

## Распознавание и классификация посевов сельскохозяйственных культур Хабаровского края с использованием NDVI и EVI

Л. В. Илларионова<sup>1</sup>, А. С. Степанов<sup>2</sup>, Е. А. Фомина<sup>1</sup>

<sup>1</sup> *Вычислительный центр ДВО РАН, Хабаровск, 680000, Россия  
E-mail: illarionova\_l@list.ru*

<sup>2</sup> *Дальневосточный научно-исследовательский институт сельского хозяйства  
Хабаровск, 680521, Россия  
E-mail: stepanfx@mail.ru*

Разработка подходов к идентификации и классификации посевов сельскохозяйственных культур на региональном уровне относится к одной из важнейших задач цифрового земледелия. Решение этой задачи с использованием данных дистанционного зондирования Земли особенно актуально для южной части Дальнего Востока, что обусловлено различиями в сроках посева, уборки, длительности фаз вегетации сельхозкультур макрорегиона в сравнении с западной частью РФ и, соответственно, снижением точности существующих алгоритмов. Для проведения классификации использовались данные о севообороте сельскохозяйственных полей Хабаровского р-на общей площадью более 4000 га, исследовались 8 классов пахотных земель. Были рассмотрены 37 мультиспектральных снимков с разрешением 20 м, полученных со спутников Sentinel-2A/B в период с апреля по октябрь 2021 г. Для каждого пикселя были сформированы временные ряды значений NDVI (*англ.* Normalized Difference Vegetation Index) и EVI (*англ.* Enhanced Vegetation index), а также построены аппроксимированные кривые сезонного хода с использованием функции Фурье. Наилучшие результаты классификации на основе машинного обучения были достигнуты при использовании метода Random Forest, общая точность составила 95 %. При применении в качестве входных данных значений разных индексов — NDVI и EVI — значимых различий в точности классифицирования выявлено не было. По результатам кроссвалидации было установлено, что точность распознавания основных классов находится на высоком уровне: ошибка определения классов «соя», «гречиха», «залежь», «пар» и «многолетние травы» не превышала 10 %. Полученные результаты подтвердили возможность использования NDVI и EVI для классификации пахотных земель южной части Дальнего Востока, при этом в качестве базы должны рассматриваться кривые сезонного хода именно этого региона, что обусловлено особенностями вегетационных циклов сельхозкультур.

**Ключевые слова:** классификация, распознавание, пахотные земли, индекс вегетации, Дальний Восток, машинное обучение, аппроксимация

Одобрена к печати: 15.03.2023

DOI: 10.21046/2070-7401-2023-20-2-155-165

### Введение

Распознавание посевов сельскохозяйственных культур с использованием данных дистанционного зондирования Земли (ДЗЗ) относится к одной из важнейших задач цифрового земледелия. Актуальность решения задачи связана как с необходимостью уточнения севооборотов в отчётном периоде, так и с потребностью восстановления севооборота в ретроспективе. Для решения задачи классификации пахотных земель региона, а также идентификации отдельных сельскохозяйственных культур активно применяются методы машинного обучения. В качестве входных данных при этом чаще всего используются вегетационные индексы, рассчитанные по данным ДЗЗ в оптическом диапазоне (Якушев и др., 2022; Bartalev et al., 2014; Feyisa et al., 2020; Yan et al., 2021), иногда в комбинации с радарными данными (Tufail et al., 2021). Исследования, проведённые в Бразилии, заключались в выделении и описании шаблонов севооборота с последующим вычислением площадей, занятых сельскохозяйственными культурами, на основе 16-дневных композитов MODIS (*англ.* Moderate Resolution Imaging

Spectroradiometer) (Chen et al., 2018). Также для одного из штатов Бразилии с использованием данных Landsat-8 вычислялись средние значения NDVI (*англ.* Normalized Difference Vegetation Index — нормализованный разностный вегетационный индекс) для каждого поля и проводилась идентификация 10 сельскохозяйственных культур (Bellon et al., 2018). Для определения культуры, произрастающей на поле, использовались только средние значения NDVI по этому полю. В трёх провинциях Германии была проведена автоматическая классификация пахотных земель методом Random Forest с выделением 12 сельскохозяйственных культур по значениям индекса NDVI, полученным посредством обработки изображений со спутников Landsat-8 и Sentinel-2 (Griffiths et al., 2019). В южной части провинции Онтарио (Канада) для прогнозирования урожайности были созданы маски основных культур. Для классификации использовались методы нечёткой логики, в качестве данных — сезонный ход вегетационного индекса EVI (*англ.* Enhanced Vegetation Index — усовершенствованный вегетационный индекс) в 2011–2013 гг. (вычислен по данным MODIS). Точность классификации на тестовом множестве составила 75 % (Liu et al., 2019). На северо-востоке Китая группой исследователей была изучена возможность совместного применения различных вегетационных индексов и методов классификации для определения принадлежности пикселей полей тому или иному классу сельскохозяйственных культур (Zhang et al., 2020). Также был описан подход к классификации пахотных земель до достижения пика вегетации, заключавшийся в формировании эталонных временных рядов, где в качестве характеристики ряда рассматривалось отношение NDVI/EVI, полученных по данным Landsat и Sentinel-2 (Hao et al., 2020). Эти результаты очень интересны с практической точки зрения: они демонстрируют возможность использования ДЗЗ для классификации пахотных земель на Дальнем Востоке России, сходных по климатическим условиям с некоторыми провинциями Канады и северо-востока Китая.

Необходимо отметить, что теоретические разработки российских и зарубежных учёных достаточно широко используются на практике для управления пахотными землями субъектов РФ (Арзамасцева и др., 2021; Буланов и др., 2019).

Так, например, в Институте космических исследований РАН (ИКИ РАН) в последние годы решалась задача создания маски пахотных земель, т.е. распознавания пашни. В разработанных и поддерживаемых ИКИ РАН сервисах «Вега» классификация земель проводится с использованием графиков сезонного хода NDVI: выделяются пахотные земли (отдельно земли, занятые яровыми, и земли, занятые озимыми культурами), леса, травянистая растительность, болота и тундра (Лупян и др., 2011, 2014). Была отмечена высокая точность идентификации пашни в европейской части России, при этом при использовании алгоритма на Дальнем Востоке ошибка возрастала (Миклашевич и др., 2019).

Также была показана возможность идентифицировать культуры на полях посредством анализа временного хода характеристик рассеивания (по данным высокого разрешения, спутники Sentinel-1). В среде «Вега-PRO» продемонстрирована возможность интерактивной классификации земель при помощи создания обучающих выборок и собственных масок (Tolpin et al., 2014). Было отмечено, что применительно к российскому Дальнему Востоку существующие алгоритмы идентификации и классификации имеют меньшую точность. Это обусловлено особенностями циклов вегетации сельхозкультур в регионе и требует дополнительного изучения (Плотников и др., 2018).

Очевидно, что отличные от западных регионов сортовой состав, сроки сева и уборки урожая, что связано в том числе и с климатическими особенностями, затрудняют использование существующих критериев разделения сельскохозяйственных культур для регионов Дальнего Востока, и в том числе для Хабаровского края.

Таким образом, разработка и сравнительный анализ точности методов распознавания и классификации земель сельскохозяйственного назначения Хабаровского края по временным рядам значений индексов вегетации представляется актуальной задачей, имеющей как практическую, так и теоретическую ценность.

## Материалы и методы

Для проведения классификации были использованы данные о структуре землепользования для 197 сельскохозяйственных полей Хабаровского р-на. Общая площадь исследуемых полей составила более 4000 га, все поля были отнесены к одному из 8 классов: соя — 1260 га, гречиха — 238 га, залежь — 1613 га, кукуруза — 17 га, многолетние травы — 114 га, овёс — 27 га, пар — 496 га, другое — 272 га. В настоящее время на этой территории культивируются только яровые культуры, что обусловлено климатическими условиями. Многолетние травы в основном представлены тимофеевкой луговой, относящейся к злаковым культурам и районированной для Дальнего Востока. Посевные мероприятия в Хабаровском р-не проводятся с начала мая, уборка — до конца октября.

Для исследуемого района были рассмотрены 37 мультиспектральных снимков с разрешением 20 м, полученных со спутников Sentinel-2A/B в период с 19 апреля по 31 октября 2021 г. В дальнейшем формировались временные ряды NDVI и EVI, в которых значения показателей для каждого дня определялись по формулам:

$$\text{NDVI} = \frac{\text{NIR} - \text{RED}}{\text{NIR} + \text{RED}}, \quad \text{EVI} = 2,5 \frac{\text{NIR} - \text{RED}}{\text{NIR} + 6\text{RED} - 7,5\text{BLUE} + 1},$$

где NIR — отражение в ближней инфракрасной области спектра; RED, BLUE — отражения в красной и синей областях спектра.

Для формирования обучающей и тестовой выборки были использованы временные ряды значений NDVI или EVI для пикселей с наложением маски облачности, также проводилась фильтрация неоднородных значений ряда.

Подготовка оптических снимков интересующей территории, вычисление индексов, работа с маской облачности, фильтрация пикселей и некоторые другие процессы были автоматизированы с использованием библиотек Rasterio (<https://rasterio.readthedocs.io/>) и Osgo ([www.osgeo.org](http://www.osgeo.org)) языка Python.

Для классификации были использованы следующие методы: дерево решений DTC (*англ.* Decision Tree Classifier) и комбинация деревьев решений RF (*англ.* Random Forest), реализованные в библиотеке scikit-learn ([www.scikit-learn.org](http://www.scikit-learn.org)) языка Python. В обучающую выборку случайным образом включались 70 % от общего числа пикселей, в тестовую — 30 %.

Для сравнительной оценки методов проводилась трёхкратная кроссвалидация, определялись общая точность (OA), проценты и  $f_1$ -метрика для DTC и RF, которые рассчитывались по формулам:

$$\text{OA} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \quad f_1 = \frac{\text{TP}}{\text{TP} + (\text{FP} + \text{FN})/2},$$

где TP — верно предсказанная принадлежность пикселя к данному классу; TN — верно предсказанное отнесение пикселя к другому классу; FP — неверное отнесение пикселя к данному классу; FN — неверное отнесение пикселя к другому классу.

Также формировались нормализованные матрицы ошибок классификации (*англ.* Normalized confusion matrix). В них представлено отношение числа пикселей, отнесённых к классу по результатам прогнозирования, к общему количеству пикселей данного класса, выраженное в процентах.

Для оценки точности идентификации сельхозкультуры на уровне полей была сформирована следующая выборка: для каждого класса случайным образом было отобрано по одному полю (всего 8 полей), для оставшегося массива данных проводилась стандартная процедура по разделению на обучающую и тестовую выборку. На множестве пикселей отобранных 8 полей также определялись общая точность и  $f_1$ -метрика для каждого класса с использованием DTC и RF. Предложенный способ позволил смоделировать условия практической задачи по идентификации сельскохозяйственных земель с использованием созданного ранее классификатора.

## Результаты

По данным оптических снимков Sentinel-2A/B в период с апреля по октябрь 2021 г. для каждого из пикселей, соответствующих 197 полям Хабаровского р-на, были рассчитаны значения NDVI и EVI, для построения графиков сезонного хода были определены средние показатели для каждого поля. На *рис. 1* (см. с.159) представлены графики изменения средних значений индексов вегетации в 2021 г. для основных культур Хабаровского края: сои, гречихи, многолетних трав, овса и кукурузы. Также на *рис. 1* отражены аппроксимированные кривые сезонного хода, полученные с использованием функции Фурье. В принципе, полученные графики отражают вегетационный цикл культур. В частности, максимумы для сои — с середины августа до начала сентября (NDVI), 1–2-я декады августа (EVI). Более ранний максимум (середина июля) характерен для овса и тимофеевки луговой, составляющей основу многолетних трав. Напротив, максимальные значения NDVI и EVI кукурузы достигаются несколько позже, к 240–250 дням с начала года, что соответствует первой декаде сентября.

В целом анализ графиков сезонного хода посевных культур Хабаровского края показал, что визуальные отличия аппроксимированных кривых друг от друга, а также характеристики максимума делают возможным использование временных рядов NDVI и EVI для классификации.

В *табл. 1* представлены результаты проведённой трёхкратной кроссвалидации для классификации с использованием временных рядов NDVI или EVI. С применением двухфакторного анализа было установлено, что точность метода RF значимо выше, чем точность DTC ( $p < 0,05$ ). Вместе с тем значимых различий в средних значениях OA и  $f_1$  при использовании разных индексов вегетации выявлено не было ( $p > 0,05$ ). В среднем значение OA при применении RF составило 95,1–95,6 %, а  $f_1$  — 0,94–0,95 %.

В *табл. 2* и *3* представлены нормализованные матрицы ошибок для метода DTC (глубина дерева — 8, дальнейшее увеличение количества уровней не приводило к существенному увеличению точности). Значение OA для NDVI было равным 90,0 %, для EVI — 89,3 %, а общая  $f_1$ -метрика для всех классов — 0,83 (NDVI) и 0,84 (EVI). Как видно, точность определения пикселей сои и залежи превышает 90 % в обоих случаях. Для идентификации гречихи, кукурузы, овса и многолетних трав предпочтительнее использовать индекс EVI. Парующие земли, а также смешанные классы более точно распознаются с применением NDVI. В основном пиксели, которые неверно идентифицированы, классифицируются как залежь. Во многих случаях действительно либо имеет место зарастание этого участка поля сорной травой, либо эти пиксели относятся к краевым.

В *табл. 4* и *5* приведены нормализованные матрицы ошибок для метода RF. Общая точность метода (максимальная глубина дерева — 5, число деревьев — 5) для NDVI — 95,3 %, для EVI — 95,0 %. Общая метрика  $f_1$  для всех культур была равной 0,94 для каждого индекса вегетации. Точность определения классов «залежь», «соя», «гречиха», «пар» и «многолетние травы» превышала 90 % для NDVI и EVI. При этом для оставшихся классов этот показатель для овса выше при использовании EVI (84,5 % в сравнении с 76,6 %), а для кукурузы, напротив, при применении NDVI (88,1 и 80,0 %).

Таблица 1. Оценка точности методов классификации по результатам трёхкратной кроссвалидации

Итерация	NDVI				EVI			
	DTC		RF		DTC		RF	
	OA, %	$f_1$	OA, %	$f_1$	OA, %	$f_1$	OA, %	$f_1$
1	90,3	0,83	96,1	0,94	89,2	0,85	95,1	0,92
2	89,6	0,85	95,3	0,95	89,6	0,85	94,7	0,94
3	90,4	0,83	95,5		88,9	0,83	95,4	
$\bar{x}$	90,1	0,84	95,6		89,4	0,84	95,1	
$\Delta\bar{x}$	0,9	0,03	0,70	0,01	0,70	0,03	0,60	0,01

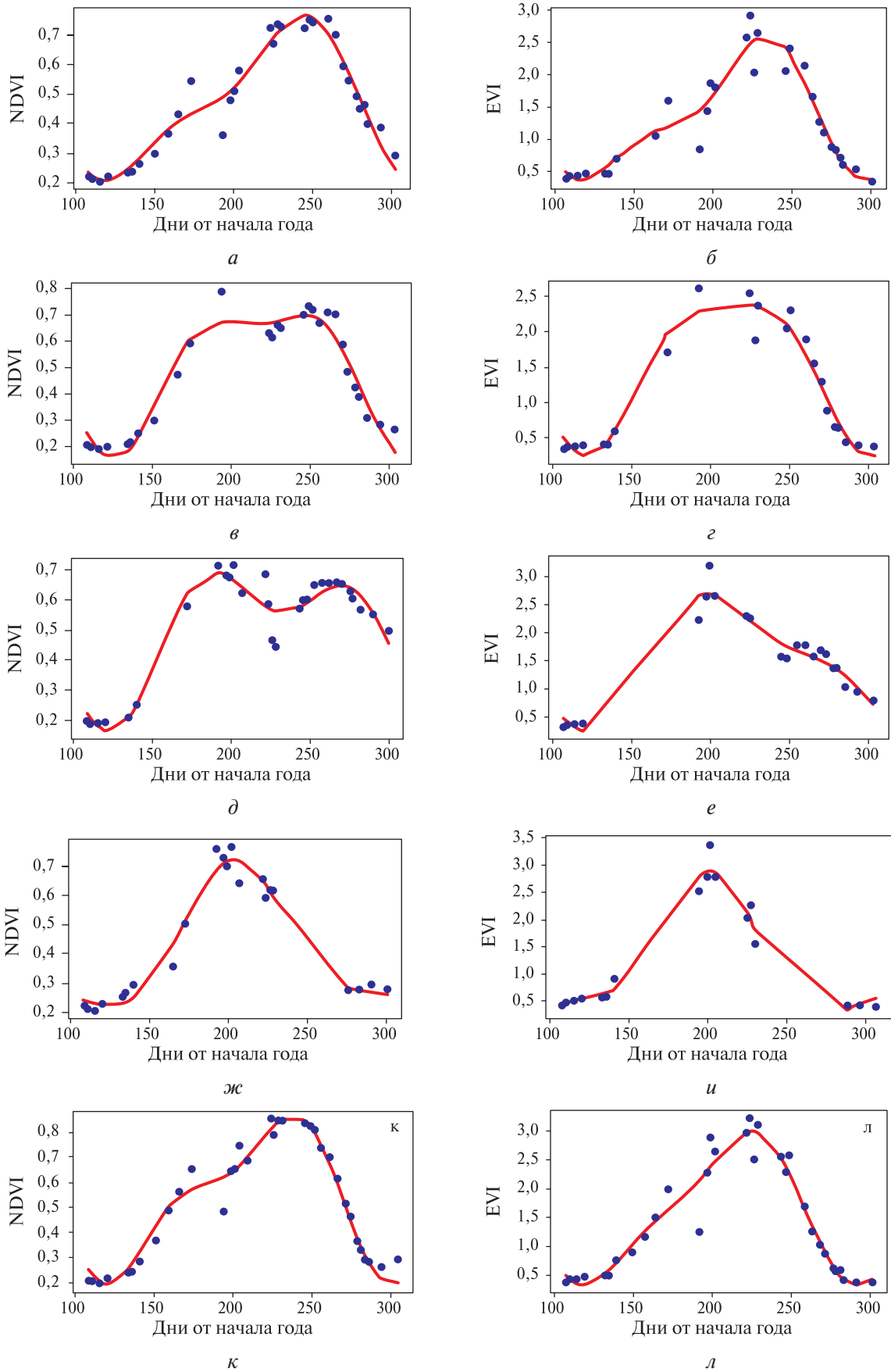


Рис. 1. Фактические (синие точки) и аппроксимированные (красные линии) значения NDVI и EVI для сельхозкультур Хабаровского р-на в 2021 г.: а, б — соя; в, г — гречиха; д, е — многолетние травы; ж, и — овёс; к, л — кукуруза

Таблица 2. Нормализованная матрица ошибок (DTC, NDVI)

	Класс	Прогноз							
		«залежь»	«soя»	«гречиха»	«овёс»	«пар»	«другое»	«кукуруза»	«травы»
Реальность	«залежь»	97,3	0,7	0,0	0,0	0,9	1,0	0,0	0,0
	«soя»	4,6	94,4	0,1	0,0	0,1	0,6	0,1	0,0
	«гречиха»	7,8	0,6	87,4	0,0	0,6	0,1	0,1	3,3
	«овёс»	3,7	5,6	0,0	59,8	30,8	0,0	0,0	0,0
	«пар»	3,3	0,2	0,8	0,1	93,9	1,2	0,0	0,1
	«другое»	49,7	0,2	0,1	0,0	10,3	39,4	0,0	0,0
	«кукуруза»	2,4	8,3	0,0	0,0	0,0	4,8	83,3	0,0
	«травы»	6,5	3,8	0,0	0,0	3,3	4,3	0,0	82,3

Таблица 3. Нормализованная матрица ошибок (DTC, EVI)

	Класс	Реальность							
		«залежь»	«soя»	«гречиха»	«овёс»	«пар»	«другое»	«кукуруза»	«травы»
Прогноз	«залежь»	98,0	0,5	0,0	0,0	0,8	0,6	0,0	0,0
	«soя»	3,8	95,8	0,2	0,0	0,1	0,0	0,0	0,0
	«гречиха»	6,2	3,3	88,9	0,0	1,5	0,0	0,0	0,1
	«овёс»	2,6	1,1	0,0	70,7	22,5	2,6	0,5	0,0
	«пар»	5,8	0,3	0,7	0,1	89,6	2,7	0,0	0,2
	«другое»	63,9	0,1	0,1	0,0	10,9	22,9	0,0	0,0
	«кукуруза»	5,8	7,3	0,0	0,0	0,0	0,0	87,0	0,0
	«травы»	6,2	1,5	0,8	0,0	1,8	0,2	0,0	89,4

Таблица 4. Нормализованная матрица ошибок (RF, NDVI)

	Класс	Реальность							
		«залежь»	«soя»	«гречиха»	«овёс»	«пар»	«другое»	«кукуруза»	«травы»
Прогноз	«залежь»	98,2	0,3	0,0	0,0	0,3	1,1	0,0	0,0
	«soя»	1,7	98,0	0,0	0,0	0,1	0,2	0,0	0,0
	«гречиха»	1,5	1,6	96,3	0,0	0,2	0,2	0,0	0,2
	«овёс»	4,7	6,5	0,0	76,6	10,3	0,9	0,0	0,0
	«пар»	1,8	0,0	0,5	0,0	96,6	0,7	0,0	0,1
	«другое»	21,8	0,3	0,0	0,0	2,7	75,0	0,0	0,0
	«кукуруза»	2,4	7,1	0,0	1,2	0,0	1,2	88,1	0,0
	«травы»	3,3	0,8	1,3	0,8	0,5	0,3	0,0	93,3

Таблица 5. Нормализованная матрица ошибок (RF, EVI)

	Класс	Реальность							
		«залежь»	«soя»	«гречиха»	«овёс»	«пар»	«другое»	«кукуруза»	«травы»
Прогноз	«залежь»	97,8	0,4	0,0	0,0	0,3	1,4	0,0	0,0
	«soя»	1,7	98,1	0,0	0,0	0,0	0,0	0,0	0,0
	«гречиха»	1,0	1,2	97,3	0,0	0,3	0,0	0,0	0,1
	«овёс»	4,0	2,5	0,0	84,5	7,5	1,5	0,0	0,0
	«пар»	3,3	0,2	0,6	0,2	94,4	1,0	0,0	0,2
	«другое»	25,7	0,1	0,0	0,0	3,8	70,2	0,0	0,0
	«кукуруза»	0,0	17,5	0,0	0,0	0,0	2,5	80,0	0,0
	«травы»	1,9	1,6	0,2	0,0	0,6	0,2	0,0	95,7

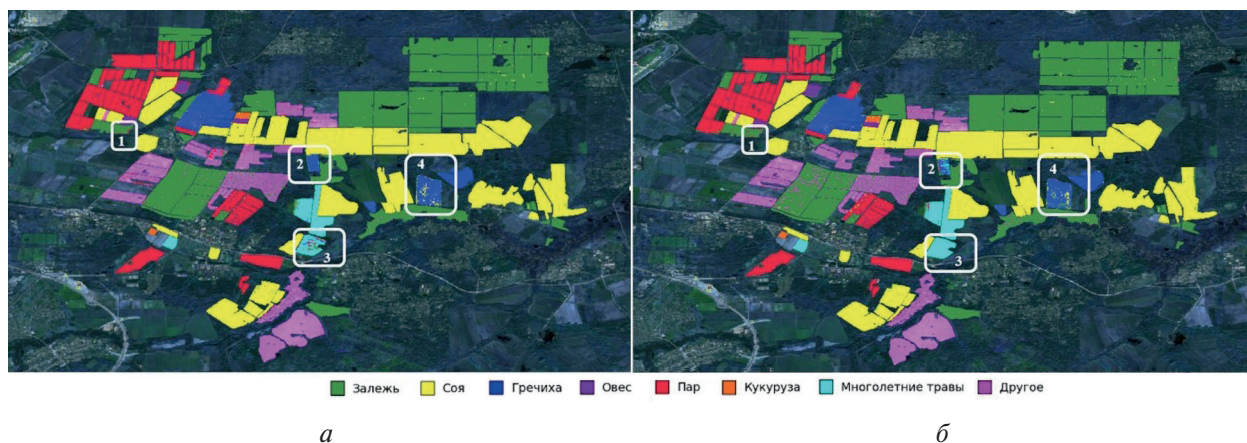


Рис. 2. Результаты попиксельной классификации (RF) сельскохозяйственных полей Хабаровского р-на в 2021 г.: *а* — на основе NDVI; *б* — на основе EVI

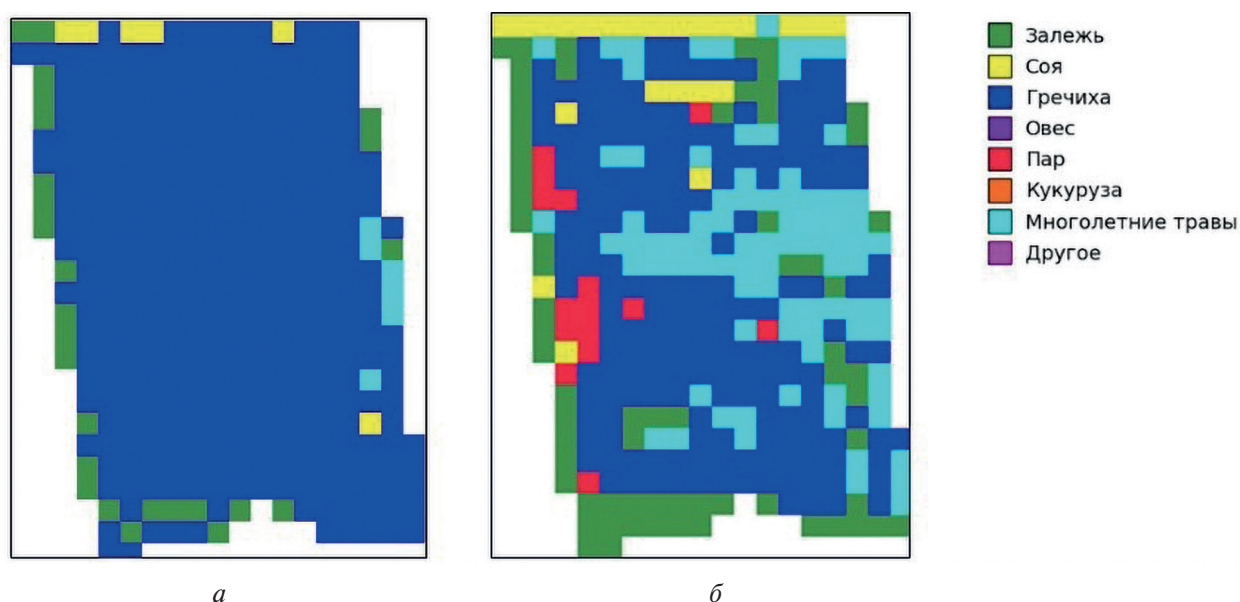


Рис. 3. Результаты попиксельной классификации (RF) поля с гречихой: *а* — на основе NDVI; *б* — на основе EVI

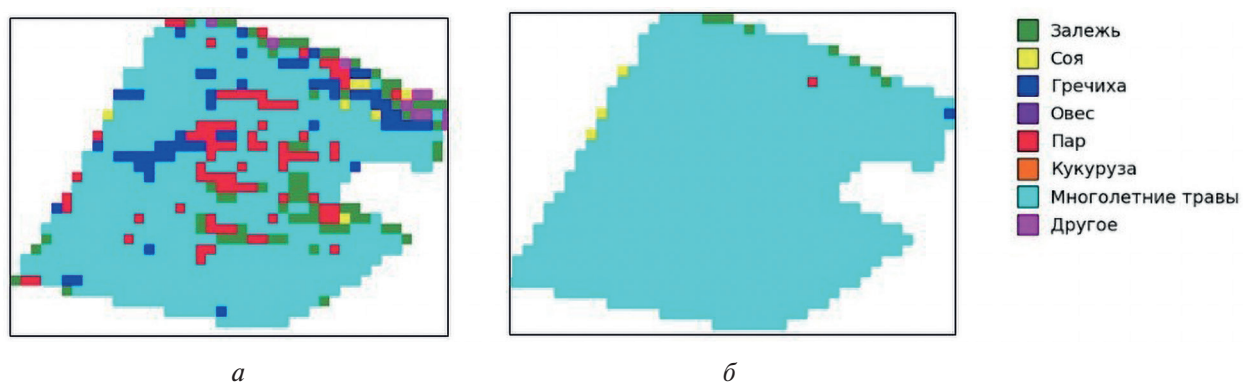


Рис. 4. Результаты попиксельной классификации (RF) поля с многолетними травами: *а* — на основе NDVI; *б* — на основе EVI

Результаты попиксельной классификации полей Хабаровского р-на на основе метода RF представлены на *рис. 2* (см. с. 161). Все поля были верно определены, при этом, как видно из рисунка, ошибочно идентифицированные пиксели могли располагаться как на краях поля (поле 1, см. *рис. 2*), так и в центре (поле 4, см. *рис. 2*). При этом фактически в 2021 г. на поле 4 с гречихой визуально действительно наблюдались не только участки с сорняками, но и с фрагментами сои и зерновых культур. В то же время сложно утверждать, что один из двух индексов — NDVI или EVI — имеет преимущество перед другим. Так, например, поле 2 (см. *рис. 2а, 3а*, см. с. 161) верно классифицировано с использованием NDVI как посеvy гречихи с небольшим вкраплением иных пикселей. С применением классификатора с EVI около 40 % пикселей на этом поле (поле 2, см. *рис. 2б, 3б*) распознано как многолетние травы. Напротив, на поле 3 (см. *рис. 2б, 4б*, см. с. 161) при использовании классификатора с EVI практически все пиксели были определены верно (многолетние травы), а классификатор с NDVI продемонстрировал определённую погрешность (поле 3, см. *рис. 2а, 4а*).

Для того чтобы избежать ошибки классификации, вызванной схожестью пикселей в рамках одного сельскохозяйственного поля и попаданием пикселей каждого поля как в тестовую, так и в обучающую выборку, дополнительно проводилась оценка классификатора на предварительно сформированной выборке из 8 отдельных полей (временные ряды пикселей этих полей не использовались при построении модели).

Таблица 6. Оценка точности (ОА, %) классификатора RF при попиксельной классификации отдельных сельскохозяйственных полей Хабаровского р-на в 2021 г.

Класс	Индекс вегетации	
	NDVI	EVI
«залежь»	95,6	98,5
«соя»	98,5	99,1
«гречиха»	92,7	79,8
«овёс»	89,3	94,9
«пар»	99,0	99,5
«другое»	52,4	88,1
«кукуруза»	86,3	84,5
«многолетние травы»	84,0	88,7

Как видно (*табл. 6*), значение ОА для сои, являющейся основной культурой юга российского Дальнего Востока, находилось в диапазоне 98,5–99,1 %. Также с высокой точностью, превышающей 95 % при использовании как NDVI, так и EVI, классифицировались пар и залежь. Все прочие культуры были классифицированы с точностью от 79,8 до 95 %, исключение составили только пиксели, отнесённые к классу «другое».

## Заключение

Исследования временных рядов NDVI и EVI показали, что разным типам растительности сельскохозяйственных полей Хабаровского р-на характерны соответствующие кривые сезонного хода, различающиеся в том числе по времени достижения и величине максимума. Наилучшие результаты классификации методами машинного обучения были достигнуты при использовании RF, при этом значение ОА составило 95,6 %. Не было выявлено значимых различий в точности классификации при применении разных индексов — NDVI и EVI — и при оценке ошибок распознавания основных классов: по результатам кроссвалидации значения ОА для классов «залежь», «соя», «гречиха», «пар» и «многолетние травы» превышали 90 %. При попиксельной классификации случайным образом отобранных полей точность классификации сельхозкультур была выше 79,8 %.



В целом полученные результаты свидетельствуют о возможности использования NDVI и EVI для классификации пахотных земель южной части Дальнего Востока, при этом в качестве базы должны рассматриваться кривые сезонного хода именно этого региона, что обусловлено особенностями вегетационных циклов сельхозкультур.

Исследования проведены с использованием ресурсов Центра коллективного пользования научным оборудованием «Центр обработки и хранения научных данных ДВО РАН», финансируемого Министерством науки и высшего образования РФ по соглашению № 075-15-2021-663.

## Литература

1. Арзамасцева Н. В., Прохорова Н. В., Хамидова Л. Л. Проблема достоверности и полноты информации о состоянии и использовании земель сельскохозяйственного назначения // Изв. Тимирязевской с.-х. акад. 2021. № 3. С. 119–128. DOI: 10.26897/0021-342X-2021-3-119-128.
2. Буланов К. А., Денисов П. В., Лупян Е. А., Мартыанов А. С., Середа И. И., Трошко К. А., Толпин В. А., Барталев С. А., Хвостиков С. А. Блок работы с данными дистанционного зондирования Земли Единой федеральной информационной системы о землях сельскохозяйственного назначения // Современные проблемы дистанционного зондирования Земли из космоса. 2019. Т. 16. № 3. С. 171–182. DOI: 10.21046/2070-7401-2019-16-3-171-182.
3. Лупян Е. А., Савин И. Ю., Барталев С. А., Толпин В. А., Балашов И. В., Плотников Д. Е. Спутниковый сервис мониторинга состояния растительности («ВЕГА») // Современные проблемы дистанционного зондирования Земли из космоса. 2011. Т. 8. № 1. С. 190–198.
4. Лупян Е. А., Барталев С. А., Толпин В. А., Жарко В. О., Крашенинникова Ю. С., Оксюкевич А. Ю. Использование спутникового сервиса ВЕГА в региональных системах дистанционного мониторинга // Современные проблемы дистанционного зондирования Земли из космоса. 2014. Т. 11. № 3. С. 215–232.
5. Миклашевич Т. С., Барталев С. А., Плотников Д. Е. Интерполяционный алгоритм восстановления длинных временных рядов данных спутниковых наблюдений растительного покрова // Современные проблемы дистанционного зондирования Земли из космоса. 2019. Т. 16. № 6. С. 143–154. DOI: 10.21046/2070-7401-2019-16-6-143-154.
6. Плотников Д. Е., Колбудаев П. А., Барталев С. А., Лупян Е. А. Автоматическое распознавание используемых пахотных земель на основе сезонных временных серий восстановленных изображений Landsat // Современные проблемы дистанционного зондирования Земли из космоса. 2018. Т. 15. № 2. С. 112–127. DOI: 10.21046/2070-7401-2018-15-2-112-127.
7. Якушев В. П., Захарян Ю. Г., Блохина С. Ю. Состояние и перспективы использования дистанционного зондирования Земли в сельском хозяйстве // Современные проблемы дистанционного зондирования Земли из космоса. 2022. Т. 19. № 1. С. 287–294. DOI: 10.21046/2070-7401-2022-19-1-287-294.
8. Bartalev S., Egorov V., Loupian E., Khvostikov S. A new locally-adaptive classification method LAGMA for large-scale land cover mapping using remote-sensing data // Remote Sensing Letters. 2014. V. 5. No. 1. P. 55–64. DOI: 10.1080/2150704X.2013.870675.
9. Bellon B., Begue A., Lo Seen D., Lebourgeois V., Evangelista B. A., Simoes M., Ferraz R. Improved regional-scale Brazilian cropping systems' mapping based on a semi-automatic object-based clustering approach // Intern. J. Applied Earth Observation and Geoinformation. 2018. V. 68. P. 127–138. DOI: 10.1016/j.jag.2018.01.019.
10. Chen Y., Lu D., Moran E., Batistella M., Dutra L., Del'Arco L., da Silva R., Huang J., Luiz A., de Oliveira M. Mapping croplands, cropping patterns, and crop types using MODIS time-series data // Intern. J. Applied Earth Observation and Geoinformation. 2018. V. 69. P. 133–147. DOI: 10.1016/j.jag.2018.03.005.
11. Griffiths P., Nendel C., Hostert P. Intra-annual reflectance composites from Sentinel-2 and Landsat for national-scale crop and land cover mapping // Remote Sensing of Environment. 2019. V. 220. P. 135–151. DOI: 10.1016/j.rse.2018.10.031.
12. Feyisa G. L., Palao L. K., Nelson A., Gumma M. K., Paliwal A., Win K. T., Nge K. H., Johnson D. E. Characterizing and mapping cropping patterns in a complex agro-ecosystem: An iterative participatory mapping procedure using machine learning algorithms and MODIS vegetation indices // Computers and Electronics in Agriculture. 2020. V. 175. Art. No. 105595. DOI: 10.1016/j.compag.2020.105595.
13. Hao P., Tang H., Chen Z., Meng Q., Kang Y. Early-season crop type mapping using 30-m reference time series // J. Integrative Agriculture. 2020. V. 19. No. 7. P. 1897–1911. DOI: 10.1016/S2095-3119(19)62812-1.

14. Liu J., Shang J., Qian B., Huffman T., Zhang Y., Dong T., Jing Q., Martin T. Crop yield estimation using time-series MODIS data and the effects of cropland masks in Ontario, Canada // *Remote Sensing*. 2019. V. 11. Art. No. 2419. DOI: 10.3390/rs11202419.
15. Tolpin V., Loupian E., Bartalev S., Plotnikov D., Matveev A. Possibilities of agricultural vegetation condition analysis with the “VEGA” satellite service // *Atmospheric and Oceanic Optics*. 2014. V. 27. P. 581–586.
16. Tufail R., Ahmad A., Javed M., Sajid R. A machine learning approach for accurate crop type mapping using combined SAR and optical time series data // *Advances in Space Research*. 2021. V. 69. Iss. 1. P. 331–346. DOI: 10.1016/j.asr.2021.09.019.
17. Yan S., Yao X., Zhu D., Liu D., Zhang L., Yu G., Gao B., Yang J., Yun W. Large-scale crop mapping from multi-source optical satellite imagery using machine learning with discrete grids // *Intern. J. Applied Earth Observation and Geoinformation*. 2021. V. 103. Art. No. 102485. DOI: 10.1016/j.jag.2021.102485.
18. Zhang H., Kang J., Xu X., Zhang L. Accessing the temporal and spectral features in crop type mapping using multi-temporal Sentinel-2 imagery: A case study of Yi'an County, Heilongjiang province, China // *Computers and Electronics in Agriculture*. 2020. V. 176. Art. No. 105618. DOI: 10.1016/j.compag.2020.105618.

## Recognition of crops in Khabarovsk Krai using NDVI and EVI

L. V. Illarionova<sup>1</sup>, A. S. Stepanov<sup>2</sup>, E. A. Fomina<sup>1</sup>

<sup>1</sup> *Computing Center FEB RAS, Khabarovsk 680000, Russia*  
*E-mail: illarionova\_l@list.ru*

<sup>2</sup> *Far Eastern Agriculture Research Institute, Khabarovsk 680521, Russia*  
*E-mail: stepanfx@mail.ru*

The development of approaches to the identification and classification of crops at the regional level is one of the most important tasks of digital farming. The solution of this problem using remote sensing data is especially important for the southern part of Far East, which is due to differences in the timing of sowing, harvesting, the duration of the vegetation phases of crops in the macroregion in comparison with the western part of the Russian Federation, and, accordingly, a decrease in the accuracy of existing algorithms. To carry out the classification, data on the crop rotation of agricultural fields of the Khabarovsk region with a total area of more than 4000 hectares were used, 8 classes of arable land were studied. We reviewed 37 multispectral images with a resolution of 20 m obtained from Sentinel-2A/B satellites from April to October 2021. For each of the pixels, time series of NDVI (Normalized Difference Vegetation Index) and EVI (Enhanced Vegetation Index) values were formed, and approximate seasonal variation curves were constructed using the Fourier function. The best machine learning classification results were achieved using Random Forest, with an overall accuracy of 95 %. There were no significant differences in the accuracy of classification when using the values of different indices as input data — NDVI and EVI. According to the results of cross-validation, it was found that the accuracy of recognition of the main classes is at a high level: the error in determining the classes of soy, buckwheat, fallow, steam and perennial grasses did not exceed 10 %. The results obtained confirmed the possibility of using NDVI and EVI to classify arable lands in the southern part of Far East, while the curves of the seasonal course of this particular region should be considered as a base, which is due to the peculiarities of the vegetation cycles of crops.

**Keywords:** classification, recognition, arable land, vegetation index, Far East, machine learning, approximation

Accepted: 15.03.2023

DOI: 10.21046/2070-7401-2023-20-2-155-165

## References

1. Arzamastseva N. V., Prokhorova N. V., Khamidova L. L., Problem of the accuracy and completeness of information on the state and use of agricultural land, *Izvestiya Timiryazevskoi sel'skokhozyaistvennoi akademii*, 2021, No. 3, pp. 119–128 (in Russian), DOI: 10.26897/0021-342X-2021-3-119-128.
2. Bulanov K. A., Denisov P. V., Lupyayev E. A., Martyanov A. S., Sereda I. I., Troshko K. A., Tolpin V. A., Bartalev S. A., Khvostikov S. A., Remote sensing unit of Integrated Federal Information System on Agricultural Lands, *Sovremennye problemy distantsionnogo zondirovaniya Zemli iz kosmosa*, 2019, Vol. 16, No. 3, pp. 171–182 (in Russian), DOI: 10.21046/2070-7401-2019-16-3-171-182.
3. Loupian E., Savin I., Bartalev S., Tolpin V., Balashov I., Plotnikov D., Satellite service for vegetation monitoring VEGA, *Sovremennye problemy distantsionnogo zondirovaniya Zemli iz kosmosa*, 2011, Vol. 8, No. 1, pp. 190–198 (in Russian).
4. Loupian E. A., Bartalev S. A., Tolpin V. A., Zharko V. O., Krashenninnikova Yu. S., Oksyukevich A. Yu., VEGA satellite service applications in regional remote monitoring systems, *Sovremennye problemy distantsionnogo zondirovaniya Zemli iz kosmosa*, 2014, Vol. 11, No. 3, pp. 215–232 (in Russian).
5. Miklashevich T. S., Bartalev S. A., Plotnikov D. E., Interpolation algorithm for the recovery of long satellite data time series of vegetation cover observation, *Sovremennye problemy distantsionnogo zondirovaniya Zemli iz kosmosa*, 2019, Vol. 16, No. 6, pp. 143–154 (in Russian), DOI: 10.21046/2070-7401-2019-16-6-143-154.
6. Plotnikov D. E., Kolbudaev P. A., Bartalev S. A., Loupian E. A., Automated annual cropland mapping from reconstructed time series of Landsat data, *Sovremennye problemy distantsionnogo zondirovaniya Zemli iz kosmosa*, 2018, Vol. 15, No. 2, pp. 112–127 (in Russian), DOI: 10.21046/2070-7401-2018-15-2-112-127.
7. Yakushev V. P., Zakharyan Yu. G., Blokhina S. Yu., Current problems and prospects for the use of remote sensing of the Earth in agriculture, *Sovremennye problemy distantsionnogo zondirovaniya Zemli iz kosmosa*, 2022, Vol. 19, No. 1, pp. 287–294 (in Russian), DOI: 10.21046/2070-7401-2022-19-1-287-294.
8. Bartalev S., Egorov V., Loupian E., Khvostikov S., A new locally-adaptive classification method LAGMA for large-scale land cover mapping using remote-sensing data, *Remote Sensing Letters*, 2014, Vol. 5, No. 1, pp. 55–64, DOI: 10.1080/2150704X.2013.870675.
9. Bellon B., Begue A., Lo Seen D., Lebourgeois V., Evangelista B. A., Simoes M., Ferraz R., Improved regional-scale Brazilian cropping systems' mapping based on a semi-automatic object-based clustering approach, *Intern. J. Applied Earth Observation and Geoinformation*, 2018, Vol. 68, pp. 127–138, DOI: 10.1016/j.jag.2018.01.019.
10. Chen Y., Lu D., Moran E., Batistella M., Dutra L., Del'Arco L., da Silva R., Huang J., Luiz A., de Oliveira M., Mapping croplands, cropping patterns, and crop types using MODIS time-series data, *Intern. J. Applied Earth Observation and Geoinformation*, 2018, Vol. 69, pp. 133–147, DOI: 10.1016/j.jag.2018.03.005.
11. Griffiths P., Nendel C., Hostert P., Intra-annual reflectance composites from Sentinel-2 and Landsat for national-scale crop and land cover mapping, *Remote Sensing of Environment*, 2019, Vol. 220, pp. 135–151, DOI: 10.1016/j.rse.2018.10.031.
12. Feyisa G. L., Palao L. K., Nelson A., Gumma M. K., Paliwal A., Win K. T., Nge K. H., Johnson D. E., Characterizing and mapping cropping patterns in a complex agro-ecosystem: An iterative participatory mapping procedure using machine learning algorithms and MODIS vegetation indices, *Computers and Electronics in Agriculture*, 2020, Vol. 175, Art. No. 105595, DOI: 10.1016/j.compag.2020.105595.
13. Hao P., Tang H., Chen Z., Meng Q., Kang Y., Early-season crop type mapping using 30-m reference time series, *J. Integrative Agriculture*, 2020, Vol. 19, No. 7, pp. 1897–1911, DOI: 10.1016/S2095-3119(19)62812-1.
14. Liu J., Shang J., Qian B., Huffman T., Zhang Y., Dong T., Jing Q., Martin T., Crop yield estimation using time-series MODIS data and the effects of cropland masks in Ontario, Canada, *Remote Sensing*, 2019, Vol. 11, Art. No. 2419, DOI: 10.3390/rs11202419.
15. Tolpin V., Loupian E., Bartalev S., Plotnikov D., Matveev A., Possibilities of agricultural vegetation condition analysis with the “VEGA” satellite service, *Atmospheric and Oceanic Optics*, 2014, Vol. 27, pp. 581–586.
16. Tufail R., Ahmad A., Javed M., Sajid R., A machine learning approach for accurate crop type mapping using combined SAR and optical time series data, *Advances in Space Research*, 2021, Vol. 69, Issue 1, pp. 331–346, DOI: 10.1016/j.asr.2021.09.019.
17. Yan S., Yao X., Zhu D., Liu D., Zhang L., Yu G., Gao B., Yang J., Yun W., Large-scale crop mapping from multi-source optical satellite imagery using machine learning with discrete grids, *Intern. J. Applied Earth Observation and Geoinformation*, 2021, Vol. 103, Art. No. 102485, DOI: 10.1016/j.jag.2021.102485.
18. Zhang H., Kang J., Xu X., Zhang L., Accessing the temporal and spectral features in crop type mapping using multi-temporal Sentinel-2 imagery: A case study of Yi'an County, Heilongjiang province, China, *Computers and Electronics in Agriculture*, 2020, Vol. 176, Art. No. 105618, DOI: 10.1016/j.compag.2020.105618.