UNIWERSYTET WARSZAWSKI

Wydział Geografii i Studiów Regionalnych

Adriana Marcinkowska-Ochtyra

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

AUTOREFERAT

Rozprawa doktorska przygotowana pod kierunkiem: dr. hab. Bogdana Zagajewskiego dr Anny Jarocińskiej

Warszawa 2017

Spis treści

1.	Wstęp	2			
1.	1. Cele i problemy badawcze pracy	3			
1.	2. Maszyny wektorów nośnych	3			
 Dotychczasowe badania z zakresu wykorzystania danych hiperspektralnych do klasyfikacji roślinności					
3.	Obszar i obiekt badań	5			
4.	. Metodyka badań				
5.	. Najważniejsze wyniki				
6.	Przydatność SVM i obrazów hiperspektralnych do klasyfikacji roślinności 1	3			
7.	Podsumowanie i wnioski 1	4			
Lite	Literatura15				

1. Wstęp

Ekosystemy górskie są ważnym wskaźnikiem zmian klimatycznych (Körner 1994). Roślinność górska zmienia się na niewielkich odległościach od lasów liściastych przez bory szpilkowe, zarośla kosodrzewiny, zbiorowiska murawowe do zbiorowisk o strukturze kępkowej ze znacznym udziałem mszaków i porostów (Kozłowska 2008). O występowaniu i rozmieszczeniu górskich zbiorowisk roślinnych decyduje wysokość nad poziomem morza, ekspozycja, spadki terenu, podłoże, a także stopień rozwoju i wilgotność gleb, długość sezonu wegetacyjnego czy zaleganie śniegu (Żołnierz, Wojtuń 2013, Piękoś-Mirkowa, Mirek 1996).

Tradycyjne kartowanie roślinności odbywa się najczęściej w terenie poprzez wykonywanie zdjęć fitosocjologicznych oraz na podstawie fotointerpretacji zdjęć lotniczych. Metody te są czasochłonne, wymagają nakładu środków oraz pracy. Badania i monitoring roślinności wysokogórskiej są utrudnione z powodu słabej dostępności terenu oraz krótkiego okresu wegetacyjnego. Teledetekcja hiperspektralna dostarcza danych oraz narzędzi stosowanych w analizach środowiska. Za pomocą sensorów lotniczych i satelitarnych można rejestrować widmo elektromagnetyczne, najbardziej przydatny do analiz roślin jest zakres 400-2500 nm (Jensen 1983). Charakterystyki spektralne odzwierciedlają stan fizjologiczny, budowę anatomiczną i morfologiczną gatunków, np. zawartość barwników fotosyntetycznie czynnych, wody, nutrientów, celulozy, czy stan struktur komórkowych. Dzięki mnogości kanałów spektralnych zarejestrowanych w wielu wąskich zakresach promieniowania, obrazy hiperspektralne przyczyniają się do wsparcia tradycyjnego kartowania roślinności przez półautomatyczne metody klasyfikacji zbiorowisk roślinnych (Zhang, Xie 2013).

Klasyfikacja w teledetekcji oznacza zautomatyzowany proces ilościowej analizy obrazu cyfrowego, polegający na przyporządkowaniu pikseli do określonych klas, na podstawie wartości liczbowych zapisanych w kanałach spektralnych obrazu (Richards, Jia 2005). Obiekty należące do tych samych klas mają zbliżone charakterystyki spektralne, tworząc klastry w przestrzeni spektralnej.

1.1. Cele i problemy badawcze pracy

W niniejszej pracy postawione zostały trzy cele:

- poznawczy, mający za zadanie opracowanie i przetestowanie algorytmu klasyfikacji 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.

Realizacja celów pracy możliwa jest dzięki zastosowaniu procedur związanych z przetwarzaniem i klasyfikacją obrazu, a także analiz poklasyfikacyjnych. Do problemów badawczych należy identyfikacja dominujących zbiorowisk roślinnych na terenie Karkonoszy (dokonana poprzez wybór powierzchni badawczych dobrze identyfikowalnych na obrazie) oraz optymalizacja procedur przetwarzania obrazu hiperspektralnego. Jako dane referencyjne do klasyfikacji wykorzystano mapę roślinności nieleśnej (Wojtuń, Żołnierz 2002) oraz badania terenowe z trzech kampanii (2012, 2013 i 2014), podczas których kartowane były płaty zbiorowisk roślinnych za pomocą odbiornika GPS. Na podstawie wyników badań terenowych opracowane zostały wzorce do klasyfikacji oraz weryfikacji uzyskanych wyników. Przygotowanie danych hiperspektralnych opiera się na wykonaniu korekcji danych (nadaniu georeferencji oraz usunięciu wpływu atmosfery). Ważnym elementem jest opracowanie algorytmu wyboru najlepszych jakościowo kanałów, przy użyciu metod redukcji przestrzeni spektralnej. Ostatnim z problemów badawczych jest wybór metody klasyfikacji dostosowanej do charakteru i skali zjawiska oraz procedur poklasyfikacyjnych, prowadzących do opracowania mapy oraz raportu dokładności względem rozpoznania terenowego.

1.2. Maszyny wektorów nośnych

Maszyny wektorów nośnych (*Support Vector Machines*, SVM) są metodą nieparametryczną pozwalającą na rozpoznawanie wzorców w sposób nadzorowany (Vapnik 1995). Do separacji klas wykorzystuje hiperpłaszczyznę, która je oddziela z odpowiednim

marginesem w przestrzeni wielowymiarowej. Dane znajdujące się najbliżej hiperpłaszczyzny nazywane są wektorami nośnymi. Głównym elementem transformacji przestrzeni jest wybór tzw. funkcji jądra, spośród czterech dostępnych: liniowej, wielomianowej, radialnej oraz sigmoidalnej. W przypadku każdej funkcji istnieje możliwość zdefiniowania parametru kary *C*, odpowiadającego za kontrolę kompromisu pomiędzy błędami w klasyfikacji a wymuszeniem marginesów. Funkcja radialna i sigmoidalna wymagają dodatkowo parametru gamma, świadczącego o szerokości funkcji Gaussa. Metoda SVM pozwala uzyskać dobre rezultaty, nawet dla zaszumionych kanałów, przy tym nie jest czasochłonna, mimo wykorzystania znacznej pojemności danych.

2. Dotychczasowe badania z zakresu wykorzystania danych hiperspektralnych do klasyfikacji roślinności

Najczęściej klasyfikowaną w literaturze roślinnością jest roślinność leśna, charakteryzująca się odmienną fizjonomią od roślinności nieleśnej, częściej również spotyka się opracowania dotyczące roślinności położonej na terenach nizinnych. Pierwsze klasyfikacje roślinności na danych hiperspektralnych przeprowadzane były przez badaczy amerykańskich na początku XXI w, przy użyciu pierwszych skanerów (AVIRIS¹, CASI²) i z wykorzystaniem różnych metod (Kokaly in. 2003, Jollineau, Howarth 2008). W Europie najcześciej klasyfikowano pokrycie terenu, obszary podmokłe, siedliska Natura 2000 (Chan, Paelinckx 2008, Delalieux i in. 2010). We wschodniej części Europy wykorzystanie danych hiperspektralnych do klasyfikacji roślinności dopiero zaczyna się rozwijać, w Polsce początkowo wykorzystywano dane ze skanera DAIS 7915³ do klasyfikacji pokrycia terenu (Hejmanowska, Głowienka 2004), typów roślinności (Sobczak i in. 2005), czy zbiorowisk roślinnych (Zagajewski 2010). Dane APEX pozyskiwane są od roku 2010 i należą do nowocześniejszych, w literaturze dotychczas opisywane były pod kątem przetworzeń wstępnych i niewiele przeprowadzono analiz tematycznych. W tym kontekście nalot hiperspektralny skanerem APEX nad obszarem Karkonoszy i wykorzystanie danych do klasyfikacji roślinności wydaje się być unikatowym.

Analizy porównawcze różnych metod klasyfikacji danych hiperspektralnych z metodą maszyn wektorów nośnych pozwalają ocenić tę metodę jako jedną z najlepszych. Uzyskiwane dokładności dla metod parametrycznych czy podpikselowych wykazują, iż dokładności dla metody SVM przeważnie są wyższe o kilka procent (Camps-Valls i in., 2004), w przypadku porównania ze sztucznymi sieciami neuronowymi bywają zbliżone, jednak autorzy wykazują,

¹ rejestracja promieniowania w 224 kanałach spektralnych z zakresu 410-2450 nm.

² rejestracja promieniowania w 288 kanałach spektralnych w zakresie 400-1000 nm.

³ rejestracja promieniowania w 79 kanałach spektralnych z zakresu 400-12600 nm, dane pozyskane w ramach eksperymentów HySens przez Niemiecką Agencję Kosmiczną.

że metoda SVM jest bardziej wydajna pod względem stabilności, prędkości, jak również możliwości prawidłowego rozróżnienia klas, przy nielicznym zestawie danych treningowych oraz dużej liczbie wykorzystanych kanałów (Melgani, Bruzzone 2004, Pal, Mather 2004).

3. Obszar i obiekt badań

Obszar badań obejmuje najwyższe partie polskich oraz czeskich Karkonoszy w granicach parków narodowych (ryc. 1), wyselekcjonowany na podstawie zasięgu linii zobrazowania APEX oraz Numerycznego Modelu Terenu (powyżej 1200 m n.p.m.).



Ryc. 1. Karkonoski Park Narodowy i Krkonošský Národní Park

Obiektem badań są zbiorowiska roślinne Karkonoszy, wybrane na podstawie *Mapy roślinności rzeczywistej ekosystemów nieleśnych Karkonoskiego Parku Narodowego* Wojtunia i Żołnierza w 2002 r. Czeska część parku nie posiada mapy zbiorowisk roślinnych, dlatego dokonano klasyfikacji na podstawie danych referencyjnych z polskiej części parku. Na podstawie oceny możliwości klasyfikowania i dostępnej rozdzielczości przestrzennej obrazu APEX, a także rozróżnienia spektralnego klas, spośród 48 wybrano 22 zbiorowiska roślinne, które poddano klasyfikacji: chinofilne porosty naskalne, chinofobne porosty naskalne, murawy halne (forma subalpejska), murawy halne (forma alpejska), zbiorowiska

źródlisk niewapiennych, zbiorowiska torfowisk przejściowych, zbiorowiska torfowisk wysokich, traworośla wysokogórskie, traworośle trzcinnika owłosionego, zarośla czeremchy i jarzębiny, zarośla wierzby lapońskiej, ziołorośle miłosny górskie, ziołorośle paprociowe, zbiorowisko trzęślicy modrej, zbiorowisko śmiałka darniowego, murawy bliźniczkowe, borówczysko bażynowe, borówczysko czernicowe, wrzosowisko wysokogórskie, sudeckie zarośla kosówki, świerczyny górnoreglowe, nitrofilne zbiorowiska ruderalne, ponadto sklasyfikowano stawy oraz obszary bez roślinności (Wojtuń, Żołnierz 2002).

4. Metodyka badań

Schemat postępowania przedstawiający kroki pracy składał się z dwóch równoległych elementów: pozyskania danych oraz przetwarzania obrazu APEX (ryc. 2). Pozyskanie danych bazowało na wyborze poligonów na podstawie mapy roślinności nieleśnej (Wojtuń, Żołnierz 2002), pozykaniu poligonów treningowych i weryfikacyjnych z badań terenowych (wrzesień 2012, sierpień 2013 i 2014) z użyciem odbiornika GPS Trimble GeoXT⁴ oraz wykorzystaniu ortofotomapy o rozdzielczości przestrzennej 12 cm. Dane ze skanera APEX pozyskano dla Karkonoszy 10 września 2012 r. w ramach projektu *HyMountEcos⁵* przez Niemiecką Agencję Kosmiczną (DLR) i belgijski instytut VITO. Dane zarajestrowano w 288 kanałach spektralnych z zakresu 380,5-2501,5 nm o rozdzielczości przestrzennej 3,12 m i 16-bitowej rozdzielczości radiometrycznej. Równocześnie z nalotem wykonano w terenie pomiary spektrometrem hiperspektralnym ASD FieldSpec 3⁶ w celu pozyskania cech spektralnych poligonów kalibracyjnych do korekcji obrazów. Następnie dane APEX skalibrowano, skorygowano atmosferycznie uzyskując dane prezentujące wartości współczynnika odbicia. Procedury te wykonano w VITO przez operatora skanera APEX.

Wizualna analiza kanałów APEX przyczyniła się do usunięcia kanałów związanych z absorpcją promieniowania przez parę wodną (1335-1422 i 1759-1954 nm) i pozostawieniu 252 kanałów do analiz. Ponieważ na badanym obszarze występują duże deniwelacje zdecydowano się na maskowanie cieni topograficznych, aby uniknąć błędu wystąpienia klasy zbliżonej spektralnie do obszaru zacienionego. Na przygotowanych danych rozpoczęto testowanie algorytmu klasyfikacji przy użyciu maszyn wektorów nośnych i różnych funkcji jądra na jednym z obszarów testowych (Mały Staw). Testowano różne parametry zmieniając ich wartości (parametr *C*, gamma), liczbę klas oraz pikseli treningowych. Modyfikowano wzorce poprzez losowanie z nich określonej liczby pikseli.

⁴ łącznie zgromadzono 812 zidentyfikowanych płatów roślinności.

⁵ Hyperspectral Remote Sensing for Mountain Ecosystem, dofinansowanie przez European Facility For Airborne Research Transnational Access (EUFAR TA).

⁶ rejestracja promieniowania w zakresie 350-2500 nm.



Ryc. 2. Schemat postępowania podczas klasyfikacji obrazów hiperspektralnych APEX

Wykorzystano trzy algorytmy kompresji danych, dwa bazujące na transformacji: *Principal Component Analysis* (PCA), *Minimum Noise Fraction* (MNF) oraz jeden bazujący na wyborze kanałów spektralnych: *backward elimination*. Dzięki algorytmom opartym na transformacji, po obliczeniu statystyk w każdym kanale obrazu, uzyskano najmniej skorelowane ze sobą kanały, świadczące o ich największej informacyjności. Na podstawie wyników transformacji PCA wybrano 40 kanałów, na podstawie MNF, uwzględniającej dekorelację szumów, 30 kanałów. Wyboru kanałów spektralnych dokonano przez ich odejmowanie, przeprowadzając 288 klasyfikacji obszaru, na którym wyznaczono poligony do klasyfikacji i weryfikacji. Podczas każdej klasyfikacji odejmowano jeden kanał, aby ocenić jego wpływ na dokładność klasyfikacji. Dzięki tej procedurze wybrano dwa zestawy kanałów o liczebności 18 oraz 70. Klasyfikację przeprowadzono także na 252 kanałach spektralnych.

Po optymalizacji procedury klasyfikacji finalne opracowanie wykonano na całej linii zobrazowania. Dla klas najbardziej licznych losowano po 200 pikseli treningowych, przy 400 weryfikacyjnych. Losowanie wykonywano jednorazowo, jak również 100-krotnie, dzięki czemu uzyskano 100 wyników dokładności.

Oceny ilościowej dokonano na podstawie macierzy błędów, wyrażającej, w jaki sposób zostały zaklasyfikowane piksele treningowe w odniesieniu do weryfikacyjnych (Congalton 1991). Na jej podstawie obliczono: a) dokładność całkowitą – liczbę pikseli sklasyfikowanych poprawnie podzieloną przez całkowitą liczbę pikseli w macierzy; b) dokładność użytkownika – liczbę pikseli poprawnie zaklasyfikowanych do danej klasy, podzieloną przez liczbę pikseli w klasie według klasyfikacji; c) dokładność producenta – liczbę pikseli zaklasyfikowanych do klasy poprawnie podzieloną przez łączną liczbę pikseli w klasie wzorcowej; d) współczynnik kappa – proporcjonalną redukcję błędów powstałych w wyniku klasyfikacji w stosunku do błędów klasyfikacji przeprowadzonej losowo.

5. Najważniejsze wyniki

Obszar testowy Małego Stawu poddano klasyfikacjom uwzględniając funkcje jądra SVM i zmienność ich parametrów, a także na danych wejściowych o różnej liczbie kanałów i liczebności pikseli. Najlepszą dokładność całkowitą osiągnięto dla funkcji liniowej (82,69%), drugą była funkcja radialna (72,82%), dla której dokładność była niższa o niemal 10%, a zbliżony wynik osiągnięto dla wielomianowej (72,77%). Najniższą dokładność odnotowano dla funkcji sigmoidalnej (59,69%). Analizę wpływu zredukowanych zestawów danych na dokładność klasyfikacji przeprowadzono przy użyciu funkcji liniowej (tab. 1).

Liczba kanałów	Dokładność całkowita (%)	Współczynnik kappa	Rozmiar pliku (MB)
252 kanały	82,69	0,83	465,0
40 kanałów PCA	81,04	0,79	64,6
30 kanałów MNF	80,76	0,79	48,4
70 kanałów	76,68	0,74	113,0
18 kanałów	68,14	0,65	30,7

Tab. 1. Dokładności klasyfikacji dla różnych zestawów kanałów APEX

Najwyższe dokładności osiągnięto dla oryginalnych 252 kanałów, następnie dla 40 kanałów PCA, najniższe dla 18 wybranych kanałów. Mimo różnic w dokładności, zauważono, że znaczna redukcja wielkości obrazu, jak w przypadku metody PCA spowodowała spadek dokładności o niecałe 2%, co pozwoliło stwierdzić dużą przydatność metod redukcji przestrzeni spektralnej, zwiększając operacyjność procesu klasyfikacji.

Analiza wrażliwości dokładności klasyfikacji wywołanej zmiennym parametrem kary *C* dla każdej funkcji SVM w przypadku kanałów oryginalnych i PCA pozwoliła zauważyć wzrost dokładności klasyfikacji wraz ze wzrostem wartości parametru kary *C*. Przy użyciu metody maszyn wektorów nośnych całej linii zobrazowania APEX pokrywającej najwyższe pasmo Karkonoszy uzyskano wyniki w postaci dokładności klasyfikacji (ryc. 3) oraz map roślinności subalpejskiej i alpejskiej (ryc. 4).

Najwyższą dokładność całkowitą klasyfikacji uzyskano dla 40 kanałów transformacji PCA oraz radialnej funkcji jądra SVM, wyniosła ona 85,5%, natomiast współczynnik kappa osiągnął wartość 0,84. Ogółem, dokładności dla większości klas były wysokie, 17 na 24 klasy i 20 na 24 klasy wyróżniono z dokładnością producenta i użytkownika wyższą od 80%.



Ryc. 3. Dokładności producenta i użytkownika dla klas, wyrażone w procentach
 (DP – dokładność producenta, DU – dokładność użytkownika)

Mapa roślinności subalpejskiej i alpejskiej Karkonoszy ARKUSZ 6 15°41E



.5 km

N.97×.09 Ryc. 4. Mapa roślinności subalpejskiej i alpejskiej Karkonoszy (wynik klasyfikacji 40 kanałów PCA, arkusz 6) 10

Najwyższe dokładności osiągnięto dla muraw halnych w formie alpejskiej, odróżniających się od pozostałych zbiorowisk rudawą barwą dzięki przebarwieniom, które dominujący sit skucina przyjmuje we wrześniu, oraz muraw bliźniczkowych obejmujących homogeniczne obszary zdominowane przez bliźniczkę psią trawkę, odróżniając się od zarośli kosówki czy borówczysk czernicowych wyższym odbiciem promieniowania w zakresie widzialnym. Wysokie dokładności uzyskano dla pozostałych rozległych zbiorowisk, dla których możliwe było zgromadzenie reprezentatywnej liczby pikseli treningowych i weryfikacyjnych (200/400), jak traworośla wysokogórskie, murawy halne w formie subalpejskiej, chinofilne porosty naskalne (powyżej 90% dokładności producenta). Ziołorośle paprociowe, mimo zaobserwowanego zjawiska zamierania i trudności w identyfikacji na obrazie APEX, dzięki inwentaryzacji terenowej osiągnęło dokładności około 80%.

Traworośla trzcinnika owłosionego, dla których dokładność producenta wyniosła 78%, przy dokładności użytkownika równej 70%, mylone były głównie ze zbiorowiskami śmiałka darniowego ze względu na podobieństwo spektralne, jak również z borówczyskami czernicowymi, przez występowanie w rozproszonych płatach. Dokładności dla borówczysk czernicowych były niższe (74% dokładności producenta i użytkownika) – głównie mylone z traworoślami, w mniejszym stopniu z ziołoroślami paprociowymi.

Świerczyny górnoreglowe, nienależące *de facto* do zbiorowisk nieleśnych, obejmujące obszary niewielkich skupisk drzew były mylone z sudeckimi zaroślami kosówki. Nitrofilne zbiorowiska ruderalne, nielicznie reprezentowane przez piksele treningowe (30) i weryfikacyjne (60) osiągnęły niższą dokładność producenta (52%), wynikającą ze zmienności tej klasy, gdyż roślinność antropogeniczna ulega dynamicznym zmianom i jej obecny zasięg nie pokrywał się całkowicie ze stanem z roku 2002. Niską dokładność producenta zaobserowano dla wrzosowisk wysokogórskich (51%, ryc. 5), były one najczęściej mylone z borówczyskami czernicowymi, ze względu na ich współwystępowanie w niewielkich płatach z ziołoroślami paprociowymi oraz murawami halnymi w formie subalpejskiej, zdominowanymi przez wrzos.



Ryc. 5. Wrzosowiska wysokogórskie współwystępujące z borówczyskami czernicowymi

Najgorzej sklasyfikowane były zarośla wierzby lapońskiej (dokładności poniżej 50%). Zostały one przeszacowane i w 21% znalazły się w miejscach występowania zarośli czeremchy i jarzębiny. Odnalezienie reprezentatywnych próbek dla tej klasy w terenie było trudne, gdyż często zarośla zajmują obszary nie większe niż 3 metry, a część z nich występuje na stromych, zacienionych stokach. Ponadto, klasa ta była reprezentowana przez mniej liczny zestaw pikseli treningowych (50) i weryfikacyjnych (100).

Aby przedstawić zróżnicowanie wyników dokładności poszczególnych klas, opracowano je w postaci wykresów pudełkowych, przedstawiających zakresy wartości przyjmowanych przez każdą z klas podczas 100-krotnego powtarzania algorytmu (ryc. 6).



SVM - Dokładność producenta

Ryc. 6. Dokładności producenta dla poszczególnych klas uzyskane podczas 100-krotnego losowania pikseli

Dokładności producenta dla większości klas (19/24) wykazały niewielkie zróżnicowanie (poniżej 15% różnicy między wartością najniższą a najwyższą), co świadczy o dobrze dobranych wzorcach do treningu i weryfikacji. Wartości mediany w 23 z 24 przypadków były wyższe od 50%. Największe zróżnicowanie w obrębie klasy reprezentowało ziołorośle miłosny górskiej, ponadto, duże zróżnicowanie zaobserwowano dla trzech klas: zarośli wierzby lapońskiej, wrzosowisk wysokogórskich oraz nitrofilnych zbiorowisk ruderalnych. Dokładności użytkownika przedstawiały wartości mediany dla klas utrzymujące się na wyższym poziomie niż producenta, dla 23 na 24 z klas wyniosły powyżej 70%. Większość klas reprezentowała wysoki poziom zgodności uzyskanych wyników, zróżnicowanie w obrębie poszczególnych zbiorowisk było niewielkie. Najszerszy rozkład zaobserwowano dla ziołorośli miłosny górskiej, zarośli wierzby lapońskiej, zbiorowisk źródlisk niewapiennych, wrzosowisk wysokogórskich i nitrofilnych zbiorowisk ruderalnych.

Analiza wpływu liczebności zestawu treningowego na dokładność klasyfikacji przeprowadzona dla obrazu z transformacji PCA (40 kanałów) wykazała wzrost dokładności całkowitej klasyfikacji wraz ze wzrostem liczebności zestawu treningowego.

6. Przydatność SVM i obrazów hiperspektralnych do klasyfikacji roślinności

Strefowość roślinności nie pozwala na bezpośrednie odniesienie wyników do literatury, gdyż zbiorowiska roślinne Karkonoszy są unikalne. Porównanie wyników klasyfikacji najbardziej zbliżonych fizjonomicznie – zbiorowisk tatrzańskich, przy użyciu sztucznych sieci neuronowych i obrazów DAIS7915 (Zagajewski 2010) wykazało dokładność całkowitą wyższą niż uzyskana w niniejszej pracy (89%), jednak dokładności poszczególnych klas były porównywalne. Np. murawy alpejskie typowe, wysokogórskie borówczyska bażynowe w Tatrach i odpowiadające im murawy halne w formie alpejskiej i borówczyska bażynowe klasyfikowane w niniejszej pracy osiągnęły dokładności producenta powyżej 90%.

W pracach dotyczących klasyfikacji główny nacisk kładziony jest na zastosowane metody oraz uzyskane wyniki, przeważnie stosowane jest jedno wybrane podejście do wyboru pikseli treningowych i weryfikacyjnych. Zestawienie różnych wyników uzyskanych dzięki wyborowi całych utworzonych poligonów, a także pikseli z nich wylosowanych na różne sposoby, pozwoliło na bardziej obiektywną ocenę wyników. Dzięki 100-krotnemu losowaniu pikseli można było ocenić trafność wylosowanych pikseli w przypadku zbiorowisk, dla których przyjmowane wartości były do siebie zbliżone.

Dzięki właściwościom spektralnym możliwe jest prawidłowe wyróżnianie większości zbiorowisk, jednak zbiorowiska znajdujące się w cieniach będą stanowiły utrudnienie dla klasyfikatora. Również zbiorowiska tworzące kompleksy są trudne do rozróżnienia.

Przedstawiony algorytm klasyfikacji danych APEX wykorzystano do klasyfikacji typów roślinności Karkonoskiego Parku Narodowego symulowanych danych satelitarnych EnMAP⁷ (Marcinkowska-Ochtyra i in. 2016). Na podstawie obrazu APEX dla Karkonoszy zasymulowany został obraz EnMAP, którego zakres spektralny jest zbliżony do zakresu APEX, jednak rozdzielczość przestrzenna jest niższa (30 m), przez to klasyfikacja niewielkich zbiorowisk roślinnych byłaby niemożliwa. Sklasyfikowano 8 typów roślinności z mapy roślinności nieleśnej Karkonoszy. Zaobserwowano podobieństwo do wyników klasyfikacji

⁷ *The Environmental Mapping and Analysis Program,* zakres spektralny 420-2450 nm, rozdzielczość przestrzenna 30 m, czas rewizyty 4 dni. W celu przyszłego rozwoju algorytmów niemiecki instytut GFZ generuje obrazy na podstawie wartości odbicia z innych danych hiperspektralnych.

zbiorowisk w niniejszej pracy: duże, homogeniczne klasy klasyfikowały się lepiej, np. zarośla kosodrzewiny, murawy. Większość klas zaklasyfikowała się na poziomie dokładności powyżej 70%. Najniższe dokładności dotyczyły klasy ziołorośli, składającej się z wrzosowisk oraz borówczysk, które były zróżnicowane i nie zajmowały 30-metrowego piksela EnMAP.

7. Podsumowanie i wnioski

Obrazy APEX, dzięki mnogości wąskich kanałów spektralnych oraz wysokiej rozdzielczości radiometrycznej i przestrzennej pozwoliły z wysoką dokładnością klasyfikować zbiorowiska roślinne. Większość klas sklasyfikowała się na poziomie powyżej 80% dokładności producenta (16/24), jak również powyżej 80% dokładności użytkownika (20/24). Poprawność wyników osiągniętą dla większości klas zapewniło użycie metody maszyn wektorów nośnych, dla której nawet niewielkie zestawy danych treningowych nie są ograniczeniem. Losowy dobór pikseli treningowych i weryfikacyjnych pozwolił na większe zobiektywizowanie wyników. Najwyższe dokładności osiągnięto dla homogenicznych zbiorowisk, niższe dla zbiorowisk niewielkich, o bardziej złożonej strukturze, bądź występujących na stromych, zacienionych stokach. Nie zawsze wzrost dokładności szedł w parze z większym zestawem treningowym, bardziej istotne dla dokładności były homogeniczność i właściwości zbiorowisk, przekładające się na charakterystyki spektralne.

Porównanie wyników klasyfikacji uzyskanych metodą maszyn wektorów nośnych oraz różnych funkcji jądra SVM pozwoliło potwierdzić uzyskane w pracy wnioski, iż najlepsze efekty uzyskano przy użyciu funkcji liniowej oraz radialnej, w zależności od zastosowanego zestawu danych wejściowych – oryginalnych lub zredukowanych spektralnie. Użycie danych hiperspektralnych czyni możliwym wydobywanie nawet niewielkich różnic między podobnymi spektralnie klasami. Z tego powodu dane składające się z wielu kanałów dają większe możliwości różnicowania klas, szczególnie jeśli przedmiotem badań jest mozaika gatunków tworząca zróżnicowane zbiorowisko.

Stosowanie metod redukcji przestrzeni spektralnej w literaturze jest popularne, dając różne rezultaty w zależności od danych i metod. Różni badacze podejmują próby sprawdzenia wpływu stosowania różnych liczebności pikseli treningowych na dokładności klasyfikacji i osiągają efekty podobne jak dla zbiorowisk Karkonoszy – wzrost dokładności spowodowany wzrostem liczebności zestawu treningowego.

Uzyskane wyniki potwierdzają założenia pracy. Przetworzone dane hiperspektralne APEX w połączeniu z odpowiednią metodą klasyfikacji pozwoliły na wyróżnienie wielu górskich zbiorowisk roślinnych, które są trudnym obiektem badawczym. Ponieważ poddany analizie obszar jest rozległy i zróżnicowany pod względem występowania zbiorowisk

roślinnych, należy uznać wyniki za satysfakcjonujące i przydatne do wykorzystania przez obydwa parki narodowe do wspierania monitoringu roślinności.

Literatura

- Camps-Valls G., Gomez-Chova L., Calpe-Maravilla J., Martin-Guerrero J. D., Soria-Olivas E., Alonso-Chorda L., Moreno J., 2004, Robust Support Vector Method for Hyperspectral Data Classification and Knowledge Discovery, *IEEE Transactions on Geoscience and Remote Sensing*, 42(7), 1530-1542.
- Chan J., Paelinckx D., 2008, Evaluation of Random Forest and Adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery, *Remote Sensing of Environment*, *112*(6), 2999-3011.
- Congalton R. G., 1991, A review of assessing the accuracy of classifications of remotely sensed data, *Remote Sensing of Environment*, 37, 35-46.
- Delalieux S., Somers B., Haest B., Kooistra L., Mücher C.A., Vanden Borre J., 2010, Monitoring heathland habitat status using hyperspectral image classification and unmixing, *Proceedings of the 2nd Whispers on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, IEEE GRSS, University of Iceland, Reykjawik, 50-54.
- Hejmanowska B., Głowienka E., 2004, Wstępne wyniki pomiarów spektrometrycznych i klasyfikacji obrazów hiperspektralnych rekultywowanego obszaru Tarnobrzeskiego Zagłębia Siarkowego, *Geoinformatica Polonica*, 6, 49-58.
- Jensen J. R., 1983, Biophysical remote sensing, Annals of the Association of American Geographers, 73(1), 111-132.
- Jollineau M. Y., Howarth P. J., 2008, Mapping an inland wetland complex using hyperspectral imagery, *International Journal of Remote Sensing*, 29(12), 3609-3631.
- Kokaly R. F., Despain D. G., Clark R. N., Livo K. E., 2003, Mapping vegetation in Yellowstone National Park using spectral feature analysis of AVIRIS data, *Remote Sensing of Environment*, 84, 437-456.
- Körner C., 1994, Impact of atmospheric changes on high mountain vegetation, *Mountain environments in changing climates*, Routledge, London, 155-166.
- Kozłowska A., 2008, Strefy przejścia między układami roślinnymi-analiza wieloskalowa (na przykładzie roslinności górskiej), Prace Geograficzne, 215, IGiPZ PAN.
- Marcinkowska-Ochtyra A., Zagajewski B., Ochtyra A., Jarocińska A., Wojtuń B., Rogass C., Mielke C., Lavender S., 2016, Subalpine and alpine vegetation classification based on hyperspectral APEX and simulated EnMAP images, *International Journal of Remote Sensing* (wersja online).
- Melgani F., Bruzzone L., 2004, Classification of hyperspectral remote sensing images with support vector machines, *IEEE Transactions on geoscience and remote sensing*, 42(8), 1778-1790.
- Pal M., Mather P. M., 2004, Assessment of the effectiveness of support vector machines for hyperspectral data, *Future Generation Computer Systems*, 20(7), 1215-1225.
- Piękoś-Mirkowa H., Mirek Z., 1996, Zbiorowiska roślinne [w:] *Przyroda Tatrzańskiego Parku Narodowego*, Mirek Z. (red.), Tatrzański Park Narodowy, Zakopane-Kraków, 237-272.
- Richards J. A., Jia X., 2005, Remote Sensing Digital Image Analysis, 3, Springer-Verlag, Berlin.
- Sobczak M., 2009, Hiperspektralna metoda badania i kartowania roślinności wysokogórskiej, *Teledetekcja środowiska*, 41, PTG KTŚ, 79-103.
- Vapnik V. N., 1995, The Nature of Statistical Learning Theory, New York, Springer-Verlag.
- Wojtuń B., Żołnierz L., 2002, Plan ochrony ekosystemów nieleśnych inwentaryzacja zbiorowisk [w:] Plan Ochrony Karkonoskiego Parku Narodowego. Biuro Urządzania Lasu i Geodezji Leśnej Oddział w Brzegu, 67 + 2 mapy.
- Zagajewski B., 2010, Ocena przydatności sieci neuronowych i danych hiperspektralnych do klasyfikacji roślinności Tatr Wysokich, *Teledetekcja środowiska*, 43, PTG KTŚ.
- Zhang C., Xie Z., 2013, Object-based vegetation mapping in the Kissimmee River watershed using HyMap data and machine learning techniques, *Wetlands*, *33*(2), 233-244.
- Żołnierz L., Wojtuń B., 2013, Roślinność subalpejska i alpejska [w:] Przyroda Karkonoskiego Parku Narodowego, Knapik R., Raj A. (red.), Karkonoski Park Narodowy, Jelenia Góra, 241-278.