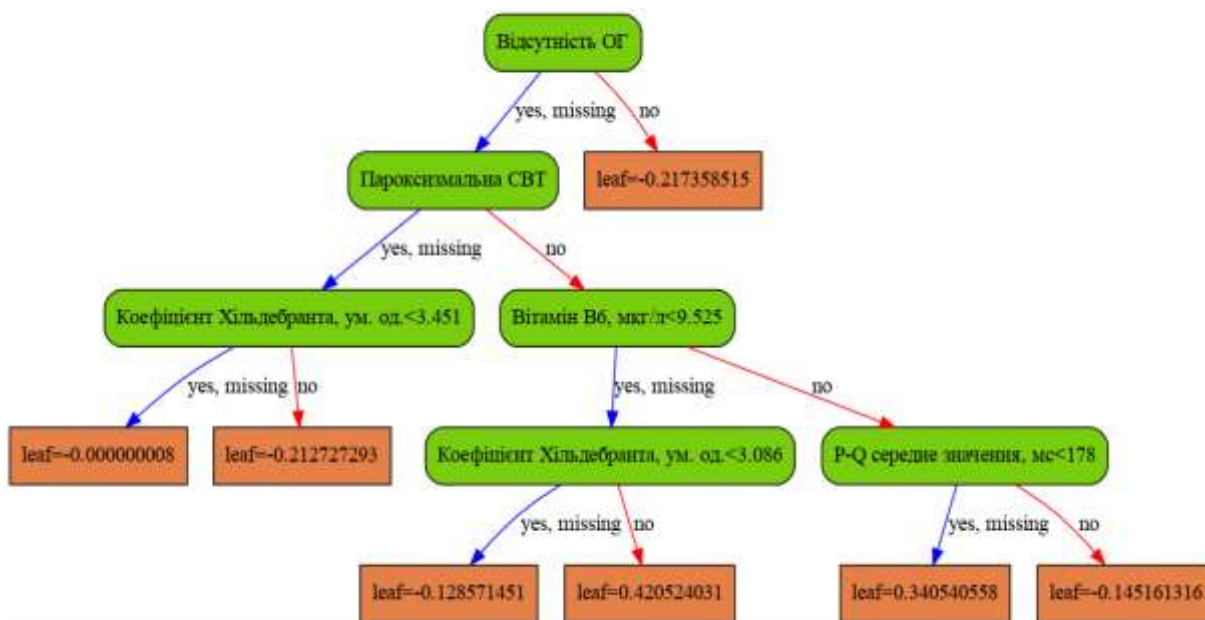


Рис. 4. Перше дерево моделі машинного навчання для диференційної діагностики ВС, СВОГ і КС.
 Умовні позначки: ОГ- ортостатична гіпотензія; СВТ – суправентрикулярна тахікардія
 Fig. 4. The first tree of the machine learning model for the differential diagnosis of vasovagal syncope, syncope due to orthostatic hypotension and cardiac syncope

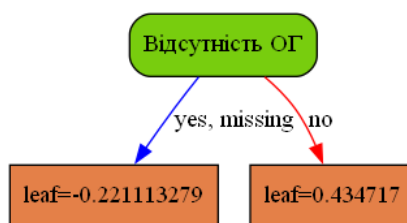


Рис. 5. Друге дерево моделі машинного навчання для диференційної діагностики ВС, СВОГ і КС. Умовні позначки: ОГ- ортостатична гіпотензія

Fig. 5. The second tree of the machine learning model for the differential diagnosis of vasovagal syncope, syncope due to orthostatic hypotension and cardiac syncope

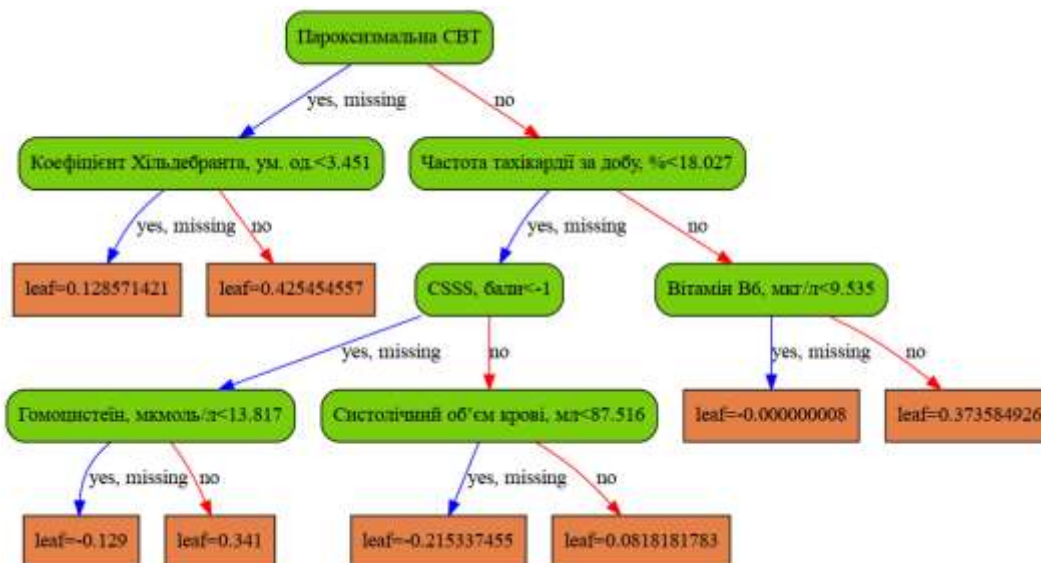


Рис. 6. Третє дерево моделі машинного навчання для диференційної діагностики ВС, СВОГ і КС. Умовні позначки: СВТ – суправентрикулярна тахікардія; CSSS - Calgary Syncope Seizure Score

Fig. 6. The third tree of the machine learning model for the differential diagnosis of vasovagal syncope, syncope due to orthostatic hypotension and cardiac syncope

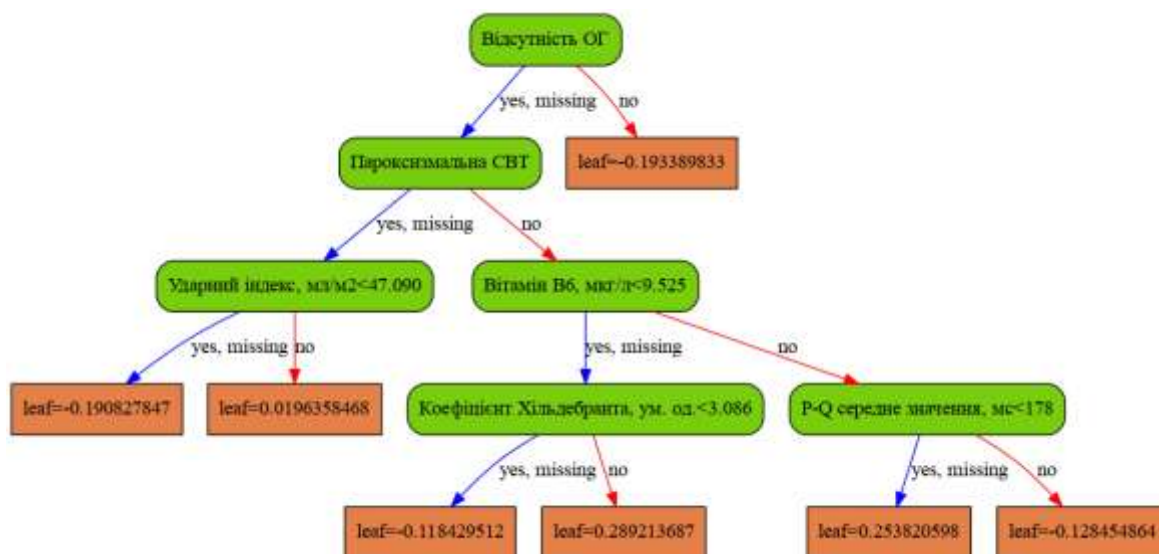


Рис. 7. Четверте дерево моделі машинного навчання для диференційної діагностики ВС, СВОГ і КС. Умовні позначки: ОГ – ортостатична гіпотензія; СВТ – суправентрикулярна тахікардія

Fig. 7. The fourth tree of the machine learning model for the differential diagnosis of vasovagal syncope, syncope due to orthostatic hypotension and cardiac syncope

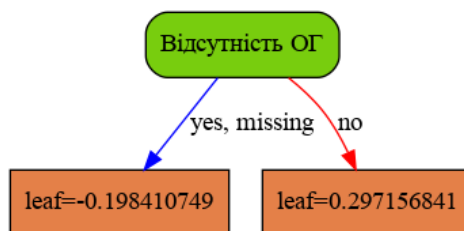


Рисунок 8 – П'яте дерево моделі машинного навчання для диференційної діагностики ВС, СВОГ і КС. Умовні позначки: ОГ – ортостатична гіпотензія

Figure 8. The fifth tree of the machine learning model for the differential diagnosis of vasovagal syncope, syncope due to orthostatic hypotension and cardiac syncope

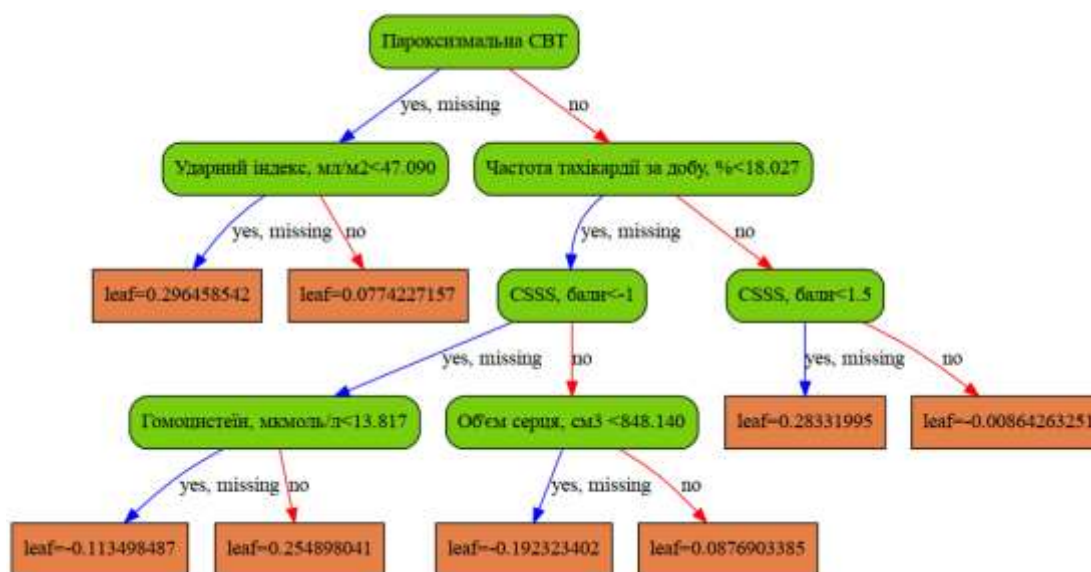


Рис. 9. Шосте дерево моделі машинного навчання для диференційної діагностики ВС, СВОГ і КС. Умовні позначки: СВТ – суправентрикулярна тахікардія; CSSS – Calgary Syncope Seizure Score

Fig.9. The sixth tree of the machine learning model for the differential diagnosis of vasovagal syncope, syncope due to orthostatic hypotension and cardiac syncope

ОБГОВОРЕННЯ

За останні 2 десятиліття було досягнуто суттєвого успіху у розумінні різних аспектів проблеми діагностики та лікування синкопе. Значна частина цього прогресу полягає у формуванні чітких критеріїв, які є необхідними для постановки діагнозу синкопе відповідно до міжнародних протоколів діагностики та лікування Європейської асоціації кардіологів перегляду 2018 року [11] й Американського коледжу кардіологів разом з Американською асоціацією серця 2017 року розгляду [17]. Однак, зважаючи на різноманіття клінічних симптомів та труднощі збору анамнезу у пацієнтів, які

втрачали свідомість, постановка виду синкопе – ВС, СВОГ чи КС, досі залишається складною задачею для цілого ряду клініцистів на всіх етапах моніторингу дитини із непритомністю. Пошук нових методів диференційної діагностики видів синкопе є одним із актуальних та пріоритетних напрямків сучасної педіатрії, який дозволить не лише покращити діагностику ВС, СВОГ та КС у дитячому віці, але й слугуватиме основою для прийняття своєчасних діагностичних та терапевтичних рішень.

Для побудови даної програми машинного навчання за основу було взято алгоритм XGBoost для мультикласової класифікації, який є сучасним та

ефективним методом у побудові алгоритмів диференційної діагностики та прогнозування ризику розвитку захворювань у дорослих та дітей [18–19]. Незважаючи на це, існують лише поодинокі повідомлення щодо використання програм машинного навчання для диференційної діагностики синкопе. В одному з таких досліджень було показано, що класифікатор XGBoost може диференціювати пацієнтів, у яких розвинеться ВС під час інвазивних медичних маніпуляцій з використанням голки від побічної на основі даних про температуру обличчя з чутливістю 0,87, специфічністю 0,84, f1-мірою 0,86 і PR-AUC 0,93 [20]. Changjian Li, et al. розробили модель диференційної діагностики психогенних псевдо-синкопе та ВС, базуючись на оцінці клінічних симптомів з чутливістю 73,7 %, специфічністю 93,0 % і точністю 87,7 % [21].

Запропонована модель машинного навчання є ефективною у диференціації ВС, СВОГ та КС за показниками метрик точності (0,93), влучності (0,93 у групі ВС 1,00 у групі СВОГ; 0,80 у групі КС), чутливості (0,96 для ВС; 1,00 для СВОГ; 0,67 у групі КС), f1-міри (0,95 у групі ВС; 1,00 у групі СВОГ; 0,73 у групі КС), ROC AUC (0,95 у дітей із ВС; 1,00 у дітей із СВОГ; 0,89 у дітей із КС), PR AUC (0,96 для ВС; 1,00 для СВОГ; 0,79 у групі КС), коефіцієнту Каппа (0,85) та коефіцієнту

кореляції Метьюса (0,85). Отримані результати є свідченням високого рівня диференціації для класів захворювань ВС та СВОГ за показниками точності, влучності, чутливості, F1-міри, ROC AUC та PR AUC. Однак, для класу КС модель підтверджує помірні результати диференціації з нижчими показниками метрик оцінки ефективності моделі, але все ще демонструє здатність розрізняти клас КС від інших класів синкопе. Узагальнюючи, модель показує добрі результати класифікації, особливо для класів ВС та СВОГ, але може потребувати додаткових даних для точної диференціації КС від інших видів синкопе.

В даній моделі найбільшою інформативністю володіють показники ОГ, пароксизмальності СВТ, коефіцієнту Хільдебранта, шкали CSSS та вітаміну В6. Дещо менш інформативними, проте важливими показниками диференціації видів синкопе є середня тривалість інтервалу P-Q за добу, частота тахікардії за добу, ударний індекс, гомоцистеїн, об'єм серця, систолічний об'єм крові.

Для прикладу, застосуємо нашу модель для диференційної діагностики синкопе у трьох дітей із діагнозами ВС, СВОГ та КС. Вихідні дані цих пацієнтів відтворені у таблиці 1.

На основі вхідних отримуємо наступні листкові значення дерев (табл. 2).

Таблиця 1

Table 1

Вихідні дані пацієнтів із діагнозами ВС, СВОГ і КС, які використовуються для прикладу обрахунку запропонованої моделі машинного навчання

Initial data of patients with diagnoses of vasovagal syncope, syncope due to orthostatic hypotension and cardiac syncope, used for the calculation example of the proposed machine learning model

Показник	Діагноз пацієнта		
	ВС	СВОГ	КС
Відсутність ОГ	так	ні	так
Пароксизмальна СВТ	ні	ні	так
Коефіцієнт Хільдебранта, ум. од.	4,6	3,75	4,90
CSSS, бали	3,0	1,0	0,0
Вітамін В6, мкг/л	8,52	9,51	16,33
Середня тривалість P-Q за добу, мс	203,0	179,0	179,0
Частота тахікардії за добу, %	28,90	4,00	46,20
Ударний індекс, мл/м ²	40,2	42,8	41,4
Гомоцистеїн, мкмоль/л	15,3	13,8	12,9
Об'єм серця, см ³	751,8	768,4	524,4
Систолічний об'єм крові, мл	66,9	73,0	37,3

Листкові значення дерев моделі диференційної діагностики синкопе у кожному окремо взятому випадку ВС, СВОГ, КС

The trees leaf values of the model of differential diagnosis syncope for vasovagal syncope, syncope due to orthostatic hypotension and cardiac syncope

Номер дерева моделі	Діагноз		
	ВС	СВОГ	КС
1	0,420524031	-0,217358515	-0,212727293
2	-0,221113279	0,434717000	-0,221113279
3	-0.000000008	-0,215337455	0,425454557
4	0,289213687	-0,193389833	-0,190827847
5	-0,198410749	0,297156841	-0,198410749
6	-0,008642632	-0,192323402	0,296458542

Розрахунок першого прикладу для дитини з діагнозом ВС.

$$\sigma(x_1) = e^{(0,420524031 + 0,289213687)} / e^{(0,420524031 + 0,289213687)} + e^{(-0,221113279 - 0,198410749)} + e^{(-0.000000008 - 0,008642632)} = 0,5522.$$

$$\sigma(x_2) = e^{(-0,221113279 - 0,198410749)} / e^{(0,420524031 + 0,289213687)} + e^{(-0,221113279 - 0,198410749)} + e^{(-0.000000008 - 0,008642632)} = 0,1785.$$

$$\sigma(x_3) = e^{(-0.000000008 - 0,008642632)} / e^{(0,420524031 + 0,289213687)} + e^{(-0,221113279 - 0,198410749)} + e^{(-0.000000008 - 0,008642632)} = 0,2692.$$

Застосування моделі диференційної діагностики синкопе свідчить про вірогідність діагнозу ВС – 56 %, СВОГ – 18 %, КС – 27 %.

Розрахунок другого прикладу для пацієнта зі СВОГ.

$$\sigma(x_1) = e^{(0,434717000 + 0,297156841)} / e^{(0,434717000 + 0,297156841)} + e^{(-0,217358515 - 0,193389833)} + e^{(-0,215337455 - 0,192323402)} = 0,6101.$$

$$\sigma(x_2) = e^{(-0,217358515 - 0,193389833)} / e^{(0,434717000 + 0,297156841)} + e^{(-0,217358515 - 0,193389833)} + e^{(-0,215337455 - 0,192323402)} = 0,1946.$$

$$\sigma(x_3) = e^{(-0,215337455 - 0,192323402)} / e^{(0,434717000 + 0,297156841)} + e^{(-0,217358515 - 0,193389833)} + e^{(-0,215337455 - 0,192323402)} = 0,1952.$$

Таким чином, вірогідність діагнозу ВС складає 20 %, СВОГ – 61 %, КС – 19 %.

Розрахунок другого прикладу для дитини з діагнозом КС.

$$\sigma(x_1) = e^{(0,425454557 + 0,296458542)} / e^{(0,425454557 + 0,296458542)} + e^{(-0,212727293 - 0,190827847)} + e^{(-0,221113279 - 0,198410749)} = 0,6083.$$

$$\sigma(x_2) = e^{(-0,212727293 - 0,190827847)} / e^{(0,425454557 + 0,296458542)} + e^{(-0,212727293 - 0,190827847)} + e^{(-0,221113279 - 0,198410749)} = 0,1974.$$

$$\sigma(x_3) = e^{(-0,221113279 - 0,198410749)} / e^{(0,425454557 + 0,296458542)} + e^{(-0,212727293 - 0,190827847)} + e^{(-0,221113279 - 0,198410749)} = 0,1943.$$

За результатами даного підрахунку ймовірність діагнозу ВС – 20 %, СВОГ – 19 %, КС – 61 %.

Отже, дана модель машинного навчання володіє достатньою ефективністю та може використовуватися педіатрами та спеціалістами з дитячої кардіоревматології для диференційної діагностики синкопе у дитячому віці як додатковий інструмент поряд із класичними діагностичними критеріями ВС, СВОГ і КС на першому етапі діагностики.

ВИСНОВКИ

Запропонована модель машинного навчання на основі алгоритму XGBoost є ефективним додатковим інструментом диференційної діагностики ВС, СВОГ та КС у дітей із нетравматичними транзиторними втратами свідомості синкопального походження, про що свідчать метрики

точності (0,93), влучності (0,93 у групі ВС; 1,00 у групі СВОГ; 0,80 у групі КС), чутливості (0,96 для ВС; 1,00 для СВОГ; 0,67 у групі КС), f1-міри (0,95 у групі ВС; 1,00 у групі СВОГ; 0,73 у групі КС), ROC AUC (0,95 у дітей із ВС; 1,00 у дітей із СВОГ; 0,89 у дітей із КС), PR AUC (0,96 для ВС; 1,00 для СВОГ; 0,79 у групі КС), коефіцієнт Каппа (0,85) та коефіцієнт кореляції Метьюса (0,85). В алгоритмі диференційної діагностики видів синкопе найінформативнішими показниками є ОГ, пароксизмальна СВТ, коефіцієнт Хільдебранта, шкала CSSS, вітамін В6, середня тривалість інтервалу P-Q за добу,

частота тахікардії за добу, ударний індекс, гомоцестин, об'єм серця, систолічний об'єм крові. Використання даних показників у запропонованій моделі підвищує точність діагностики ВС, СВОГ та КС у дитячому віці.

Інформація про фінансування. Дане дослідження було проведено виключно з ініціативи авторів та не було фінансоване жодною установою чи грантом. The study was conducted on the initiative of the author and wasn't covered by any funding source.

Подяки. Автори щиро вдячні усім дітям та їхнім батькам за участь у цьому дослідженні.

СПИСОК ЛІТЕРАТУРИ

1. Leibetseder A, Eisermann M, LaFrance WC, et al. How to distinguish seizures from non-epileptic manifestations. *Epileptic disorders : international epilepsy journal with videotape*. 2020;22(6):716–738. DOI: <https://doi.org/10.1684/epd.2020.1234>
2. Wang X, Wang S, Xiao H, et al. The value of QT interval in differentiating vasovagal syncope from epilepsy in children. *Ital J Pediatr*. 2022;48(1):197. DOI: <https://doi.org/10.1186/s13052-022-01388-2>
3. Belhassen B, Tovia-Brodie O. Red Flags in Syncope: Clues for the Diagnosis of Idiopathic Ventricular Fibrillation. *The American journal of medicine*. 2022;135(12):1434–1436. DOI: <https://doi.org/10.1016/j.amjmed.2022.06.020>
4. Albassam OT, Redelmeier RJ, Shadowitz S, et al. Did This Patient Have Cardiac Syncope?: The Rational Clinical Examination Systematic Review. *JAMA*. 2019;321(24):2448–2457. DOI: <https://doi.org/10.1001/jama.2019.8001>
5. Mascia G, Bona RD, Ameri P, et al. Brugada syndrome and syncope: a practical approach for diagnosis and treatment. *Europace : European pacing, arrhythmias, and cardiac electrophysiology : journal of the working groups on cardiac pacing, arrhythmias, and cardiac cellular electrophysiology of the European Society of Cardiology*. 2021;23(7):996–1002. DOI: <https://doi.org/10.1093/europace/euaa370>
6. Von Alvensleben JC. Syncope and Palpitations: A Review. *Pediatr Clin North Am*. 2020;67(5):801-810. DOI: <https://doi.org/10.1016/j.pcl.2020.05.004>
7. Zavala R, Metais B, Tuckfield L, et al. Pediatric Syncope: A Systematic Review. *Pediatr Emerg Care*. 2020;36(9):442-445. DOI: <https://doi.org/10.1097/PEC.0000000000002149>
8. Yeom JS, Woo HO. Pediatric syncope: pearls and pitfalls in history taking. *Clin Exp Pediatr*. 2023;66(3):88–97. DOI: <https://doi.org/10.3345/cep.2022.00451>
9. Pecci R, Ceccofiglio A, Ungar A. Syncope-presyncope and dizziness, differential diagnosis. *Minerva medica*. 2022;113(2):273–280. DOI: <https://doi.org/10.23736/S0026-4806.21.07808-3>
10. Fedorowski A. Postural orthostatic tachycardia syndrome: clinical presentation, aetiology and management. *J Intern Med*. 2019;285(4):352-366. DOI: <https://doi.org/10.1111/joim.12852>
11. Brignole M, Moya A, de Lange FJ, et al. 2018 ESC Guidelines for the diagnosis and management of syncope. *Eur Heart J*. 2018;39(21):1883-1948. DOI: <https://doi.org/10.1093/eurheartj/ehy037>
12. Zou R, Wang S, Zhu L, et al. Calgary score and modified Calgary score in the differential diagnosis between neurally mediated syncope and epilepsy in children. *Neurol Sci*. 2017;38(1):143–149. DOI: <https://doi.org/10.1007/s10072-016-2740-5>
13. Tanaka H, Fujita Y, Takenaka Y, et al. Japanese clinical guidelines for juvenile orthostatic dysregulation version 1. *Pediatr Int*. 2009;51(1):169-79. DOI: <https://doi.org/10.1111/j.1442-200X.2008.02783.x>
14. Ковальчук ТА, Лучишин НЮ. Рівень функціонування адаптаційних механізмів серцево-судинної системи в дітей із синкопе різного генезу. *Сучасна педіатрія. Україна*. 2022;3(123):16-26. DOI: <https://doi.org/10.15574/SP.2022.123.16>
15. Gan L. XGBoost-Based E-Commerce Customer Loss Prediction. *Comput Intell Neurosci*. 2022;2022:1858300. DOI: <https://doi.org/10.1155/2022/1858300>
16. Xia Y, Li X, Chen X, et al. Inferring Retinal Degeneration-Related Genes Based on Xgboost. *Front Mol Biosci*. 2022;9:843150. DOI: <https://doi.org/10.3389/fmolb.2022.843150>

17. Shen WK, Sheldon RS, Benditt DG, et al. 2017 ACC/AHA/HRS Guideline for the Evaluation and Management of Patients With Syncope: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines and the Heart Rhythm Society. *Circulation*. 2017;136(5):e60–e122. DOI: <https://doi.org/10.1161/CIR.0000000000000499>
18. Song X, Zhu J, Tan X, et al. XGBoost-Based Feature Learning Method for Mining COVID-19 Novel Diagnostic Markers. *Front Public Health*. 2022;10:926069. DOI: <https://doi.org/10.3389/fpubh.2022.926069>
19. Li Q, Yang H, Wang P, et al. XGBoost-based and tumor-immune characterized gene signature for the prediction of metastatic status in breast cancer. *J Transl Med*. 2022;20(1):177. DOI: <https://doi.org/10.1186/s12967-022-03369-9>
20. Rudokaite J, Ong LLS, Onal Ertugrul I, et al. Predicting vasovagal reactions to needles with anticipatory facial temperature profiles. *Sci Rep*. 2023;13(1):9667. DOI: <https://doi.org/10.1038/s41598-023-36207-z>
21. Li C, Zhang Y, Liao Y, et al. Differential Diagnosis Between Psychogenic Pseudosyncope and Vasovagal Syncope in Children: A Quantitative Scoring Model Based on Clinical Manifestations. *Front Cardiovasc Med*. 2022;9:839183. DOI: <https://doi.org/10.3389/fcvm.2022.839183>

Отримано: 11.08.2023

Прийнято до друку: 16.10.2023

Конфлікт інтересів: відсутній.

Tetiana Kovalchuk^{A, B, C, D, F}, **Oksana Boyarchuk**^{A, B, E, F}, **Sviatoslav Bogai**^{A, C, F}

tetianakovalchuk@gmail.com

XGBOOST MACHINE LEARNING ALGORITHM FOR DIFFERENTIAL DIAGNOSIS OF PEDIATRIC SYNCOPE

A – research concept and design; B – collection and/or assembly of data; C – data analysis and interpretation; D – writing the article; E – critical revision of the article; F – final approval of the article

Abstract. The search for new methods of differential diagnosis of syncope types will allow to improve the diagnosis of vasovagal syncope (VVS), syncope due to orthostatic hypotension (OH) and cardiac syncope (CS) in childhood in order to make timely adequate diagnostic and therapeutic decisions.

The aim of the study was to develop an effective machine learning model for the differential diagnosis of VVS, syncope due to OH and CS in children.

Materials and Methods. 140 patients with syncope, aged 8–17 years, were examined: 92 children with a diagnosis of VVS, 28 children with syncope due to OH and 20 children with CS. A machine learning model was built using XGBoost algorithm for multiclass classification based on input clinical, laboratory and instrumental patient data.

Results. The developed machine learning model based on the XGBoost algorithm is effective in the differential diagnosis of VVS, syncope due to OH and CS, which is confirmed by the metrics of accuracy (0.93), precision (0.93 for VVS; 1.00 for syncope due to OH; 0.80 for CS), recall (0.96 for VVS; 1.00 for syncope due to OH; 0.67 for CS), f1 (0.95 for VVS; 1.00 for syncope due to OH; 0.73 for CS), ROC AUC (0.95 for VVS; 1.00 for syncope due to OH; 0.89 for CS), PR AUC (0.96 for VVS; 1.00 for syncope due to OH; 0.79 for CS), Cohen's Kappa (0.85), and Matthews correlation coefficient (0.85). The most informative parameters of the syncope types differential diagnosis model are OH, paroxysmal supraventricular tachycardia, Hildebrandt coefficient, Calgary Syncope Seizure Score, vitamin B6, average duration of the P-Q interval during 24 hours, duration of tachycardia during 24 hours, stroke index, homocysteine, heart volume, and systolic blood volume.

Conclusions. The proposed machine learning model has sufficient efficiency and can be used by pediatricians and pediatric cardiologists for the differential diagnosis of VS, syncope due to OH, and CS in childhood.

KEY WORDS: *vasovagal syncope, syncope due to orthostatic hypotension, cardiac syncope, differential diagnosis, XGBoost model, children*

INFORMATION ABOUT AUTHORS

Kovalchuk Tetiana – MD, PhD, Associate Professor of the Department of Pediatrics № 2 at I. Horbachevsky Ternopil National Medical University, Ternopil, Ukraine; e-mail: tetianakovalchuk@gmail.com. <http://orcid.org/0000-0003-2455-3278>.

Boyarchuk Oksana – MD, PhD, DSc, Professor, Head of the Department of Pediatrics and Pediatric Surgery at I. Horbachevsky Ternopil National Medical University, Ternopil, Ukraine. e-mail: boyarchuk@tdmu.edu.ua. <http://orcid.org/0000-0002-1234-0040>.

Bogai Sviatoslav - it-specialist. e-mail: slavabogaj@gmail.com

For citation:

Kovalchuk TA, Boyarchuk OR, Bogai SYe. XGBOOST MACHINE LEARNING ALGORITHM FOR DIFFERENTIAL DIAGNOSIS OF PEDIATRIC SYNCOPE. The Journal of V.N. Karazin Kharkiv National University. Series «Medicine». 2023;47:33–46. DOI: **10.26565/2313-6693-2023-47-04** (in Ukrainian)

REFERENCES

1. Leibetseder A, Eisermann M, LaFrance W C, Jr, Nobili, L., von Oertzen, T. J. How to distinguish seizures from non-epileptic manifestations. *Epileptic disorders : international epilepsy journal with videotape.* 2020;22(6):716–738. DOI: <https://doi.org/10.1684/epd.2020.1234>
2. Wang X, Wang S, Xiao H, Zou R, Cai H, Liu L, Li F, Wang Y, Xu Y, Wang C. The value of QT interval in differentiating vasovagal syncope from epilepsy in children. *Ital J Pediatr.* 2022;48(1):197. DOI: <https://doi.org/10.1186/s13052-022-01388-2>
3. Belhassen B, Tovias-Brodie O. Red Flags in Syncope: Clues for the Diagnosis of Idiopathic Ventricular Fibrillation. *The American journal of medicine.* 2022;135(12):1434–1436. DOI: <https://doi.org/10.1016/j.amjmed.2022.06.020>
4. Albassam OT, Redelmeier RJ, Shadowitz S, Husain AM, Simel D, Etchells EE. Did This Patient Have Cardiac Syncope?: The Rational Clinical Examination Systematic Review. *JAMA.* 2019 Jun 25;321(24):2448–2457. DOI: <https://doi.org/10.1001/jama.2019.8001>
5. Mascia G, Bona RD, Ameri P, Canepa M, Porto I, Parati G, Crotti L, Brignole, M. Brugada syndrome and syncope: a practical approach for diagnosis and treatment. *Europace : European pacing, arrhythmias, and cardiac electrophysiology : journal of the working groups on cardiac pacing, arrhythmias, and cardiac cellular electrophysiology of the European Society of Cardiology.* 2021;23(7):996–1002. DOI: <https://doi.org/10.1093/europace/eaab370>
6. Von Alvensleben JC. Syncope and Palpitations: A Review. *Pediatr Clin North Am.* 2020;67(5):801–810. DOI: <https://doi.org/10.1016/j.pcl.2020.05.004>
7. Zavala R, Metais B, Tuckfield L, DelVecchio M, Aronoff S. Pediatric Syncope: A Systematic Review. *Pediatr Emerg Care.* 2020;36(9):442–445. DOI: <https://doi.org/10.1097/PEC.0000000000002149>
8. Yeom JS, Woo HO. Pediatric syncope: pearls and pitfalls in history taking. *Clin Exp Pediatr.* 2023;66(3):88–97. DOI: <https://doi.org/10.3345/cep.2022.00451>
9. Pecci R, Ceccofiglio A, Ungar A. Syncope-presyncope and dizziness, differential diagnosis. *Minerva medica.* 2022;113(2):273–280. DOI: <https://doi.org/10.23736/S0026-4806.21.07808-3>
10. Fedorowski A. Postural orthostatic tachycardia syndrome: clinical presentation, aetiology and management. *J Intern Med.* 2019;285(4):352–366. DOI: <https://doi.org/10.1111/joim.12852>
11. Brignole M, Moya A, de Lange FJ, et al. 2018 ESC Guidelines for the diagnosis and management of syncope. *Eur Heart J.* 2018;39(21):1883–1948. DOI: <https://doi.org/10.1093/eurheartj/ehy037>
12. Zou R, Wang S, Zhu L, et al. Calgary score and modified Calgary score in the differential diagnosis between neurally mediated syncope and epilepsy in children. *Neurol Sci.* 2017;38(1):143–149. DOI: <https://doi.org/10.1007/s10072-016-2740-5>
13. Tanaka H, Fujita Y, Takenaka Y, et al. Japanese clinical guidelines for juvenile orthostatic dysregulation version 1. *Pediatr Int.* 2009;51(1):169–79. DOI: <https://doi.org/10.1111/j.1442-200X.2008.02783.x>
14. Kovalchuk TA, Luchyshyn NYu. The level of functioning of adaptive mechanisms of the cardiovascular system in children with syncope of various genesis. *Modern pediatrics. [Ukraine].* 2022;3(123):16–26. DOI: <https://doi.org/10.15574/SP.2022.123.16>
15. Gan L. XGBoost-Based E-Commerce Customer Loss Prediction. *Comput Intell Neurosci.* 2022;2022:1858300. DOI: <https://doi.org/10.1155/2022/1858300>
16. Xia Y, Li X, Chen X, Lu C, Yu X. Inferring Retinal Degeneration-Related Genes Based on Xgboost. *Front Mol Biosci.* 2022;9:843150. DOI: <https://doi.org/10.3389/fmolb.2022.843150>
17. Shen WK, Sheldon RS, Benditt DG, Cohen MI, Forman DE, Goldberger ZD, Grubb BP, Hamdan MH, Krahn AD, Link MS, Olshansky B, Raj SR, Sandhu RK, Sorajja D, Sun BC, Yancy CW. 2017 ACC/AHA/HRS Guideline for the Evaluation and Management of Patients With Syncope: A Report of the American College

- of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines and the Heart Rhythm Society. *Circulation*. 2017;136(5):e60-e122. DOI: <https://doi.org/10.1161/CIR.0000000000000499>
18. Song X, Zhu J, Tan X, Yu W, Wang Q, Shen D, Chen W. XGBoost-Based Feature Learning Method for Mining COVID-19 Novel Diagnostic Markers. *Front Public Health*. 2022;10:926069. DOI: <https://doi.org/10.3389/fpubh.2022.926069>
19. Li Q, Yang H, Wang P, Liu X, Lv K, Ye M. XGBoost-based and tumor-immune characterized gene signature for the prediction of metastatic status in breast cancer. *J Transl Med*. 2022;20(1):177. DOI: <https://doi.org/10.1186/s12967-022-03369-9>
20. Rudokaite J, Ong LLS, Onal Ertugrul I, Janssen MP, Huis In 't Veld EMJ. Predicting vasovagal reactions to needles with anticipatory facial temperature profiles. *Sci Rep*. 2023;13(1):9667. DOI: <https://doi.org/10.1038/s41598-023-36207-z>
21. Li C, Zhang Y, Liao Y, et al. Differential Diagnosis Between Psychogenic Pseudosyncope and Vasovagal Syncope in Children: A Quantitative Scoring Model Based on Clinical Manifestations. *Front Cardiovasc Med*. 2022;9:839183. DOI: <https://doi.org/10.3389/fcvm.2022.839183>

Received: 08.11.2023

Accepted: 10.16.2023

Conflicts of interest: author has no conflict of interest to declare.
