

dr hab. inż. Beata Hejmanowska prof. n. AGH

Kraków, dnia 5.02.2017 r.

Katedra Geoinformacji, Fotogrametrii i Teledetekcji Środowiska

Wydział Geodezji Górniczej i Inżynierii Środowiska

Akademia Górniczo-Hutnicza w Krakowie

Al. Mickiewicza 30

RECENZJA

rozprawy doktorskiej mgr Adrianny Marcinkowskiej-Ochyra

**"Ocena przydatności obrazów hiperspektralnych APEX oraz maszyn wektorów
nośnych (SVM) do klasyfikacji roślinności subalpejskiej i alpejskiej
Karkonoszy "**

1. Ogólna charakterystyka zakresu i treści rozprawy

Rozprawa, której promotorem jest dr hab. Bogdan Zagajewski i dr Anna Jarocińska obejmuje na 156 stronach 9 rozdziałów w tym bibliografię, spis rysunków i tabel. Tematem rozprawy jest ocena przydatności obrazów hiperspektralnych APEX oraz metody klasyfikacji SVM do klasyfikacji roślinności subalpejskiej i alpejskiej Karkonoszy. Temat jest jasno sformułowany i praca jest zgodna z tematem.

2. Ocena celu i zakresu pracy

We wstępie zostały wymienione trzy cele rozprawy doktorskiej:

- „poznawczy, mający za zadanie opracowanie i przetestowanie algorytmu klasyfikacji wysokorozdzielczych hiperspektralnych danych APEX do klasyfikacji roślinności,
- aplikacyjny, którego efektem jest opracowanie mapy rozkładu dominujących zbiorowisk roślinnych piętra subalpejskiego i alpejskiego Karkonoszy,
- metodyczny, pozwalający na określenie przydatności lotniczych zobrazowań hiperspektralnych i metody maszyn wektorów nośnych (Support Vector Machines – SVM) do klasyfikacji roślinności subalpejskiej i alpejskiej Karkonoszy”.

Został również zarysowany zakres pracy obejmujący przygotowanie danych referencyjnych oraz przetwarzanie danych hiperspektralnych w tym, przetwarzanie wstępne, ekstrakcja informacji i ocena dokładności.

Analizując cele pracy wydaje się, że można by połączyć cel pierwszy i trzeci.

W pracy brakuje postawienia tezy, którą w ramach badań zwykle próbuje się udowodnić.

3. Merytoryczna ocena rozprawy doktorskiej

We wstępie znajduje się krótki przegląd literatury na temat wykorzystania teledetekcji do kartowania roślinności. Autorka przytacza różne publikacje, pokazując pewne zastosowania i metody, nie konkluduje jednak tego wywodu. Pod koniec wstępu, po zaprezentowaniu celu badań krótko przedstawia wybrane elementy metodyki.

Rozdział pierwszy dotyczy podstaw teledetekcji hiperspektralnej, w którym Autorka podaje definicje, omawia sposoby pozyskiwania i przetwarzania danych hiperspektralnych. W rozdziale tym znajduje się podrozdział 1.2.3. - redukcja wymiarowości danych. W przyszłości Autorka mogłaby skonfrontować swoje doświadczenia z doświadczeniem Dr Ewy Głowienki w tym zakresie, której rozprawa doktorska (2014) dotyczyła między innymi podobnego tematu: „Analiza porównawcza metod przetwarzania danych hiperspektralnych o zróżnicowanej rozdzielczości”.

Rozdział drugi zawiera informacje na temat metod klasyfikacji danych teledetekcyjnych. Na stronie 32 Autorka stwierdza: „W pracach Foody'ego i Mathura (2006), Waske i in. (2010), Camps-Valls'a (2004) wykazano, że metoda maszyn wektorów nośnych, zastosowana w pracy, pozwala na stosowanie nawet niewielkich zestawów treningowych osiągając wysokie dokładności”, co jest argumentem za zastosowaniem tej metody w rozprawie doktorskiej.

W następnym rozdziale 3 znajduje się opis metody klasyfikacji opartej o maszyny wektorów nośnych (SVM). Na stronie 41 Autorka stwierdza, że: „dokładności osiągnięte przy klasyfikacji SVM są wyższe niż w przypadku sztucznych sieci neuronowych, szczególnie

przy większej liczbie klas” co należy uznać za istotne z uwagi na przyjęcie w pracy właśnie metody SVM.

Rozdział 4 zawiera przegląd literatury na temat wykorzystania danych hiperspektralnych do klasyfikacji roślinności. Rozdział ten jest obszerny, zawiera odwołania do wielu pozycji literatury. Brak jest jednak syntezy i konkretnych wniosków. Autorka na koniec tego rozdziału stwierdza jedynie, że: „ Jak wynika z powyższego przeglądu literatury, dane hiperspektralne pozyskiwane z różnych sensorów są z powodzeniem wykorzystywane do klasyfikacji roślinności na różnych poziomach.”. A ostatnie zdanie: „Stosowane są różne metody, jednak analizy porównawcze najczęściej wykazują najwyższą dokładność przy użyciu metody maszyn wektorów nośnych” należy uznać za nieco „na wyrost” biorąc pod uwagę w szczególności dalsze rozdziały.

Szczegółowe uwagi dotyczące rozdziałów teoretycznych znajdują się w dalszej części recenzji.

W rozdziale 5 zamieszczono szczegółowy opis obszaru i obiektu badań. Ze względu na to, że nie jestem specjalistą w dziedzinie geografii nie odnoszę się w recenzji do tej części pracy.

Rozdział 6 zatytułowano: „metodyka pracy”, właściwie powinna to być „metodyka badań”. Obrazy hierspektralne APEX zostały zarejestrowane w ramach projektu: HyMountEcos – Hyperspectral Remote Sensing for Mountain Ecosystem, dofinansowanego przez European Facility For Airborne Research Transnational Access (EUFAR TA). Wstępne przetwarzanie obrazów, obejmujące korekcję geometryczną i radiometryczną wykonane zostało przez firmę VITO. Nie udało się przeprowadzić korekcji zakłócającego wpływu ukształtowania terenu i dlatego analizy dalsze wykonano na obrazach z wykorzystaniem maski, która wyłączyła z analiz obszary zacienione. Dane referencyjne wykorzystane w analizie dokładności pochodziły z mapy referencyjnej prof. Wojtunia i Żołnierza oraz z wywiadu terenowego, który przeprowadziła Autorka wraz z innymi specjalistami w tej dziedzinie. Natomiast w ramach rozprawy doktorskiej wykonano wariantowe klasyfikacje danych APEX metodą SVM. Warianty obejmowały klasyfikację z wykorzystaniem różnej liczby kanałów (testowanie metod redukcji nadwymiarowości danych hiperspektralnych) oraz przy użyciu różnych parametrów metody SVM. Ciekawym aspektem było porównywanie wyników

klasyfikacji na całych polach treningowych oraz na zestawach losowanych pikseli z pól treningowych, co nawiasem mówiąc jest znane z literatury i Autorka przeprowadziła ciekawe porównanie własnych wyników z wynikami z literatury.

Szkoda, że Autorka nie porównała wyników klasyfikacji metodą SVM z innymi metodami, niekoniecznie bardzo zaawansowanymi, jak sieci neuronowe, random forest czy drzewa decyzyjne, MF etc. ale prostymi, takimi jak ML czy SAM. Bardzo interesujące byłoby porównanie dokładności uzyskanej metodą SVM i metodami „tradycyjnymi”. Byłoby to szczególnie ciekawe w porównaniu w wynikami badań literaturowych, w których właśnie tak postępowano.

Innym zagadnieniem jest istotność statystyczna w różnicy dokładności wyników klasyfikacji, tzn. odpowiedź na pytanie: na ile rzeczywiście poprawiana jest dokładność klasyfikacji?

Na końcu podrozdziału 6.1.2. Redukcja przestrzeni spektralnej znajduje się opis wyników badań nad redukcją nadwymiarowości danych. Na Ryc. 30, zaprezentowano: spadek dokładności całkowitej klasyfikacji wywołany brakiem określonego kanału. Pytanie: jaki z tego wniosek? Poza tym ten fragment jest opisem wyników, a nie metodyki, zatem nie powinien znajdować się w tym rozdziale.

Podrozdziały: 6.2.2. - wybór wzorców do klasyfikacji i weryfikacji, 6.2.3. - klasyfikacja przy użyciu maszyn wektorów nośnych (SVM) i 6.2.4. - weryfikacja i ocena jakości klasyfikacji zawierają opis właściwej metodyki badawczej. Do weryfikacji „łącznie zgromadzono 812 zidentyfikowanych płatów roślinności zapisanych w formacie wektorowym *.shp (shapefile)”. „Klasyfikację SVM wykonano w ENVI i w oprogramowaniu R, pakiet e1071”. „Wybrano różne zestawy kanałów do klasyfikacji. Były to kolejno: 40 kanałów PCA, 30 kanałów MNF, 18 i 70 kanałów z analizy informacyjności, a także 252 oryginalne kanały APEX. Po wyborze zestawu kanałów wskazano wszystkie utworzone klasy, a następnie wybrano odpowiednie opcje SVM, uwzględniając typ funkcji jądrowej i związane z nią parametry. Zdecydowano się na przetestowanie czterech funkcji jądrowych, aby wybrać tę, która daje najlepsze rezultaty”. Pojawiają się w tym momencie 2 pytania. Dlaczego wybrano takie zestawy kanałów, czy wynika to z wcześniejszych badań nad redukcją nadwymiarowości danych? Jakie były przesłanki dotyczące doboru parametrów funkcji jądrowej SVM?

W rozdziale 7 zamieszczono wyniki. W tabeli 7 znajduje się dokładności klasyfikacji analizowanych obszarów. Dokładność całkowita oscyluje wokół 80%. Czy to są najlepsze uzyskane wyniki?

Poniżej tabeli 7 na stronie 87 znajduje się stwierdzenie: „Obrazy przetworzone do poziomu 2, które otrzymano w dalszym etapie prac, pozbawione były wpływu atmosfery, co pozwoliło ocenić wyniki jako bardziej wiarygodne”. To w ogólności jest prawdą, ale po pierwsze nie wynika to z badań Autorki, a po drugie ma znaczenie głównie w innych metodach klasyfikacji opartych na porównywaniu krzywych spektrometrycznych z krzywymi z obrazu, a nie w metodzie SVM. Nie należy wysuwać wniosków nie wynikających bezpośrednio z badań. Ale może było to badane przez Autorkę tylko nie wystarczająco jasno opisane.

Rozdział 7.1. - wyniki optymalizacji i testowania algorytmu klasyfikacji zawiera wyniki wariantowych analiz z wykorzystaniem różnej liczby kanałów i różnego sposobu próbkowania. W tabeli 8 znajdują się ciekawe wyniki analizy dokładności, z której wynika konkretny wniosek, że „najlepsza dokładność całkowitą ogółem osiągnięto dla funkcji liniowej przy użyciu całych poligonów do klasyfikacji”, ponad 80%. Tabela 9 zawiera również ciekawe, konkretne wyniki dotyczące dokładności klasyfikacji dla różnych zestawów kanałów APEX. Najwyższą dokładność całkowitą wraz ze współczynnikiem kappa osiągnięto dla oryginalnych 252 kanałów (82,69%), a najniższą wartość dokładności przyjął zaś obraz złożony z 18 kanałów z analizy informacyjności (68,14%)”. Tutaj pojawia się niewłaściwe używanie przez Autorkę strony czynnej czasownika: „obraz przyjął wartości dokładności”. Bardziej zgrabne byłoby użycie strony biernej: „dla obrazu uzyskano...”. Nie podoba mi się poza tym sformułowanie: „analiza informacyjności”.

Dalej Autorka testuje tzw. parametr kary: C. Dla mnie niejasne są wykresy: Ryc. 35 i 36, na których rośnie dokładność wraz ze wzrostem kary? W dalszej części analizowane są parametry gamma w Tab. 10. Niewystarczająco jasno opisany jest wpływ tego parametru na wynik klasyfikacji oraz wnioski wynikające z jego testowania.

W podrozdziale 7.2. Wyniki klasyfikacji roślinności wysokogórskiej Karkonoszy zamieszczono mapy z wykorzystaniem SVM przy przyjęciu najlepszych parametrów. „Najwyższą dokładność całkowitą (85,49%) uzyskano dla 40 kanałów PCA przy zastosowaniu funkcji radialnej”. W dalszej części znajduje się szczegółowy opis analizy

dokładności poszczególnych klas. Ciekawe są wykresy na ryc. 53 i 54 przedstawiające: dokładności producenta/użytkownika dla poszczególnych klas uzyskane podczas 100-krotnego losowania pikseli. Dokładność zaprezentowana w ten sposób pozwala na określenie analizę jej wiarygodności. Również Ryc. 55. - wzrost dokładności klasyfikacji spowodowany zwiększeniem liczebności pikseli treningowych jest ciekawa i pokazuje: „wzrost dokładności całkowitej klasyfikacji wraz ze wzrostem liczebności zestawu treningowego, uzyskany dla 13 klas” od poniżej 74% do 80%.

Następne rozdziały mają charakter dyskusyjny. Pierwszy z nich został nazwany: „Przydatność SVM i obrazów hiperspektralnych do klasyfikacji roślinności”. Po pierwsze Autorka porównała własne wyniki z wynikami uzyskanymi wcześniej przez Promotora (Zagajewski 2010) w wyniku klasyfikacji obrazów DAIS metodą sieci neuronowych na obszarze testowym w Tatrach. Dokładność całkowita uzyskana w Tatrach było wyższa (89%) niż dokładność uzyskana przez Autorkę (80%). W tabeli 13 zamieszczono ponadto dokładność uzyskaną dla poszczególnych klas. Autorka rozprawy przeprowadziła analizę porównawczą dla poszczególnych klas. Brak natomiast syntetycznych wniosków. Na przykład, że dla 5 klas uzyskano lepszą dokładność producenta, a dla 7 gorszą, a w przypadku dokładności użytkownika na odwrót w 8 przypadkach wyniki Autorki są lepsze, a w 3 gorsze. Pojawia się pytanie czy te różnice są znaczące i na ile (w porównaniu do wiarygodności dokładności określonej w rozdziale poprzednim). Po drugie co z tego wynika?

W dalszej kolejności Autorka zestawia w Tab. 14. Porównanie dokładności klasyfikacji roślinności na danych hiperspektralnych, z które nawiasem mówiąc nie wynika, że metoda SVM jest najlepsza. Poza tym czytelnik jest nieco zdezorientowany widząc w tabeli 14 różne sensory i jedną publikację APEX (Autorki), a potem tabelę 15 już tylko z danymi APEX. Dlaczego Autorka porównała w tabeli 14 tylko swoje wyniki ? Niezależnie od tego obie tabele są ciekawe. Zakładając kompletność przeglądu literatury na temat klasyfikacji danych APEX to po pierwsze Autorka wraz z zespołem kolegów ma zdecydowanie najwięcej publikacji na ten temat (6), 2 inne publikacje. Dominującą metodą jest SVM (tym bardziej szkoda, że nie Autorka nie porównała jej z żadną inną metodą). Nawiasem mówiąc nie ma w spisie literatury publikacji (Raczko i in. 2015). Trudno powiedzieć jakie metody są najczęściej wykorzystywane do klasyfikacji danych APEX, ponieważ w „innych” publikacjach wykorzystano ML, SVM, Random Forest, AdaBoost i sieci neuronowe.

W dalszej części dyskusji na stronach 121-123 Autorka analizuje wpływ redukcji kanałów oraz „kompresji” metodą PCA i MNF. Ta część jest trochę mało syntetycznie napisana, wynika to być może z braku jednoznacznych wniosków, ponieważ badacze uzyskują sprzeczne wyniki.

Dalsza analiza dotyczy dokładności klasyfikacji uzyskanych przy zastosowaniu różnych liczebności zestawów treningowych i różnych metod klasyfikacji danych hiperspektralnych (Tab. 17). W tym przypadku wnioski są jednoznaczne tzn. „we wszystkich badaniach występuje wzrost dokładności klasyfikacji wraz ze zwiększeniem się zestawu treningowego, jednak w każdym przypadku najwyższe dokładności osiągnięte są z użyciem metody SVM”. Na zakończenie tej części Autorka porównuje własne wyniki do wyników klasyfikacji wykonanej w czeskiej części Karkonoszy.

W dyskusji Autorka odnosi się również do własnych badań innego rodzaju, mianowicie z wykonanych na symulowanych danych EnMAP, o czym Autorka pisze dopiero w dalszej części pracy. Moim zdaniem należało opisać w części metodycznej metodykę wszystkich wykonanych badań co zapewne spowodowałaby większą przejrzystość pracy.

Na końcu podrozdziału 8.1 znajduje się podsumowanie: „porównując wyniki własnych badań do przedstawionych powyżej można stwierdzić, że zastosowanie wysokorozdzielczych spektralnie obrazów APEX o wielkości piksela 3,12 m do klasyfikacji dużej liczby zbiorowisk roślinnych obszarów nieleśnych wraz z dającą wysokie wyniki dokładności metodą maszyn wektorów nośnych było właściwe i uzasadnione”. Ten wniosek jest bardzo ogólny i ewentualnie mógłby znaleźć się na końcu pracy. Poza tym jest nieprecyzyjny, dlaczego Autorka podkreśla akurat wielkość piksela: 3.12 m? Czy dla innych rozmiarów piksela będzie inaczej, czy Autorka to badała?

Rozdział 8.2 opisuje, poprawnie z merytorycznego punktu widzenia, trudności w klasyfikacji zbiorowisk roślinnych. Tekst jest jasny i dobrze zilustrowany. Jednak Autorka niepotrzebnie zamieszcza również wnioski niewynikające z jej badań. „Niniejsza praca zakładała wykorzystanie do identyfikacji algorytmu maszyn wektorów nośnych, które lepiej sprawdziły się w tym przypadku dla zbiorowisk o większych powierzchniach, natomiast

do wykrycia zbiorowisk niewielkich, homogenicznych zajmujących pojedyncze płyty, czy zbiorowisk występujących w kompleksach złożonych z mniejszych płatów, można byłoby zastosować metody bazujące na pikselach czystych spektralnie (np. Spectral Angle Mapper) lub podpikselowe (np. Spectral Unmixing), uprzednio gromadząc dla nich charakterystyki spektralne”. Czy Autorka to badała?

I dalej: „w obszarach o dużych deniwelacjach analiza bardziej wysokorozdzielczych przestrzennie materiałów, np. ortofotomap, również nie przyniosłaby informacji o lokalizacji zbiorowisk, gdyż na nich cienie także byłyby widoczne”. Należy w tym miejscu zauważyć, że strefa cienia zależy od położenia słońca, a zatem od godziny rejestracji zdjęć lotniczych i może być inna niż na obrazach APEX.

Podrozdział 8.3. - zalety zastosowanych badań zawiera różne informacje w tym opis analiz wykonanych na danych symulowanych EnMap. Na początku tego podrozdziału Autorka opisuje zalety „algorytmu szczegółowej klasyfikacji”, ale nie bardzo wiadomo o jaki algorytm chodzi. W poprzednich rozdziałach znajduje się opis wyników badań różnych wariantów, ale brak wystarczająco jasnego opisu proponowanego „algorytmu”. „Szczegółowy dobór parametrów poszczególnych funkcji jądra SVM pozwolił wyselekcjonować optymalny algorytm, dopasowany do danych wejściowych” ? W dalszej części znajduje się stwierdzenie: „Tradycyjne klasyfikatory nie byłyby w stanie poprawnie zidentyfikować klas na obrazie APEX”, czy wynika to z badań własnych?

Dalej Autorka analizując przydatność danych APEX odnosi się do daty wykonania zobrazowania (wrzesień). „We wrześniu roślinność zaczyna się przebarwiać, co pozwala ją bardzo dobrze różnicować”. Pojawia się w związku z tym refleksja: przecież mamy obrazy hiperspektralne, w zakresie spektralnym znacznie wykraczającym poza pasmo widzialne, czy mimo to najlepsza jest jesień kiedy to roślinność bardzo się różnicuje, o czym oczywiście wszyscy wiemy? Czy badano krzywe spektralne różnych roślin i analizowano ich zróżnicowanie?

Rozdział 9 zawiera podsumowanie i wnioski.

Autorka stwierdza: „Na podstawie uzyskanych dokładności klasyfikacji możliwe jest stwierdzenie, że lotnicze zobrazowania hiperspektralne oraz metoda maszyn wektorów

nośnych są przydatne do klasyfikacji roślinności subalpejskiej i alpejskiej Karkonoszy”. Można by powiedzieć, że udowodniona została taka teza, pomimo że tezy w pracy nie postawiono. Tyle tylko, że nie potrzeba było wykonywać badań, żeby taka tezę postawić. Moim zdaniem należy we wnioskach unikać stwierdzeń ogólnych odnosząc się bezpośrednio do własnych wyników odnosząc je oczywiście do literatury.

„Dane poziomu 1, skorygowane geometrycznie oraz atmosferycznie, z wykorzystaniem szczegółowego Numerycznego Modelu Terenu, pozwoliły na pełniejsze i bardziej wierne odwzorowanie rzeczywistości niż dane nieskorygowane, co potwierdzają niezależne badania porównawcze wyników klasyfikacji uzyskanych na danych bez korekcji atmosferycznej, jak i na danych skorygowanych (wzrost dokładności całkowitej klasyfikacji z 70,01% do 78,62%)”. Gdzie znajduje się opis tych badań?

Na zakończenie Autorka formułuje 9 wniosków. Szczególnie cenny jest wniosek 3 i 4 wynikający bezpośrednio z badań Autorki. Wniosek 3 dotyczy wiarygodności dokładności klasyfikacji: „losowanie pikseli do treningu klasyfikatora oraz weryfikacji wyników, szczególnie jeśli było przeprowadzane iteracyjnie, pozwoliło na uniknięcie subiektywizmu wynikającego z jednorazowego doboru danych”. Należy to uznać za bardzo cenne.

Wniosek 4 definiuje właściwie proponowany algorytm: „optymalny dobór parametrów funkcji jądrowych metody SVM i analiza ich wpływu na dokładność klasyfikacji pozwoliła na osiągnięcie najbardziej satysfakcjonujących rezultatów. Ponadto zauważono, na jakim zestawie danych które funkcje sprawdzają się najlepiej. Testowanie czterech funkcji pozwoliło zauważyć najwyższe dokładności dla funkcji liniowej przy wszystkich oryginalnych kanałach APEX, natomiast dla danych przetransformowanych (PCA) najwyższe dokładności uzyskano przy użyciu funkcji radialnej”.

Podsumowując, recenzowana rozprawa doktorska jest ciekawa, napisana właściwie poprawnie, chociaż mam uwagi do metodyki. Mianowicie, całość metodyki powinna być opisana w jednym miejscu i należy wyraźnie oddzielać metodykę od wyników. W rozprawach naukowych powinna to być metodyka badawcza, a nie metodyka wykonanych prac.

Autorka wykazała się umiejętnością przeprowadzania studiów literaturowych, wykonywania badań, przetwarzaniem danych, wyciągania wniosków wraz z dyskusją wyników. Cenne są w pracy szczególnie właśnie rozdziały dyskusyjne. Przy czym należy stwierdzić, że bardzo dobrze zaprezentowano w pracy umiejętności analityczne, natomiast w niektórych miejscach brakuje większego podejścia syntetycznego. Praca stanowi interesujący przyczynek do badań dotyczących wykorzystania obrazów hiperspektralnych dla potrzeb kartowania roślinności na przykładzie roślinności subalpejskiej i alpejskiej.

4. Uwagi szczegółowe

Podczas czytania wstępu nasuwają się dwa pytania. Pierwsze dotyczy danych referencyjnych i stwierdzenia, że „kartowane były dominujące płaty poszczególnych zbiorowisk roślinnych z dokładnością do około 60-80 cm za pomocą odbiornika GPS”. Pytanie czemu przyjęto taką dokładność, czy jest to dokładność GPS w trybie autonomicznym, czy sumarycznie dokładność GPS i dokładność wydzielenia/generalizacji w terenie? I dalej stwierdzenie: „Ważny jest sposób przygotowania danych treningowych tak, aby wyniki były wiarygodne statystycznie.” Co to oznacza?

Strona 18, przypis 10 – kompozycja barw RGB - „złożenie trzech kanałów zarejestrowanych w różnych zakresach spektralnych oraz przyporządkowanie im filtrów barwnych: czerwonego (Red), zielonego (Green) i niebieskiego (Blue). Zrzutowanie na wspólny ekran pozwala na utworzenie obrazu barwnego”. Jest to anachroniczny opis i brak powołania na literaturę.

Strona 21 - „Parametryczną metodą przetwarzania obrazu jest także ortorektyfikacja (parametric orthorectification), prowadząca do przekształcenia obrazu z rzutu środkowego na rzut ortogonalny, przy użyciu informacji dotyczącej rzeźby terenu, pochodzącej z danych wysokościowych, np. Numerycznego Modelu Terenu (Schlaepfer, Richter 2002). Na podstawie wymienionych informacji rejestrowany jest każdy piksel obrazu, po czym

nadawany jest mu odpowiedni układ współrzędnych i określana jest wielkość piksela wynikowego, a także interpolacja wartości pikseli wynikowych na podstawie jasności pikseli obrazu nieprzekształconego (resampling).” - To jest nieściśle, ponieważ ortorektyfikację można wykonać metodą zarówno parametryczną jak i nieparametryczną.

Strona 22 - W przypadku korekcji danych hiperspektralnych bardziej dokładne i wiarygodne są metody bezwzględne, wykorzystujące informacje pozyskane w trakcie nalotu. Na jakiej podstawie sformułowano ten wniosek?

Strona 23- „wybór kanałów najbardziej informacyjnych” - słowo „informacyjny” wydaje się tutaj niezbyt trafne. I dalej na stronie 25: „poprzez wskazanie wypoziomowania się linii informacyjności”.

Strona 23 – powinno się unikać takich stwierdzeń jak : „zwiększanie liczby kanałów nie wpływa na dokładność klasyfikacji lub nawet powoduje jej spadek” bez powoływania się na literaturę (co Autorka ostatecznie zrobiła w dalszej części pracy).

Strona 26 – początek podrozdziału 1.2.3.3. Wybór kanałów spektralnych: „W literaturze opisanych jest wiele metod wyboru zmiennych (kanałów) do dalszych analiz. John i in. 1994 przedstawiają podział na dwa modele: filter – podstawowy, bazujący na selekcji danych przed treningiem na podstawie jak najlepszego rozróżnienia klas, niezależny od używanego klasyfikatora oraz wrapper – "opakowany", oparty na ocenie dokładności klasyfikacji na podstawie zestawu danych weryfikacyjnych oraz zależny od klasyfikatora.” jest zupełnie niejasny.

Strona 28 – Ryc. 8 – dziwny schemat, brak połączeń pomiędzy niektórymi elementami, czy one są niezależne?

Strona 35 – niejednorodne odwoływanie się do literatury: „A. Samuel (1959)”, podczas gdy normalnie Autorka cytując podaje tylko nazwisko i rok.

Strona 36 - „Uczenie bez nadzoru wykorzystywane jest w klasteryzacji (clustering) i redukcji wymiarów (dimensionality reduction).” - proces ten nie zawsze jest nienadzorowany.

Strona 36 - „Stosowana jest do klasyfikacji bądź regresji.”, analizy regresji?

Strona 37 - „rozszerzono działanie algorytmu na regresję (Vapnik 1995)”, o analizę regresji.

Strona 37 - „Pierwotnie algorytm SVM używany był do klasyfikacji danych binarnych. Zaproponowano jednak rozwiązania, dzięki którym możliwa była klasyfikacja danych ciągłych, złożonych z wielu klas”. Co Autorka miała na myśli?

Strona 68 - . „W celu przyspieszenia procesu klasyfikacji i wyboru danych weryfikacyjnych do badań terenowych wykorzystano trzy algorytmy kompresji danych”. Jakie to były algorytmy? Strona 77 – niepotrzebne powtórki o maskowaniu

5. Wniosek końcowy

W konkluzji wyrażam opinię, że recenzowana rozprawa doktorska **mgr Adrianny Marcinkowskiej-Ochyra: "Ocena przydatności obrazów hiperspektralnych APEX oraz maszyn wektorów nośnych (SVM) do klasyfikacji roślinności subalpejskiej i alpejskiej Karkonoszy "** ma charakter poznawczy i użyteczny, stanowi kompleksowe i oryginalne rozwiązanie problemu naukowego sformułowanego w temacie rozprawy, co potwierdza, że Autorka posiada pełne umiejętności samodzielnego wykonywania pracy naukowej. Analizowana rozprawa doktorska spełnia kryteria zawarte w ustawie z dnia 14 marca 2003 r. "O stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki", dla kandydata do stopnia doktora nauk o Ziemi, zatem przedkładam wniosek Radzie Wydziału Geografii i Studiów Regionalnych Uniwersytetu Warszawskiego o dopuszczenie jej do publicznej obrony.

Kraków, dnia 5 lutego 2017 r.

