

Prof. dr hab. inż. Beata Hejmanowska
Katedra Fotogrametrii Teledetekcji Środowiska i Inżynierii przestrzennej
Wydział Geodezji Górniczej i Inżynierii Środowiska
Akademia Górniczo-Hutnicza im. St. Staszica w Krakowie
al. Mickiewicza 30
30-059 Krakow
tel. +4812 6173826
kom. +48 605061510
URL: http://home.agh.edu.pl/~galia/index_en.html

**Recenzja rozprawy doktorskiej mgr. inż. Krzysztofa Dyby pt.
Interpretation of geospatial data using explainable machine learning methods**

Podstawa recenzji

Podstawą przygotowania recenzji była uchwała nr 49-2023/2023 rady naukowej dyscypliny nauki o Ziemi i środowisku Uniwersytetu im. Adama Mickiewicza w Poznaniu z dnia 30 kwietnia 2024. Przedmiotem recenzji jest rozprawa doktorska mgr. inż. Krzysztofa Dyby pt.: **Interpretation of geospatial data using explainable machine learning methods**, przygotowana pod opieką naukową promotora dr. hab. Jarosława Jasiewicza, prof. UAM oraz promotora pomocniczego dr. hab. Cezarego Kazimierskiego, prof. UAM. Recenzję opracowano zgodnie z ustawą z dnia 20 lipca 2018 roku - Prawo o szkolnictwie wyższym i nauce (Dz.U. 2018 poz. 1668 z późn. zm.).

Wstęp

Przedstawiono do recenzji:

- manuskrypt składający się z 30 stron, który zawiera 7 rozdziałów
- 3 publikacje
 - Dyba, K., Ermida, S., Ptak, M., Piekarczyk, J., Sojka, M. (2022) Evaluation of Methods for Estimating Lake Surface Water Temperature Using Landsat 8. Remote Sens., 14, 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, Volume 413, 2022, 108373, ISSN 0169-555X, <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>

Ocena merytoryczna i formalna treści pracy

Z formalnego punktu widzenia rozprawa spełnia wymagania ustawowe. Rozprawę doktorską stanowi bowiem w tym przypadku zbiór 3 opublikowanych i powiązanych tematycznie artykułów naukowych, co jest zgodne z Art. 187 punkt 3 Ustawy z dnia 20 lipca 2018 roku - Prawo o szkolnictwie wyższym i nauce (Dz.U. 2018 poz. 1668 z późn. zm.).

Udział doktoranta został określony w wymaganych oświadczeniach w tekście publikacji (*authorship contribution statement*).

W pierwszej publikacji:

- K.D.: *methodology, software, validation, formal analysis, data curation, writing, original and draft preparation, review, editing, and visualization,*
- S.E.: *software and writing, review, editing,*
- M.P.: *writing original and draft preparation, review, and editing,*
- J.P.: *writing original and draft preparation, review, and editing,*
- M.S.: *formal analysis, writing and original draft preparation, review, editing, and visualization.*

W drugiej publikacji:

- Krzysztof Dyba: *Conceptualization, Data curation, Formal analysis, Investigation, Software, Validation, Writing – original draft, Visualization, Funding acquisition.*
- Jarosław Jasiewicz: *Conceptualization, Methodology, Writing – original draft, Writing – review & editing, Supervision.*

W trzeciej publikacji Doktorant jest jedynym autorem.

Publikacje zostały poprzedzone 30 stronicowym opracowaniem stanowiącym podsumowanie badań opublikowanych w ww. 3 publikacjach. Opracowanie to zawiera 7 rozdziałów: *Introduction with Objectives, Materials and methods, Summary of articles with 3 subparagraphs, Discussion, Conclusion, Bibliography and Attachments with 2 subparagraphs (Academic achievements and Copies of articles).*

W pierwszej części recenzji odnoszę się do tego opracowania.

Ocena opracowania stanowiącego syntetyczne omówienie problematyki badawczej będącej przedmiotem 3 artykułów

We wstępie Doktorant podaje, że głównym powodem wykorzystywania obecnie w teledetekcji metod maszynowego uczenia jest ogrom danych do przetworzenia i niewystarczający w tym kontekście potencjał metod, ogólnie mówiąc eksperckich, z czym trudno się nie zgodzić. W dalszej kolejności omawia różne metody stosowane w uczeniu maszynowym: nienadzorowane i nadzorowane, *white-box* i *black-box*. Słusznie stwierdza, że zbyt dużo uwagi poświęca się pracom nad efektywnością modeli i zwiększaniem ich dokładności (moim zdaniem często zwiększenie dokładności nie jest znaczące), a zbyt mało tzw. „wyjaśnialności” modeli *black-box* (modele *white-box* są z natury rzeczy wyjaśnialne).

W tym momencie należy zauważyć, że rozprawa jest napisana po angielsku, zarówno autoreferat jak i publikacje. W związku z tematyką pracy

pojawia się szereg pojęć, których tłumaczenie na język polski nie jest ani łatwe, ani jednoznaczne (być może dlatego Doktorant nie zamieścił streszczenia w języku polskim).

Na stronie 3 autoreferatu, we wstępie, pojawia się zdanie, które można uznać za tezę (jeśli pisana po wykonaniu badań) lub hipotezę (jeśli stawiana na początku badań).

„*It can be hypothesized that interpretable machine learning can be a tool to support the process of interpreting spatial data*”, czyli po polsku:

„Można postawić hipotezę, że interpretowalne uczenie maszynowe może być narzędziem wspomagającym proces interpretacji danych przestrzennych”.

Prace badawcze, które są przedmiotem dysertacji miały na celu sprawdzenie tej hipotezy i wypełnienie istniejącej luki:

„*A current research gap is the lack of interpretation of the impact of explanatory variables (for example, spectral bands or geomorphometric variables) on the results of modeling spatial data using machine learning.*”.

Czyli: „Obecną luką badawczą jest brak interpretacji wpływu zmiennych objaśniających (na przykład pasm spektralnych lub zmiennych geomorfometrycznych) na wyniki modelowania danych przestrzennych za pomocą uczenia maszynowego”.

Wprowadzenie kończy się stwierdzenie:

„*The current state of research requires intensified work on the interpretive approach in machine learning, to answer the question whether the discussed methods will actually bring benefits in the process of acquiring new knowledge*”.

„Obecny stan badań wymaga zintensyfikowania prac nad podejściem interpretacyjnym w uczeniu maszynowym, aby **odpowiedzieć na pytanie, czy omawiane metody rzeczywiście przyniosą korzyści w procesie zdobywania nowej wiedzy**”.

W podrozdziale: „*Objectves*” Doktorant zdefiniował cel prac badawczych, które stanowią dysertację:

„Celem tej rozprawy jest wykazanie, czy wyjaśnialne metody uczenia maszynowego mogą być narzędziem wspomagającym proces modelowania danych geoprzestrzennych, gdzie miarą sukcesu nie jest wytrenowanie efektywnego modelu, ale możliwość uzyskania dodatkowej wiedzy o czynnikach prowadzących do wyniku. Tym samym **celem rozprawy jest przesunięcie kwestii wyłącznie oceny dokładności modeli uczenia maszynowego w kierunku zrozumienia ich działania i powiązania z innymi metodami przetwarzania danych.**” W mojej ocenie cel pracy jest jasny i poprawnie sformułowany. Rysunek 3 przedstawia w sposób klarowny podsumowanie zakresu omawianych w dalszej części 3 artykułów.

Rozdział 2: „*Material and methods*” zawiera opis wykorzystywanych metod badawczych oraz wykorzystane dane. Ogólnie rozdział jest napisany poprawnie, z wyjątkiem dyskusji nad różnymi metodami wykorzystywanymi do „wyjaśniania” modeli (strona 8). Moim zdaniem należało ten fragment zamieścić we wprowadzeniu.

Rozdział 3: *Summary of article* zawiera krótkie (1.5 strony) streszczenie każdego artykułu. Rozdział ten należy uznać za niewystarczający do oceny wartości naukowej badań przedstawionych w artykułach.

Rozdział 4: *Discussion* w teorii powinien być dyskusją ale zawiera w pierwszej części informacje bardziej nadające się do wprowadzenia jako przegląd literatury, a w drugiej luźne wnioski z publikacji.

Rozdział 5: *Conclusion* zawiera bardzo ogólne wnioski niewynikające bezpośrednio z przeprowadzonych prac badawczych, szczególnie w drugim akapicie.

Rozdział 6: *Bibliography* zawiera spis literatury.

Rozdział 7: *Attachemnts* zawiera inne osiągnięcia Doktoranta oraz kopie 3 artykułów.

Ocena dorobku Doktoranta, nawet w ograniczonym stopniu nie jest możliwa w oparciu o autoreferat, a szkoda bo publikacje są wartościowe, rzeczywiście stanowią spójną całość i pokazują rozwój naukowy w przeciągu ostatnich 4 lat. Pierwsza publikacja dotyczy wykorzystania modeli *white-box* i *black-box* do określania temperatury powierzchni wody w jeziorach z wykorzystaniem obrazów satelitarnych Landsat 8. Przedmiotem drugiej i trzeciej publikacji jest zagadnienie „wyjaśniania” modeli typu *black-box*, a konkretnie badanie wpływu zmiennych geomorfometrycznych na klasyfikację form terenu.

Artykuły opublikowane zostały w renomowanych czasopismach (*top 10 percent*) i zawierają opis badań, które stanowią znaczący przyczynek do nauk o Ziemi i środowisku:

- Remote Sensing (IF 4.2), Geomorphology (IF 4.2), 100 pkt,
- Scientific Reports (IF 3.8), 140 pkt.

Wszystkie artykuły są recenzowane, zatem rolę recenzenta rozprawy doktorskiej nie powinna być recenzja artykułów. Czytając publikacje naukowe często nasuwają się różne pytania i wątpliwości, które rzadko można zadać czy przedyskutować z autorami. Jest to możliwe na konferencji oraz właśnie w przypadku pisania recenzji dorobku naukowego. W związku z tym w niniejszej recenzji przedstawię swoje uwagi dotyczące opublikowanych badań, do dyskusji na obronie,.

Uwagi do pierwszej publikacji:

Dyba, K., Ermida, S., Ptak, M., Piekarczyk, J., Sojka, M. (2022) Evaluation of Methods for Estimating Lake Surface Water Temperature Using Landsat 8. *Remote Sens.*, 14, 3839. <https://doi.org/10.3390/rs14153839>

Tematyka badawcza poruszana w niniejszej publikacji jest ważna z praktycznego punktu widzenia i ciekawa pod względem naukowym. Celem było modelowanie temperatury kinetycznej powierzchni wody z wykorzystaniem obrazów satelitarnych Landsat-8. Testowano 3, a właściwie 4 podejścia: analizę regresji (*simple linear*), lasy losowe (*random forest RF*) oraz 2 modele dedykowane *obliczaniu temperatury powierzchni z wykorzystaniem obrazów Landsat 8 (land surface temperature models LST)*. W ocenie wykorzystanych w analizach metod kluczowa jest zawsze walidacja modelu, która w niniejszym artykule metodycznie została opisana w podrozdziale 2: *Model Validation Procedure*. Niestety są to tylko 3 zdania. Autorzy użyli metryk: *root mean square error (RMSE)*, *mean bias error (MBE)*, *coefficient of Pearson correlation (R)*. *Additionally, we calculated the standard deviation (SD) of the predicted values*. W artykule Autorzy nie podali wzorów, co utrudnia analizę wyników w tabeli 1. Interpretacja wartości MBE, RMSE i COR, czyli R nie stwarza problemu. Natomiast wartości SD wymagają wyjaśnienia, o jakie odchylenie standardowe chodzi. Zwykle odchylenie standardowe jest mniejsze niż RMSE, który zwiera zarówno bias i SD. W tym podrozdziale brak opisu sposobu podziału zbioru danych referencyjnych na treningowy i walidacyjny oraz testowy (pewne

informacje na ten temat znajdują się w rozdziale *Results*. Proszę o wyjaśnienie tego zagadnienia na obronie.

W rozdziale 3: *Results* przedstawione zostały wyniki porównania testowanych modeli. Na początku znajduje się opis dokładności uzyskanych podczas uczenia na zbiorze treningowym (dokładność walidacji?). Podano różne wyniki dla różnych zbiorników, czy to oznacza, że uczenie przeprowadzono oddzielnie dla każdego zbiornika, czy dla całości, a jedynie wyniki podano odrębnie dla zbiorników? Wydaje się, że tak, ponieważ poniżej znajduje się już zbiorcze porównywanie na zbiorze testowym. Prosiłabym o dokładne wyjaśnienie kwestii doboru zbioru uczącego i testowego oraz pojęć „walidacja” i dokładność modelu. Zwykle zbiór danych referencyjnych jest dzielony na 2 części: zbiór uczący i testowy. Tak wydaje się być w tym przypadku, wybrano losowo 10 zbiorników z 38 i 538 pomiarów temperatury in-situ (wszystkie pomiary z tych 10 zbiorników? Co stanowi 23% wszystkich danych?). Często podawane wyniki dokładności walidacji dotyczą procesu uczenia (wzór 3 str. 7), a nie dokładności obliczonej w oparciu o zbiór testowy, czyli danych których model „nie widział” w procesie uczenia. Wydaje się, że Autorzy podali w tabeli 1 wyniki dla danych testowych.

„Considerably more credible from the point of view of further application of the model are results of the validation of statistical models on independent data not used at the stage of model development. The results are obviously worse than those obtained at the stage of model development. In this study, for LM, the obtained values R and RMSE were lower than in the case of RF model (Table 1).”

Znacznie bardziej wiarygodne z punktu widzenia dalszego stosowania modelu są wyniki walidacji modeli statystycznych na niezależnych danych niewykorzystanych na etapie budowy modelu. Sformułowanie, że „bardziej wiarygodne” jest oczywistością. Tak naprawdę, jaka jest dokładność modelu okazują się dopiero na zbiorze testowym.

Wyniki czterech analizowanych metod przedstawione zostały na rysunku 5. Wszystkie wykresy na osi y mają pomiar temperatury in-situ, a na osi x temperaturę modelowaną za pomocą poszczególnych podejść (temperatura modelowana jest zmienną wyjaśnianą we wszystkich przypadkach). W metodzie LM temperatura przewidywana jest z wykorzystaniem regresji liniowej i dwóch zmiennych wyjaśniających: B10 i B11. W metodzie LST wykorzystano standardowe podejście. Proszę o uszczegółowienie, czy w artykule nazwa modelu LST oznacza przetworzenie dostępne na platformie Google Earth Engine bazujące na kanale B10 oraz innych danych dodatkowych, a LST-L2 standardowy produkt dostępny np. na earthexplorer oraz jaka jest różnica między nimi w omawianych badaniach. W metodzie RF wykorzystano 12 obrazów (zmiennych wyjaśniających: 9 kanałów Landsat 8, miesiąc, NDVI i NDWI).

Biorąc pod uwagę RMSE najlepszy wynik modelu temperatury uzyskano dla metody RF (1.73 °C, dla porównania LST-L2 3.38 °C). Natomiast jeśli chodzi o korelacje to różnice między czterema badanymi metodami były mniejsze i wynosiły średnio ok. 0.75.

W dalszej części artykułu Autorzy poprawnie zauważyli i skorygowali błąd systematyczny w metodach LST (2.42 °C). Skuteczność korekcji jest widoczna na histogramach na rysunku 9. Natomiast wykresy po lewej są niejasne. Dlaczego po korekcji jest mniejsza liczba punktów? Być może jest to złudzenie.

Zaprezentowana w dalszej części analiza *feature importance using impurity algorithm* w RF jest poprawna i zgodna z oczekiwaniami (największy wpływ mają kanały termalne i miesiąc).

Rozdział dyskusja składa się z 3 części. Pierwsza zawiera porównanie wybranych artykułów pod kątem dokładności określania temperatury z wykorzystaniem kanałów termalnych satelity Landsat. W mojej ocenie z tej dyskusji nic nie wynika poza tym, że prezentowane w artykule wyniki są podobne do wyników publikowanych przez innych badaczy. Druga część dyskusji dotyczy ekstrapolacji przestrzennej zaprezentowanej na rysunku 10. Jest ona moim zdaniem ryzykowna. Ostatnia część dyskusji dotyczy zmienności temperatury na powierzchni zbiornika wodnego. Autorzy prezentują wyniki dla dwóch wybranych jezior i porównują z wynikami innych badaczy. Nie jestem specjalistą w tej dziedzinie i nie mogę się do tego odnieść.

W rozdziale *Conclusions* Autorzy formułują wnioski dotyczące korekcji atmosferycznej, które moim zdaniem nie wynikają z przeprowadzonych badań (cały akapit: *Many papers recommend the application of atmospheric correction...*). Natomiast następujący wniosek należy uznać za poprawny, rzeczywiście korekcja poprawiła wynik o 30%.

„This paper proposes an appropriate calibration correction for data from Landsat Level-2 Surface Temperature Science Product (LST-L2) that considerably improves results (root mean square error reduced by 30%) of estimation of lake temperature in the territory of Poland. Nevertheless, despite the correction, the random forest model provided the best results.”

„W artykule tym zaproponowano odpowiednią korektę kalibracji dla danych z Landsat Level-2 Surface Temperature Science Product (LST-L2), która znacznie poprawia wyniki (średni błąd kwadratowy zmniejszony o 30%) szacowania temperatury jezior na terenie Polski. Niemniej jednak, pomimo korekty, model lasu losowego zapewnił najlepsze wyniki.”

Uwagi do drugiej publikacji:

Dyba K., Jasiewicz J. (2022) Toward geomorphometry of plains - Country-level unsupervised classification of low-relief areas (Poland), *Geomorphology*, Volume 413, 2022, 108373, ISSN 0169-555X, <https://doi.org/10.1016/j.geomorph.2022.108373>

Przedmiotem badań w publikacji nr 2 była geomorfometria obszarów o umiarkowanych deniwelacjach terenu. Cel badań został jasno sformułowany we wprowadzeniu, po obszernym przeglądzie literatury. *„The overview of previous studies concludes that despite the wide variety of plains, especially in post-glacial areas, no general methods have been made to consider the diversity of low-relief areas. In this article, we aim to present the quantitative variation of the geomorphometric features of the plains including several morphogenetic zones as a step toward developing assumptions for the automatic classification of plains.”* Przegląd literatury prowadzi do wniosku, że pomimo dużej różnorodności równin, zwłaszcza na obszarach postglacjalnych, **nie opracowano żadnych ogólnych metod, do opisu różnorodności obszarów o mało urozmaiconej rzeźbie**. W artykule przedstawiono ilościowo zmienność cech geomorfometrycznych równin, w tym kilku stref morfogenetycznych, co należy uznać za krok w kierunku **opracowania założeń dla automatycznej klasyfikacji równin**. Ponadto Autorzy nadmieniają, że nie tylko nie ma metod do opisu morfologii obszarów równinnych, ale nawet **nie ma przejrzystej definicji tych obszarów**.

Z uwagi na fakt, że nie jestem specjalistą ani w zakresie geografii ani geomorfologii w mojej recenzji ograniczę się do części metodycznej związanej z analizami przestrzennymi z wykorzystaniem numerycznych modeli terenu (NMT) oraz klasyfikacją nienadzorowaną. Autorzy wybrali klasyfikację nienadzorowaną chcąc uniknąć wpływu na wydzielane klasy. Z tą argumentacją zgadzam się całkowicie. *The most significant limitation of supervised methods is that they cannot reveal new patterns in the data and the measure of success is the degree of compliance with existing, usually manually created maps.* Najpoważniejszym ograniczeniem metod nadzorowanych jest to, że nie są w stanie ujawnić nowych wzorców w danych, a miarą sukcesu jest stopień zgodności z istniejącymi, zazwyczaj ręcznie tworzonymi mapami.

Autorzy wybrali jako metodę klasyfikacji nienadzorowanej: *Gaussian Mixture Model (GMM)*. *The GMM is a clustering system that provides so-called soft clusters, dividing an area into discrete units and determining the probability of belonging to each cluster.* GMM to system klastrowania zapewniający tzw. miękkie klastry, dzielący obszar na odrębne jednostki i określający prawdopodobieństwo przynależności do każdego klastra. Wybór metody miękkiej uzasadniono ciągłym charakterem morfologii terenu i problematycznym wykorzystaniem najpopularniejszej metody twardej k-means. Z tym uzasadnieniem również się zgadzam.

The results were compared with other automatic methods and with existing expert driven geomorphological maps in selected areas. Wyniki porównano z wynikami innych automatycznych metod, a także z istniejącymi, opracowanymi przez ekspertów mapami geomorfologicznymi na wybranych obszarach. Jest to również uzasadnione podejście.

W dalszej części artykułu Autorzy szeroko omawiają wybór 10 cech morfometrycznych, które wyznaczają z NMT. W podrozdziale 2.3 *Clustering* opisują różne metody klasteringu, ich zalety i wady, szczegółowo metodę GMM oraz *Bayesian Information Criterion (BIC)* do wyboru optymalnej liczby klas. Analizowano warianty od 4 do 20 klas, a ostatecznie zdecydowano się na 20 klas. Rysunek 3 pokazuje schematycznie przebieg analiz. Ciekawe było wykorzystanie metody PCA do przypisania etykiet klastrom, omówione w podrozdziale 3.2.1 *Surface variability inside the principal components space.*

Rozdział 3.3 *Uncertainty of the clustering* zawiera krótką analizę niepewności klastrowania. Metoda GMM określa prawdopodobieństwo należenia piksela do danej klasy (po etykietowaniu do danego typu morfologicznego) co umożliwia utworzenie obrazu rozkładu niepewności (Rys. 9). Wyjaśnienia dotyczące występowania większych błędów w obszarach „przejściowych”, które w obszarach o niezbyt urozmaiconej rzeźbie są rozległe należy uznać za przekonujące.

Artykuł jest zakończony obszernym rozdziałem 4. *Discussion.*

The detailed regional analysis goes beyond the aims of this work; therefore, in the discussion, we will analyze to what extent the spatial distribution of surface types is related to the geomorphological processes and landforms taking place in a given zone. Regardless of compliance with the extent of morphogenetic zones,

*we also observe a strong relation to the physico-geographical units (Solon et al., 2018) at the mezo-region level (Fig. 10). Thus, it becomes the best source for comparing the convergence of the classification results with the expert understanding of this phenomenon. Such a comparison **confirms that the concept of classification, including the selection of variables and the method of***

grouping and labeling the variables, creates a pattern recognizable to those familiar with the geographic division of the country.

„Szczegółowa analiza regionalna wykracza poza cele tej pracy, dlatego w dyskusji przeanalizujemy, w jakim stopniu przestrzenny rozkład typów powierzchni jest powiązany z procesami geomorfologicznymi i formami terenu zachodzącymi w danej strefie. Niezależnie od zgodności z zasięgiem stref morfogenetycznych, obserwujemy również silny związek z jednostkami fizyczno-geograficznymi na poziomie mezoregionu. Staje się to zatem najlepszym źródłem do porównania zbieżności wyników klasyfikacji z eksperckim rozumieniem tego zjawiska. **Takie porównanie potwierdza, że koncepcja klasyfikacji, w tym dobór zmiennych oraz sposób grupowania i etykietowania zmiennych, tworzy wzorzec rozpoznawalny dla osób znających podział geograficzny kraju.**”

W tym rozdziale porównano wynik klasyfikacji z ekspercką mapą geomorfologiczną dla czterech obszarów testowych. Autorzy pokazali słabe i mocne strony obu metod.

Podsumowując, artykuł prezentuje dojrzałe i głębokie studia nad wykorzystaniem klasyfikacji nienadzorowanej do wydzielenia cech morfometrycznych powierzchni terenu. Autorzy skupili się, z sukcesem, na wyjaśnianiu wyników automatycznej klasyfikacji.

Uwagi do trzeciej 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, <https://doi.org/10.1038/s41598-024-56066-6>

Trzeci artykuł stanowi kontynuację tematyki z artykułu drugiego i dotyczy wyjaśniania wpływu zmiennych geomorfometrycznych na klasyfikację form terenu na przykładzie wybranych obszarów Polski, w tym przypadku z wykorzystaniem narzędzia AI ALE (*accumulated local effects*). Ponadto zamiast klasyfikacji nienadzorowanej Autor wykorzystał klasyfikację nadzorowaną z wykorzystaniem maszynowego uczenia.

W rozdziale *Morphometric variables* Autor dyskutuje różne zmienne morfometryczne i ostatecznie prezentuje wybranych 9 zmiennych w tabeli 1.

Autor wykorzystał 3 popularne modele bazujące na uczeniu maszynowym: lasy losowe (*RF - Random Forests*), *gradient boosting XGBoost (eXtreme Gradient Boosting)* i *LightGBM (Light Gradient-Boosting Machine)* i konwolucyjne sieci neuronowe (*CNN - convolutional neural network*) U-Net.

Rozdział dotyczący walidacji jest krótki i niewystarczająco jasno prezentujący zagadnienia związane z walidacją. Z ostatniego zdania w tym artykule można się domyślić, że analiza dokładności dotyczy walidacji modelu, a nie dokładności na zbiorze testowym, którego model „nie widział” podczas uczenia. Ponadto podano metryki bez wzorów, co w przypadku „accuracy” może oznaczać różne wielkości. W maszynowym uczeniu ACC (accuracy) oblicza się ze wzoru $(TP+TN)/(TP+TN+FP+FN)$. Natomiast w klasyfikacji używa się overall accuracy (OA), które nie równa się ACC w przypadku wielu klas. Jest istotne z jakiego wzoru oblicza się dokładność, ponieważ ACC w znacznym stopniu zawyża dokładność klasyfikacji (w przypadku wielu klas).

W rozdziale *Results* podano dokładność klasyfikacji z wykorzystaniem metryki accuracy, przy czym najwyższą dokładność uzyskano dla modelu XGBoost 92.8% a najniższą dla U-Net 59.8%. Pozostaje pytanie z jakiego wzoru korzystano.

Wynik klasyfikacji można porównać z mapą referencyjną na rysunku 6 (ocena wizualna jest bardzo dobra).

Wyjaśnialność modelu analizowano z wykorzystaniem metody ALE, która pozwala na analizę wpływu zmiennych na wynik modelu. Metoda ALE jest podobna do *partial dependence plot*. Wynikiem analizy z wykorzystaniem tej metody jest Rys. 7, na którym związek między użytymi zmiennymi geomorfometrycznymi, a prawdopodobieństwem formy terenu określono za pomocą skumulowanych efektów lokalnych.

Rozdział *Discussion* jest ciekawy chociaż pierwszy akapit zawiera właściwie wnioski, a nie dyskusje. Wnioski są ciekawe. Autor podaje, że według jego najlepszej wiedzy metoda ALE nie była do tej pory stosowana w kartowaniu geomorfologicznych. Metoda ta pozwala sprawdzić, dlaczego klasyfikator wyróżnił daną formę terenu, co jest szczególnie ważne w przypadku błędnych klasyfikacji (możemy zinterpretować na podstawie zmiennych objaśniających, co powoduje błąd). Co więcej, tę metodę można stosować nawet w tradycyjnym mapowaniu; jeśli geomorfolog nie jest pewien, czy rozpoznaje daną formę terenu w terenie, może wykorzystać wykresy ALE. Główną zaletą metody ALE jest względna łatwość interpretacji, ponieważ zapewnia ona przejrzystą wizualizację tego, jak każda cecha geomorfometryczna wpływa na przewidywania modelu. ALE jest z pewnością przydatnym narzędziem, ma również pewne wady. Główne ograniczenia są związane z małymi zbiorami danych, niską zmiennością cech i wrażliwością samego modelu. Jakość zbioru danych wejściowych i dokładność modelu powinny być dokładnie rozważone przed wyciągnięciem wniosków. Te ostatnie dwa stwierdzenia są ogólne i nie wynikają z przeprowadzonych badań. Należy unikać takich wniosków.

W dalszej części tego rozdziału znajduje się rzeczywiście dyskusja, która obejmuje 3 zagadnienia. Pierwsze dotyczy porównania wyników klasyfikacji CNN z danymi literaturowymi, a drugie dyskusji różnych wersji modeli CNN (głównie jeśli chodzi o liczbę parametrów oraz możliwości ekstrapolacji wytrenowanych modeli). Dokładność modeli CNN raportowana w literaturze jest wyższa niż Autor uzyskał w swoich badaniach. Wyjaśnieniem tego faktu jest mniejsza liczba wydzielanych parametrów geomorfologicznych oraz „łatwiejszy” obszar. Można się zgodzić z tym wytłumaczeniem.

Jeśli chodzi o analizę różnych modeli CNN to w mojej opinii są to luźne rozważania na ten temat bez bezpośredniego związku z przeprowadzonymi badaniami.

Ostatnie zagadnienie dotyczy „przenoszalności” modelu tzn. odpowiedzi na pytanie czy można nauczyć model na jakimś zbiorze treningowym i wykonać klasyfikacje na dowolnym innym obszarze np. dla całej Polski. Odpowiedz Autora jest negatywna i należy ją uznać za uzasadnioną na tym etapie badań. Docelowo jednak idea AI jest takie wytrenowanie modelu, żeby można go było potem używać już bez żadnych danych treningowych. Rozumiem obiekcje Doktoranta, bo ja je również mam, ale z założenia oczekujemy, że AI jest w stanie tak się nauczyć, żeby potem nie potrzebować danych uczących.

Na koniec parę uwag:

1. Recenzowane 3 publikacje są bardzo wartościowe i nowatorskie szczególnie artykuł drugi i trzeci, pozwalają zauważyć niewątpliwy rozwój Doktoranta.
2. W mojej ocenie należy jednak poprawić 2 elementy:

- zawsze dokładnie opisywać metodykę (każdy etap), szczególnie analizę dokładności: podział danych referencyjnych, zagadnienie korelacji danych, wzory, metryki dokładności,
- nie zamieszczać wniosków ogólnych, nie wynikających z przeprowadzonych badań.

Podsumowanie

Konkludując, należy stwierdzić, że dysertacja magistra inżyniera Krzysztofa Dyby odpowiada w pełni wymogom formalnym stawianym rozprawom doktorskim oraz spełnia wymagania ustawowe: Prawo o szkolnictwie wyższym i nauce (Dz.U. 2018 poz. 1668 z późn. zm.). Stanowi oryginalne rozwiązanie ważnego problemu naukowego, świadczy o dużej wiedzy Autora z zakresu nauk o Ziemi i środowisku oraz o umiejętnościach niezbędnych do samodzielnego prowadzenia badań. Wnioskuje o dopuszczenie Doktoranta do dalszych etapów postępowania.

B. Hojmanowicz