

К.Ю. Островська, В.Г. Порохнявий

ДОСЛІДЖЕННЯ НЕЙРОМЕРЕЖЕВИХ МОДЕЛЕЙ ДЛЯ ВІДСТЕЖЕННЯ ПОГЛЯДУ ТА ФІКСАЦІЇ ОБ'ЄКТІВ

Анотація. В роботі представлено порівняння нейромережевих моделей для відстеження погляду та фіксації об'єктів.

У цій роботі були вивчені алгоритми виявлення подій з різних областей: I-VT і I-DT, засновані на порогових значеннях, модель випадкового лісу в галузі машинного навчання і модель CNN в галузі глибокого навчання.

Була проведена оцінка ефективності їхньої класифікації з використанням одного і того ж набору даних для всіх методів. Також обговорювався взаємозв'язок між програмами та алгоритмами. Результати показали критичний вплив порогових значень класифікацію алгоритмів I-VT і I-DT. У зв'язку з цим пошук оптимального порогу алгоритмах, заснованих на порогових значеннях, є складним завданням. Алгоритми RF і CNN перевершують алгоритми, засновані на порогових значеннях, за всіма показниками продуктивності та здатні класифікувати дані щодо кількох класів.

Проведено тестування та порівняння різних АВП (алгоритмів виявлення подій). Дано теоретичні та експериментальні обґрунтування отриманим результатам. За результатами проведених експериментів було обрано найкращий наших завдань АВП (алгоритмів виявлення подій).

Також описані засоби реалізації до створення докладання. Описано переваги та недоліки обраних рішень. Вирішено технічні проблеми, що виникли. Спроектовано архітектуру, розроблено макет програми виходячи з необхідних завдань.

Ключові слова: машинне навчання, random forest, алгоритмів виявлення подій, додаток, саккада. плавне переслідування, постсаккадичні коливання, глісади згорткова нейронна мережа.

Вступ. Відстеження руху очей використовується у різноманітних сферах досліджень, таких як когнітивна наука, психологія, неврологія, інженерія, медицина та маркетинг.

Відстеження руху очей також використовується для керування автомобілем. Це робить цю технологію вкрай міждисциплінарною та використовується у різних сферах. Це також відображається в тому, як з часом розроблялися апаратне та програмне забезпечення для трекінгу очей.

В останні роки інтерес до алгоритмів виявлення подій (АВП) відновився у зв'язку із застосуванням до проблеми методів машинного навчання та появи баз даних. Розро-

бка програми, що використовує АВП дозволить розширити можливості застосування та вивчення алгоритмів.

Метою роботи є дослідження нейромережових моделей для відстеження погляду та фіксації об'єктів.

Опис набору даних. Для перевірки ефективності АВП буде використано загальнодоступний набір даних, записаний за допомогою високошвидкісного очного трекера 1250, фірми SensoMotoric Instruments (Teltow, Німеччина) на частоті 500 Гц.

При створенні датасета випробуванним пред'являлися статичні зображення, тексти, відеокліпи і прості точкові подразники, що рухаються. Дані були позначені вручну двома братами Маркусом Нюстромом та Річардом Андерсоном. Вони були розділені на наступні категорії: фіксації, саккади, пост-саккадичні коливання, плавне слідування, моргання та невизначені. Для цього дослідження використовувалися дані про перегляд зображень, позначені фіксаціями, саккадами та постсаккадичними коливаннями. Дані датасета відображені на рисунку 1 і є «сирими» даними. На рисунку 2 дані поділені на фіксації.

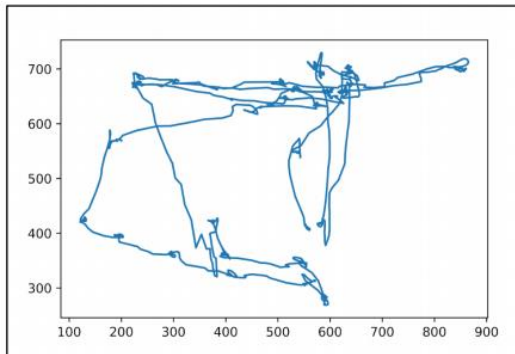


Рисунок 1 – Візуалізація даних датасету

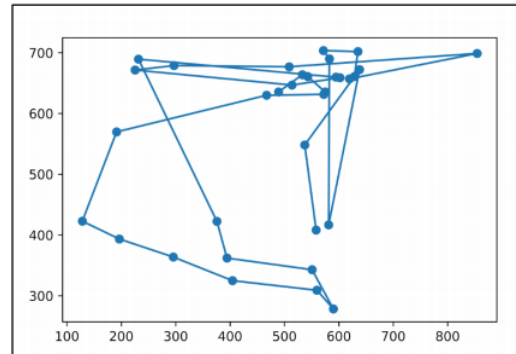


Рисунок 2 – Розмічені дані, на яких відображено фіксацію

Алгоритми. Алгоритми, що використовують граничне значення і які використовувались в роботі це:

- I-DT (визначення порогового значення дисперсії). Особливістю даного алгоритму є те, що він демонструє низьку ефективність при виявленні фіксацій та саккад, коли сигнал сильно зашумлений. У зв'язку з цим вибір порогових значень є найважливішим і безпосередньо впливатиме на ефективність класифікації.

Отримані результати відображено на рисунку 3. За ними можна говорити, що збільшення порогового значення дисперсії збільшує точність визначення фіксацій, але зменшує точність визначення саккади. При цьому зменшення порогового значення зменшує точність виявлення фіксації та збільшує виявлення саккади.

При значенні порогової дисперсії, що дорівнює 7-ми, точність виявлення фіксацій становить 99%, при цьому точність визначення саккади становить 82%. При пороговому значенні, що дорівнює 1-му, вже точність визначення саккади дорівнює 99%, але точність виявлення фіксації всього 39%.

Було вибрано оптимальне значення порогового значення, що дорівнює 3,5. При даному значенні можна отримати точність визначення фіксації 95% і точність виявлення саккади 93%.

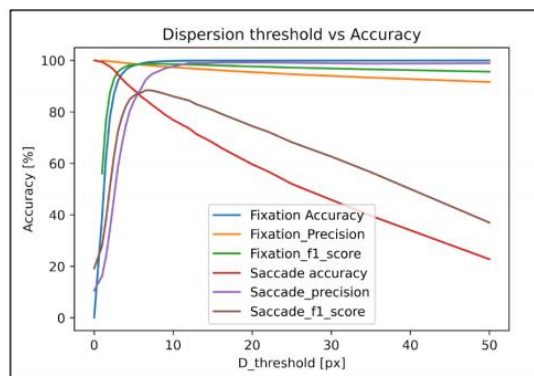


Рисунок 3 – Залежність точності визначення фіксацій та саккад для алгоритму I-DT від порогового значення дисперсії

• I-VT - Алгоритм визначення порогового значення швидкості (velocity threshold algorithm) є ще один алгоритм і основу для автоматизованого і об'єктивного стандартного алгоритму виявлення подій. Цей підхід використовувався в багатьох дослідженнях. Він заснований на тому факті, що саккадичні рухи очей характеризуються вищими значеннями швидкості, порівняно з рухами при фіксації. Профілі швидкостей рухів очей демонструють два основні розподіли швидкостей: низькі швидкості для фіксації та високі швидкості для саккад. Метод I-VT ідентифікує події шляхом обчислення швидкості руху від точки до точки, а потім класифікує подію як фіксацію або саккаду на основі значення цієї швидкості. Класичний метод I-VT призначений для класифікації тільки фіксацій та саккад на основі вхідних даних, отриманих при відстеженні руху очей.

На рисунку 4 показано вплив зміни порогових значень швидкості ефективність класифікації алгоритму I-VT. Точності класифікації кожного класу вимірюються з використанням коефіцієнтів запам'ятовування, точності та F1-балів, розрахованих на основі матриці помилок.

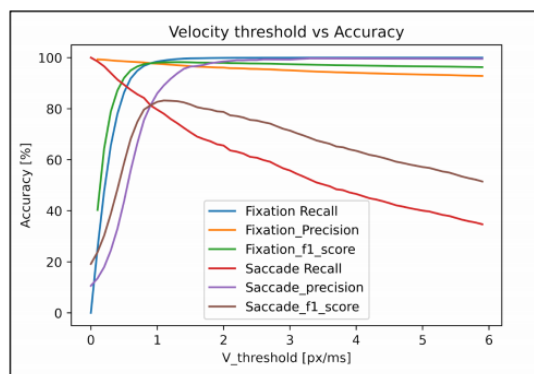


Рисунок 4 – Залежність точності визначення фіксацій та саккад алгоритму I-VT від порогових значень

У прикладі I-VT забезпечує максимальну точність фіксації на рівні 99% при пороговій швидкості 3,5 пікселів в секунду, а точність відтворення саккади трохи знижується при збільшенні порогового значення швидкості. Повторення саккад досягає 98%, а повторення фіксації - 25% при найменшому пороговому значенні швидкості, що дорівнює 0,1 пікселя в секунду, так як при цьому пороговому значенні більшість точок класифікуються як саккади. Через вплив граничного значення на точність класифікації алгоритму I-VT важливо визначити оптимальне граничне значення для обох методів фіксації та точності виконання саккади. Таким чином, у даному випадку оптимальне граничне значення швидкості для I-VT становить 0,5 пікселів в секунду.

Також в роботі використовувались алгоритми машинного навчання:

Random Forest (RF) – випадковий ліс та Convolutional neural network (CNN) – згортоква нейронна мережа.

Random Forest (RF) – випадковий ліс.

Повністю автоматизована класифікація подій, пов'язаних із рухом очей, із використанням класифікатора Random Forest для класифікації фіксацій, саккад та постсаккадичних коливань. Ефективність класифікації порівняли із сучасними алгоритмами та експертами, які працюють вручну. Результати показали, що алгоритм машинного навчання перевершує сучасні алгоритми та майже досягає продуктивності експертів. Однак, ця продуктивність була досягнута тільки для високоякісних даних з низьким рівнем шуму.

На рисунку 5 показано матрицю помилок для оцінки вибірки за вибіркою. Фіксації правильно позначені у 97 % випадків, тоді як ПСК і саккад є тенденція до того що, щоб їх позначали як фіксації. Саккада і ПСК правильно ідентифікуються у 91% та 76% кадрів відповідно. Близько 7% ПСК помилково класифікуються як фіксація, а 17% ПСК класифікуються як саккада. Це тому, що більшість навчальних подій були фіксаціями, і модель, природно, прагне класифікувати всі неоднозначні вибірки як фіксації.

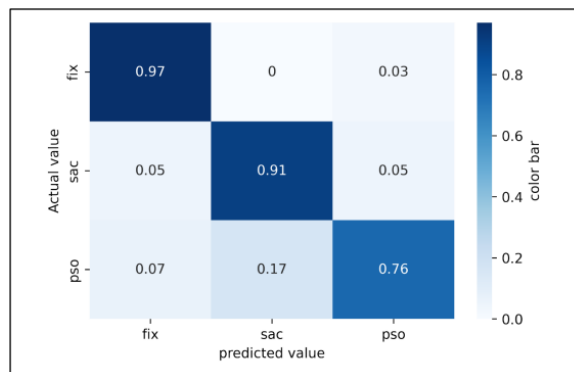


Рисунок 5 – Матриця помилок класифікатора

У таблиці 1 подано метрики роботи класифікатора Random Forest. Результати класифікації показують, що RF добре справляється з класифікацією фіксацій та саккад. У цьому ефективність виявлення ПСК далекою від ідеалу.

Метрики класифікатора

	Accuracy	Precision	Recall	F1-Score
Фіксація	97%	99%	97%	98%
Саккада	92%	87%	91%	89%
ПСК	76%	64%	76%	69%

• Convolutional neural network (CNN) – згортова нейронна мережа - добре підходить для аналізу закономірностей даних, тому їх можна використовувати для виявлення подій, пов'язаних з рухом очей.

Ця мережа кожної вибірки передбачає ймовірності приналежності до фіксації, саккади чи плавному переслідування з урахуванням послідовності вибірок погляду. Цей метод спрямований на усунення недоліків попередніх методів, які використовували форму та амплітуду сигналу для визначення чи класифікації руху очей, що може бути проблематично, наприклад, при плавному русі.

Щоб перевірити здатність мережі CNN класифікувати події, пов'язані з рухом очей, створено просту мережу, представлену на рисунку 6. Як вхідні дані мережа використовує безперервний потік двовимірних вибірок погляду.

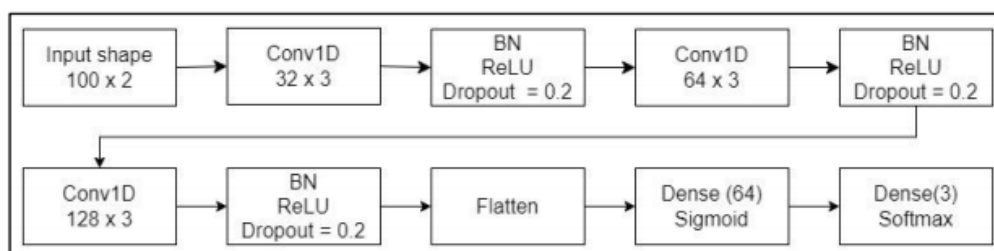


Рисунок 6 - Архітектура згорткової нейронної мережі

На рисунку 7 представлено матрицю помилок для класифікації CNN.

Фіксації правильно класифікуються у 99% випадків, саккади – у 88%, а ПСК – у 76% випадків. Загалом 4% саккад помилково класифікуються як фіксації, а 16% ПСК – як хибні фіксації. Крім того, 8% саккад класифікуються як ПСК.

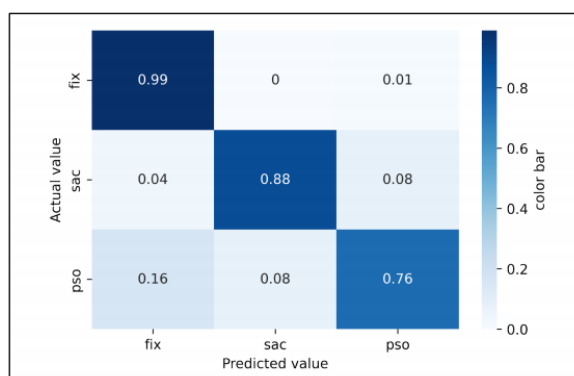


Рисунок 7 – Матриця помилок згорткової нейронної мережі

У таблиці 2 представлені метрики роботи згорткової нейронної мережі. Результати класифікації показують, що CNN добре справляється з фіксацією та класифікацією саккад. Проте ефективність класифікації для ПСК далека від ідеалу.

Таблиця 2

Метрики для згорткової нейронної мережі

	Accuracy	Precision	Recall	F1-Score
Фіксація	99%	98%	99%	99%
Саккада	89%	93%	89%	91%
ПСК	75%	83%	75%	79%

Аналіз результатів. Для порівняння роботи різних алгоритмів було проведено оцінку продуктивності чотирьох різних алгоритмів класифікації подій з областей порогових значень, машинного навчання та глибокого аналізу. Також було враховано спільну роботу двох оцінювачів-людей.

Результати показали, що I-DT працює краще, ніж I-VT, за всіма показниками продуктивності. Проте алгоритми RF і CNN перевершили алгоритми з урахуванням порогових значень (I-VT і I-DT) за всіма показниками продуктивності, крім повноти для саккади (recall метрика). У разі класифікаційних моделей RF і CNN суттєвої різниці у класифікації фіксацій та саккад не виявлено.

Результати для I-VT та RF представлені на рисунках 8 та 9 відповідно. Алгоритм I-VT виявив 189 фіксацій із середньою тривалістю 121 мс, тоді як алгоритм CNN виявив лише 64 фіксації із середньою тривалістю 121 мс. Середня тривалість становила 264 мс. Це добре видно на рисунках 8 та 9.

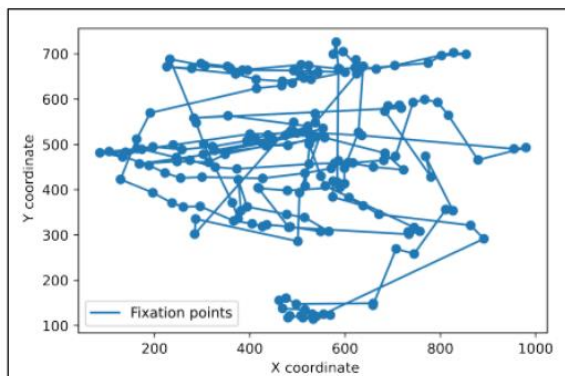


Рисунок 8 – Результат роботи алгоритму I-VT (при граничному значенні 3.5 px/ms)

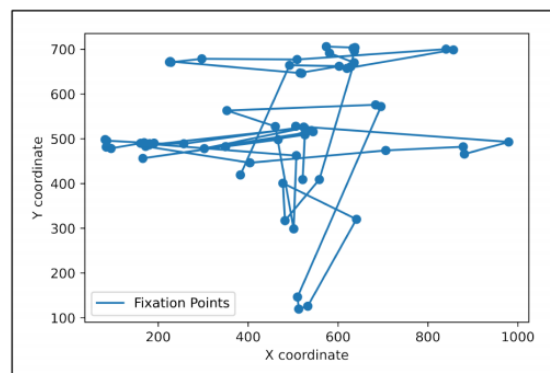


Рисунок 9 – Результат роботи алгоритму CNN

Основу для функціонування системи складатиме середовище розробки Unity, алгоритм класифікації подій, написаний мовою Python, а також API, яке пов'язуватиме середовище Unity та алгоритм для обробки даних.

Порівняльна таблиця

	I-VT	I-DT	RF	CNN
Fixation F1-score	94%	96%	99%	99%
Saccade F1-score	60%	66%	87%	91%
PSO F1-Score	–	–	64%	79%

Висновки. У цій роботі були вивчені алгоритми виявлення подій з різних областей: I-VT і I-DT, засновані на порогових значеннях, модель випадкового лісу в галузі машинного навчання і модель CNN в галузі глибокого навчання.

Була проведена оцінка ефективності їхньої класифікації з використанням одного і того ж набору даних для всіх методів. Також обговорювався взаємозв'язок між програмами та алгоритмами. Результати показали критичний вплив порогових значень класифікацію алгоритмів I-VT і I-DT. У зв'язку з цим пошук оптимального порогу алгоритмів, заснованих на порогових значеннях, є складним завданням. Алгоритми RF і CNN перевершують алгоритми, засновані на порогових значеннях, за всіма показниками продуктивності та здатні класифікувати дані щодо кількох класів.

ЛІТЕРАТУРА / REFERENCES

1. Komogortsev O.V., Gobert D.V., Jayarathna S., Koh D. H., Gowda S.M. // IEEE Transactions on Biomedical Engineering, 2010. - 2635-2645 pp.
2. Birawo B., Kasproski P. Review and Evaluation of Eye Movement Event Detection Algorithms. // Sensors, 2022. - 1-10 pp.
3. Startsev M., Zemblys R. Evaluating Eye Movement Event Detection: A Review of the State of the Art. // Behavior research methods, 2023. - 1653-1714 pp.
4. Andrew Duchowski. Eye Tracking Methodology Theory and Practice. // Springer Science & Business Media, 2007. - 55-57 pp.
5. Holmqvist K., Nyström M., Andersson R., Dewhurst R., Jarodzka H., Weijer J. Eye Tracking. A Comprehensive Guide to Methods and Measures. // OUP Oxford, 2011. - 45-47 pp.
6. Bojko A. Eye Tracking the User Experience. A Practical Guide to Research. // Rosenfeld Media, 2013. - 89-92 pp.
7. Wedel M., Pieters R. Eye Tracking for Visual Marketing // Now Publishers, 2008. - 101-106 pp.
8. Conklin K., Pellicer A., Carrol G. Eye-Tracking Guide for Applied Linguistics Research. // Cambridge University Press, 2018. - 143-145 pp.

Received 21.03.2025.
Accepted 24.03.2025.

Research on neural network models for eye tracking and object fixation

The paper presents a comparison of neural network models for eye tracking and object fixation.

In this paper, event detection algorithms from different domains were studied: I-VT and I-DT, based on threshold values, a random forest model in the field of machine learning, and a CNN model in the field of deep learning.

Their classification performance was evaluated using the same dataset for all methods. The relationship between programmers and algorithms was also discussed. The results showed the critical impact of threshold values on the classification of I-VT and I-DT algorithms. In this regard, finding the optimal threshold for threshold-based algorithms is a difficult task. RF and CNN algorithms outperform threshold-based algorithms in all performance indicators and are capable of classifying data into multiple classes.

Various ABCs (event detection algorithms) have been tested and compared. Theoretical and experimental justifications for the results obtained have been given. Based on the results of the experiments, the best ABC for our tasks (event detection algorithms) has been selected.

The means of implementation for creating the application are also described. The advantages and disadvantages of the selected solutions are described. The technical problems that have arisen have been solved. The architecture has been designed, and a program layout has been developed based on the required tasks.

Keywords: machine learning, random forest, event detection algorithms, application, saccade. smooth pursuit, post-saccadic oscillations, glide paths, convolutional neural network.

Островська Катерина Юріївна – доцент, кандидат технічних наук, доцент кафедри Інформаційних технологій і систем ННІ «ДМетІ» УДУНТ.

Порохнявий Вячеслав Григорович - магістр кафедри Інформаційних технологій і систем ННІ «ДМетІ» УДУНТ.

Ostrovskaya Kateryna – Associate Professor, Candidate of Technical Sciences, Associate Professor of the Department of Information Technologies and Systems of the Scientific Research Institute "DMetI" UDUNT.

Porokhnyavyy Vyacheslav Grigorovich - Master of the Department of Information Technologies and Systems of the Scientific Research Institute "DMetI" UDUNT.