

УДК 633/635:528.854(575.2-17)
DOI: 10.36979/1694-500X-2023-23-12-130-138

**КЛАССИФИКАЦИЯ СЕЛЬСКОХОЗЯЙСТВЕННЫХ КУЛЬТУР
В ЧУЙСКОЙ ОБЛАСТИ КЫРГЫЗСКОЙ РЕСПУБЛИКИ
С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ ДИСТАНЦИОННОГО ЗОНДИРОВАНИЯ
ЧЕРЕЗ GOOGLE EARTH ENGINE**

Н.А. Алимбекова, А.А. Земсков, Р.А. Рыскулов, Б.А. Джусупбеков

Аннотация. Представлена электронная система, разработанная с использованием Google Earth Engine для определения 9-ти сельскохозяйственных культур в Чуйской области Кыргызской Республики на основе спутниковых снимков Sentinel-2. Использование этой электронной системы обеспечило возможность получения карт о состоянии посевных площадей в вегетационном периоде, отражающих различные типы культур, доли полей урожайности и возможное повышение урожайности по районам Чуйской области. Поскольку главной задачей сельского хозяйства является обеспечение продовольственной безопасности страны, применение ГИС-технологий в этой отрасли позволяет изучать и разрабатывать методы продовольственных поставок, планировать посевные площади и др.

Ключевые слова: Google Earth Engine; Кыргызстан; сельскохозяйственные культуры; Sentinel-2; NDVI; дерево классификации и регрессии; метод случайного леса.

**АРАЛЫКТАН ЗОНДДОО ЫКМАЛАРЫН КОЛДОНУУ МЕНЕН
GOOGLE EARTH ENGINE АРКЫЛУУ КЫРГЫЗ РЕСПУБЛИКАСЫНЫН
ЧҮЙ ОБЛУСУНДАГЫ АЙЫЛ ЧАРБА ӨСҮМДҮКТӨРҮН КЛАССИФИКАЦИЯЛОО**

Н.А. Алимбекова, А.А. Земсков, Р.А. Рыскулов, Б.А. Джусупбеков

Аннотация. Макалада Sentinel-2 спутниктик сүрөттөрдүн негизинде Кыргыз Республикасынын Чүй облусундагы 9 айыл чарба өсүмдүгүн аныктоо үчүн Google Earth Engine аркылуу иштелип чыккан электрондук система көрсөтүлдү. Бул электрондук системаны колдонуу өсүмдүктөрдүн ар кандай типтерин, түшүмдүүлүк талааларынын үлүшүн жана Чүй облусунун райондору боюнча түшүмдүүлүктүн мүмкүн болгон жогорулашын чагылдырган вегетациялык мезгилдеги айдоо аянттарынын абалы жөнүндө карталарды алуу мүмкүнчүлүгүн камсыз кылды. Айыл чарбасынын негизги милдети өлкөнүн азык-түлүк коопсуздугун камсыз кылуу болгондуктан, бул тармакта ГИС-технологияларын колдонуу азык-түлүк менен камсыздоо ыкмаларын изилдөөгө жана иштеп чыгууга, айдоо аянттарын пландаштырууга ж. б. мүмкүндүк берет.

Түйүндүү сөздөр: Google Earth Engine; Кыргызстан; айыл чарба өсүмдүктөрү; Sentinel-2; NDVI; классификация жана регрессия дарагы; туш келди токой ыкмасы.

**CLASSIFICATION OF AGRICULTURAL CROPS IN THE CHUI REGION
OF THE KYRGYZ REPUBLIC USING REMOTE SENSING METHODS
THROUGH GOOGLE EARTH ENGINE**

N.A. Alimbekova, A.A. Zemskov, R.A. Ryskulov, B.A. Dzhusupbekov

Abstract. The article presents developed an electronic system using Google Earth Engine to determine 9 agricultural crops in the Chui region of the Kyrgyz Republic based on Sentinel-2 satellite images. Due to the developed electronic system, it became possible to obtain maps on the state of sowing areas in the growing season, reflecting various types

of crops, a share of yield fields and a possible increase in productivity in the areas of the Chui region. Since the main task of agriculture is to ensure the country's food security, the use of GIS technologies in this industry allows us to study and develop methods of food supply, plan acreage, etc.

Keywords: Google Earth Engine; Kyrgyzstan; agricultural crops; Sentinel-2; NDVI; classification and regression trees; random forest.

Введение. Сельское хозяйство является одним из приоритетных направлений экономического развития Кыргызстана в прошлой, ближайшей и в будущей перспективе. Это подтверждается функционированием национальной экономики, которое раскрывает потенциал и большие возможности сельского хозяйства страны, что выражается в повышении темпов роста сельскохозяйственного производства и экспортного потенциала отрасли [1].

Государственная политика сельского хозяйства страны предусматривает рациональное, эффективное управление и учет сельскохозяйственных земель. Точное определение и мониторинг сельскохозяйственных культур являются критически важными для эффективного управления ресурсами и принятия обоснованных решений. В последние годы развитие геопространственных технологий и доступность обширных архивов спутниковых данных стали основой для разработки новых методов классификации культур.

Целью данного исследования является разработка и применение методологии классификации сельскохозяйственных культур в Чуйской области Кыргызской Республики (КР) с использованием методов дистанционного зондирования на основе платформы Google Earth Engine. Платформа Google Earth Engine предоставляет доступ к множеству спутниковых снимков и инструментам их обработки, что позволяет проводить анализ крупных территорий и детектировать растительность и ее изменения.

По данным Статистического комитета КР в 2021 году по Чуйской области были отмечены следующие показатели урожайности сельхозкультур, тыс. тонн [2]:

Производство зерновых культур:

Пшеница – 116,3 (в 2020 – 300,9)
Ячмень – 111,3 (в 2020 – 297,5)
Кукуруза на зерно – 206,4
(в 2020 – 227,9)

Корнеплоды:

Картофель – 173,7 (в 2020 – 197,6)
Сахарная свекла – 352 794,1
(в 2020 – 434 074,6)

Масличные культуры:

Сафлор – 3 295,0 (в 2020 – 6 274,4)

По данным земельного учета страны, который проводится каждый год, на 1 января 2022 года в Чуйской области общая площадь земель сельскохозяйственного назначения составляла 1 276 638 га. Из общего числа сельскохозяйственных угодий: пашня – 408 070 га, многолетние насаждения – 7 061 га, залежь – 10 703 га, сенокос – 20 438 га, пастбища – 590 514 га [3].

В данной статье авторами представлены результаты разработок, которые включают: сбор и обработку спутниковых данных с использованием платформы Google Earth Engine, разработку методов классификации сельскохозяйственных культур и оценку точности полученных результатов. Проведен сравнительный анализ существующих методов классификации с обсуждением их преимуществ и недостатков в контексте Чуйской области.

Инструмент, разработанный на платформе Google Earth Engine, будет полезен для Министерства сельского хозяйства КР, сельскохозяйственных организаций и управленцев, инвесторов, поможет им принимать более обоснованные решения и оптимизировать использование земельных ресурсов в Чуйской области.

Исследуемая область. Чуйская область является зоной развитого орошаемого земледелия, этому способствуют почвенные и климатические условия. Орошаемые земли пригодны для технических, зерновых, кормовых, овощных культур, садоводства и виноградарства [4].

Чуйская область находится в северной части страны и является самой развитой областью (рисунк 1). Область охватывает около 20,2 тыс. км², и около 419 517 га составляют посевные площади сельскохозяйственных культур [5].

Исследуемая область по климатическому районированию относится к Северному и Северо-Западному Кыргызстану. Количество осадков на местности зависит от высоты над уровнем моря. Количество осадков возрастает от 370 мм на севере долины – до 425 вверх по долине в восточном направлении и до 450–500 мм в южном направлении к предгорьям. В нижней зоне и в предгорьях до высоты 2000–2300 м над уровнем моря максимум осадков приходится на апрель-май, в пригребневой зоне смещается на май-июнь, минимум приходится на август.

В теплый период количество осадков в нижней зоне составляет от 53–60 до 70 % в предгорьях [6].

Материалы и методология. В качестве исходных данных использовали:

- Административные границы Чуйской области и районов, которые имеют систему координат WGS_1984_UTM_Zone_43N в формате shp. Административные границы получены от Государственного предприятия «Государственный проектный институт по землеустройству “Кыргызгипрозем”» при Земельной службе.
- Информацию о сельскохозяйственных культурах по площадям по районам Чуйской области. Информация Статистического комитета КР.
- Разработка модели проведена на платформе Google Earth Engine. Для расчета индекса NDVI (Normalized Difference Vegetation Index) использовали имеющиеся в Google Earth Engine данные спутника Harmonized Sentinel-2 MSI: MultiSpectral Instrument, Level-1C [7]. Индекс NDVI рассчитывается по формуле:

$$NDVI = \frac{NIR - Red}{NIR + Red},$$

где NIR – отражение в ближней инфракрасной области спектра; Red – отражение в красной области спектра.

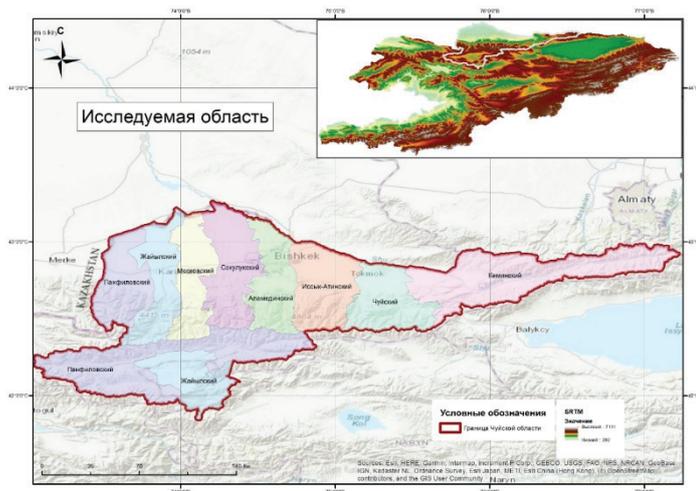


Рисунок 1 – Территория Чуйской области

Выбор пилотных участков проводили по следующим сельскохозяйственным культурам:

1. Корнеплоды (сахарная свекла, картофель).
2. Зерновые (пшеница, ячмень, кукуруза).
3. Масличные (сафлор).

Одновременно с этим, для анализа дополнительно были выбраны и другие культуры в определенных районах (соя, люцерна).

В каждом районе Чуйской области были определены 362 пилотных участка с необходимыми культурами для последующего наблюдения и анализа. При выборе пилотных участков полевыми специалистами были опрошены фермеры – владельцы пилотных участков, были зафиксированы такие данные по растениеводству основных культур, как:

- длительность вегетационного периода;
- координаты участка;
- сроки посева и сбора урожая выращиваемых сортов с указанием их особенностей;
- особенность сорта;
- вид полива;
- применение химикатов и удобрений;
- посев сельскохозяйственных культур на участках за предыдущие годы.

Одновременно с этим был произведен мониторинг отобранных пилотных участков в Кеминском, Чуйском, Иссык-Атинском и Аламудунском районах с целью уточнения реально выращиваемых культур, выявления имеющихся аномалий и выяснения причин их возникновения.

Построение модели. На первом этапе работы выполнены измерения координат полей с различными культурами, выращиваемыми на территории Чуйской области. Всего было измерено 362 поля с восьмью различными типами выращиваемых культур.

Результаты измерения полей были оформлены в виде шейп-файлов с информацией о локации, сроках посева и уборки, культуре, сорте, применении удобрений. Однако в результате проведенной очистки данных полевых данных часть полей была отсеяна и в модели были использованы только 294 поля, а поля с пшеницей были разделены на озимые и яровые.

На предварительном этапе построения модели машинного обучения были построены бинарные модели для сахарной свеклы и мультиклассовая модель по всем культурам с использованием двух классических алгоритмов машинного обучения из всех имеющихся в Google Earth Engine: Classification and Regression Trees (Дерево классификации и регрессии – один из вариантов алгоритма дерева решений) и Random Forest (метод случайного леса). Для оценки точности моделей, имеющийся массив полей по каждой культуре был случайным образом разделен на обучающий и тестовый в пропорции 70 на 30 %.

Точность для бинарных моделей оказалась очень высокой: Classification and Regression Trees, точность на тестовом наборе – 0.97975, Random Forest: точность на тестовом наборе – 0.98570. Точность мультиклассовой модели оказалась существенно ниже: Classification and Regression Trees, точность на тестовом наборе – 0.68483, Random Forest точность на тестовом наборе – 0.73779.

Более низкие значения точности мультиклассовых моделей связаны в первую очередь с тем, что мультиклассовая модель предполагает наличие в тестовом наборе данных образцов всех типов культур на исследуемой площади. В нашем случае были измерены тестовые поля только по 9 культурам. Для повышения точности мультиклассовой классификации необходимо иметь тестовые поля по всем типам культур, выращиваемым на исследуемой территории, в том числе разделять некоторые культуры на два и более отдельных классов. Для этого необходимо проведение дополнительных исследований.

Для достижения высокой точности определения выращиваемых культур будем использовать комбинацию бинарных моделей Random Forest по каждой культуре. На рисунке 2 представлены усредненные по каждой культуре значения NDVI в зависимости от номера недели 2022 года. Поскольку платформа Google Earth Engine в бесплатной версии предоставляет ограниченные ресурсы, в качестве параметров модели будем использовать не все недельные значения вегетационного индекса NDVI,

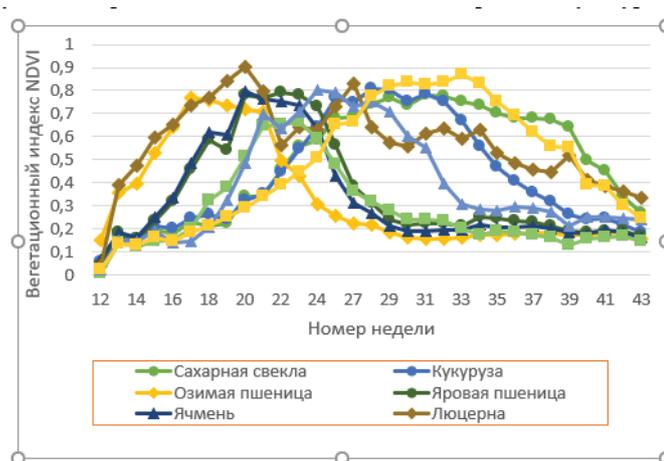


Рисунок 2 – Усредненная зависимость NDVI от номера недели 2022 года для культур, участвующих в проекте

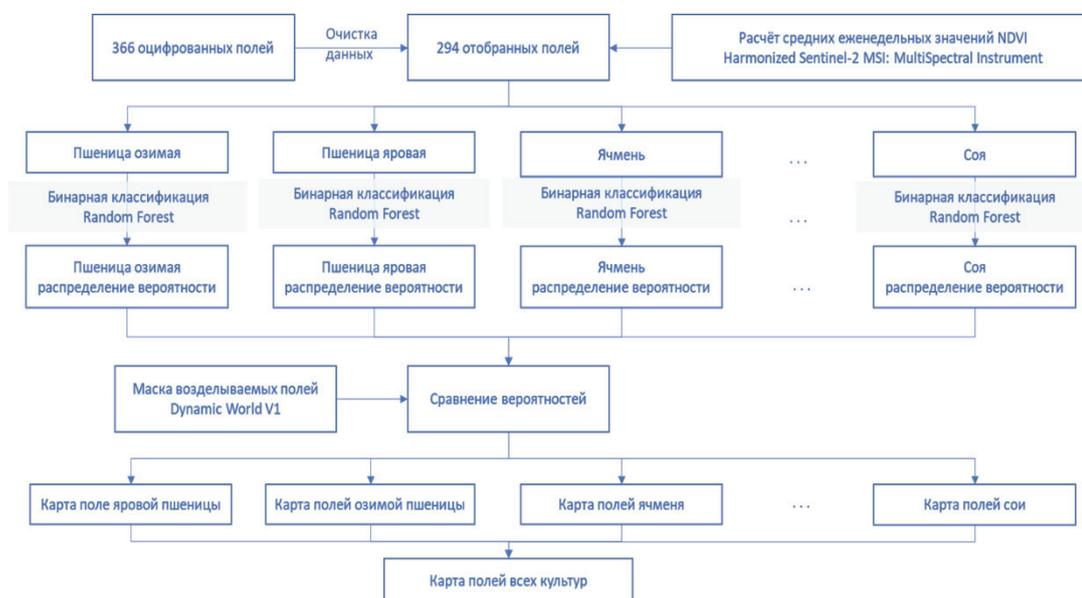


Рисунок 3 – Модель машинного обучения

а только для недель с номерами 13, 15, 19, 24, 26, 29, 34, 38, 41. Этим значениям достаточно, поскольку они отображают основные отличительные признаки культур.

По тестовым полям каждой из 9-ти культур были построены бинарные вероятностные модели классификации Random Forest. В результате были получены 9 отдельных слоев, соответствующих расположению каждой из 9-ти исследуемых культур. Для построения единой карты, содержащей расположение всех культур, полученные 9 слоев по каждой культуре были объединены.

Общая схема построения модели машинного обучения показана на рисунке 3.

Оценка точности полученных результатов. Для проверки точности полученной модели на тестовых наборах полей рассчитали значения Точности (Accuracy) и Сбалансированной точности (Balanced Accuracy). Для расчёта использовали стандартные формулы [8]:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN},$$

где TP (true positive) – количество точек на тестовых полях, правильно предсказанных моделью как являющиеся исследуемой культурой; TN (true negative) – количество точек на тестовых полях, правильно определенных моделью как не являющейся исследуемой культурой; FP (false positive) – количество точек на тестовых полях, ложно определённые моделью как относящиеся к исследуемой культуре; FN (false negative) – количество точек на тестовых полях, ложно определённые моделью как не относящиеся к исследуемой культуре.

$$Balanced Accuracy = \frac{TRP + TNR}{2},$$

где $TRP = \frac{TP}{TP + FN}$ – доля истинно положительных результатов,

$TNR = \frac{TN}{TN + FP}$ – доля истинно отрицательных результатов.

Результаты полученных точностей, а также площадь каждой из культур определенная моделью в Чуйской области, представлены в таблице 1.

Как видно из данных таблицы 1, полученная модель имеет достаточно высокие значения точности и сбалансированной точности по всем культурам, за исключением сбалансированной точности сои. Этот факт необходимо учитывать при дальнейшей работе.

Определение полей с низкой урожайностью. При оценке чистой первичной продукции (Net Primary Production – NPP) с помощью методов дистанционного зондирования земли считается доказанной пропорциональность значений NPP и NDVI. Это позволяет оценить относительную урожайность исследуемых культур.

В качестве примера на рисунке 4 показаны поля с кукурузой в Московском районе Чуйской области. В верхней части рисунка синим цветом показаны поля, определенные моделью как кукуруза. В нижней части – те же поля с суммарным значением NDVI ниже среднего. Цветовой заливкой полей в нижней части указано значение отклонения суммарного NDVI от среднего. Зеленый цвет соответствует среднему значению суммарного NDVI, желтый – 50 % от среднего, красный – минимальному значению NDVI для данной культуры на территории Чуйской области.

Таблица 1 – Точность, сбалансированная точность и площадь, занимаемая культурами в Чуйской области, определенная моделью

№ п/п	Культура	Точность, %	Сбалансированная точность, %	Расчетная площадь, га
1	Пшеница озимая	96,3	96,5	91 844
2	Пшеница яровая	96,0	97,7	2 396
3	Ячмень	94,3	87,5	44 626
4	Кукуруза	98,1	92,8	48 764
5	Сахарная свекла	97,6	98,0	10 168
6	Люцерна	97,0	95,4	83 958
7	Сафлор	94,9	96,3	26 575
8	Каргофель	99,3	94,4	2 140
9	Соя	99,3	68,8	3 656

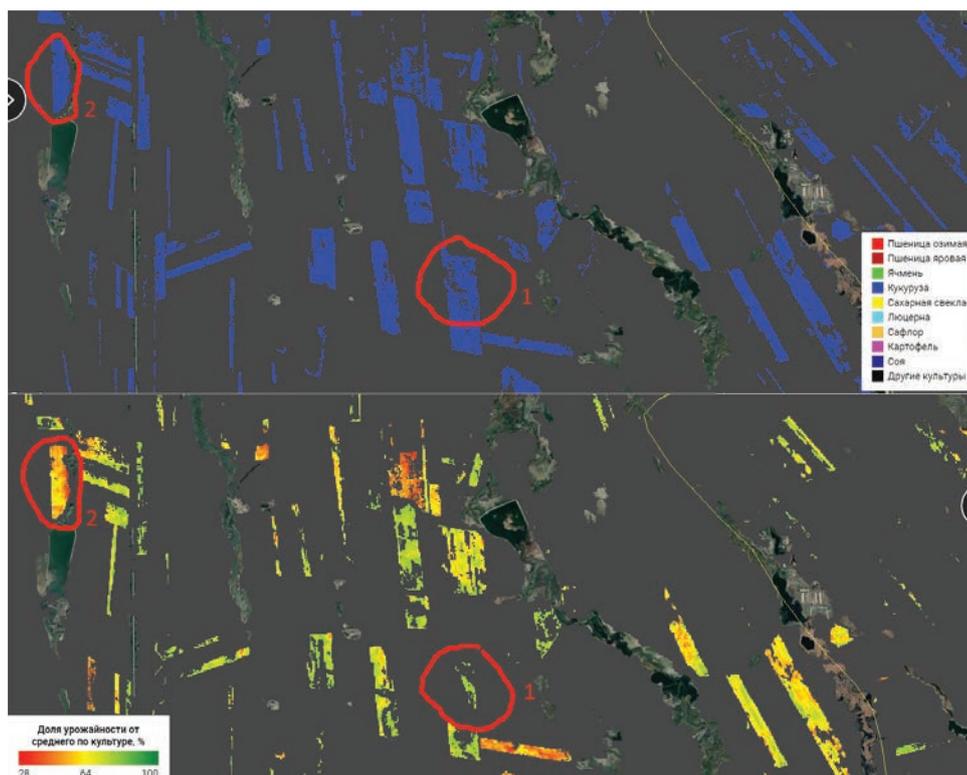


Рисунок 4 – Пример определения полей с урожайностью ниже средней

Таблица 2 – Повышение урожайности и валового сбора по всем культурам

№ п/п	Культура	Доля полей с урожайностью ниже средней	Возможное повышение урожайности на проблемных полях	Увеличение валового урожая
1	Пшеница озимая	30,8 %	14,9 %	4,6 %
2	Пшеница яровая	36,8 %	26,8 %	9,9 %
3	Ячмень	57,9 %	17,4 %	10,1 %
4	Кукуруза	38,6 %	14,3 %	5,5 %
5	Сахарная свекла	66,9 %	16,8 %	11,2 %
6	Люцерна	45,2 %	17,6 %	8,0 %
7	Сафлор	70,8 %	26,4 %	18,7 %
8	Картофель	34,3 %	16,9 %	5,8 %
9	Соя	76,7 %	21,0 %	16,1 %



Рисунок 5 – Интерфейс электронной системы

Далее определены доля полей с урожайностью ниже средней, возможное повышение урожайности на проблемных полях и увеличение валового урожая. Результаты такого расчета приведены в таблице 2.

Полученные значения повышения валового сбора продукции по исследуемым культурам основываются на допущении того, что среднее значение урожайности по каждой культуре является близким к максимально возможному в текущих условиях в Чуйской области. Однако такие же расчеты можно провести для любого другого значения урожайности. Например, на 20 процентов выше среднего. Например, средняя урожайность пшеницы в 2022 году в Чуйской области составила 27,6 ц/га. Можно провести расчет увеличения валового сбора пшеницы при доведении средней урожайности до 35 ц/га, то есть при увеличении урожайности на 26,8 %.

Закключение. В ходе выполнения работы успешно применены: спутниковые данные Sentinel 2, методы обработки дистанционного зондирования с использованием машинного обучения на платформе Google Earth Engine и полевые данные сельскохозяйственных культур, разработана электронная система (рисунок 5).

Результаты анализа позволили выявить основные тренды в сельскохозяйственной деятельности на изучаемой территории. Информационная система обнаружила изменения в использовании земель, сезонные колебания в выращиваемых культурах, а также идентифицировала участки сельскохозяйственной деятельности, требующие дополнительного внимания и мер по улучшению устойчивости.

На основе полученных результатов можно рекомендовать использование электронной системы на базе платформы Google Earth Engine в дальнейших исследованиях и проектах, связанных с анализом сельскохозяйственных культур.

Система позволяет выполнять следующие работы:

- проводить мониторинг сельскохозяйственных культур в режиме реального времени. Сотрудники Министерства сельского хозяйства могут дистанционно наблюдать за состоянием полей в вегетационный период;
- выступать как инструмент для анализа севооборота;

- планировать посевные культуры (семена, удобрения и др.), посмотрев в системе состояние по коэффициенту NDVI, определять состояние растительности и необходимость применения удобрений;
- обеспечить рациональное использование земель (выявление неиспользуемых земель и др.).
- анализировать экономические и экологические последствия природных катаклизмов и деятельности человека;
- обмениваться информацией и картами с внутренними отделами и сторонними организациями, где бенефициаром системы является Министерство сельского хозяйства.

Рекомендации. Для улучшения работы электронной системы необходимо продолжение проведения работ по данному проекту. Рекомендуется расширить географический охват исследования, включив другие сельскохозяйственные регионы страны и культуры. Это поможет получить более общую картину и сравнить различные аспекты сельскохозяйственной деятельности.

Использование дополнительных источников данных, таких как метеорологические, кадастровые данные, данные о почвенном состоянии, гидрологию и др. Это позволит более детально изучить взаимосвязи между различными факторами и сельскохозяйственной продуктивностью.

Для более полного понимания сельскохозяйственной деятельности рекомендуется интегрировать социально-экономические данные: о производстве и потреблении сельскохозяйственных продуктов, данные о ценах на рынке и др. Это поможет оценить экономическую эффективность и устойчивость сельскохозяйственного сектора.

Для дальнейшей работы важно вовлечь заинтересованные стороны, такие как сельскохозяйственные производители, органы власти и научные институты. Это поможет уточнить потребности в аналитических данных, обеспечить взаимодействие.

Поступила: 24.10.2023; рецензирована: 09.11.2023; принята: 13.11.2023.

Литература

1. *Okenova A.O.* Tsifrovizatsiya selskogo khozyaystva v Kyrgyzskoy Respublike [Digitization of agriculture in the Kyrgyz Republic]. *Ekonomicheskie otnosheniya*. 9 (1). 2019. С. 97–106. DOI:10.18334/eo.9.1.39675 (дата обращения: 05.05.2023).
2. Национальный статистический комитет Кыргызской Республики. Валовой сбор и урожайность сельскохозяйственных культур в Чуйской области, 2021 г. URL: <http://www.stat.kg/> (дата обращения: 05.05.2023).
3. Отчетность земельного кадастра Кыргызской Республики. 2022 г.
4. *Акималиев Д.А.* Пути повышения плодородия почв в Киргизии / Д.А. Акималиев. Фрунзе, 1990. 101 с.
5. Национальный статистический комитет Кыргызской Республики. Посевные площади сельскохозяйственных культур и многолетних насаждений по всем категориям хозяйств Чуйской области, 2021 г. URL: <http://stat.kg/ru/opendata/category/4892/> (дата обращения: 05.05.2023).
6. Мониторинг, прогнозирование опасных процессов и явлений на территории Кыргызской Республики. Изд. 18-е с изм. и доп. Бишкек: МЧС КР, 2021. 565 с.
7. Harmonized Sentinel-2 MSI: MultiSpectral Instrument, Level-1C. URL: https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S2_HARMONIZED (дата обращения: 06.05.2023).
8. *García V., Mollineda R.A., Sánchez J.S.* Index of Balanced Accuracy: A Performance Measure for Skewed Class Distributions. In: Araujo H., Mendonça A.M., Pinho A.J., Torres M.I. (eds.). *Pattern Recognition and Image Analysis*. IbPRIA 2009. Lecture Notes in Computer Science. Vol. 5524. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-02172-5_57 (дата обращения: 05.05.2023).