

Abstract-Polish

Analiza ruchów gałek ocznych odgrywa kluczową rolę w badaniach nad ludzką uwagą wzrokową, percepcją oraz procesami poznawczymi. Dokładna detekcja podstawowych zdarzeń okulomotorycznych fiksacji, sakkad, płynnych ruchów podążających (smooth pursuits, SP) oraz oscylacji postsakkadowych (post-saccadic oscillations, PSO) jest niezbędna do rzetelnej interpretacji danych okulograficznych w takich dziedzinach jak psychologia, neuronauka czy interakcja człowiek–komputer. Detekcja zdarzeń pozostaje jednak zadaniem trudnym ze względu na dużą zmienność wzorców spojrzeń pomiędzy zadaniami, trudności w rozróżnianiu krótkotrwałych zdarzeń, takich jak PSO, oraz brak ustandaryzowanych protokołów ewaluacyjnych. Niniejsza rozprawa doktorska podejmuje te wyzwania poprzez kompleksową analizę metod detekcji zdarzeń okulomotorycznych, - od metod progowych, przez algorytmy uczenia maszynowego, po modele głębokiego uczenia, - ze szczególnym uwzględnieniem ich uogólnialności, doboru cech oraz metodologicznego znaczenia detekcji PSO.

Praca rozpoczyna się od systematycznej ewaluacji szerokiego spektrum algorytmów detekcji zdarzeń z wykorzystaniem danych z wysokorozdzielczych systemów śledzenia wzroku. Tradycyjne metody progowe, takie jak IVT i IDT, zostały porównane z podejściami opartymi na uczeniu maszynowym (Random Forest) oraz modelami głębokiego uczenia (CNN i LSTM). W przeciwieństwie do wcześniejszych badań porównawczych, które często opierały się na różnych zbiorach danych lub odmiennych kryteriach oceny, wszystkie algorytmy w niniejszej pracy były testowane w identycznych warunkach, z wykorzystaniem ręcznie anotowanych danych referencyjnych oraz spójnych miar jakości. Wyniki potwierdziły, że metody progowe są wydajne w binarnej klasyfikacji fiksacji i sakkad, jednak cechują się dużą wrażliwością na dobór parametrów i ograniczoną zdolnością do uogólniania. Z kolei metody uczenia maszynowego i głębokiego osiągnęły wyższą odporność oraz lepszą zgodność z anotacjami eksperckimi, przy czym sieci konwolucyjne uzyskały najwyższą dokładność ogólną. Wyniki te ustanawiają solidną podstawę metodologiczną dla przewagi podejść opartych na danych oraz podkreślają kluczowe znaczenie procedur ewaluacyjnych w szczególności strategii walidacji krzyżowej dla uzyskania wiarygodnych rezultatów.

Na tej podstawie zaproponowano hybrydową architekturę 2D-CNN-LSTM do jednoczesnej klasyfikacji fiksacji, sakkad, PSO oraz SP. Warstwy konwolucyjne pełnią rolę

automatycznych ekstraktorów cech, natomiast warstwy LSTM modelują zależności czasowe, co czyni architekturę szczególnie odpowiednią do analizy sekwencyjnych danych okulograficznych. W celu oceny wpływu reprezentacji wejściowej, systematycznie analizowano cztery cechy kinematyczne: prędkość, przyspieszenie, jerk oraz kierunek ruchu, łącząc je w różne zestawy cech. Analiza wykazała, że skuteczność klasyfikacji jest silnie zależna od doboru cech. Kombinacje obejmujące prędkość i kierunek wraz z przyspieszeniem lub jerkiem dawały najbardziej stabilne wyniki, natomiast pełny zestaw czterech cech nie prowadził do jednoznacznej poprawy dokładności. Szczególną uwagę poświęcono detekcji PSO, która pozostaje istotnym wyzwaniem. Chociaż PSO były często mylone z krótkimi sakkadami lub fiksacjami, uwzględnienie cechy jerk pozwoliło zwiększyć dokładność ich detekcji do poziomu 67%, przy znacznie wyższej skuteczności dla pozostałych klas zdarzeń. Wyniki te podkreślają trudność klasyfikacji PSO, ale jednocześnie wskazują na potencjał modeli hybrydowych w radzeniu sobie z tym problemem.

Kolejna część pracy koncentruje się na praktycznym znaczeniu detekcji PSO w badaniach nad czytaniem. Wyniki klasyfikacji opartej na sieciach CNN zostały porównane z rezultatami uzyskanymi przy użyciu komercyjnego oprogramowania, które nie uwzględnia PSO i klasyfikuje wyłącznie fiksacje oraz sakkady. Analiza wykazała, że pomijanie PSO prowadzi do zawyżenia czasów trwania fiksacji oraz modyfikuje powszechnie stosowane miary czytania, takie jak średni czas fiksacji, kluczowe dla oceny złożoności tekstu i obciążenia poznawczego. Uwzględnienie PSO pozwoliło uzyskać bardziej precyzyjne i zniuansowane miary zachowania wzrokowego, potwierdzając, że PSO nie są jedynie artefaktami, lecz istotnymi elementami dynamiki spojrzeń. Stanowi to ważny wkład metodologiczny, wskazujący, że ignorowanie PSO może prowadzić do systematycznych błędów interpretacyjnych w badaniach okulograficznych nad czytaniem i pokrewnymi dziedzinami.

W końcowej części rozprawy rozszerzono analizę na zagadnienie uogólniania modeli detekcji zdarzeń pomiędzy różnymi zadaniami wizualnymi. Dane okulograficzne analizowano w kontekstach takich jak czytanie, oglądanie statycznych obrazów, obserwacja materiałów wideo oraz śledzenie poruszającego się punktu. Analizy statystyczne potwierdziły, że zachowanie okulomotoryczne jest silnie zależne od rodzaju zadania: czytanie charakteryzuje się szybkimi fiksacjami i sakkadami, bodźce dynamiczne wywołują płynne ruchy podążające, a PSO wykazują zmienność zależną od kontekstu zadaniowego. Model CNN trenowany i testowany w obrębie jednego zadania osiągał wysoką dokładność, jednak jego skuteczność znacząco spadała w przypadku transferu między zadaniami, zwłaszcza w odniesieniu do PSO. Wyniki te uwidaczniają ograniczenia obecnych modeli poza domeną treningową oraz wskazują na potrzebę opracowania metod detekcji zdarzeń o większej zdolności uogólniania, dostosowanych do zróżnicowanych kontekstów wizualnych i poznawczych.

Podsumowując, niniejsza rozprawa wnosi następujący wkład: (i) rygorystyczną ewalu-

ację klasycznych, opartych na uczeniu maszynowym i głębokim algorytmów detekcji zdarzeń w ustandaryzowanych warunkach; (ii) wprowadzenie hybrydowego modelu CNN–LSTM wraz z systematyczną analizą doboru cech; (iii) empiryczne potwierdzenie kluczowego znaczenia detekcji PSO w badaniach nad czytaniem; oraz (iv) jedną z pierwszych kompleksowych analiz uogólniania modeli detekcji zdarzeń pomiędzy zadaniami. Łącznie uzyskane wyniki wzmacniają metodologiczne podstawy badań nad ruchami gałek ocznych i mają istotne implikacje dla projektowania adaptacyjnych, precyzyjnych i kontekstowo świadomych systemów analizy okulograficznej w szerokim spektrum zastosowań naukowych i praktycznych.