



UNIwersytet  
Warszawski

Wydział Geografii i Studiów Regionalnych  
Katedra Geomatyki i Systemów Informatycznych

Warszawa, 08.07.2024

### Recenzja rozprawy doktorskiej mgr. inż. Krzysztofa Dyby

pod tytułem:

#### **Interpretation of geospatial data using explainable machine learning methods**

przygotowanej pod kierunkiem:

dr. hab. Jarosława Jasiewicza, prof. ucz. i dr. hab. inż. Cezarego Kaźmierowskiego, prof. ucz.  
na Wydziale Nauk Geograficznych i Geologicznych UAM w Poznaniu

Formalną podstawą niniejszej recenzji jest decyzja Rady Naukowej Dyscypliny nauk o Ziemi i środowisku Uniwersytetu im. Adama Mickiewicza w Poznaniu (Uchwała nr 49-2023/2024 z dnia 30.04.2024) bazująca na Uchwale Senatu UAM nr 133/2020/2021 z dnia 28.06.2021 oraz Prawie o szkolnictwie wyższym i nauce z dnia 20.07.2018 (Dz.U. 2018 poz. 1668 z późniejszymi zmianami).

Rozprawa doktorska przygotowana jest w języku angielskim, liczy 84 strony tekstu maszynopisu bazującego na cyklu trzech artykułów opublikowanych w anglojęzycznych czasopismach: *Remote Sensing* (IF<sub>2022</sub> = 5,0; MEiN<sub>2022</sub> = 100 pkt), *Geomorphology* (IF<sub>2022</sub> = 3,9; MEiN<sub>2022</sub> = 100 pkt), *Scientific Reports* (IF<sub>2023</sub> = 3,8; MEiN<sub>2023</sub> = 140 pkt). Skumulowany Impact Factor wyniósł 12,7, sumaryczna liczba punktów MEiN 340, a także 9 cytowań na Scopus oraz 8 na Web of Science na dzień 05.07.2024.

Układ pracy jest typowy dla rozpraw doktorskich i obejmuje następujące części:

- strona tytułowa, podziękowania (1/3 strony) oraz spis treści rozprawy doktorskiej.
- Rozdział 1: *Introduction*; zajmuje 6 stron, na których Doktorant przedstawił problem znaczącego wzrostu regularnie pozyskiwanych dużych zbiorów danych geoprzestrzennych, które opisują złożone zależności pomiędzy komponentami środowiska, a wiedza ekspercka okazuje się niewystarczająca do opisanie obserwowanych złożonych i nieliniowych problemów występujących w naukach przyrodniczych. Stąd też wykorzystuje się nieparametryczne metody przetwarzania regularnie aktualizowanych zbiorów danych (big data computing), integrując je z lokalnymi pomiarami weryfikacyjnymi. Do tego celu wykorzystuje się różne grupy algorytmów, część z nich może być modyfikowana (otwarte kody), natomiast część funkcjonuje jako „czarne skrzynki”, których procedury przetwarzania danych są zamknięte. Doktorant słusznie zauważył lukę i problem badawczy polegający na tym, że część algorytmów jest tworzona przez programistów, którzy nie mają podstawowej wiedzy przyrodniczej, w związku z tym daje się zaobserwować brak właściwych interpretacji wpływu zmiennych objaśniających (np. poszczególne zakresy widma



elektromagnetycznego, czy też właściwości przyrodniczych) na wyniki modelowania danych przestrzennych za pomocą uczenia maszynowego, czyli deterministycznego przypisania wielowymiarowych cech obiektów do zachodzących procesów środowiskowych w celu właściwego uchwycenia determinant. Pozwoliłoby to ocenić zgodność wiedzy eksperckiej z przetwarzaniem dużych zestawów wieloczasowych danych w modelowaniu środowiska. Hipotezą badawczą jest stwierdzenie, że interpretowalne uczenie maszynowe może być narzędziem wspierającym proces interpretacji danych przestrzennych (w oryginale: interpretable machine learning can be a tool to support the process of interpreting spatial data). Osobiście zgadzam się z tym kierunkiem myślenia, uważam że jest to właściwy kierunek badań naukowych, ale hipoteza jest sformułowana w sposób zbyt ogólny, bo już w przytoczonym przeglądzie literatury można znaleźć przypadki potwierdzające, że są takie rozwiązania. Niemniej intencja jest jak najbardziej słuszna i w pełni uzasadniona, jest to tym bardziej cenne, iż mgr inż. Krzysztof Dyba posiada zarówno właściwą wiedzę przyrodniczą oraz algorytmiczną. W podrozdziale *1.1 Objective* Doktorant zaprezentował cel rozprawy doktorskiej jakim była ocena zgodności metody uczenia maszynowego i pozyskania nowych informacji na temat czynników determinujących uzyskany wynik, czyli czy uczenie maszynowe może być narzędziem wspierającym proces modelowania danych geoprzestrzennych, gdzie miarą sukcesu nie jest tylko wytrenowanie efektywnego modelu, ale także możliwość uzyskania dodatkowej wiedzy na temat czynników prowadzących do tego wyniku. Zatem Doktorant nie koncentrował się wyłącznie na uzyskaniu optymalnego wyniku poprzez optymalizację architektury przetwarzania danych, ale na ich implementacji w celu powiązania z nowymi jakościowo informacjami o zachodzących zmianach w środowisku. Cel ten został osiągnięty poprzez trzy eksperymenty badawcze: analizę regresji, klasyfikację nie-, nadzorowaną, a także analizę metody uczenia maszynowego do interpretacji modelu globalnego. Wyniki tych prac zostały opublikowane w postaci trzech publikacji stanowiących monotematyczny cykl naukowy tworzący podstawę niniejszej rozprawy:

- Dyba K., Ermida S., Ptak M., Piekarczyk J., Sojka M. (2022). Evaluation of methods for estimating lake surface water temperature using Landsat 8. *Remote Sensing*, 14(15), 3839. <https://doi.org/10.3390/rs14153839>.
- Dyba K., Jasiewicz J. (2022). Toward geomorphometry of plains - Country-level unsupervised classification of low-relief areas (Poland). *Geomorphology*, 413, 108373. <https://doi.org/10.1016/j.geomorph.2022.108373>.
- Dyba K. (2024). Explanation of the influence of geomorphometric variables on the landform classification based on selected areas in Poland. *Scientific Reports*, 14(1), 5447. <https://doi.org/10.1038/s41598-024-56066-6>.
- Rozdział 2: *Materials and methods* obejmuje 3 strony. W tej części Autor przedstawił najważniejsze koncepcje badawcze z zakresu uczenia maszynowego, prezentując szeroką gamę modeli bazujących na różnych algorytmach zaprojektowanych do rozwiązywania konkretnych problemów, Doktorant wskazał konkretne przykłady zastosowań klasyfikatorów w geomorfologii. Jest to ciekawy rozdział, ale mógłby być bardziej rozbudowany, gdyż koncepcja jest właściwa, a zawartość zbyt ogólna jak dla osób chcących porównać mocne i słabe strony różnych rozwiązań metodycznych.



- Rozdział 3: *Summary of articles* składa się z 3 podrozdziałów, każdy z nich zajmuje około 1 1/3 strony i charakteryzuje kolejne publikacje:
  - 3.1. *Regression analysis* koncentruje się na artykule: Dyba K., Ermida S., Ptak M., Piekarczyk J., Sojka M. (2022). Evaluation of methods for estimating lake surface water temperature using Landsat 8. *Remote Sensing*, 14(15), 3839, DOI: 10.3390/rs14153839. Autorzy porównali cztery scenariusze szacowania temperatury powierzchni wody na podstawie termalnych danych satelitarnych LST (Land Surface Temperature Model) i LST-L2 (Landsat Level-2 Surface Temperature Science Product) bazując na wielokrotnej regresji liniowej (LM) i lasach losowych (RF). Najlepsze wyniki uzyskane zostały dla modelu lasów losowych (stwierdzono błędy systematyczne w modelach atmosfery w strefie przybrzeżnej). Natomiast analiza 9 zakresów spektralnych Landsat i teledetekcyjnych wskaźników: NDVI (Normalized Difference Vegetation Index) i NDWI (Normalized Difference Water Index) potwierdziła, że kanał B10 (10,6 – 11,19  $\mu\text{m}$ ) powoduje mniejszy błąd prognozowania temperatury wody powierzchniowej niż zakres B11 (11,5 – 12,51  $\mu\text{m}$ ), a pozostałe zakresy nie okazały się istotne (nawet w zakresie SWIR). Autorzy potwierdzili, że dominujący wpływ ma górna warstwa wody, która silnie absorbuje promieniowanie, stąd objętość i średnia głębokość jezior nie wpływają na rejestrowany obraz. Ciekawe spostrzeżenia dotyczą efektywności poszczególnych modeli, gdyż RF (las losowy) zapewnił większą przydatność w modelowaniu, ale interpretacja wpływu zmiennych jest ograniczona, podczas gdy model liniowy precyzyjniej określił zależność pomiędzy zmienną zależną, a zmiennymi objaśniającymi, także dlatego, że w modelach liniowych analizy bazują na prostych relacjach. Jest to niewątpliwa zaleta tego typu modelowania, ale trzeba pamiętać, że zjawiska przyrodnicze najczęściej cechują się nieparametrycznymi zależnościami.
  - 3.2. *Supervised classification* obejmuje najważniejsze elementy publikacji: Dyba K. (2024). Explanation of the influence of geomorphometric variables on the landform classification based on selected areas in Poland. *Scientific Reports*, 14(1), 5447, DOI: 10.1038/s41598-024-56066-6. Publikacja ta koncentruje się na porównaniu trzech algorytmów opartych na uczeniu maszynowym: lasy losowe (RF), wzmacnianie gradientu (XGBoost) oraz konwolucyjne sieci neuronowe (U-Net) do kartowania form geomorfologicznych. Najlepszy wynik mgr Krzysztof Dyba uzyskał dla modelu XGBoost, natomiast najniższy dla modelu U-Net. Doktorant wskazał, że główną przyczyną różnic w działaniu tych modeli był stosunkowo mały zbiór danych wejściowych i nie zrównoważona reprezentatywność form terenu, gdyż głębokie uczenie wymaga znacznie większej ilości danych do dostrojenia parametrów. Temu celowi służyły: entropia, zagregowana mediana wysokości w komórce 1000 i 500 m, wysokość bezwzględna, odchylenie standardowe wysokości, wypukłość powierzchni, otwartość topograficzna, nachylenie i wieloskalowy wskaźnik położenia topograficznego, wskaźnik zbieżności, faktura powierzchni, krzywizny styczne i profilowe oraz wielokierunkowe zacienienie zboczy. W drugim etapie analizy



Doktorant przeanalizował nieliniowy wpływ poszczególnych zmiennych morfometrycznych na prawdopodobieństwo klasyfikacji jednostek geomorfologicznych poprzez zastosowanie skumulowanych efektów lokalnych; analizę przeprowadził dla wszystkich form terenu, ponadto przeanalizował jak powierzchnia jednostek geomorfologicznych wpływa na prawdopodobieństwo prawidłowej klasyfikacji. W wyniku analizy uzyskał wartość progową wynoszącą 45 km<sup>2</sup>, co oznacza, że powyżej tego progu wielkość wpływu zmiennych morfometrycznych stała się znacząca. Interesującym wynikiem analizy jest obserwacja, że wpływ cech morfometrycznych na klasyfikację zależy od strefy morfogenetycznej (zależnie od obszar badań zmienne morfometryczne są przydatne do identyfikacji występujących form terenu). Analizy te pozwoliły Autorowi ocenić wpływ zmiennych objaśniających na decyzje klasyfikacyjne modelu, przekładając się na interpretację funkcjonowania modelu czarnej skrzynki.

- 3.3. *Unsupervised classification* koncentruje się na artykule: Dyba K., Jasiewicz J. (2022). Toward geomorphometry of plains – Country-level unsupervised classification of low-relief areas (Poland). *Geomorphology*, 413, 108373, DOI: 10.1016/j.geomorph.2022.108373. Autorzy zaprezentowali procedurę klasyfikacji nienadzorowanej, która nie wymaga wzorców treningowych, dlatego przedmiotem artykułu była interpretacja klastrów utworzonych podczas kolejnych iteracji procesu klasyfikacji, a weryfikacja wyniku polegała na dopasowaniu uzyskanych wzorców przestrzennych do podziału fizyczno-geograficznego Polski. Autorzy przetestowali różną liczbę klastrów pikseli (4–24) porównując je z zasięgiem stref morfogenetycznych oraz układem przestrzennym zmiennych geomorfometrycznych; bayesowskie kryterium informacyjne potwierdziło, że optymalnym wyborem było 20 klastrów, co stanowiło konsensus, gdyż według Autorów większa liczba skupień budziła wątpliwości co do jednorodności powstałych poligonów, natomiast mniejsza liczba nie ujawniała ważnych form ukształtowania terenu, takich jak dna dolin rzecznych. Koncentrując się procedurze wyodrębniania klastrów pikseli, Autorzy zauważyli, że analiza składowych głównych (pierwsze trzy kanały PCA) odpowiadały za wyjaśnienie prawie 70% zmienności danych, w tym PCA1 generowała 39,7% wynik. Ponadto były one ujemnie powiązane z wysokością bezwzględną, szorstkością i wzniesieniem nad podstawą erozyjną, natomiast dodatnio z płaskością terenu. Drugi główny składnik (PCA2) odpowiadał za 15,9% zmienności danych i był negatywnie powiązany ze zmiennymi reprezentującymi właściwość tekstury, w tym średnią zbieżnością, szum powierzchniowy, lokalnego położenia topograficznego i położeniem zbocza. Trzecia składowa główna (PCA3) stanowiła 14,1% wariacji i była dodatnio powiązana z właściwościami teksturalnymi (zbieżność średnia, szum powierzchniowy) i ujemnie z lokalnym położeniem topograficznym. Dodatkowo do interpretacji wykorzystano cztery zmienne geomorfometryczne, tj. średnią zbieżność, płaskość, położenie zbocza oraz rzeźbę terenu, które miały największy potencjał do różnicowania form terenu na obszarach nizinnych.



# UNIwersytet Warszawski

Autorzy opisali 15 jednostek nizinnych w 4 podgrupach i 5 jednostek wyżynnych. Uzyskane wyniki wykazały silną zbieżność z podziałem fizyczno-geograficznym w skali mezoregionów. Jest to cenne opracowanie, gdyż pozwala weryfikować i aktualizować opracowane mapy typów powierzchni terenu stanowiąc podstawę do weryfikacji istniejących opracowań geomorfologicznych, znajdując szereg praktycznych zastosowań w planowaniu przestrzennym i budownictwie.

- Rozdział 4: *Discussion* obejmuje 1,5 strony na których Doktorant przedstawił aktualne trendy rozwoju uczenia maszynowego w tym algorytmów działających na zasadzie białych i czarnych skrzynek, niemniej Autor słusznie wskazał, że kluczem do sukcesu procedury optymalizacji jest konieczność znajomości procedury przetwarzania danych i świadoma integracja procesów oraz uwarunkowań przyrodniczych z doбором danych i ich analizą. W kolejnych akapitach rozdziału Doktorant podsumował najważniejsze osiągnięcia w poszczególnych artykułach cytując kilka publikacji. Rozdział ten treścią przypomina podsumowanie badań, a nie Dyskusję wyników, w której najważniejsze osiągnięcia powinny być kolejno odniesione do wyników innych badaczy w celu wskazania innowacyjnych, autorskich rozwiązań z racji na świadomy dobór algorytmów, a także przewagi w jakości uzyskanych wyników przeprowadzonych analiz. Oczywiście w poszczególnych artykułach jest to właściwie zaprezentowane.
- Rozdział 5: *Conclusion* zajmuje prawie stronę tekstu maszynopisu, który potwierdza najważniejsze osiągnięcia Doktoranta, czyli analizę danych wejściowych i znajomość procesów przyrodniczych pozwalających na interpretację właściwie dobranych modeli, szczególnie czarnoskrzynkowych, np. stosowanych w klasyfikacji i analizie regresji. Ma to istotne znaczenie w optymalizacji architektury modeli i predykcji uzyskanych wyników. Procedura ta zwiększa przejrzystość i dostępność procesu modelowania. Doktorant zauważył, że pomimo znaczących korzyści płynących z wyjaśnienia procesu uczenia maszynowego, należy uwzględnić szereg ograniczeń, które dotyczą danych wejściowych (błędy w danych, niewystarczająca reprezentatywność lub niewłaściwy dobór zmiennych objaśniających), gdyż generują błędy zarówno w wynikach, jak i interpretacji uzyskanych danych. Innym źródłem problemów może być sam model, czyli zastosowanie modelu nieodpowiedniego do danego zadania, skutkując niską dokładnością predykcji i małą wrażliwością na cechy. Doktorant wskazał, że niewystarczająca wydajność modelu często prowadzi do błędnych interpretacji, dlatego przed wyciągnięciem wniosków należy dokładnie rozważyć jakość danych wejściowych i dokładność modelu. W efekcie mgr K. Dyba rekomenduje szersze wykorzystanie metod objaśniających w badaniach wykorzystujących uczenie maszynowe oraz dalszy rozwój metod wyjaśniania decyzji modelowych z uwzględnieniem kontekstu przestrzennego.
- 6. *Bibliography* obejmuje 67 pozycji zacytowanej literatury w rozprawie doktorskiej (nie licząc literatury z poszczególnych artykułów), z czego 30 pozycji literatury opublikowanych zostało w ciągu ostatnich 5 lat (2025-2019), a 37 jest starszych niż 5 lat (opublikowane przed 2019 roku).
- 7. *Attachments* rozdział ten prezentuje dokonania Doktoranta: 7.1. *Academic achievements* (other publications, conferences, grants, teaching, other achievements).



# UNIwersytet Warszawski

Zaprezentowane osiągnięcia potwierdzają, że Doktorant jest aktywnym młodym badaczem działającym na wszystkich polach zdobywania wiedzy, publikowania uzyskanych wyników, a także promocji osiągnięć. 7.2. *Copies of articles* podrozdział zawiera kopie trzech artykułów, które stanowią podstawę rozprawy doktorskiej.

Do rozprawy doktorskiej nie zostały dołączone oświadczenia autorów, w których współautorzy dwóch publikacji określiliby zaangażowanie Doktoranta i pozostałych współautorów w powstanie poszczególnych publikacji. Znacząco utrudnia to ocenę wkładu Doktoranta w publikację 1 i 3 (w drugiej publikacji jest mgr K. Dyba samodzielnym autorem). Niemniej, we wszystkich przypadkach Doktorant jest pierwszym autorem, co potwierdza, że odegrał znaczącą rolę w przygotowaniu i opublikowaniu artykułów (w publikacji nr 3 autorem korespondencyjnym jest Promotor rozprawy).

## Ocena merytoryczna rozprawy doktorskiej

Doktorant podjął się ambitnego zadania jakim było powiązanie procedury przetwarzania danych z procesami zachodzącymi w przyrodzie, trudność ta wynika z konieczności posiadania rozległej wiedzy zarówno z nauk przyrodniczych, jak i algorytmiki, geo- i informatyki. Niewątpliwie jest to cenne podejście, które wskazało nowe spostrzeżenia ilościowej i jakościowej analizy zachodzących procesów przyrodniczych. Założone cele badawcze zostały zrealizowane, gdyż Doktorant bazując na aktualnym przeglądzie literatury zaimplementował najnowsze dane teledetekcyjne oraz algorytmy uczenia maszynowego do wygenerowania map prezentujących komponenty środowiska przyrodniczego analizując:

- skuteczność czterech scenariuszy szacowania temperatury wód powierzchniowych jezior w oparciu o dane satelitarne. Analiza porównawcza wykazała najlepsze wyniki metody RF oraz danych Landsat Level-2 Surface Temperature Science Product (LST-L2; najwyższy współczynnik korelacji i najniższy RMSE). Ponadto Doktorant określił warunki brzegowe stosowania poszczególnych metod (np. przy dużej zawartości pary wodnej w atmosferze), a także konieczność stosowania danych empirycznych w doskonaleniu wyników predykcji. Cennym osiągnięciem są precyzyjne rekomendacje co do dalszych kierunków modelowania temperatury wód powierzchniowych, gdyż ma to szczególne znaczenie dla oceny zawartości nutrientów, tlenu i mikrofitów w wodach, a tym samym stabilności ekosystemu wodnego;
- potencjał modeli uczenia maszynowego (w tym splotowych sieci neuronowych) do automatycznego kartowania geomorfologicznego oraz przydatność poszczególnych zmiennych geomorfometrycznych na wyniki klasyfikacji ukształtowania terenu. Na podstawie wyników tego eksperymentu Doktorant ocenił skuteczność wyniku klasyfikacji, w tym ekstrapolacji procesu na nowe obszary, cechujące się odmienną charakterystyką występujących form terenu stwierdzając, że metody automatyczne nie mogą zastąpić podejścia tradycyjnego, mogą jednak wesprzeć kartowanie, jeśli cechy geomorfologiczne na nowym obszarze są podobne do tych w zbiorze uczącym. To stwierdzenie nie jest zaskakujące, bo wskazane klasyfikatory wymagają wzorców treningowych, stąd moje pytanie do Doktoranta, jakie byłyby efekty wykorzystania



danych 3D, np. pochodnych DEM jako składowe danych wejściowych? Czy Doktorant próbował stosować algorytmy o rozmytej logice, np. fuzji ArtMap, które pozwalają wygenerować nowe klasy, które nie zostały uwzględnione w we wzorach treningowych? Pewnym rozwiązaniem może być zastosowanie algorytmów nienadzorowanych, czyli czy próbował zastosować algorytmy z trzeciego artykułu na obszarze badań przedstawionym w drugim artykule? Jakie były efekty tych klasyfikacji? Jak wyodrębnione poligony przedstawiały się na tle przygotowanych wzorców treningowych przygotowanych do klasyfikacji nadzorowanej?

- koncepcję nienadzorowanej klasyfikacji typów powierzchni na obszarach o niewielkich deniwelacjach terenu bazując na analizie morfometrycznej na podstawie różnych procesów morfogenetycznych. Autorzy zaproponowali nowe zmienne opisujące wtórne cechy powierzchni: Mean Convergence (MCON) i Surface Noise (SNIS), które odegrały ważną rolę w rozróżnianiu równin pochodzenia lodowcowego, peryglacialnego lub rzeczno-glacialnego. Jako algorytm tworzenia klastrów Autorzy wykorzystali Gaussian Mixture Model, który zapewnia elastyczne tworzenie klastrów bazując na ocenie niepewności przypisanej do danego klastra. Uzyskane klastry pozwoliły zidentyfikować cztery typy równin polodowcowych Niżu Europejskiego oraz typy powierzchni terenu częściowo odpowiadających zasięgowi głównych stref morfogenetycznych.

Każda z publikacji stanowi koncepcyjnie cenne opracowanie, widać wyraźną kontynuację badań zainicjowanych w poprzedniej publikacji, ale nierozstrzygniętą kwestią pozostaje problem, czy klasyfikacja nienadzorowana (publikacja nr 3) pozwoliłaby osiągnąć podobne wyniki dokładności do efektów klasyfikacji nadzorowanej (publikacja nr 2)? Czy zaobserwowane dokładności w klasyfikacji 3 pokryłyby się z dokładnościami uzyskanymi metodą nienadzorowaną na danych wykorzystanych w publikacji nr 2?, czy także potwierdziłoby się, że optymalna liczba klastrów wyniosłaby około 20? Tego typu analiza pozwoliłaby potwierdzić czy zależność pomiędzy optymalną liczbą klastrów bardziej zależy od ilości i jakości danych wejściowych, czy też od zróżnicowania analizowanej powierzchni?

Osobiście bardzo pozytywnie oceniam niniejszą rozprawę doktorską, ponieważ Doktorant skoncentrował się na trudnym obiekcie badań, gdyż zasięg form geomorfologicznych jest na bieżąco modyfikowany przez działalność człowieka, np. poprzez zmianę pokrycia terenu i prowadzenie prac inżyniersko-budowlanych, a Doktorant podjął się metodycznej oceny powszechnie wykorzystywanych i skutecznych algorytmów badawczych uzyskując cenne wyniki. Efekty badań mgr. Krzysztofa Dyby nie zostały jeszcze wyczerpane, gdyż dostęp do nowych, wysokorozdzielczych danych pozwala modyfikować, ulepszać i dostosowywać poszczególne modele do konkretnych wyzwań społeczno-gospodarczych, np. podczas zmieniającego się klimatu, modelowania powodzi, czy wzrostu temperatur powietrza, spadku rozpuszczalności tlenu w wodzie i zmniejszonego dopływu wód rzecznych do jezior, istnieje realne zagrożenie rozwoju mikrofitów (w tym powszechnie znane zagrożenia Odry, czy strefy przybrzeżnej Bałtyku), śnięcia ryb słodkowodnych i zamierania bezkręgowców. Rozprawa doktorska bazuje na dużej ilości aktualnych danych, które w należyty sposób zostały pozyskane, zweryfikowane i przetworzone przez szereg optymalnych algorytmów oferując cenne wyniki ilościowe i jakościowe, które potwierdziły, że przyjęta metodyka badawcza jest



## UNIwersytet Warszawski

właściwa. Praca doktorska jest bardzo dobrze osadzona w literaturze przedmiotu, Doktorant cytuje 287 pozycji aktualnej literatury, w sposób swobody potwierdza, że posiada bardzo dobry warsztat badawczy, nie tylko potrafi zaproponować interesującą procedurę badawczą, ale ją należycie zrealizować, umieszczając uzyskane wyniki we właściwym nurcie literatury przedmiotu. Mgr Krzysztof Dyba opracował i skutecznie przetestował integrację różnych modeli przetwarzania danych, a uzyskane wyniki opublikował w prestiżowych, międzynarodowych czasopismach z zakresu poszczególnych modułów pracy. Jest to o tyle cenne, że profesjonalne zespoły redakcyjne wraz z recenzentami umożliwiły dopracowanie koncepcji badawczych do międzynarodowych standardów oferując wysoką jakość publikacji. Co będzie skutkować zainteresowaniem wśród czytelników zarówno zajmujących się środowiskiem przyrodniczym, jak i przetwarzaniem danych (algorytmika, informatyka).

Silną stroną rozprawy doktorskiej jest integracja najnowszych metod rozwijanych przez geomorfologów, hydrologów i programistów, pozwala to połączyć najważniejsze wyzwania poszczególnych dyscyplin w ujęciu interdyscyplinarnym. Sprawia to, że niniejsza rozprawa doktorska wpisuje się w kanon aktualnie rozwijanych zagadnień naukowych, a pomimo że składa się z cyklu artykułów opublikowanych w czasopismach o różnym profilu zainteresowań przedmiotowych, to wyraźnie zarysowuje ciągłość koncepcji metodycznej, ale także dostarcza aktualnych rozwiązań technologicznych dla geomorfologów, hydrologów, przyrodników. Metody uczenia maszynowego, obrazy satelitarne i dane 3D potwierdzają swój ogromny potencjał nie tylko w kartowaniu pokrycia terenu, ale też form geomorfologicznych, co pozwala w większym stopniu automatyzować procedury weryfikacji i uszczegółowienia istniejących map. Prowadząc analizy geoinformatyczne Doktorant miał do dyspozycji duże zbiory danych pozwoliło to zachować odpowiednio liczne próbki przetwarzanych danych, a z drugiej strony implementacja różnych modeli wymagała odpowiedniej procedury przetwarzania danych, by usystematyzować i zobiektywizować analizę poszczególnych etapów i weryfikować uzyskane wyniki. Uzyskane mapy są cenne w monitoringu środowiska i rozwoju społeczno-gospodarczym Polski. Jest to ważny aspekt pracy, gdyż opracowane narzędzia monitoringu pozwalają w sposób powtarzalny i porównywalny dla różnych obszarów analizować zachodzące zmiany, a tereny polodowcowe należą do ważnych wskaźników globalnych zmian klimatu (np. bezodpływowe zlewnie). Powyższe elementy są dobrze udokumentowane źródłami literaturowymi cytowanymi nie tylko w rozprawie, ale także w publikacjach stanowiących główny trzon analizowanej pracy doktorskiej.

Doktorant wykazał się bardzo dobrą znajomością: a) tła teoretycznego przygotowanej rozprawy, ponieważ w publikacjach oraz w dysertacji zacytował 287 pozycji literaturowych (zdecydowana większość to aktualne artykuły naukowe, a część to odnośniki do użytych pakietów oprogramowania lub algorytmów); b) nowoczesnych metod geoinformatycznych i geomorfologicznych; c) alternatywnych rozwiązań technologicznych stosując różne pakiety oprogramowania. Metodycznie są to cenne elementy, a uzyskane wyniki potwierdziły trafność proponowanych rozwiązań. Mocną stroną opracowania jest fuzja danych rastrowych i wektorowych kwantyfikujących środowisko przyrodnicze.

Z punktu widzenia recenzenta pewnym wyzwaniem jest ocena wkładu autorskiego mgr Krzysztofa Dyby w publikacje współautorskie, tym bardziej, że nie dostarczone zostały oświadczenia poszczególnych współautorów, ale na uwagę zwraca fakt, iż Doktorant jest





pierwszym autorem wszystkich publikacji, a w ostatniej publikacji jest współautorem ze swoim promotorem, co potwierdza, że zachowana została tradycja rozwoju młodego badacza pod okiem doświadczonego pracownika nauki.

Pracę czyta się dobrze, jest ona zrozumiała i ciekawa merytorycznie. Pojedyncze błędy stylistyczne, językowe czy też edycyjne nie wpływają na jakość pracy.

Rozwinięciem zaprezentowanych badań powinna być integracja metod nadzorowanych i nienadzorowanych na identycznym zestawie danych referencyjnych, by uchwycić czy i w jaki sposób można modyfikować liczbę klastrów oraz iteracji procesu klasyfikacji, by uzyskać możliwie dokładną mapę form geomorfologicznych/morfometrycznych.

W recenzowanej pracy nie znajdują słabych punktów, natomiast kilka rzeczy powinno być przedmiotem dyskusji podczas publicznej obrony:

- przydatność danych 3D (pochodne DEM) w modelowaniu morfometrycznym oraz geomorfologicznym;
- jaka jest zgodność klasyfikatorów nie-, nadzorowanych w klasyfikacji form geomorfologicznych? Jakie dane odgrywają kluczową rolę w prawidłowej klasyfikacji?
- czy wyniki klasyfikacji nienadzorowanej pokrywają się z wynikami klasyfikacji nadzorowanej? Jaka jest dokładność obu rodzajów klasyfikatorów?

Uwagi do rozprawy doktorskiej:

- stosunkowo często powtarza się sformułowanie „can be” co oznacza, że Doktorant widzi pewien potencjał, ale to sformułowanie pozostaje na bardzo ogólnym poziomie dokładności, a powinno być bardziej zdefiniowane, czyli bazując na przeglądzie literatury Doktorant powinien precyzyjnie skoncentrować się na konkretnych rozwiązaniach/wyzwaniach, szczególnie, wyrażając to w celach badawczych (1.1 Objective) prezentując bardziej konkretne elementy badawcze, które stały się przedmiotem pracy doktorskiej;
- Doktorant stosuje tzw. citation trains, czyli długie ciągi cytowań bez podania szczegółów/znaczenia/wkładu merytorycznego danych publikacji w prezentowanym wywodzie, np. „*significant number of publications using machine learning to solve problems in remote sensing (Chang & Bai, 2018; Hänsch et al., 2018; Maxwell et al., 2018), soil science (Heung et al., 2016; Heuvelink et al., 2021; Padarian et al., 2020; Wadoux et al., 2020), archaeology (Bickler, 2021; Cacciari & Pocobelli, 2022), or ecology (Crisci et al., 2012; Lorena et al., 2011; Rubbens et al., 2023; Shan et al., 2006; Stupariu et al., 2022)*”. Nie jest jasne dlaczego Doktorant wybrał te publikacje? czy posiadają jakąś unikatową metodykę, czy też uzyskane wyniki są cenne? Wystarczyło opisać kluczowe osiągnięcia prezentowanych publikacji.
- Dyskusja rozprawy doktorskiej jest zbyt ogólna, Doktorant powinien zdecydowanie dokładniej odnieść własne wyniki badań do osiągnięć innych autorów, by wykazać innowacyjność zastosowanych metod, a także uzyskanych wyników. W Dyskusji uzyskanych wyników Doktorant cytuje aktualne publikacje, tj. z ostatnich pięciu lat (2020–2024) co jest cenne, potwierdza, że dobrze zna aktualne osiągnięcia nauki, ale powinno być zdecydowanie więcej bezpośrednich odniesień do publikacji z lat 2023-



2024 w celu podkreślenia innowacyjności zaproponowanych metod, a także uzyskanych wyników, w stosunku do innych autorów, którzy stosują podobne metody lub też analizują podobne problemy naukowe. Jest to dobrze widoczne w rekomendacji publikacji nr 3, czyli Doktorant sugeruje rozwój koncepcji zaproponowanych ponad 10 lat temu (“Future work will include the implementation of recognized principles in cartographic-oriented research using object-oriented analyses (Drągut and Blaschke, 2006) and computer vision approach (Jasiewicz and Stepinski, 2013)”), ma to o tyle istotne znaczenie, że obecnie powszechnym elementem jest wykorzystanie wysokorozdzielczych modeli terenu, co jest kluczem w analizach morfometrycznych.

- Doktorant zaproponował jeden podrozdział: *1.1 Objective*, co nie powinno mieć miejsca, lepiej było zintegrować tę część w całość lub zaakcentować inne ważne elementy Wprowadzenia, by uporządkować całość tekstu w różnych podrozdziałach, np. założeniach pracy. Zabrakło mi bardziej rozbudowanego opisu problemu badawczego oraz hipotezy pracy badawczej, w ten sposób Doktorant wskazałby precyzyjnie kluczowe bariery w nauce oraz koncepcję rozwiązania tych problemów.
- Brakuje szczegółowego podsumowania, w którym Doktorant odniósłby się precyzyjnie do założeń rozprawy doktorskiej, czyli precyzyjnie powinny zostać przedstawione odpowiedzi na problemy badawcze, cele oraz hipotezy badawcze (co było zaplanowane, a co i jak zrealizowane).
- W spisie literatury przy prezentacji cytowanych książek brakuje szczegółów danych bibliograficznych, np. Anderson, J. A. (1995). *An introduction to neural networks*. MIT press.; Biecek, P., & Burzykowski, T. (2021). *Explanatory model analysis: Explore, explain, and examine predictive models* (1st ed.). CRC Press.; Menard, S. W. (2008). *Applied logistic regression analysis* (2. ed., [Nachdr.]). Sage Publ.; Molnar, C. (2022). *Interpretable machine learning: A guide for making black box models explainable* (Second edition).

### **Wniosek wraz z uzasadnieniem co do spełnienia przez recenzowaną rozprawę doktorską warunków określonych w przepisach prawa**

Osobiście bardzo pozytywnie oceniam rozprawę mgr. inż. Krzysztofa Dyby, pt.: *Interpretation of geospatial data using explainable machine learning methods*, ponieważ Doktorant w sposób metodyczny łączy różne dane teledetekcyjne z różnymi algorytmami uczenia maszynowego w celu optymalizacji architektury i parametrów uczenia maszynowego. Pokazuje to biegłość w posługiwaniu się narzędziami, a wszystko służy analizom przyrodniczym w celu uchwycenia prawidłowości zmian środowiska i odzwierciedleniu tych zmian odpowiednio dostosowanym aparatem informatycznym. Opracowana metoda bazuje na dużej liczbie danych wejściowych pozwalając na satysfakcjonujące wyodrębnienie poszczególnych form geomorfologicznych i morfometrycznych. Całość pracy jest bardzo dobrze osadzona w teorii przedmiotu badań (hydrologia, geomorfologia i teledetekcja) a także automatycznego przetwarzania dużej ilości danych. Przygotowane publikacje potwierdziły, że Doktorant bardzo dobrze zna wiedzę teoretyczną z zakresu nauk o Ziemi i środowisku, gdyż właściwie wybrał i zacytował 287 odniesień do literatury przedmiotu. Ta wiedza pozwoliła



# UNIwersytet Warszawski

właściwie zaplanować i zrealizować proces badawczy, gdyż Doktorant zebrał bogaty materiał dokumentacyjny (obejmujący znaczną część Polski), który został gruntownie przeanalizowany pozwalając osiągnąć zamierzone cele; potwierdza to zarówno proces badawczy, bardzo dobre wyniki, których ocena dokładności nie budzi zastrzeżeń, ale także publikacja osiągnięć badawczych w międzynarodowych czasopismach o wysokich parametrach bibliometrycznych, co podnosi rangę osiągnięcia Doktoranta oraz macierzystej jednostki. Warto podkreślić, że Doktorant był zaangażowany w realizację trzech projektów badawczych co podnosi umiejętności oraz doświadczenie Doktoranta w pracach zespołowych.

Uzyskane wyniki znajdują praktyczne zastosowania w badaniach środowiska (wody powierzchniowe należą do najbardziej wrażliwych obszarów na zachodzące zmiany klimatu, a to bezpośrednio wpływa na dostępność tlenu rozpuszczonego w wodzie i funkcjonowanie hydrobiosferę). Ponadto elementy geomorfologiczne i morfometryczne są niezbędnym komponentem planowania przestrzennego, modelowania środowiska przyrodniczego, a także rozwoju gospodarczego Polski (rozwój uczenia maszynowego i optymalizacji algorytmów pod względem zachodzących zmian w przyrodzie oraz zarządzaniu i optymalizacją przestrzeni).

Reasumując, jednoznacznie potwierdzam, że niniejsza rozprawa doktorska prezentuje bogatą wiedzę teoretyczną mgr. Krzysztofa Dyby w dyscyplinie nauk o Ziemi i środowisku oraz potwierdza umiejętność samodzielnego prowadzenia pracy naukowej; dotyczy to realizacji dobrze zaplanowanych badań bazujących na prawidłowo zdefiniowanym problemie naukowym, a także oryginalności stosowanych rozwiązań. Warto podkreślić, iż zaproponowana metodyka przetwarzania danych teledetekcyjnych i geoinformatycznych ma ogromny potencjał w rozwoju społeczeństwa informacyjnego bazującego na najnowszych osiągnięciach teledetekcyjnych oraz przetwarzaniu dużych ilości danych (cloud computing), znajdujących zastosowania w sferze naukowej, gospodarczej i społecznej Polski

Zgodnie z odpowiednimi dokumentami, w tym przepisami ustawy *Prawo o szkolnictwie wyższym i nauce* z dnia 20 lipca 2018 r. (z kolejnymi zmianami) w pełni popieram kontynuację procedury przewodu doktorskiego zmierzającego do publicznej obrony i nadania stopnia doktora w dziedzinie nauk ścisłych i przyrodniczych, dyscyplinie nauk o Ziemi i środowisku.

Z poważaniem

Dr. hab. Bogdan Zagajewski