

Oleksii Avilov, Anton Popov, Vladimir Timofeiev, Laurent Bougrain, Patrick
Henaff

► **To cite this version:**

Oleksii Avilov, Anton Popov, Vladimir Timofeiev, Laurent Bougrain, Patrick Henaff.
, . Microsystems, Electronics and Acoustics, Oleksii
Viktorovych Bogdanov, 2018, 23 (5), pp.25-31. 10.20535/2523-4455.2018.23.5.134021 . hal-02153795

HAL Id: hal-02153795

<https://hal.inria.fr/hal-02153795>

Submitted on 12 Jun 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Електронні системи та сигнали

УДК 004.67

DOI: [10.20535/2523-4455.2018.23.5.134021](https://doi.org/10.20535/2523-4455.2018.23.5.134021)

Оцінка якості уявних рухів на основі машинного навчання для застосування в нейрокомп'ютерних інтерфейсах

Авілов О.О., ORCID [0000-0001-6802-0663](https://orcid.org/0000-0001-6802-0663)e-mail: lesha.avilov@gmail.comПопов А.О., к.т.н., доцент, ORCID [0000-0002-1194-4424](https://orcid.org/0000-0002-1194-4424)e-mail: anton.popov@ieee.orgТимофеев В.І., д.т.н., проф., ORCID [0000-0003-0515-1580](https://orcid.org/0000-0003-0515-1580)e-mail: v.timofeyev@kpi.ua

Національний технічний університет України

"Київський політехнічний інститут ім. Ігоря Сікорського" <https://kpi.ua>,

Київ, Україна

Бугрен Л., к.т.н. доц., ORCID [0000-0001-6794-0505](https://orcid.org/0000-0001-6794-0505)e-mail: laurent.bougrain@loria.frЕнаф П., д.т.н. проф., ORCID [0000-0002-5186-2466](https://orcid.org/0000-0002-5186-2466)e-mail: patrick.henaff@loria.frУніверситет Лотарингії <http://univ-lorraine.fr>

Нансі, Франція

Реферат—У роботі представлено застосування машинного навчання для виявлення «викидів» (випадкових відхилень) у даних та підвищення точності класифікації уявних рухів нейрокомп'ютерними інтерфейсами.

Використана база даних сигналів електроенцефалограм 29 суб'єктів, що виконували уявні рухи правою та лівою рукою (NIRx GmbH, Берлін). Представлений підхід на основі методів машинного навчання без вчителя, за допомогою яких відбувається детектування викидів у даних. В результаті отримується співвідношення параметру насичення даних викидами до точності класифікації. Обраний параметр насичення даних викидами фактично є показником якості вхідних даних і використовується для очищення даних перед тренуванням нейрокомп'ютерного інтерфейсу. У результаті використання запропонованого підходу точність класифікації уявних рухів збільшилась в середньому на 14.9% для восьми суб'єктів. Для інших суб'єктів точність лишилась незмінною, або зменшилась. В результаті виявлено, що запропонований підхід є суб'єкто-специфічним і потребує налаштування для окремого користувача та вдосконалення для широкого застосування, проте вже зараз дозволяє суттєво підвищити точність класифікації уявних рухів окремих суб'єктів.

Бібл. 28, рис. 2, табл. 2.

Ключові слова — машинне навчання; нейрокомп'ютерний інтерфейс; штучний інтелект; виявлення аномалій.

I. ВСТУП

Машинне навчання — підгалузь штучного інтелекту, яка вивчає алгоритми, що можуть навчатися на сукупності даних і виконувати передбачувальний аналіз на них [1]. Моделі, побудовані завдяки цим алгоритмам, дозволяють дослідникам, інженерам та аналітикам створювати надійні рішення та розкривати приховані залежності шляхом навчання з історичних співвідношень та тенденцій у даних.

Так завдяки технікам машинного навчання з високою точністю розраховуються такі параметри антен, як підсилення, смуга пропускання, ефективність випромінювання, резонансна частота [2].

Підходи машинного навчання ефективні у зниженні ступеню складності даних і швидко стали трендом в обробці багатовимірних даних. У роботі [3] автори розглянули проблему спрощення процесу окреслення структурної еволюції, що дозволило підвищити точність класифікації структурних фаз через



потрійні фазові простори. Моделювання за допомогою методів машинного навчання в зонній інженерії дозволило розрізнити графенові наночастиці з різними енергетичними щільностями на основі їх молекулярної топології [4], розробляти нові оксиди металів для елементів сонячних батарей [5], аналізувати енергію зв'язку молекул [6], що є основою для моделювання і прогнозування маси лікарських препаратів у твердих ліпідних наночастичках. Техніки машинного навчання також можна використовувати для пошуку та виявлення дефектів кристалічних ґраток матеріалів [7, 8, 9].

Методи та техніки машинного навчання широко використовуються в задачах біомедицини. Так останні технологічні досягнення дозволили надійно виявляти і отримувати характеристики циркулюючих пухлинних клітин у крові хворих на рак [10]. Також машинне навчання використовується у задачах молекулярної біології, а саме для розпізнавання тунелювання, для ідентифікації захоплених між тунельними електродами цільових молекул, заміщених молекулами розпізнавання, які служать у якості специфічних хімічних зв'язків між металевими електродами і захопленими цільовими молекулами [11]. Кількісний опис шару адсорбції білка використовується для вивчення впливу на з'єднання клітин, викликаного поверхнево-модифікованими наночастичками золота різних розмірів [12].

Ще однією областю, де широко використовується машинне навчання, є область нейрокомп'ютерних інтерфейсів, адже завдяки технікам машинного навчання можна моделювати складні процеси у мозку людини, розпізнавати, оцінювати, класифікувати об'єкти. Нейрокомп'ютерний інтерфейс — це система, створена для обміну інформацією між мозком та електронним пристроєм. Дослідження та розробка нейрокомп'ютерних інтерфейсів зосереджена, головним чином, на застосуванні у нейропротезуванні, на відновленні пошкоджень слуху, зору та руху. Це можливе завдяки пластичності нервової системи, коли здорові нейрони з сусідніх областей мозку після навчання і адаптації можуть переймати на себе функції пошкоджених.

Нейрокомп'ютерні інтерфейси поділяють на інвазивні, частково інвазивні та неінвазивні. Переважна більшість систем є неінвазивними та створюються на основі електроенцефалографа. Перевагою таких систем є те, що вони не потребують хірургічного втручання.

Нейрокомп'ютерні інтерфейси, засновані на руховій активності, спрямовані на відновлення рухових навичок у людей з паралічем або надання пристроїв для допомоги таким людям (інтерфейси з комп'ютерами, роботизовані руки, інвалідні крісла) [13, 14]. Основними напрямками розробки в даній сфері є: підвищення точності розпізнавання рухів шляхом розробки нових методів, архітектур машинного навчання та наборів ознак, скорочення часу прийняття рішення, збільшення кількості ступенів свободи систем. У таких інтерфейсах широко використовуються

уявні рухи. Такі рухи відбуваються в результаті ментального процесу, під час якого людина уявляє певний рух [15], в наслідок чого модулюється нейронна активність в моторних областях кори головного мозку, яка може спостерігатись за допомогою електроенцефалограм [16,17]. Завдяки машинному навчанню рухи розпізнаються і кодуються в певні команди для керованого приладу і далі слідує певна дія або ж зворотній відгук для користувача.

Для визначення уявних рухів використовують методи електроенцефалографії, магнітоенцефалографії, функціональної магніто-резонансної томографії. За допомогою кожного з цих методів вимірюється зміна нейронної активності в мю-, альфа-, бета-ритмах у моторних областях кори головного мозку та безпосередньо у первинній моторній корі, яка відповідає за свідомі рухи.

Комбіновані уявні рухи використовуються для управління роботизованою рукою. Зазвичай у нейрокомп'ютерних інтерфейсах використовуються 2-3 уявні рухи (права рука, ліва рука, ноги). Проте, комбінуючи ці рухи, можна отримати вісім різних команд. Завдяки збільшенню кількості команд можливо збільшити кількість ступенів свободи пристрою, що управляється. Точність класифікації для такої 8-класової задачі, суттєво нижче ніж при бінарних задачах, — 30-40% [18]. Такий рівень точності суттєво вищий ніж випадкова величина (12,5% для 8-класової задачі). Проте даний рівень точності наразі унеможливує застосування даного підходу для реальних пристроїв і потребує доопрацювання.

Ще одним напрямком, в якому відбуваються розробки, є пришвидшення роботи нейрокомп'ютерних інтерфейсів. Останні дослідження підтверджують, що використання дискретних (одиничних) уявних рухів замість тривалих уявлень даних рухів (впродовж 2-3 секунд) для нейрокомп'ютерних інтерфейсів може прискорити швидкість детектування рухів та підвищити швидкодію інтерфейсу в цілому. При цьому точність детектування дискретних уявних рухів знаходиться на тому ж рівні, що і тривалих уявних рухів — близько 70% для бінарної задачі [19].

Оскільки для тренування нейрокомп'ютерних інтерфейсів використовуються уявні рухи, часто неможливо оцінити якість виконання суб'єктом поставленого завдання. Процес виконання уявного руху не є легким завданням і потребує попередніх тренувань. Тому часто точність класифікації для одних суб'єктів у рази вища ніж у інших суб'єктів.

Всі вище описані методи можна покращити, якби існувала методика оцінки якості виконання суб'єктом уявних рухів. Тоді дані експериментів можна було б відфільтрувати і залишити лише коректні дані для тренування. У даному дослідженні показано, як шляхом виявлення викидів у даних можна відбрати лише якісні дані для тренування нейрокомп'ютерного інтерфейсу, що дозволить підвищити точність детектування уявних рухів.



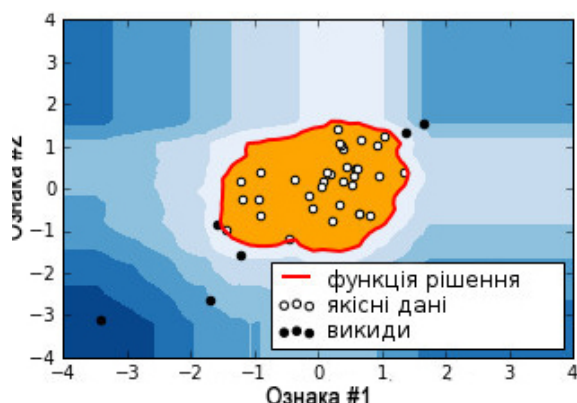


Рис. 1 Графічне представлення результатів роботи алгоритму «Ізоляційний ліс» для відокремлення якісних даних від викидів.

У цій роботі пропонується використовувати методи виявлення викидів у даних, зареєстрованих при виконанні уявних рухів, для аналізу їх якості.

II. МЕТОДИ

Виявлення аномалій або виявлення викидів – це ідентифікація елементів, подій або спостережень, які не відповідають очікуваному шаблону, чи іншим елементам набору даних. Широко застосовуються у виявленні шахрайств серед банківських операцій, помилок у тексті, структурних дефектів та медичних проблем.

У роботі було досліджено три методи виявлення викидів: метод еліптичної обвідної, однокласовий метод опорних векторів та ізоляційний ліс.

Одним із поширених способів виявлення викидів є припущення, що дані надходять з відомого розподілу (наприклад, Гауссового розподілу). З цього припущення визначається форма густини розподілу даних, а віддалені спостереження, які виходять далеко за межі форми густини розподілу, визначаються як викиди.

Метод еліптичної обвідної визначає коваріаційну оцінку даних таким чином, щоб центральні точки даних знаходились в еліпсі, ігноруючи точки поза центром (викиди) [20].

Однокласова реалізація методу опорних векторів навпаки працює краще, коли розподіл даних сильно не Гауссовий, оскільки цей алгоритм здатний фіксувати форму набору даних. Однокласовий метод опорних векторів виявляє м'яку межу заданого набору даних і визначає, чи нові спостереження належать цьому набору даних, чи ні [21]. Для визначення кордону потрібен вибір ядра та параметр, що відповідає за імовірне насичення даних викидами. В цій роботі використане ядро RBF (radial basis function).

Алгоритм "Ізоляційний ліс" виділяє (ізолює) спостереження шляхом випадкового вибору ознаки та випадковим вибором роздільного значення між максимальним та мінімальним значеннями обраної ознаки. Оскільки рекурсивне розділення може бути

представлене як дерево, кількість розщеплень, необхідних для виділення спостереження, еквівалентна довжині шляху від кореневого вузла до кінцевого вузла. Тоді оцінка розраховується як довжина шляху для ізоляції спостереження. Довжина цього шляху, усереднена по всім таким випадковим деревам, визначає функцію рішення. Випадкове розділення створює помітно коротші шляхи для аномалій. Отже, коли ліс випадкових дерев часто виявляє коротші довжини шляху для одних і тих самих спостережень, вони, ймовірно, і є аномаліями [22].

В цій роботі використано алгоритм "Ізоляційний ліс" з кількістю дерев, рівній 100. Для тренування кожного дерева використовувалась повна тренувальна вибірка, адже величина вибірки є незначною і зменшення її не призведе до пришвидшення роботи алгоритму, а лише до зниження точності. Кількість ознак для побудови кожного дерева була обрана рівною п'яти.

III. ПІДВИЩЕННЯ ЯКОСТІ УЯВНИХ РУХІВ НА ОСНОВІ ВИДАЛЕННЯ ВИКИДІВ

У дослідженні була використана база даних сигналів електроенцефалограм для нейрокомп'ютерних інтерфейсів [23] записана NIRx GmbH, Берлін, Німеччина. В експерименті взяло участь 29 суб'єктів, яким було запропоновано виконувати уявні рухи правою та лівою рукою. Сигнали ЕЕГ були записані з частотою дискретизації 1000 Гц та стиснені до частоти дискретизації 200 Гц. Сигнали містять 30 каналів ЕЕГ згідно з міжнародною системою 10-5 (AFp1, AFp2, AFF1h, AFF2h, AFF5h, AFF6h, F3, F4, F7, F8, FCC3h, FCC4h, FCC5h, FCC6h, T7, T8, Cz, CCP3h, CCP4h, CCP5h, CCP6h, Pz, P3, P4, P7, P8, PPO1h, PPO2h, POO1, POO2 та Fz). Суб'єкти мали уявно стискати праву чи ліву руку за звуковим сигналом впродовж трьох сесій. В кожній сесії міститься по 10 уявних рухів кожного виду.

Далі сигнали з восьми електродів (FCC3h, FCC4h, FCC5h, FCC6h, CCP3h, CCP4h, CCP5h, CCP6h), що відповідають моторним зонам кори головного мозку, фільтруються в альфа- та бета-діапазонах частот. Використовується смуговий фільтр Баттерворта 5-го порядку в діапазонах альфа (8-12 Гц), нижня альфа (7-10 Гц), верхня альфа (11-13 Гц), бета (17-25 Гц) та альфа + бета (7-30 Гц). Головна особливість фільтра Баттерворта [24] полягає в тому, що частотна характеристика є максимально плоскою (не має пульсації) в смузі пропускання і знижується до нуля в смузі затримання [25].

За отриманим фільтрованим сигналам розраховується набір ознак на основі «десинхронізації» та «синхронізації» (event-related desynchronization (ERD) and synchronization (ERS) [26]. Використовуючи цей показник, можна побачити спадання та зростання коливань в альфа- та бета-ритмах, які мають місце в локалізованих ділянках моторної кори головного мозку [27, 28].



Дані діляться на тестову та тренувальну вибірки (дві сесії на тренування, одна сесія на тест) для кожного суб'єкта окремо. Обсяги вибірок склали 40 спостережень для тренувальної та 20 спостережень для тестової вибірки. Для кожного класу тренувальної вибірки відбувається виявлення та видалення викидів (для різного параметру насичення даних викидами, цей параметр відповідає за відсоток даних, які будуть відібрані як викиди). Виявлення викидів відбувається на основі одного з алгоритмів кластеризації (наприклад, «Ізоляційний ліс» (рис. 1).

Після кластеризації відбувається класифікація (використовується метод логістичної регресії) та крос-валідація методом Leave-One-Out, коли для тесту залишається одне спостереження, а всі інші використовуються для тренування. У результаті класифікації отримується співвідношення параметру насичення даних викидами до точності класифікації. Обирається параметр з максимальною точністю класифікації. Після цього відбувається процес навчання моделей кластеризації (детектори аномалій) з обраним кращим параметром на попередньому кроці для кожного класу даних. Зберігаються навчені на тренувальній вибірці детектори аномалій та очищені дані.

З використанням навчених детекторів аномалій, відбувається очищення тестового набору даних. Оскільки істинний клас тестових даних невідомий, то невідомо, який з детекторів аномалій використовувати. Тому, в роботі, запропоновано використовувати метод логічної диз'юнкції: щоб спостереження було позначене як аномалія, воно має бути позначене як аномалія кожним детектором аномалій. Після видалення аномалій отримується очищений тестовий набір даних. Відбувається тренування класифікаторів і розрахунок точностей (табл. 1). Можна побачити, що і однокласовий метод опорних векторів, і ізоляційний ліс мають однаковий рівень точності. Результати для методу еліптичної обвідної не наведені у таблиці, оскільки метод перенавчався (англ. overfitting) і не генералізував дані (не опишував взаємозв'язок, що лежить в основі даних) через високу кількість ознак. Це призвело до того, що модель, навчена цим методом, видала близько 90% даних з тестової вибірки як викиди.

У табл. 2 приведені результати класифікації уявних рухів до та після виявлення викидів методом «Ізоляційний ліс». Суб'єкти 5, 22, 23 були видалені через низьку якість даних. Як можемо бачити, з наведених даних, у середньому для всіх суб'єктів точність лишилась на тому самому рівні. Проте для деяких суб'єктів точність значно збільшилась (наприклад суб'єкти 1, 11, 18, 19, 21, 24, 27, 28), для яких підвищення точності склало у середньому 14.9% (рис. 2). Це говорить про те, що даний підхід є суб'єктно специфічним, і може бути успішно застосований для певної групи суб'єктів. Водночас, методика виявлення викидів у початкових даних потребує вдосконалення для застосування в ширшому колі суб'єктів.

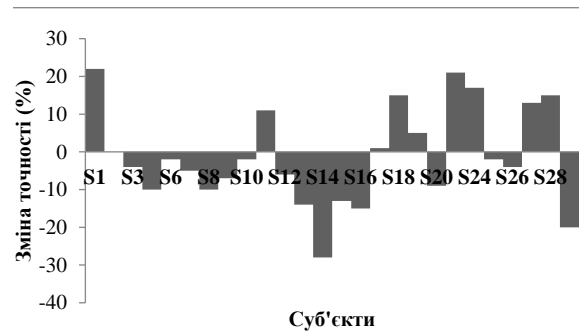


Рис. 2 Зміна точності класифікації уявних рухів після виявлення викидів для різних суб'єктів

ТАБЛИЦЯ 1 РЕЗУЛЬТАТИ КЛАСИФІКАЦІЇ ПІСЛЯ ВИЯВЛЕННЯ ВИКИДІВ

Метод	Точність	Кращий параметр насичення даних викидами (%)
Ізоляційний ліс	0.6 ± 0.13	34.82 ± 14.31
Однокласовий метод опорних векторів	0.6 ± 0.12	27.9 ± 13.72

ТАБЛИЦЯ 2 Точність до та після виявлення викидів методом «Ізоляційний ліс»

Суб'єкт	Точність до виявлення викидів	Точність після виявлення викидів	Кращий параметр насичення даних викидами (%)
S1	0.73 ± 0.06	0.97 ± 0.07	45.33 ± 8.22
S2	0.75 ± 0.0	0.75 ± 0.09	45.33 ± 15.08
S3	0.5 ± 0.04	0.46 ± 0.04	24.0 ± 21.42
S4	0.58 ± 0.02	0.48 ± 0.08	24.0 ± 22.86
S6	0.48 ± 0.08	0.46 ± 0.12	38.67 ± 19.96
S7	0.52 ± 0.05	0.47 ± 0.41	57.33 ± 3.77
S8	0.6 ± 0.08	0.5 ± 0.06	16.0 ± 8.64
S9	0.67 ± 0.09	0.6 ± 0.02	37.33 ± 19.69
S10	0.62 ± 0.09	0.6 ± 0.15	28.0 ± 17.28
S11	0.58 ± 0.06	0.69 ± 0.1	28.0 ± 16.33
S12	0.47 ± 0.05	0.41 ± 0.43	45.33 ± 17.99
S13	0.55 ± 0.11	0.41 ± 0.29	46.67 ± 4.99
S14	0.48 ± 0.02	0.2 ± 0.18	48.0 ± 11.31
S15	0.7 ± 0.11	0.57 ± 0.19	38.67 ± 16.11
S16	0.75 ± 0.07	0.6 ± 0.09	24.0 ± 11.78
S17	0.47 ± 0.05	0.48 ± 0.06	42.67 ± 12.36
S18	0.52 ± 0.09	0.67 ± 0.12	40.0 ± 17.28
S19	0.78 ± 0.08	0.83 ± 0.06	2.67 ± 1.89
S20	0.6 ± 0.04	0.51 ± 0.07	33.33 ± 9.98
S21	0.42 ± 0.06	0.63 ± 0.16	28.0 ± 16.33
S24	0.47 ± 0.05	0.64 ± 0.08	32.0 ± 14.24
S25	0.93 ± 0.09	0.91 ± 0.08	25.33 ± 25.37
S26	0.58 ± 0.06	0.54 ± 0.03	34.67 ± 10.5
S27	0.75 ± 0.04	0.88 ± 0.1	41.33 ± 10.5
S28	0.38 ± 0.05	0.53 ± 0.02	45.33 ± 13.2
S29	0.72 ± 0.02	0.52 ± 0.13	33.33 ± 24.94
Середнє	0.6 ± 0.06	0.6 ± 0.13	34.82 ± 14.31

ВИСНОВКИ

Оцінка якості уявних рухів являється новою і не вирішеною задачею, яка дозволить підвищити точність класифікації та удосконалив процеси тренування та калібрування нейрокомп'ютерних інтерфейсів. Застосовані в роботі алгоритми кластеризації (однокласовий метод опорних векторів та «ізоляційний ліс») дозволяють виявляти та видаляти викиди в даних та завдяки цьому виконувати класифікацію очищених даних. Проте, отримані точності класифікації уявних рухів, до та після виявлення викидів дозволяють зробити висновок, що даний підхід є суб'єктивно специфічним і потребує додаткового налаштування параметрів окремо для кожного суб'єкта. У 8 з 26 суб'єктів точність класифікації уявних рухів виросла в середньому на 14.9 %.

ПЕРЕЛІК ПОСИЛАНЬ

- [1] R. Kohavi and F. Provost, *Glossary of terms. Machine Learning*, 30th ed. Boston: Kluwer Academic Publishers, 1998, pp. 271–274, DOI: [10.1023/A:1017181826899](https://doi.org/10.1023/A:1017181826899).
- [2] C. Gianfagna, H. Yu, M. Swaminathan, R. Pulugurtha, R. Tummala and G. Antonini, "Machine-Learning Approach for Design of Nanomagnetic-Based Antennas", *Journal of Electronic Materials*, vol. 46, no. 8, pp. 4963–4975, 2017, DOI: [10.1007/s11664-017-5487-8](https://doi.org/10.1007/s11664-017-5487-8).
- [3] A. Kusne, T. Gao, A. Mehta, L. Ke, M. Nguyen, K. Ho, V. Antropov, C. Wang, M. Kramer, C. Long and I. Takeuchi, "On-the-fly machine-learning for high-throughput experiments: search for rare-earth-free permanent magnets", *Scientific Reports*, vol. 4, no. 1, 2014, DOI: [10.1038/srep06367](https://doi.org/10.1038/srep06367).
- [4] M. Fernandez, J. Abreu, H. Shi and A. Barnard, "Machine Learning Prediction of the Energy Gap of Graphene Nanoflakes Using Topological Autocorrelation Vectors", *ACS Combinatorial Science*, vol. 18, no. 11, pp. 661–664, 2016, DOI: [10.1021/acscombsci.6b00094](https://doi.org/10.1021/acscombsci.6b00094).
- [5] A. Yosipof, O. Nahum, A. Anderson, H. Barad, A. Zaban and H. Senderowitz, "Data Mining and Machine Learning Tools for Combinatorial Material Science of All-Oxide Photovoltaic Cells", *Molecular Informatics*, vol. 34, no. 6–7, pp. 367–379, 2015, DOI: [10.1002/minf.201400174](https://doi.org/10.1002/minf.201400174).
- [6] R. Hathout and A. Metwally, "Towards better modelling of drug-loading in solid lipid nanoparticles: Molecular dynamics, docking experiments and Gaussian Processes machine learning", *European Journal of Pharmaceutics and Biopharmaceutics*, vol. 108, pp. 262–268, 2016, DOI: [10.1016/j.ejpb.2016.07.019](https://doi.org/10.1016/j.ejpb.2016.07.019).
- [7] N. Laanait, Z. Zhang and C. Schlepütz, "Imaging nanoscale lattice variations by machine learning of x-ray diffraction microscopy data", *Nanotechnology*, vol. 27, no. 37, p. 374002, 2016, DOI: [10.1088/0957-4484/27/37/374002](https://doi.org/10.1088/0957-4484/27/37/374002).
- [8] W. Cheng, "Identify problematic layout patterns through volume diagnosis", *2015 International Symposium on VLSI Technology, Systems and Applications*, 2015, DOI: [10.1109/VLSI-TSA.2015.7117551](https://doi.org/10.1109/VLSI-TSA.2015.7117551).
- [9] A. Hughes, Z. Liu, M. Raftari and M. Reeves, "A workflow for characterizing nanoparticle monolayers for biosensors: Machine learning on real and artificial SEM images", *PeerJ PrePrints*, 2015, DOI: [10.7287/peerj.preprints.671v2](https://doi.org/10.7287/peerj.preprints.671v2).
- [10] Y. Ozaki, H. Yamada, H. Kikuchi, T. Murakami, T. Matsumoto, T. Kawabata, Y. Hiramatsu, M. Ohta, K. Kamiya, M. Baba, T. Yamauchi, K. Goto, Y. Ueda, S. Okazaki and H. Konno, "Label-free imaging identification of WBCs based on the features of quantitative phase microscope images for negative selection of CTCs", *Cancer Research*, vol. 76, no. 14, pp. 3952–3952, 2016, DOI: [10.1158/1538-7445.AM2016-3952](https://doi.org/10.1158/1538-7445.AM2016-3952).
- [11] P. Krstić, B. Ashcroft and S. Lindsay, "Physical model for recognition tunneling", *Nanotechnology*, vol. 26, no. 8, p. 084001, 2015, DOI: [10.1088/0957-4484/26/8/084001](https://doi.org/10.1088/0957-4484/26/8/084001).
- [12] E. Papa, J. Doucet, A. Sangion and A. Doucet-Panaye, "Investigation of the influence of protein corona composition on gold nanoparticle bioactivity using machine learning approaches", *SAR and QSAR in Environmental Research*, vol. 27, no. 7, pp. 521–538, 2016, DOI: [10.1080/1062936X.2016.1197310](https://doi.org/10.1080/1062936X.2016.1197310).
- [13] J. Wolpaw, *Brain-computer interfaces*. New York: Oxford University Press, 2012, ISBN: 9780195388855.
- [14] A. Butler and S. Page, "Mental Practice With Motor Imagery: Evidence for Motor Recovery and Cortical Reorganization After Stroke", *Archives of Physical Medicine and Rehabilitation*, vol. 87, no. 12, pp. 2–11, 2006, DOI: [10.1016/j.apmr.2006.08.326](https://doi.org/10.1016/j.apmr.2006.08.326).
- [15] M. Jeannerod, "Mental imagery in the motor context", *Neuropsychologia*, vol. 33, no. 11, pp. 1419–1432, 1995, DOI: [10.1016/0028-3932\(95\)00073-C](https://doi.org/10.1016/0028-3932(95)00073-C).
- [16] M. Lotze and U. Halsband, "Motor imagery", *Journal of Physiology-Paris*, vol. 99, no. 4–6, pp. 386–395, 2006, DOI: [10.1016/j.jphysparis.2006.03.012](https://doi.org/10.1016/j.jphysparis.2006.03.012).
- [17] C. Neuper, R. Scherer, M. Reiner and G. Pfurtscheller, "Imagery of motor actions: Differential effects of kinesthetic and visual-motor mode of imagery in single-trial EEG", *Cognitive Brain Research*, vol. 25, no. 3, pp. 668–677, 2005, DOI: [10.1016/j.cog-brainres.2005.08.014](https://doi.org/10.1016/j.cog-brainres.2005.08.014).
- [18] C. Lindig-Leon, S. Rimbart, O. Avilov and L. Bougrain, "Scalp EEG activity during simple and combined motor imageries to control a robotic ARM", *IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON)*, pp. 322–327, 2017, DOI: [10.1109/UKRCON.2017.8100502](https://doi.org/10.1109/UKRCON.2017.8100502).
- [19] S. Rimbart, O. Avilov, L. Bougrain, "Discrete motor imageries can be used to allow a faster detection", *7th Graz Brain-Computer Interface Conference*, 2017, HAL Id: hal-01512407.
- [20] P. Rousseeuw and K. van Driessen, "A Fast Algorithm for the Minimum Covariance Determinant Estimator", *Technometrics*, vol. 41, no. 3, p. 212, 1999.
- [21] B. Schölkopf, J. Platt, J. Shawe-Taylor, A. Smola and R. Williamson, "Estimating the Support of a High-Dimensional Distribution", *Neural Computation*, vol. 13, no. 7, pp. 1443–1471, 2001.
- [22] Liu, Fei Tony, Ting, Kai Ming and Zhou, Zhi-Hua, "Isolation forest", *Data Mining. ICDM'08. Eighth IEEE International Conference*, 2008, DOI: [10.1109/ICDM.2008.17](https://doi.org/10.1109/ICDM.2008.17).
- [23] J. Shin, A. von Luhmann, B. Blankertz, D. Kim, J. Jeong, H. Hwang and K. Muller, "Open Access Dataset for EEG+NIRS Single-Trial Classification", *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 10, pp. 1735–1745, 2017, DOI: [10.1109/TNSRE.2016.2628057](https://doi.org/10.1109/TNSRE.2016.2628057).
- [24] S. Butterworth, "On the Theory of Filter Amplifiers", *Experimental Wireless and the Wireless Engineer*, vol. 7, p. 536–541, 1930.
- [25] G. Bianchi and R. Sorrentino, "Electronic filter simulation design", *McGraw-Hill Professional*, p. 17–20, 2007, ISBN: 0071494677 / 9780071494670.
- [26] G. Pfurtscheller and F. Lopes da Silva, "Event-related EEG/MEG synchronization and desynchronization: basic principles", *Clinical Neurophysiology*, vol. 110, no. 11, pp. 1842–1857, 1999, DOI: [10.1016/S1388-2457\(99\)00141-8](https://doi.org/10.1016/S1388-2457(99)00141-8).
- [27] G. Pfurtscheller and A. Berghold, "Patterns of cortical activation during planning of voluntary movement", *Electroencephalography and Clinical Neurophysiology*, vol. 72, no. 3, p. 250–258, 1989, DOI: [10.1016/0013-4694\(89\)90250-2](https://doi.org/10.1016/0013-4694(89)90250-2).
- [28] W. Klimesch, G. Pfurtscheller, W. Mohl, and H. Schimke, "Event-related desynchronization, erd-mapping and hemispheric differences for words and numbers", *International Journal of Psychophysiology*, vol. 8, no. 3, p. 297–308, 1990, DOI: [10.1016/0167-8760\(90\)90020-E](https://doi.org/10.1016/0167-8760(90)90020-E).



Надійшла до редакції 18 червня 2018 р.

УДК 004.67

Оценка качества мнимых движений на основании машинного обучения для применения в нейрокомпьютерных интерфейсах

Авилов А.А., ORCID [0000-0001-6802-0663](https://orcid.org/0000-0001-6802-0663)

e-mail: lesha.avilov@gmail.com

Попов А.А., к.т.н. доцент, ORCID [0000-0002-1194-4424](https://orcid.org/0000-0002-1194-4424)

e-mail: anton.popov@ieee.org

Тимофеев В.И., д.т.н. проф., ORCID [0000-0003-0515-1580](https://orcid.org/0000-0003-0515-1580)

e-mail: v.timofeyev@kpi.ua

Национальный технический университет Украины

“Киевский политехнический институт имени Игоря Сикорского” <https://kpi.ua>,

Киев, Украина

Бугрэн Л., к.т.н. доц., ORCID [0000-0001-6794-0505](https://orcid.org/0000-0001-6794-0505)

e-mail: laurent.bougrain@loria.fr

Энаф П., д.т.н. проф., ORCID [0000-0002-5186-2466](https://orcid.org/0000-0002-5186-2466)

e-mail: patrick.henaff@loria.fr

Университет Лотарингии <http://univ-lorraine.fr>

Нанси, Франция

Реферат—В работе представлено применение машинного обучения для обнаружения выбросов (случайных отклонений) в данных и повышения точности классификации воображаемых движений в нейрокомпьютерных интерфейсах.

Использована база данных сигналов электроэнцефалограмм 29 субъектов, которые выполняли мнимые движения правой и левой рукой (NIRx GmbH, Берлин). Представлен подход на основе методов машинного обучения без учителя, с помощью которых происходит детектирование выбросов в данных. В результате получено соотношение параметра насыщения данных выбросами к точности классификации. Выбранный параметр насыщения данных выбросами фактически является показателем качества входных данных и используется для очистки данных перед тренировкой нейрокомпьютерного интерфейса. В результате использования предложенного подхода точность классификации воображаемых движений увеличилась на 14.9% для 8 субъектов. Для других субъектов точность осталась неизменной, или уменьшилась. Данный подход является субъектно-специфичным и требует настройки для конкретного пользователя и совершенствования для широкого применения, однако уже сейчас позволяет существенно повысить точность классификации воображаемых движений отдельных субъектов.

Библ. 28, рис. 2, табл. 2.

Ключевые слова — машинное обучение; нейрокомпьютерный интерфейс; искусственный интеллект; обнаружение аномалий.



UDC 004.67

Estimation of Imaginary Movements Quality Based on Machine Learning for Brain Computer Interface Applications

O. O. Avilov, ORCID [0000-0001-6802-0663](https://orcid.org/0000-0001-6802-0663)

e-mail: lesha.avilov@gmail.com

A. O. Popov, PhD Assoc.Prof., ORCID [0000-0002-1194-4424](https://orcid.org/0000-0002-1194-4424)

e-mail: anton.popov@ieee.org

V. I. Timofeyev, Dr.Sc.(Eng.) Prof., ORCID [0000-0003-0515-1580](https://orcid.org/0000-0003-0515-1580)

e-mail: v.timofeyev@kpi.ua

National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute” <https://kpi.ua>,
Kyiv, Ukraine

L. Bougrain, PhD Assoc.Prof., ORCID [0000-0001-6794-0505](https://orcid.org/0000-0001-6794-0505)

e-mail: laurent.bougrain@loria.fr

P. Henaff, Dr.Sc.(Eng.), Professor, ORCID [0000-0002-5186-2466](https://orcid.org/0000-0002-5186-2466)

e-mail: patrick.henaff@loria.fr

The University of Lorraine <http://univ-lorraine.fr>

Nancy, France

Abstract—Brain-computer interfaces based on motor activity aim at restoring motor skills for people with paralysis or providing devices to help such people (interfaces with computers, robotic hands, wheelchairs). In such interfaces, imaginary movements are widely used. Imaginary movement it is the mental process, during which a person imagines a certain movement, as a result of which the neural activity in the motor regions of the cerebral cortex is modulated. These oscillations can be observed in electroencephalograms. Imaginary movements are used to train neurocomputer interfaces, and it is often impossible to evaluate the quality of the subject's performance of the task. The process of performing imaginary movement is not an easy task and requires prior training. Therefore, the accuracy of classification for some subjects can be higher than for other. All methods for imagine movement classification could be improved if there would be a methodology for assessing the quality of subject performance of imaginary movements. Then the data of the experiments could be filtered, and only valid data for training could be used.

The paper presents the application of machine learning to detect anomalies (outliers) in data and improve the accuracy of the imaginary movement classification for brain-computer interfaces. The database with electroencephalogram signals of 29 subjects that performed right and left hand imaginary movements (NIRx GmbH, Berlin) was used. The signals from the eight channels corresponding to the motor zones of the cerebral cortex are filtered in alpha and beta frequency ranges. For feature extraction was used parameter event related synchronization and desynchronization which was obtained from different frequency bands after filter the raw EEG. Using this index, the decreasing and increasing of oscillations in alpha and beta rhythms that occur in the localized areas of the motor cortex of the brain can be seen. Support Vector Machine with linear kernel used for classification. An approach is presented on the basis of unsupervised machine learning methods for anomaly detection. As a result, the outlier fraction is obtained. This parameter shows the percentage of saturation data with anomalies. This parameter is actually an indicator of the quality of the input data and is used to clear data before training the brain computer interface. As a result of proposed approach application, the accuracy of the imaginary movement classification increased by 14.9% for 8 subjects. For other subjects, the accuracy remained unchanged, or decreased. This approach is subject-specific and requires customization for a specific user and improvement before further wide usage, but now it is already possible to significantly improve the accuracy of imaginary movement classification for individual subjects. Perspective use of anomaly detection for estimation of imaginary movement quality established.

Ref. 28, Fig. 2, Tabl. 2.

Keywords — machine learning; brain-computer interface; artificial intelligence; data science, anomaly detection.

