

EDWIN RACZKO, BOGDAN ZAGAJEWSKI, ADRIAN OCHTYRA, ANNA JAROCIŃSKA,
ADRIANA MARCINKOWSKA-OCHTYRA, MAREK DOBROWOLSKI

Określenie składu gatunkowego lasów Góry Chojnik (Karkonoski Park Narodowy) z wykorzystaniem lotniczych danych hiperspektralnych APEX*

Identification of tree species in Mt Chojnik (Karkonoski National Park)
forest using airborne hyperspectral APEX data

ABSTRACT

Raczko E., Zagajewski B., Ochtyra A., Jarocińska A., Marcinkowska-Ochtyra A., Dobrowolski M. 2015. Określenie składu gatunkowego lasów Góry Chojnik (Karkonoski Park Narodowy) z wykorzystaniem lotniczych danych hiperspektralnych APEX. Sylwan 159 (7): 593-599.

We used hyperspectral data from APEX scanner (288 spectral bands in 380-2500 nm spectral range; 3,5 m spatial resolution) to classify five tree species occurring in the area of Mt. Chojnik in the Karkonoski National Park (south-western Poland). Data used to delimit learning and verification polygons were acquired during field research in August 2013, when ground truth polygons were acquired using device equipped with GPS receiver. Raw APEX data went through radiometric and geometric correction at VITO office. To reduce processing time, 40 most informative bands were selected using information content analysis. The Support Vector Machines (SVM) algorithm was used for classification of the following tree species: *Fagus sylvatica* L., *Betula pendula* Roth, *Pinus sylvestris* L., *Picea alba* L. Karst and *Larix decidua* Mill. Final classification had 78.66% overall accuracy with Kappa coefficient equal to 0.71. The best classified species included beech (87.09%) and pine (83.96%), while the worst results were obtained for larch (60.29%). Low accuracy for larch could be caused by the fact that most of larch trees in the research area grow in small patches, which made it hard to specify large enough sample of training data. All classified tree species had producer's accuracy of at least 60%, with the highest value reaching 87%. User's accuracies were from 53% for pine to 85% for beech. It is possible to classify tree species using hyperspectral data with moderate to high accuracy even if the data used lacked atmospheric correction. Further work will focus on improving the classification accuracy and use of neural networks based classification methods. Results from this paper will serve as basis for tree species map of the Karkonoski National Park.

KEY WORDS

SVM classification, APEX hyperspectral data, species structure

ADDRESSES

Edwin Raczko ⁽¹⁾ – e-mail: edwin.raczko@student.uw.edu.pl

Bogdan Zagajewski ⁽¹⁾ – e-mail: bogdan@uw.edu.pl

Adrian Ochtyra ^(1, 2) – e-mail: adrian.ochtyra@uw.edu.pl

Anna Jarocińska ⁽¹⁾ – e-mail: ajarocinska@uw.edu.pl

*Praca finansowana ze środków budżetowych na naukę w latach 2013-2017 jako projekt badawczy w ramach programu „Diamantowy Grant”.

Adriana Marcinkowska-Ochtyra ⁽¹⁾ – e-mail: adriana.marcinkowska@uw.edu.pl
 Marek Dobrowolski ⁽³⁾ – e-mail: marek.dobrowolski@kpnmab.pl

⁽¹⁾ Zakład Geoinformatyki, Kartografii i Teledetekcji, Uniwersytet Warszawski;
 ul. Krakowskie Przedmieście 30, 00-927 Warszawa

⁽²⁾ Kolegium Międzywydziałowych Indywidualnych Studiów Matematyczno-Przyrodniczych, Uniwersytet Warszawski; ul. Żwirki i Wigury 93, 02-089 Warszawa

⁽³⁾ Karkonoski Park Narodowy; ul. Chałubińskiego 23, 58-570 Jelenia Góra

Wstęp

Możliwość zdalnego określenia składu gatunkowego drzewostanu otwiera nowe możliwości w monitoringu środowiska oraz ocenianiu zmian na dużych obszarach leśnych. Dane hiperspektralne, ze względu na wąskie i liczne zakresy widma elektromagnetycznego, a co się z tym wiąże – ilość danych w nich zawartych, pozwalają na określenie z dużą dokładnością elementów ekosystemu, które są słabo rozróżnialne w zbiorze danych wielospektralnych. Obszar Karkonoskiego Parku Narodowego został pod koniec lat 80. XX wieku dotknięty klęską ekologiczną [Jadczyk 2009]. Znaczne skażenie środowiska wywołane kwaśnymi deszczami naraziło drzewostany na atak szkodników. Spowodowało to wielkoskalowe wylesienie obszarów, którego skutki można obserwować do dnia dzisiejszego. Ocena zniszczeń drzewostanu na tak dużym obszarze jest bardzo trudnym zadaniem, a monitorowanie zmian przy pomocy klasycznych metod może okazać się niewystarczające. Zdalne monitorowanie i klasyfikacja drzewostanu może być więc bardzo pomocne w zarządzaniu obszarami chronionymi. Pierwsze prace dotyczące klasyfikacji drzewostanu z wykorzystaniem danych hiperspektralnych powstały na początku XXI wieku w USA. Kokaly i in. [2003], wykorzystując dane ze skanera lotniczego AVIRIS, z powodzeniem sklasyfikowali zbiorowiska leśne sosny, świerka, jodły i osiki Parku Narodowego Yellowstone. Dokładność ogólna klasyfikacji 16 zbiorowisk leśnych i nieleśnych wynosiła 74,10%, przy współczynniku Kappa 0,62. W pracy Shena i in. [2010] wykorzystano dane AISA do wykonania klasyfikacji zbiorowisk dębu, sosny, cynamonowca japońskiego, bambusowca i kasztanowca o dokładności ogólnej ponad 85%. W trakcie badań stwierdzono przewagę klasyfikatora SVM nad innymi algorytmami w przypadku kartowania roślinności na różnych etapach cyklu fenologicznego.

Dane teledetekcyjne często wykorzystywane są w monitoringu obszarów chronionych i cennych przyrodniczo, do jakich bez wątplenia należy Karkonoski Park Narodowy. Prezentowana praca ma charakter pilotażowy i może być traktowana jako wstęp do skartowania drzewostanu całego obszaru Parku.

Materiał i metody

Obszar badań obejmował eksklawę Karkonoskiego Parku Narodowego – Górę Chojnik, zlokalizowaną na północ od głównego obszaru Parku. Jest to teren o bardzo zróżnicowanej rzeźbie, gęsto porośnięty drzewostanem. Obszar został wybrany ze względu na relatywnie dużą różnorodność gatunkową w porównaniu do innych części Parku o podobnym areale oraz ze względu na dostępność punktów pomiarowych. Ponadto, z uwagi na względnie małą powierzchnię terenu, klasyfikacje oraz przetwarzanie danych mogły odbyć się dużo szybciej. Na obszarze Góry Chojnik do klasyfikacji wybrano pięć gatunków: buk pospolity *Fagus sylvatica* L., brzozę brodawkowatą *Betula pendula* Roth, sosnę zwyczajną *Pinus sylvestris* L., świerk pospolity *Picea alba* L. Karst oraz modrzew europejski *Larix decidua* Mill.

W pracy wykorzystano dane pochodzące ze skanera hiperspektralnego APEX (Airborne Prism EXperiment). Jest to jedno z nowocześniejszych urządzeń tego typu. Skaner powstał we współpracy ESA, DLR, Uniwersytetu w Zurychu oraz firmy VITO, przy wsparciu środków Unii Europejskiej. Skaner charakteryzuje się wysoką rozdzielczością spektralną oraz terenową. Wykorzystane dane posiadały 288 kanałów spektralnych w zakresie 380-2501,5 nm. Rozdzielczość terenowa danych hiperspektralnych wynosiła 3,5 m, natomiast rozdzielczość spektralna od 0,6 do 11 nm. Dane pobrano podczas nalotu 10 września 2012 roku. W czasie kampanii lotniczej wykonano zobrazowania dla polskiego i czeskiego parku narodowego. Następnie dane zostały przetworzone przez firmę VITO, gdzie poddano je korekcji radiometrycznej oraz geometrycznej [Raczko 2014].

Uzyskane zobrazowanie hiperspektralne po wstępnych przetworzeniach przycięto do obszaru badań. W celu zgromadzenia danych o obszarze badań przeprowadzono pomiary terenowe, w trakcie których określono lokalizację dużych obszarów porośniętych jednym z pięciu wymienionych gatunków drzew. Prace te wykonano przy pomocy urządzenia wyposażonego w odbiornik GPS. Następnie, wykorzystując dane pozyskane podczas badań terenowych [Raczko 2014] uzupełnione o dane pozyskane z KPN, wyznaczono poligony wzorcowe i walidacyjne klasyfikacji. Jako materiału rozpoznawczego obszaru badań przed badaniami terenowymi użyto mapy drzewostanów Biura Urządzania Lasu i Geodezji Leśnej.

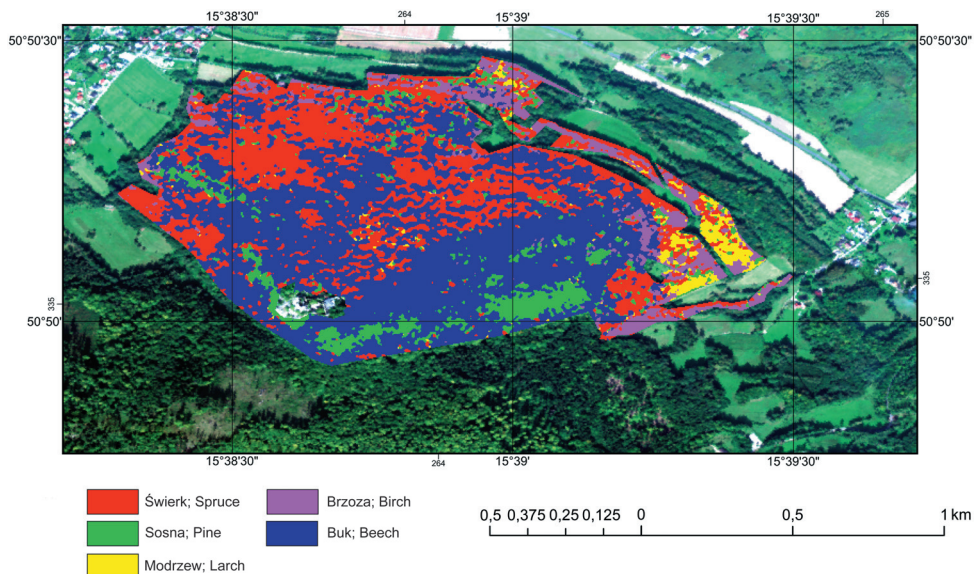
Do klasyfikacji wykorzystano algorytm SVM (Support Vector Machines). Jest to dobrze zbadany algorytm, osiągający bardzo dobre wyniki, a jednocześnie niewymagający bardzo dużych zasobów obliczeniowych czy czasowych [Bigdeli i in. 2013]. W pracy został użyty liniowy kernel SVM, ponieważ poprzednie badania dotyczące zbiorowisk nieleśnych na obszarze parku dowiodły, że daje on najlepsze wyniki [Marcinkowska i in. 2014]. Przetwarzanie obrazu zawierającego wszystkie kanały spektralne jest wymagające sprzętowo oraz powoduje znaczne wydłużenie czasu klasyfikacji (mimo wykorzystania SVM). Z tego powodu przeprowadzono analizę informacyjności obrazu [Raczko 2014]. Obraz źródłowy sklasyfikowano 288 razy, wykorzystując odrębny zestaw danych, za każdym razem wyłączając jeden kanał zobrazowania, aby określić jego wpływ na dokładność klasyfikacji. Obraz źródłowy posłużył jako materiał testowy. Na jego podstawie określono najbardziej informacyjne kanały w klasyfikacji. W wyniku analizy informacyjności otrzymano 40 najlepszych kanałów zobrazowania. Taka liczba kanałów jest najlepszym kompromisem między dokładnością klasyfikacji a czasem jej wykonywania [Pal, Mather 2006]. Następnie przeprowadzono klasyfikację obrazu pokrywającego obszar badań oraz wymaskowano na obrazie poklasyfikacyjnym obszary leżące poza granicą KPN lub niebędące lasem.

Końcowym etapem procedury klasyfikacji było określenie statystyk, czyli dokładności całkowitej, producenta, użytkownika oraz wskaźnika Kappa [Zagajewski, Olesiuk 2009]. Dokładność producenta to procent obszaru poprawnie zaklasyfikowanego w stosunku do całości obszaru wzorcowego, natomiast dokładność użytkownika określa, jaki procent pikseli sklasyfikowanych do danej klasy rzeczywiście do niej należy. W przypadku klasy brzoza dokładność użytkownika oznacza, że 82,72% pikseli sklasyfikowanych jako brzoza rzeczywiście należy do tej klasy, natomiast dokładność producenta mówi, że 73,63% pikseli wzorcowych dla tej klasy zostało poprawnie do niej przypisane.

Wyniki

W wyniku przeprowadzonych prac uzyskano klasyfikację gatunków drzewiastych obszaru badań (ryc. 1). Obszary nienależące do KPN zostały wymaskowane. Uzyskano dokładność całkowitą klasyfikacji wynoszącą 78,66% oraz otrzymano współczynnik Kappa klasyfikacji wynoszący 0,71

(tab.). Najlepiej sklasyfikowanymi gatunkami były świerk (83,96%) oraz buk (87,09%). Powyższe gatunki należą do drzewostanu dominującego na obszarze badań. Najgorzej sklasyfikowano modrzew (dokładność 60,29%). Niska dokładność jest spowodowana faktem, iż modrzewie na obszarze badań rosną sporadycznie, często w linii po kilka drzew, co sprawia, że wyznaczenie dużego poligonu treningowego jest trudne. W wyniku tego część pikseli, które w rzeczywistości powinny należeć do tej klasy, zaklasyfikowano do innych klas (13,24% do klasy świerk oraz 11,76% do klasy sosna). Również duże przemieszanie z innymi gatunkami (świerk i sosna) powoduje spadek dokładności klasyfikacji dla tej klasy. Klasę brzoza sklasyfikowano z dokładnością 73,63%, natomiast klasę sosna z dokładnością 68,57%. Warto zauważyć, że część pikseli,



Ryc. 1.

Mapa składu gatunkowego lasów porastających Górę Chojnik uzyskana w wyniku klasyfikacji danych hiperspektralnych z wykorzystaniem algorytmu SVM

Tree species composition of Mt. Chojnik based on the hyperspectral data classification using SVM algorithm

Tabela.

Macierz błędów [%] klasyfikacji z wykorzystaniem SVM dla obszaru Góra Chojnik

Confusion matrix [%] for SVM classification for Mt Chojnik area

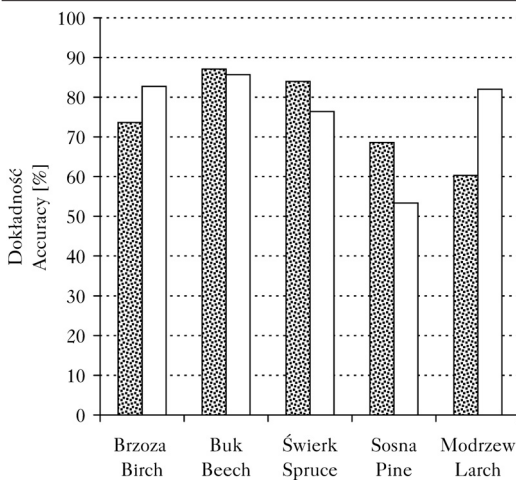
	Brzoza Birch	Buk Beech	Świerk Spruce	Sosna Pine	Modrzew Larch
Brzoza Birch	73,63	0,00	2,36	0,00	6,62
Buk Beech	6,59	87,09	4,25	25,71	8,09
Świerk Spruce	2,20	11,59	83,96	0,00	13,24
Sosna Pine	10,99	0,99	6,13	68,57	11,76
Modrzew Larch	6,59	0,33	3,30	5,71	60,29

które należą do klasy brzoza, zaklasyfikowano do klasy sosna. Tak duży błąd, jakim jest pomylenie drzew liściastych z iglastymi, można wyjaśnić brakiem korekcji atmosferycznej i topograficznej obrazów pokrywających obszar, co przy bardzo urozmaiconej rzeźbie terenu Góry Chojnik miało duży wpływ na odpowiedź spektralną obiektów na ziemi. Zadaniem korekcji topograficznej jest normalizacja wpływu ukształtowania terenu na odpowiedź spektralną z danego piksela, natomiast korekcja atmosferyczna usuwa wpływ warstwy atmosfery znajdującej się między detektorem a powierzchnią ziemi na sygnał pochodzący od obiektu badań. Wynik dla klasy sosna został spowodowany przez relatywnie niewielką liczbę dobrych homogenicznych poligonów sosny na obszarze badań, a także faktem, że rośnie ona na stromym stoku Chojnika, co powoduje, że jej spektralna odpowiedź dla tego obszaru jest mieszaniną odpowiedzi spektralnej dla skał oraz drzew. Może to wyjaśniać, dlaczego część pikseli należących w rzeczywistości do klasy sosna zostało niepoprawnie zaklasyfikowanych do klasy buk (25,71%).

Analiza dokładności użytkownika wykazała, że najlepszy wynik uzyskały klasy brzoza, buk i modrzew (ryc. 2). Ciekawy jest fakt uzyskania wysokich dokładności przez klasę modrzew, która pomimo dokładności producenta o wartości 60% uzyskała jedną z najwyższych dokładności użytkownika. Najmniejszą dokładność użytkownika uzyskano dla klasy sosna (53,33%), a największą dla klasy buk (85,67%). Współczynnik Kappa klasyfikacji wskazuje na stopień podobieństwa uzyskanych wyników do klasyfikacji powstałej w wyniku losowego przypisania klas do pikseli obrazu. Wynik na poziomie 0,71 wskazuje na małe podobieństwo, co oznacza, że uzyskanie podobnych wyników metodą losowania jest bardzo mało prawdopodobne.

Dyskusja

Ze względu na swoją przydatność dane teledetekcyjne są powszechnym sposobem uzyskiwania informacji na temat powierzchni ziemi. Informacja o przestrzennym położeniu poszczególnych gatunków drzew w lesie jest cenna nie tylko w celach monitoringu środowiska, ale również bardzo przydatna w zarządzaniu lasem. Rajee i in. [2014] wykorzystali dane z sensora Hyperion, charakteryzujące się niższą rozdzielczością przestrzenną (30 metrów/piksel) niż dane użyte w tej pracy, do klasyfikacji drzewostanów zachodnich Himalajów. Używając klasyfikatora SVM, sklasyfikowano dęby, sosny, cedry i jodły. Mimo stosunkowo niskiej rozdzielczości przestrzennej, znacznie gorszego stosunku sygnału do szumu oraz innych problemów z danymi Hyperiona ogólna dokładność klasyfikacji wynosiła 82%. Stwierdzono, że do klasyfikacji gatunków drzewiastych



Ryc. 2.

Dokładność producenta (szary) i użytkownika (biały) klasyfikacji SVM dla poszczególnych gatunków drzew

Producer's (grey) and user's (white) accuracy of SVM classification for individual species

potrzeba danych o dużej liczbie kanałów spektralnych. Feret i Asner [2013] sklasyfikowali roślinność Hawajów, wykorzystując sensor CAO (Carnegie Airborne Observatory Alpha Sensor). Dane z tego sensora charakteryzują się 24 kanałami spektralnymi w zakresie 390-1044 nm. Przy użyciu klasyfikatora SVM sklasyfikowano w pracy 17 gatunków drzewiastych z dokładnością ogólną sięgającą 83% dla 400 pikseli treningowych dla każdego sklasyfikowanego gatunku. Dowiedziono, że skupiając się na zakresach spektralnych, którymi najbardziej różnią się badane gatunki drzew, możliwe jest uzyskanie wysokiej dokładności klasyfikacji, mimo wykorzystania skanera o mniejszym zakresie spektralnym. Poza wykorzystaniem samych danych hiperspektralnych część badaczy łączy je z danymi pomocniczymi. Dalponte i in. [2012] wykorzystali dane hiperspektralne pochodzące ze skanera AISA Eagle oraz dane lidarowe do klasyfikacji gatunków drzewiastych południowych Alp. Dane AISA wykorzystane w tej pracy charakteryzowały się dość wąskim zakresem spektralnym w porównaniu do APEX, wynoszącym 402,9-989,1 nm oraz 126 kanałami spektralnymi. Dane lidarowe posłużyły do wykonania DTM (Digital Terrain Model) o rozdzielczości przestrzennej wynoszącej 1 metr. Następnie wykorzystano SVM do wyróżnienia 8 typów pokrycia terenu, spośród których 6 dotyczyło poszczególnych gatunków drzew, a pozostałe dwa zawierały w sobie obszary porośnięte gatunkami drzew nieuwzględnionymi w klasyfikacji oraz obszary nieleśne. Wykorzystując oba źródła danych, osiągnięto do 75,4% dokładności całkowitej w przypadku klasyfikacji wszystkich wyznaczonych klas. Dodatek danych lidarowych poprawiłby znacznie wyniki otrzymane w tej pracy. Połączenie danych APEX oraz lidarowych pozwoliłoby na włączenie do klasyfikacji danych dotyczących budowy korony, co z kolei umożliwiłoby poprawną identyfikację pojedynczych drzew, nawet tych, które występują sporadycznie na obszarze badań.

Innym przykładem wykorzystania danych hiperspektralnych może być klasyfikacja zbiorowisk roślinnych. Praca, której celem było sklasyfikowanie roślinności Tatrzańskiego Parku Narodowego przy wykorzystaniu danych hiperspektralnych ze skanera DAIS 7915, uzyskała dokładność dla dość ogólnej klasy bór dolnoregłowy w wysokości 81,7% dokładności producenta i 96,4% dokładności użytkownika [Zagajewski 2010]. Porównując wyniki tej pracy z niniejszymi badaniami, można stwierdzić, że dokładność całkowita na poziomie 78,66% jest zadowalająca przy braku korekacji atmosferycznej i topograficznej danych.

Również w innej pracy wykorzystującej dane ze skanera DAIS 7915 udało się sklasyfikować z dużą dokładnością klasy: las iglasty, las liściasty oraz las mieszany. Wymienione klasy wykazywały ponad 80% dokładności użytkownika i ponad 74% dokładności producenta w przypadku 13 kanałów po transformacji MNF oraz ponad 85% dokładności użytkownika i co najmniej 80% dokładności producenta dla klasyfikacji wykorzystującej 40 kanałów [Olesiuk 2008]. Wykorzystanie danych hiperspektralnych o tak szerokim zakresie nie jest szeroko rozpowszechnione, a część badaczy woli skupić się na wykorzystaniu sensorów o mniejszej liczbie kanałów spektralnych, które są ewentualnie uzupełniane o dodatkowe dane. Dane ze skanera APEX, ze względu na bardzo dobrą rozdzielczość przestrzenną oraz spektralną, są w stanie z dużą dokładnością klasyfikować gatunki drzewiaste nawet w tak trudnym terenie jak obszary górskie.

Podsumowanie

Klasyfikowanie gatunków drzewiastych przy wykorzystaniu danych hiperspektralnych z sensora APEX jest możliwe. Uzyskane wyniki mają dokładność na poziomie 78,66%. Dwa z pięciu gatunków sklasyfikowano z dokładnością ponad 83%. Biorąc pod uwagę brak korekacji atmosferycznej i topograficznej danych wykorzystanych w pracy, możliwe jest poprawienie dokładności klasyfikacji poprzez wykorzystanie danych po wszystkich przetworzeniach. Również wysoka

rozdzielczość przestrzenna danych (3,5 m) pozwala na wykonanie map gatunków drzewiastych o bardzo dużej dokładności. Opracowane wyniki oraz przyszłe prace posłużą do wykonania mapy drzewostanów całego obszaru KPN, a także do określenia stopnia kondycji roślinności po ponad 30 latach od klęski ekologicznej początku lat 80. W toku dalszych badań można spróbować wyróżnić obszary lasu w różnym wieku oraz określić ogólną kondycję roślinności KPN.

Literatura

- Bigdeli B., Samadzadegen F., Reinartz P. 2013. A Multiple SVM System for Classification of Hyperspectral Remote Sensing Data. *Journal of the Indian Society of Remote Sensing* 41 (4): 763-776.
- Dalponte M., Bruzzone L., Gianelle D. 2012. Tree species classification in the Southern Alps based on the fusion of very high geometrical resolution multispectral/hyperspectral images and LiDAR data. *Remote Sensing of Environment* 123: 258-270.
- Feret J. B., Asner G. P. 2013. Tree Species Discrimination in Tropical Forests Using Airborne Imaging Spectroscopy. *IEEE Transactions on Geoscience and Remote Sensing* 51 (1): 73-84.
- Jadczyk P. 2009. Natural Effects Of Large-Area Forest Decline In The Western Sudeten. *Environment Protection Engineering* 35 (1): 49-56.
- 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.
- Marcinkowska A., Zagajewski B., Ochtyra A., Jarocińska A., Raczek E., Kupkova L., Stych P., Meuleman K. 2014. Mapping vegetation communities of the Karkonosze National Park using APEX hyperspectral data and Support Vector Machines. *Miscellanea Geographica* 18 (2): 23-29. doi: 10.2478/mgrsd-2014-0007.
- Olesiuk D. 2008. Wykorzystanie obrazów hiperspektralnych do klasyfikacji pokrycia terenu zlewni Bystrzanki. Praca magisterska wykonana w Katedrze Geoinformatyki i Teledetekcji, Warszawa.
- Pal M., Mather P. M. 2006. Some issues in the classification of DAIS hyperspectral data. *International Journal of Remote Sensing* 27 (14): 2895-2916.
- Raczko E. 2014. Klasyfikacja gatunków drzewiastych Góry Chojnik (KPN) na podstawie lotniczych obrazów hiperspektralnych APEX i sztucznych sieci neuronowych. Praca magisterska wykonana w Zakładzie Geoinformatyki i Teledetekcji, Warszawa.
- Rajee G., Hitendra P., Kushwaha S. P. S. 2014. Forest tree species discrimination in western Himalaya using EO-1 Hyperion. *International Journal of Applied Earth Observation and Geoinformation* 28: 140-149.
- Shen G., Sakai K., Hoshino Y. 2010. High Spatial Resolution Hyperspectral Mapping for Forest Ecosystem at Tree Species Level. *Agricultural Information Research* 19: 71-78.
- Zagajewski B., Olesiuk D. 2009. SNNS classification of hyperspectral data of extensively used agricultural areas. *EARSeL eProceedings* 8 (2): 107-113.
- Zagajewski B. 2010. Ocena przydatności sieci neuronowych i danych hiperspektralnych do klasyfikacji roślinności Tatr Wysokich. *Teledetekcja Środowiska* 43: 113.