

СРАВНЕНИЕ НА КЛАСИФИКАЦИОННИ АЛГОРИТМИ ЗА ИДЕНТИФИЦИРАНЕ И КАРТОГРАФИРАНЕ НА ПОЛЕТА СЪС ЗИМНА ПШЕНИЦА ПО ДАННИ ОТ САТЕЛИТА RAPIDEYE

Илина Каменова

*Институт за космически изследвания и технологии – Българска академия на науките
e-mail: ilina.kamenova@space.bas.bg*

Ключови думи: класификационни алгоритми, метод на опорните вектори, метод на гора от дървета на решенията, метод на максималното правдоподобие, зимна пшеница

Резюме: Наблюдението на земеделските култури е от жизнено важно значение за ефективното производство на храни и устойчивото развитие на земеделския сектор. Основа цел на настоящото изследване е да се оценят възможностите за разпознаване на посеви от зимна пшеница в различни периоди на вегетационния сезон чрез класифициране на сателитни изображения от RapidEye и техни продукти по различни класификационни методи. В изследването са използвани класификационните алгоритми: Метод на опорните вектори (SVM), Метод на случайна гора на решенията (RF) и Метод на максималното правдоподобие (MXL). Тези методи са приложени върху сателитни многоканални изображения заснети от сателитите RapidEye, за тестови район в близост до квартал Требич, София. Заснети са 5 изображения за периода 31.03.2015 - 14.06.2015 г. Тези изображения са класифицирани поотделно и като разновременни изображения. Резултатите показват, че най-висока точност и за трите класификационни алгоритми се постига при разновременни изображения, включващи изображения от всички налични дати: SVM(91.4%); RF(91.14%); MXL(90.03%).

COMPARISON OF CLASSIFICATION METHODS FOR IDENTIFICATION AND MAPPING OF WINTER WHEAT FIELDS USING RAPIDEYE SATELLITE DATA

Ilina Kamenova

*Space research and technology institute – Bulgarian academy of sciences
e-mail: ilina.kamenova@space.bas.bg*

Keywords: classification methods, support vector machines, random forest, maximum likelihood method, winter wheat

Abstract: The monitoring of agricultural crops is of vital importance for efficient food production and the sustainable development of the agricultural sector. The primary objective of this study is to assess the possibilities of identifying winter wheat crops at different stages of the growing season through the classification of satellite images from RapidEye and their products using various classification methods. The study employed the following classification algorithms: Support Vector Machine (SVM), Random Forest (RF), and Maximum Likelihood (MXL). These methods were applied to satellite images acquired by RapidEye satellites in a test area near the Trebich district, Sofia. Five images were taken between March 31, 2015, and June 14, 2015. These images were classified both individually and as temporal stacks. The results indicate that the highest accuracy for all three classification algorithms is achieved when using temporal stack that include images from all available dates: SVM (91.4%); RF (91.14%); MXL (90.03%).

Въведение

В България, зимната пшеница е основна зърнено-житна култура и широко разпространено селскостопанско растение, с площ от около 1.2 милиона декара. Зимната пшеница се отглежда основно по конвенционални методи с използване на азотни торове, хербициди и други средства, за да се гарантират по-високи добиви. Поради значителното

въздействие върху околната среда, което земеделието и по-специално конвенционалното отглеждане на пшеница оказват, има нужда да се разработят ефективни методи за наблюдение на посевите от зимна пшеница [1]. В България информация за разпределението и обхвата на пшеничните посеви се извлича от „Интегрирана система за администриране и контрол (ИСАК)“. Тази система е основа за управлението на директните плащания за всички държави-членки на ЕС. Въпреки наличната информация от ИСАК, е важно да се въведат други независими източници на информация, особено в началото на вегетационния сезон на зимната пшеница, когато данните от ИСАК могат да бъдат непълни, тъй като все още не всички парцели са декларирани от фермерите. Данните за състоянието на пшеничните посеви в България на национално равнище не са своевременни, достатъчно надеждни и не обхващат големи площи. Важно е да се разработи метод, който може да генерира карти на пространственото разпределение на посевите от зимна пшеница. Тези карти могат да бъдат използвани за по-ефективната обработка на селскостопанските земи и съответно за намаляване на екологичният натиск от земеделските практики [2].

Множество изследвания подчертават ползите от използването на многоканални сателитни данни, включително за картографиране на земи, заети от селскостопански култури [3, 4]. Методите за дистанционно заснемане предоставят добра възможност за моделиране на биофизични параметри на културите с висока пространствена и времева резолюция. Настоящата статия разглежда приложението на сателитни изображения с фокус върху разпределение на посевите. Целта на изследването е да се идентифицира и картографира пространственото разпространение на посеви от зимна пшеница, базирано на многоканални сателитни изображения то сателита RapidEye.

За постигане на тази цел се налага решаването на следните научни задачи:

1. Оценка на възможностите за разпознаване на посеви от зимна пшеница през различни периоди на вегетационния сезон, чрез класификация на сателитни изображения от сателита Rapid Eye и техните продукти с използване на три метода за класификация: Метод на опорните вектори (от англ. support vector machines, SVM), метод на гора от дървета на решенията (от англ. random forest, RF) и метода на максималното правдоподобие (от англ. maximum likelihood classification, MXL).

2. Създаване на карта на полетата със зимна пшеница на базата на най-ефективния метод за класификация.

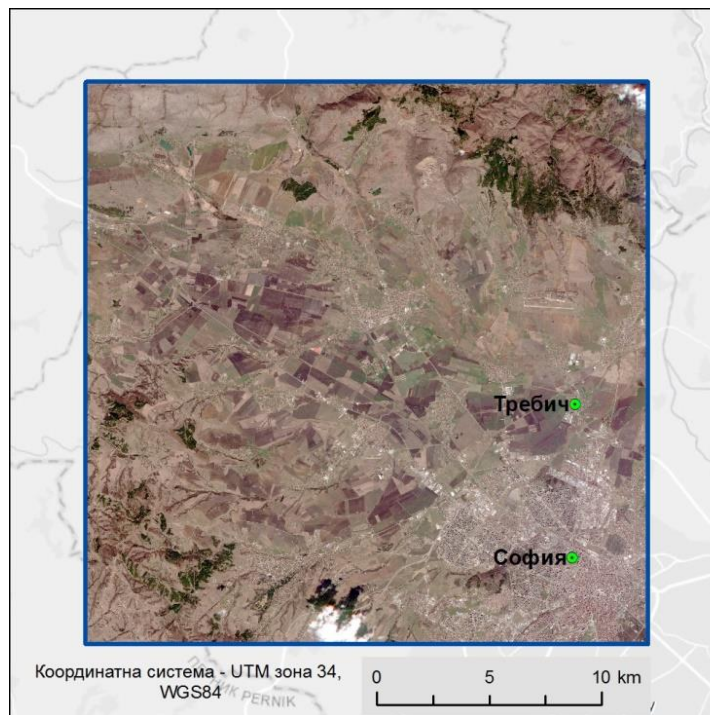
Материали и методи

Сателитни данни

Тестовият район „Требич“ (Фигура 1) обхваща част от Софийската котловина, заема части от град София и части от оградящите я планини. Този район е част от аерокосмически полигон на територията на България, на чиято територия са проведени редица подспътникови експерименти [5]. Тестовият район обхваща територия припокриваща се с една сцена от сателита RapidEye (със син цвят на Фигура 1). За изследванията в тестови район „Требич“ са ползвани сателитни данни от RapidEye. Съзвездие RapidEye се състои от 5 сателита, всеки от които носи на борда си инструмента JSS-56 (Jena-Optronik Spaceborne Scanner-56). Този сензор събира информация в 5 спектрални канала, с пространствена разделителна способност 5 m. Заснети са седем изображения за нуждите на изследването (Таблица 1).

Таблица 1. Източник на сателитни данни и дата на заснемане

Източник на сателитни данни	Дата на заснемане
RapidEye - 4	31 март 2015
RapidEye - 5	20 април 2015
RapidEye - 3	12 май 2015
RapidEye - 4	18 май 2015
RapidEye - 2	21 юни 2015



Фиг. 1. Тестови район "Требича"

Референтни данни

За обучение на класификационните алгоритми са необходими обучаващи множества. За нуждите на настоящото изследване обучаващи данни са набавени от „Интегрирана система за администриране и контрол (ИСАК)“ (Фигура 2). Той е генериран и подържан от Министерство на земеделието. ИСАК съдържа и информация за границите на земеделските парцели (обработваеми полета, пасища и трайни насаждения), придружена от информация за типа култура или земно покритие във всеки парцел, съгласно декларациите подадени от фермерите. Тази база данни се поддържа на национално ниво и е инструмент за контрол на субсидиите, в рамките на Общата селскостопанска политика (ОСП) на Европейския съюз.



Фиг. 2. Векторен слой от системата ИСАК с данни за земеделските парцели

Методи

Разпознаването на посевите и тяхното наблюдение е от голяма важност, особено в ранните фенологични фази на развитие на пшеницата, тъй като тогава най-ефективно и навременно могат да се вземат решения при земеделските практики и да се намалят загуби от

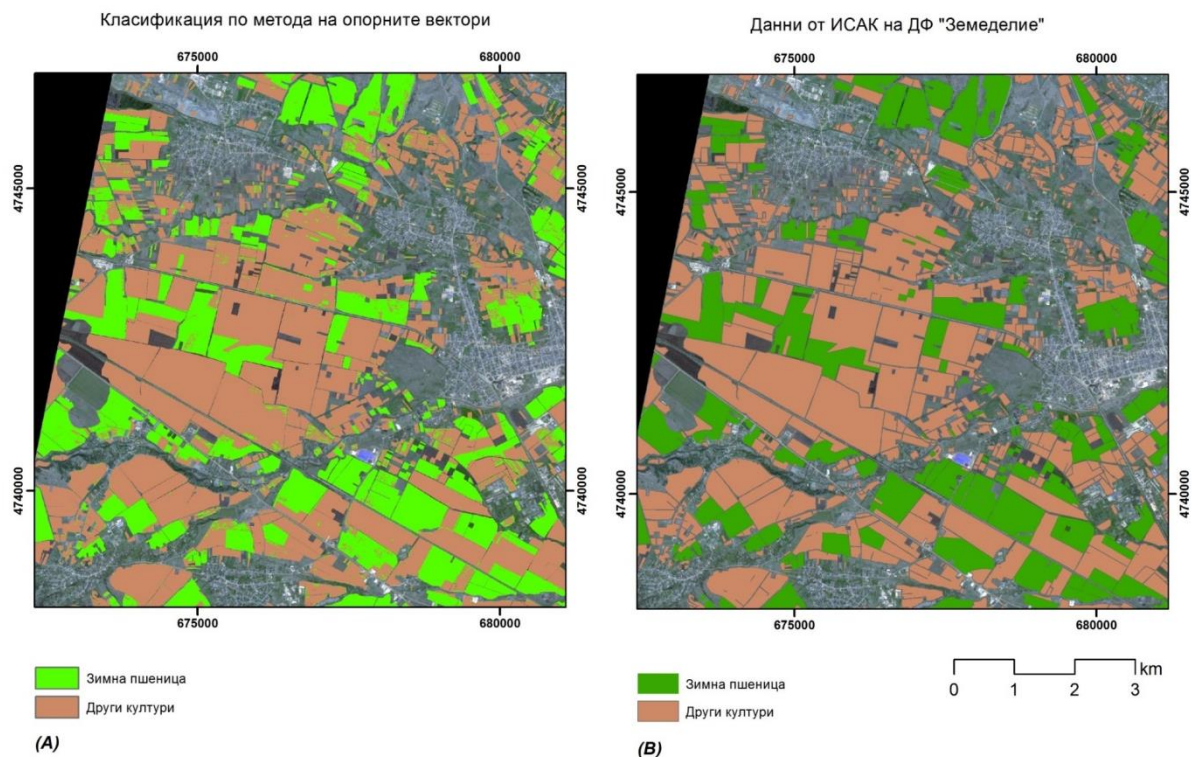
евентуални нарушения на посева. За да се оценят възможностите за разпознаване на посеви от зимна пшеница в различни периоди на вегетационният сезон, посредством класификации на сателитни изображения от RapidEye са приложени и оценени следните класифициращи алгоритми: метод на опорните вектори (SVM), метод на гора от дървета на решенията (RF) и метод на максималното правдоподобие (MXL). И трите класификационни метода са използвани за класифициране и картографиране на полетата със зимна пшеница и в двата района на изследване. За картографиране на площите заети със зимна пшеница в тестови район „Требич“ са използвани пет изображения от сателита RapidEye заснети през периода март - юни 2015 г. Съставени са и три разновременни изображения, включващи данните от няколко дати на заснемане (Таблица 2).

Резултати

Резултатите от класификациите са обобщени в Таблица 2. Постигнатите общи точности след валидация на класификациите имат обща тенденция за повишаване с напредване на сезона. Между трите класификационни алгоритма има системна разлика, с най-висока точност са класификациите по метода SVM, след това е RF и с най-ниска точност се представя MXL. Въпреки това, разликите не са драстични и не може да бъде изведен метод, който значително да превъзхожда останалите. Например, класификацията на изображението от 31.03.2015 г., постига обща точност от 79.9% за SVM, а изображението от 14.06.2015 г. има обща точност от 84.01% за SVM. При другите класификационни алгоритми, RF и MXL, резултатите показват същата тенденция. При разновременните изображения се наблюдава очакваната тенденция с добавянето на данни от повече дати точността да се повишава, като най-високата постигната точност бива тази за разновремененото изображение, включващо данни от 4 дати: 31.03.2015 г., 20.04.2015 г., 21.05.2015 г. и 14.06.2015 г. Постигнатата обща точност по метода SVM е от 91.4% и е най-висока измежду трите класификационни алгоритми. За сравнение, разновременното изображение включващо данни от 2 дати от началото на вегетационният сезон (31.03.2015 г. и 20.04.2015 г.) е класифицирано с обща точност от 82.3%.

Таблица 2. Обща точност, карра коефициент (*карра*) и F1 точност за класа “Зимна пшеница” (F1), за трите класификационни метода

Дати на заснемане	Метод на опорните вектори (SVM)			Метод на гора от дървета на решенията (RF)			Метод на максималното правдоподобие (MXL)		
	Обща точност, %	<i>карра</i> , %	F1, %	Обща точност, %	<i>карра</i> , %	F1, %	Обща точност, %	<i>карра</i> , %	F1, %
31.03.2015	79.9	54.59	68.95	78.25	52.07	67.63	71.46	39.38	59.8
20.04.2015	78.65	50.26	65.43	75.74	46.5	63.87	71.95	33.58	53.43
18.05.2015	82.65	61.53	73.95	83.73	63.05	74.66	78.93	53.7	68.78
21.05.2015	82.32	60.54	73.19	83.26	61.88	73.82	76.79	50.9	67.49
14.06.2015	84.01	63.43	74.83	82.67	60.78	73.16	82.86	60.91	73.14
31.03.2015 20.04.2015	82.3	59.93	72.57	81.21	58.01	71.45	77.03	48.09	64.5
31.03.2015 20.04.2015 21.05.2015	88.86	73.93	81.84	88.65	73.48	81.54	85.82	67.59	77.7
31.03.2015 20.04.2015 21.05.2015 14.06.2015	91.4	79.52	85.59	91.14	78.82	85.07	90.03	76.35	83.4



Фиг. 3. Извадка от карта на посевите със зимна пшеница в район „Требич“, класифицирани по метода на опорните вектори от разновремено изображение; (B) Извадка от карта на посевите със зимна пшеница в район „Требич“ по данни от "Интегрирана система за администриране и контрол" на ДФ "Земеделие" 2015

Заклучение

Трите тествани класификационни метода показват сходни точности. Макар че не може да бъде открит само един от тях се забелязва, че непараметричните методи RF и SVM като правило превъзхождат параметричния MXL. С оглед на този резултат, както и на факта, че методите RF и SVM на практика вече са широко въведени в областта на дистанционните изследвания може да се направи изводът, че тези методи следва да се предпочитат пред класическия MXL за решаване на задачи като разглежданата тук. Методът на максималното правдоподобие дава приемлив резултат в тестовия район, но предвид изискванията към разпределението на данните които поставя трябва да се прилага внимателно в други райони, където доминират култури и фенологията на посевите са различни от този тестови район. Използването на разновременни изображения подобрява резултатите от класификацията, като обикновено повишението на точността е около 10%. Това означава, че имайки предвид допълнителния обем данни, който трябва да бъде съхраняван и/или обработван, класификационният подход използващ разновременни изображения може не винаги да е оправдан. От друга страна недостатък на подхода за картографиране на посевите от зимна пшеница разчитащ на едно конкретно изображение е, че е необходимо да е налично подходящо изображение без облачна покривка над интересувания ни район. Когато този район е голям това не винаги е възможно. Може да се наложи облаците да бъдат маскирани (чрез автоматична процедура или ръчно) и част от пикселите да бъдат обозначени като райони с липса на данни, което означава, че информацията в картата ще е непълна. Друг проблем се появява, когато интересуваният ни район не се покрива изцяло от дадено изображение и е необходимо мозайкиране на две или повече изображения заснети в различни дати, за да бъде картографиран целият район. Резултатите получени в настоящото изследване могат да бъдат подобрени в бъдеще усъвършенствайки приложените методи. От особено значение е по-доброто разграничаване на зимната пшеница от останалите зимни култури, като зимната рапица. В литературата има примери в това отношение, основаващи се на използването на радарни данни [6], на разликите в отражението в червения и зеления канал между двете култури в определени фенофази [7] и др.

Благодарности

Авторът изказва благодарности на държавен фонд „Земеделие“ за предоставените данни от системата ИСАК и на Европейска Космическа Агенция (ЕКА), за безвъзмездното заснемане на изображенията с RapidEye. Благодаря на научния си ръководител доц. д-р Петър Димитров за менторството и подкрепата.

Литература:

1. Robert, P. C. Precision agriculture: A challenge for crop nutrition management. *Plant and Soil*, 2002, Vol. 247, Issue 1, pp. 143–149.
2. Samaniego, L., K. Schulz. Supervised Classification of Agricultural Land Cover Using a Modified k-NN Technique (MNN) and Landsat Remote Sensing Imagery. *Remote Sensing*, 2009, Vol. 1, pp. 875–895.
3. Jin, Z., G. Azzari, C. You, S. Di Tommaso, S. Aston, M. Burke, D.B. Lobell. Smallholder maize area and yield mapping at national scales with Google Earth Engine. *Remote Sensing of Environment*, 2019, Vol. 228, pp. 115–128.
4. Phiri, D., M. Simwanda, S. Salekin, V. R. Nyirenda, Y. Murayama, M. Ranagalage. Sentinel-2 data for land cover/use mapping: A review. *Remote Sensing*, 2020, Vol. 12(14), 2291.
5. Руменина, Е., Г. Желев. Развитие на дистанционните изследвания, провеждани на аерокосмическите полигони в България. *Proceedings of Sixteenth international scientific conference SES 2020*, pp. 187–206.
6. Dong, J., Y. Fu, J. Wang, H. Tian, S. Fu, Z. Niu, W. Han, Y. Zheng, J. Huang, W. Yuan. Early-season mapping of winter wheat in China based on Landsat and Sentinel images. *Earth System Science Data*, 2020, Vol. 12(4), pp. 3081–3095.
7. Yang, G., W. Yu, X. Yao, H. Zheng, Q. Cao, Y. Zhu, W. Cao, T. Cheng. AGTOC: A novel approach to winter wheat mapping by automatic generation of training samples and one-class classification on Google Earth Engine. *International Journal of Applied Earth Observation and Geoinformation*. 2021, Vol. 102, 102446.