

Application of Machine Learning Methods for Classification of Agricultural Crops

A.N. Badalova, S.H. Guliyeva

Azerbaijan National Academy of Aviation (Mardakan ave. 30, Baku, AZ1045, Azerbaijan)

For correspondence:

Guliyeva Sona / e-mail: guliyeva.s.h@gmail.com

Abstract

The article demonstrates the significant role of the impact of training samples on the processing of satellite images for the accurate classification of crops. To achieve the goal of the study, three machine learning methods are tested by classifying various satellite data with high spatial resolution. As a result of the application of machine learning methods, the results are compared and a method with a high classification accuracy is selected. Based on the obtained data, crop maps have been developed and an assessment of the vegetation state for the research period is carried out.

Keywords: satellite image, training sample, machine learning, classification, assessment, agricultural crop.

DOI: 10.52171/2076-0515_2022_14_02_106_116

Revised 10.01.2022

Accepted 17.06.2022

Received 22.06.2022

For citation:

Badalova A.N., Guliyeva S.H.

[Application of machine learning methods for classification and assessment of agricultural crops]

Herald of the Azerbaijan Engineering Academy, 2022, vol. 14, no. 2, pp. 106-116 (in Russian)

Kənd təsərrüfatı bitkilərinin təsnifatlandırılması və qiymətləndirilməsi üçün maşın öyrənmə üsullarının tətbiqi

A.N. Bədəlova, S.H. Quliyeva

Azərbaycan Milli Aviasiya Akademiyası (Mərdəkan pr. 30, Bakı, AZ1045, Azərbaycan)

Yazışma üçün:

Quliyeva Sona / e-mail: guliyeva.s.h@gmail.com

Xülasə

Məqalədə kənd təsərrüfatı bitkilərinin dəqiq təsnifatlandırılması üçün peyk təsvirlərinin emalına təlim nümunələrinin təsirinin mühüm rolu nümayiş etdirilmişdir. Tədqiqat məqsədinə nail olmaq üçün yüksək məkan ayırdetməli müxtəlif peyk məlumatlarının təsnifatlandırılmasında üç maşın öyrənmə metodu sınaqdan keçirilmişdir. Maşın öyrənmə üsullarının tətbiqi nəticəsində alınan məlumatlar müqayisə edilmiş və yüksək təsnifat dəqiqliyinə malik metod qeyd olunmuşdur. Alınan nəticələr əsasında kənd təsərrüfatı bitkilərinin növlərinin xəritələri hazırlanmış və tədqiq olunan dövr üçün vegetasiya vəziyyətinin qiymətləndirilməsi aparılmışdır.

Açar sözlər: peyk təsviri, təlim nümunəsi, maşın öyrənmə, təsnifatlandırma, qiymətləndirmə, kənd təsərrüfatı.

DOI: 10.52171/2076-0515_2022_14_02_106_116

УДК: 528.8

Применение методов машинного обучения для классификации сельскохозяйственных культур

А.Н. Бадалова, С.Г. Кулиева

Азербайджанская Национальная академия авиации (Мардакянский пр. 30, Баку, AZ1045, Азербайджан)

Для переписки:

Кулиева Сона / e-mail: guliyeva.s.h@gmail.com

Аннотация

В статье подчеркивается роль влияния обучаемых выборок на обработку спутниковых изображений для точной классификации сельскохозяйственных культур. Для достижения цели исследования было апробировано три метода машинного обучения при классификации различных спутниковых данных с высоким пространственным разрешением. В результате применения методов машинного обучения проведено сравнение результатов и выбран метод с высокой точностью классификации. По полученным данным разработаны карты типов сельскохозяйственных культур и проведена оценка вегетационного состояния за исследуемый период времени.

Ключевые слова: спутниковое изображение, обучаемая выборка, машинное обучение, классификация, оценка, сельскохозяйственная культура.

Введение

Как и в других сферах человеческой деятельности, в сельском хозяйстве ведется постоянный поиск новых технологий, которые позволяют улучшить процесс производства продукции и рационального использования природных ресурсов. Развитие компьютерной индустрии, появление на рынке точной сельскохозяйственной техники с приемниками глобальной системы позиционирования (GPS), всевозможными датчиками и бортовыми компьютерами с соответствующим программным обеспечением, а также возможностью использования географической информационной системы (ГИС) для обработки и визуализации пространственно-атрибутивных данных привели к появлению новой концепции, называемой *точным земледелием* [1].

Для решения задач мониторинга сельскохозяйственных полей большое значение имеет выбор оптимального метода классификации спутниковых снимков и идентификации сельхозкультур. За последние десятилетия были разработаны различные методы для контролируемой и неконтролируемой классификации спутниковых изображений [2].

Обучающие выборки являются одним из важных элементов контролируемой классификации, поскольку в зависимости от характеристик данных обучения и их адаптации с алгоритмом обучения они могут повлиять на ожидаемые результаты как отрицательно, так и положительно [3].

Основной **целью** исследования является определение влияния обучаемых выборок на классификацию спутниковых изображений высокого разрешения для идентификации типов сельскохозяйственных культур.

Постановка задачи

Для достижения цели исследования были поставлены следующие задачи: применение методов машинного обучения для классификации спутниковых изображений; изучение влияния объема данных на точность классификации с использованием разного количества классов и обучающих выборок; сравнение результатов разных снимков высокого разрешения в одинаковых условиях; определение точности классификаций спутниковых изображений; оценка состояния сельскохозяйственных культур в течение вегетационного периода.

Решение задачи

В исследовании были использованы спутниковые изображения Azersky и PlanetScore, карта посевов и земельного покрова, наземные данные и веб-карты. Территория исследования, занимающая площадь 119 км², представлена на рис. 1. Село Гозлубулаг расположено в предгорной местности юго-восточной части Шекинского района. На исследуемой территории встречаются такие объекты, как реки, поселки, поля сельскохозяйственных культур, луга, леса, горы и т. д.



Рисунок 1 - Территория исследования (село Гозлубулаг, Шекинский район)

Figure 1 - Study area (Gozlubulag district, Sheki region)

Спутниковые изображения были получены со спутника Azersky по следующим датам: 2 мая, 22 мая, 3 июня и 23 июня 2019 года. Для сравнения результатов классификации с различных спутников были выбраны мультиспектральные изображения с пространственным разрешением 6 м и 3 м соответственно для Azersky и PlanetScope. Даты и время получения снимков соответствуют для Azersky 2 мая 2019 г., 07:21:49,5 и PlanetScope 2 мая 2019 г., 06:15:37 соответственно.

Панхроматические изображения Azersky с пространственным разрешением 1.5 м использовались в качестве вспомогательных только для визуального анализа, идентификации объектов и участков посевных площадей, расположенных на исследуемой территории.

Основные обучающие выборки для дальнейших процедур были извлечены из карты посевов и земельного покрова, представленной на рис. 2.

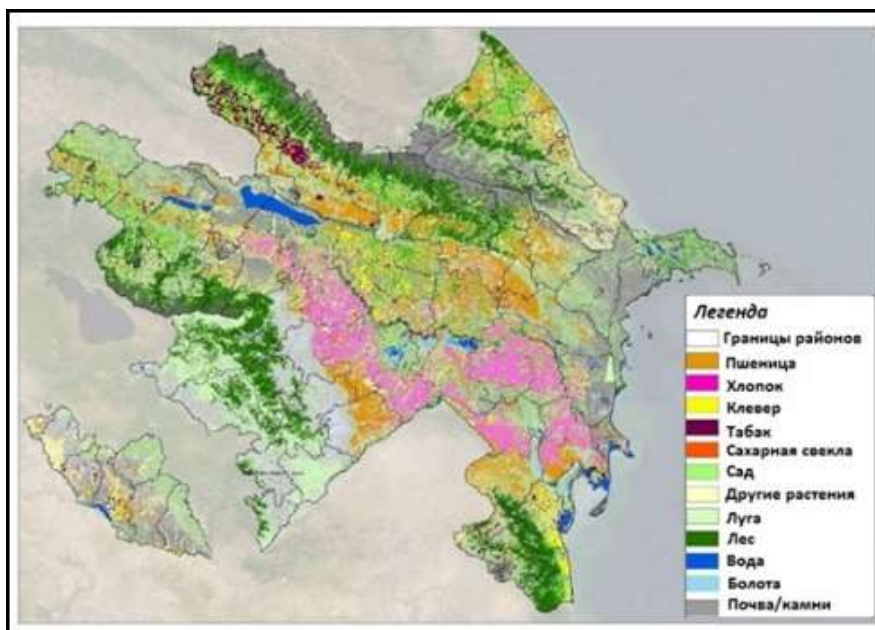


Рисунок 2 – Карта сельскохозяйственных культур и земного покрова Азербайджана (Азеркосмос, 2019)
Figure 2 – Crop and land cover map of Azerbaijan (Azercosmos, 2019)

На карте сельскохозяйственных культур и земного покрова имеется информация о различных категориях, таких как инфраструктура, пшеница, хлопок, табак и др.

Дополнительные обучающие выборки были созданы на основе веб-карт и наземных данных. Координаты полей с определенными типами культур были сняты с помощью прибора GPS E-Trex 10.

Методология исследования представляет собой эффективный метод классифи-

кации спутниковых снимков с полубучением, не требующий наличия обучающих выборок для всех классов в связи с основной целью исследования.

Методология исследования включает в себя следующие 5 основных этапов:

1. получение и предварительная обработка изображений;
2. извлечение необходимых классов из карты сельскохозяйственных культур и земельного покрова;

3. создание обучающих выборок и классификация изображений;
4. оценка точности и сравнение результатов классификации различных спутниковых изображений;
5. верификация результатов и мониторинг тестовых сельскохозяйственных полей.

Классификация спутниковых изображений рассматривается как комплексный подход на основе применения трех методов классификации машинного обучения: максимальное правдоподобие (ML), машина опорных векторов (SVM) и случайный лес (RF).

Перед классификацией изображений была выполнена атмосферная коррекция, чтобы улучшить качество изображения и преобразовать значения изображений из излучения в отражение. Отражение, измеренное в верхней части атмосферы, чрезвычайно полезно, потому что это сопоставимое число с любого спутника в любом месте, которое не меняется в зависимости от времени суток или времени года, пока не изменится содержимое территории. При вычислении спектральных индексов в ГИС часто используются такие индексы, как NDVI или NDWI [4]. После этапа преобразования изображения были обрезаны по интересующим областям (AOI) и преобразованы в единую систему координат.

Классификация может быть определена из множества источников, начиная от фактических наземных исследований мест отбора проб, использования аэрофотосъемки или спутниковых изображений [5].

Карта сельскохозяйственных культур и земельного покрова для исследуемой территории была извлечена из карты для всей территории Азербайджана. Данная

карта была преобразована из растрового в векторный формат. Из 10 представленных классов 6 классов были исключены, а 4 класса, такие как пшеница, луга, лес и почва / камни, использовались для дальнейшего анализа. К 4 классам, полученным из карты, были также добавлены 2 класса: вода и ячмень. Класс воды был определен визуально и подтвержден Open Street Map – картой, которая показана в виде слоя карты ESRI на рис. 3.

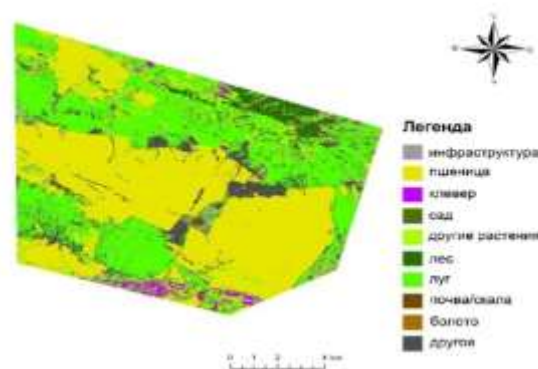


Рисунок 3 – Карта сельскохозяйственных культур и земельного покрова исследуемой территории
Figure 3 – Crop and land cover map of the study area

Обучающие выборки для сорта ячмень были получены на основании наземных данных, полученных с помощью GPS. Найденный класс воды имеет важное значение, так как для обеспечения урожайности сельскохозяйственных культур в период их вегетации имеется острый дефицит воды [6].

Применение обучающих выборок осуществлялось тремя разными способами.

В первом эксперименте обучающие выборки были сгенерированы случайным образом на основе 6 классов без соблюдения фиксированного числа объектов. Для каждого класса было примерно создано следующее количество образцов, показанное в табл. 1.

Таблица 1 - Обучающие выборки для 1-го эксперимента

Table 1 - Training samples for the 1st experiment

Класс	Обучающие выборки
Пшеница	87
Ячмень	8
Луг	42
Почва, скала	20
Лес	15
Вода	21

Таблица 2 - Обучающие выборки для 2-го эксперимента

Table 2 - Training samples for the 2nd experiment

Класс	Обучающие выборки
Пшеница	50
Луг	30
Почва, скала	25
Лес	20
Ячмень	10
Вода	20

Таблица 3 - Обучающие выборки для 3-го эксперимента

Table 3 - Training samples for the 3rd experiment

Класс	Обучающие выборки
Пшеница	50
Луг	30
Почва, скала	25
Лес	20

Таблица 4 - Точность классификации изображений Azersky во время 1-го эксперимента

Table 4 - Overall accuracy of classified Azersky images during the 1st experiment

	2 мая %	22 мая %	3 июня %	23 июня %
ML	67	41	58	54
SVM	72	61	66	67
RF	75	66	68	68

Во втором эксперименте два класса были добавлены к классам, полученным из

карты и классам, по типам сельскохозяйственных культур и идентифицированным водным ресурсам (табл. 2).

В третьем эксперименте использовались обучающие выборки, полученные только из карты, и было создано определенное количество выборок для каждого класса (табл. 3).

Проведена классификация всех четырех изображений Azersky и оценена точность классификации (табл. 4).

Наилучший результат с высоким значением точности классификации был выбран из классифицированных изображений Azersky для сравнения этого результата с результатом со снимка PlanetScope.

Анализ для проведения различных экспериментов по реализации обучающих выборок проводился на основе изображения Azersky, полученного 2 мая 2019 года, по этой дате также было выбрано изображение, полученное со спутника PlanetScope. Точно такая же операция была произведена для изображения PlanetScope, полученного в тот же день, что и изображение Azersky. По результатам оценки точности классификации была составлена таблица 5.

Таблица 5 – Точность классификации изображений во время 1-го эксперимента

Table 5 – Overall accuracy of classified images during the 1st experiment

	Azersky, %	PlanetScope, %
ML	67	65
SVM	72	70
RF	75	72

По высокому значению точности была составлена карта (рис. 4).

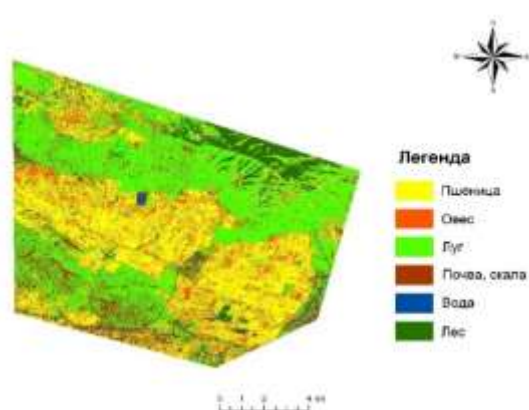


Рисунок 4 – Карта сельскохозяйственных культур и земного покрова, разработанная в ходе 1-го эксперимента

Figure 4 - Crop and land cover map developed during the 1st experiment

Оценка точности классифиции изображений в ходе второго эксперимента показана в таблице 6. Высокий результат был получены с помощью метода RF и соответствуют 73% для изображения Azersky.

Таблица 6 – Точность классификации изображений во время 2-го эксперимента

Table 6 – Overall accuracy of classified images during the 2nd experiment

	Azersky,%	PlanetScope,%
ML	65	57
SVM	72	60
RF	73	65

В конце второго эксперимента был выбран также результат с высоким значением точности классификации и на основании их получена карта, представленная на рис. 5.

Для третьего эксперимента оценки точности классифиции изображений приведены в таблице 7.

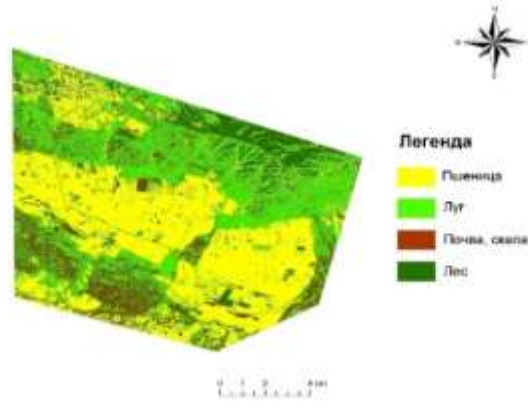


Рисунок 5 – Карта сельскохозяйственных культур и земного покрова, разработанная в ходе 2-го эксперимента

Figure 5 – Crop and land cover map developed during the 2nd experiment

Таблица 7 – Точность классификации изображений во время 3-го эксперимента

Table 7 – Overall accuracy of classified images during the 3rd experiment

	Azersky,%	PlanetScope,%
ML	84	77
SVM	83	76
RF	85	79

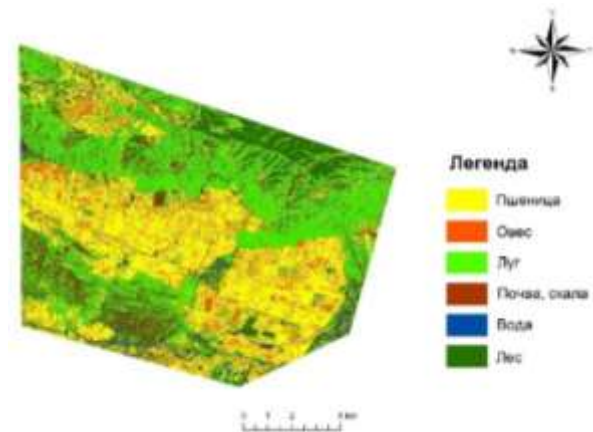


Рисунок 6 – Карта сельскохозяйственных культур и земного покрова, разработанная в ходе 3-го эксперимента.

Figure 6 – Crop and land cover map developed during the 3rd experiment

В третьем эксперименте наивысшая точность классификации изображений Azersky и PlanetScope, соответствующая

85% и 79%, была получена методом RF. В результате получена карта, соответствующая наивысшей точности по третьему эксперименту (рис. 6).

В ходе этого исследования было обнаружено, что при обучении изображений в случае применения наименьшего количества классов было получено высокое значение точности классификации.

Несмотря на то, что коэффициент каппа широко распространен и используется в качестве меры тематической точности в дистанционном зондировании, в данном исследовании как основной метод оценки точности классификации была использована общая точность, так как коэффициент каппа является неподходящим показателем для описания точности классификации. С другой стороны, общая точность, хотя и некорректна, но имеет четкое значение и по сравнению с каппа ее легче получить [7].

Таким образом, наиболее высокая оценка точности классифицированных изображений была обнаружена с помощью метода классификации случайного леса (RF). Разница между этими результатами для спутниковых изображений Azersky и PlanetScore во время первого эксперимента составила 3%, второго – 8% и 6% в третьем эксперименте.

Несмотря на то, что существует прямая связь, которая была установлена многими исследователями между высоким пространственным разрешением и высокой точностью классификации, результат этого исследования показывает немного более низкую точность для изображения PlanetScore, которое имеет вдвое большее пространственное разрешение, чем изображение Azersky.

Комбинированное использование методов классификации для мультиспектральных изображений и стандартных методов интерпретации является очевидным способом повышения надежности мониторинга и классификации сельскохозяйственных культур различных типов в соответствии с их вегетационными периодами.

Далее, для верификации полученных результатов классификации были рассчитаны индексы растительности, а также для помощи в классификации при разграничении растительности или других классов почвенного покрова, а также было исследовано состояние сельскохозяйственных полей на основе изображений Azersky.

Различия в характере реакции растительности в диапазонах спектра позволяют рассчитать индекс (NDVI) [8], который является одним из наиболее широко используемых индексов, в основном в исследованиях растительности, так как является показателем здоровья растения и рассчитывается по тому, как растение отражает и поглощает различные световые волны.

Вегетационный индекс был рассчитан для всех доступных изображений Azersky и на основании индексных изображений была проведена оценка состояния вегетации для исследуемой территории (рис. 7).

На этой карте A, B, C и D указывают даты снимков Azersky 2 мая, 22 мая, 3 июня и 23 июня 2019-го года соответственно.

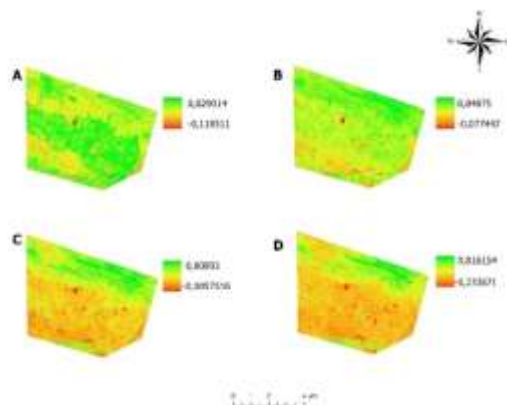


Рисунок 7 – Карты NDVI, разработанные в вегетационный период

Figure 7 – NDVI maps developed in the growing period

Поскольку на этих картах определены отрицательные значения NDVI (значения, приближающиеся к -1), соответствующие воде, это подтверждает существующие водные объекты в исследуемой области, которые не были отмечены на исходных данных. Далее для достижения цели точного земледелия были исследованы два испытательных поля. На основе оценки состояния этих полей были обнаружены области с наименьшими и наивысшими значениями NDVI, которые и показывают состояние растительности на исследуемый период времени. Тестовые сельскохозяйственные поля имеют следующие характеристики, указанные в табл. 8. На рис. 8 показаны выбранные сельскохозяйственные поля для дальнейшего анализа и оценки состояния сельскохозяйственных культур.

Таблица 8 – Основные характеристики тестовых полей

Table 8 – The main characteristics of testing fields

	Площадь, га	Тип культуры
Поле 1	6,52	пшеница
Поле 2	4,50	пшеница

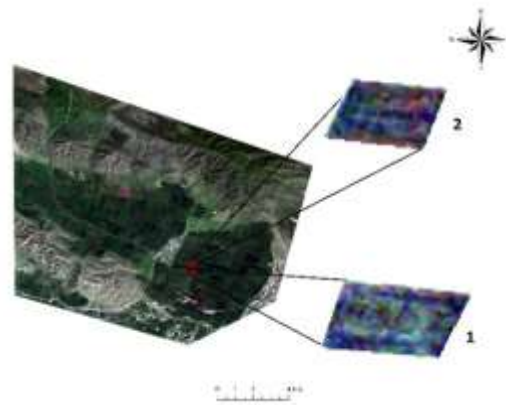


Рисунок 8 – Тестовые сельскохозяйственные поля
Figure 8 - Testing agricultural fields

Карты, описывающие динамику изменений вегетационного индекса в течение вегетационного периода для мониторинга тестовых полей, показаны на рис. 9.

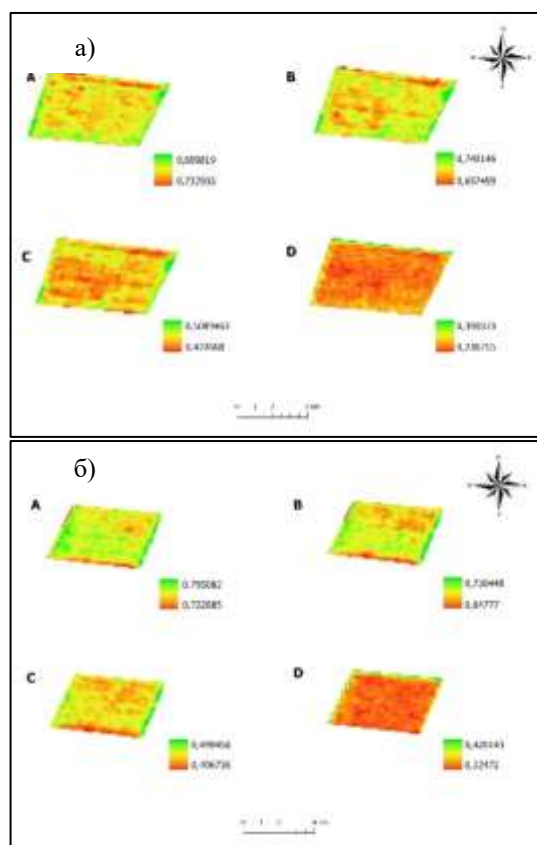


Рисунок 9 – Карты NDVI:
 (а) - первое поле, (б) - второе поле
Figure 9 – NDVI maps:
 (a) – first field, (b) – second field

В начале вегетации показатель увеличивается, в момент цветения его рост прекращается, затем по мере созревания NDVI снижается.

Как видно по карте, максимальное значение вегетационного индекса было выявлено в начале мая в период активного роста растений, которое X^2 составляет 0,809819 и 0,795082 для каждого тестового поля соответственно.

Заключение

В данном научном исследовании были продемонстрированы возможности машинного обучения для классификации сельскохозяйственных полей с целью точного земледелия. Для достижения предполагаемых результатов было опробовано три метода машинного обучения, такие как максимальное правдоподобие, машина опорных векторов и случайный лес. Исходными данными являлись спутниковые изображения Azersky и PlanetScope.

Одной из основных целей было достижение высокой точности классификации. По результатам исследования для изучаемых типов сельскохозяйственных культур была установлена точность классификации. Более высокий результат был получен в третьем эксперименте при классификации изображения Azersky с использованием метода случайного леса.

Для получения достоверных результатов были выбраны почти одинаковые условия для каждого выбранного метода обработки спутниковых данных. Было проведено три эксперимента с разным количеством классов и обучающих выборок для изображений Azersky и PlanetScope, полученных в один и тот же день, для сравнения и влияния количества этих выборок на точность классификации. Обращение к результатам с более высокой точностью показывает, что оптимизировать эту классификацию можно за счет уменьшения количества классов и увеличением количества выборок в классах.

Была использована верификация полученных результатов. После применения машинного обучения были обнаружены водные объекты на территории исследования, что является еще одним доказательством повышения точности классификации. Эти результаты были подтверждены расчетом вегетационного индекса, зафиксировавшего отрицательные значения, характерные для водных объектов.

Мониторинг двух тестовых сельскохозяйственных полей проводился на основе расчета вегетационного индекса, показатели которого соответствовали продуктивности тестовых сельскохозяйственных полей за рассматриваемый период времени.

Спутниковые изображения Azersky были предоставлены в рамках проекта ОАО Азеркосмос «Продвижение услуг дистанционного наблюдения Земли во имя устойчивого развития Азербайджана».

PlanetScope были приобретены через программу поддержки «Образование и исследования» компании Planet.

REFERENCES

1. **Shpaar D., Zakharenko A. V., Yakushev V.P., et al.** Precision Agriculture. Sankt-Peterburg: Pushkin, 2009: 397 (*in Russian*)
2. **Lupyan E.A., Tolpin V.A., Kashnitskiy A.V., Rybalko E.A., Baranova N.V.** Sovremennyye vozmozhnosti organizatsii distantsionnogo monitoringa vinogradnikov Respubliki Krym // Materialy Vserossiyskoy nauchnoy konferentsii (s mezhdunarodnym uchastiyem) "Primeneniye sredstv distantsionnogo zondirovaniya Zemli v selskom khozyaystve". Sankt-Peterburg. 16-17 sentyabrya 2015. SPb.: FGBNU AFL. 2015. S.66-69 (*in Russian*)
3. **Ustuner, M., Sanli, F. B., and Abdikan, S.** Balanced vs Imbalanced Training Data: Classifying Rapid Eye Data with Support Vector Machines, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLI-B7, 379–384, <https://doi.org/10.5194/isprs-archives-XLI-B7-379-2016>, 2016 (*in English*)
4. <https://developers.planet.com/tutorials/convert-planetscope-imagery-from-radiance-to-reflectance/> (*in English*)
5. **Pontus Olofsson, Giles M. Foody, Martin Herold, Stephen V. Stehman, Curtis E. Woodcock, Michael A. Wulde.** Good practices for estimating area and assessing accuracy of land change. Remote Sensing of Environment. Volume 148, 25 May 2014, 42-57. <https://doi.org/10.1016/j.rse.2014.02.015> (*in English*)
6. **Aliev B.H., Atababayev E.R., Zeynalova A.F.** Ecologically safe irrigation technology in the conditions of Azerbaijan. *Herald of the Azerbaijan Engineering Academy*, 2018, vol. 10, no. 4, pp. 112-118 (*in Russian*)
7. **Giles M. Foody.** Explaining the unsuitability of the kappa coefficient in the assessment and comparison of the accuracy of thematic maps obtained by image classification. Remote Sensing of Environment. Volume 239, 15 March 2020, 111630. <https://doi.org/10.1016/j.rse.2019.111630> (*in English*)
8. **Sebastian Candiago, Fabio Remondino, Michaela De Giglio, Marco Dubbini, Mario Gattelli.** Evaluating Multispectral Images and Vegetation Indices for Precision Farming Applications from UAV Images. Remote Sens. 2015, 7(4), 4026-4047; <https://doi.org/10.3390/rs70404026> (*in English*)