

УДК 528.8

## ПРИМЕНЕНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ КАРТОГРАФИРОВАНИЯ ЛЕСНОЙ РАСТИТЕЛЬНОСТИ РЕСПУБЛИКИ ТАТАРСТАН

**А. М. Гафуров, М. В. Кожевникова, В. Е. Прохоров**

*Институт экологии и природопользования, Казанский федеральный университет,  
ул. Товарищеская. 5, 42097, Казань, Россия, [busmanof@kpfu.ru](mailto:busmanof@kpfu.ru)*

В исследовании используются снимки Landsat и алгоритм WekaXMeans для картографирования лесной растительности в соответствии с системой Браун-Бланке. Кластеризация с использованием 288 вегетационных индексов и метрик позволил выделить 7 классов по системе Браун-Бланке, что было проверено на 17000 релевантных данных базы данных «Флора». Данный подход способствует проведению экологических исследований и планированию природоохранных мероприятий, позволяя получить представление об экологических характеристиках лесных сообществ.

**Ключевые слова:** неконтролируемая классификация; пространственное моделирование; мультиспектральные снимки; алгоритм кластеризации Weka X-Means; система классификации Брауна-Бланке; GoogleEarthEngine; снимки Landsat.

## MACHINE LEARNING METHODS FOR FOREST VEGETATION MAPPING IN THE REPUBLIC OF TATARSTAN

**A. M. Gafurov, M. V. Kozhevnikova, V. E. Prokhorov**

*Institute of Ecology and Environmental Management, Kazan Federal University,  
Tovarisheskaya str., 5, 42097, Kazan, Russia, [busmanof@kpfu.ru](mailto:busmanof@kpfu.ru)*

This study leverages Landsat imagery and WekaXMeans for unsupervised classification in forest vegetation mapping, aligning with the Braun-Blanquet system. A two-step clustering method using vegetation indices classifies forest types, validated against 17,000 "Flora" database relevés, identifying 44 community types into 7 Braun-Blanquet classes. This approach enhances ecological research and conservation planning, offering insights into forest communities' ecological characteristics and sequestration potential in Tatarstan.

**Keywords:** unsupervised classification; spatial modeling; multispectral imagery; Weka X-Means clustering algorithm; Braun-Blanquet classification system; google earth engine; Landsat imagery.

Исследование посвящено совершенствованию систем классификации растительности путем интеграции методов классификации без контроля со

спутниковыми снимками с целью создания пространственной модели лесной растительности в рамках системы Браун-Бланке. Данное исследование соответствует глобальной тенденции к стандартизации процедур классификации для обеспечения последовательности и объективности различных экологических исследований [1].

Основная цель — разработать более точный и эффективный метод классификации и картирования типов растительности, в частности лесных сообществ, с использованием методов классификации без контроля [2]. Исследование направлено на минимизацию субъективности исследователей и повышение надежности карт растительности, которые имеют решающее значение для экологических, природоохранных и сельскохозяйственных приложений [3].

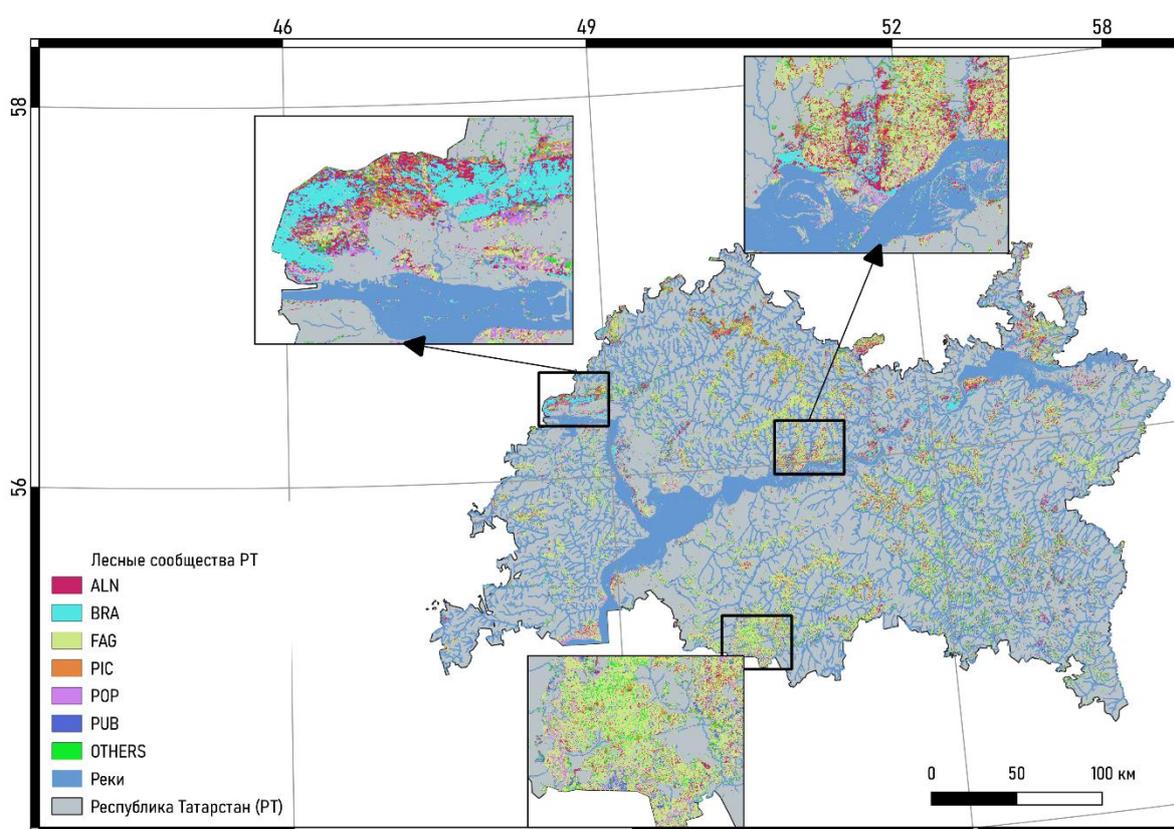
Методология включает использование спутниковых снимков Landsat 8 и 9 и применение двухэтапного подхода, сочетающего классификацию Дженкса и агломеративную иерархическую кластеризацию (АИК), реализованную в алгоритме WekaXMeans [4]. Этот подход разработан для оптимизации точности типов растительного покрова, полученных на основе спутниковых вегетационных индексов. Всего в модели было использовано 288 параметров, состоящих из 72 вегетационных индексов и их статистических метрик. Кроме того, в исследовании используется маска лесов, полученная из модели GlobalLandAnalysisandDiscovery (GLAD) [5], позволяющая сосредоточиться исключительно на лесных видах, что обеспечивает направленность анализа на основной интерес исследования.

Применение методов классификации без контроля привело к выделению 44 различных классов лесных сообществ в Республике Татарстан, Россия. Эти классы отражают не только видовой состав в группах, но и экологические условия произрастания, и его характер, например, сомкнутость кроны или рост на переувлажненных грунтах. В ходе исследования была успешно создана подробная пространственная модель, соотносящая эти классы с системой классификации Брауна-Бланке. Для этого была проведена работа по соотнесения полученных кластеров с геоботаническими описаниями из базы данных «Флора» [6], содержащая более 17000 описаний, что позволило однозначно выделить 6 классов по классификации Брауна-Бланке и один класс, требующий отдельной полевой верификации (табл. 1). Эта модель (рис. 1) является значительным шагом вперед в точном картировании и понимании динамики лесной растительности и демонстрирует потенциал неконтролируемой классификации в снижении погрешностей при классификации растительности [7].

**Матрица сопоставления классификации Брауна-Бланке (ББ)  
и классов по кластерной модели**

Сообщества по ББ	Классы кластерной модели
ALN	4, 5, 7, 17, 28, 42, 43
BRA	6, 36
FAG	11, 13, 14, 18, 19, 22, 23, 24, 29, 31, 34, 35, 37, 38, 44
PIC	2, 30
POP	1, 26, 33, 36
PUB	14, 25, 32
OTHERS	*

\* - остальные классы



*Рис. 1. Карта лесных сообществ Республики Татарстан  
по классификации Брауна-Бланке*

Так же проведена работа по определению предельного числа параметров, после которого точность классификации значительно падает. Для этого все предикторы были ранжированы по статистической важности (табл. 2) и итеративно удалялись из модели до тех пор, пока точность классификации не падала более чем на 5 %. Это позволило выявить, что точность модели начинает ухудшаться уже после удаления 9 наименее

значимых параметра (рис. 2). Полученные результаты подчеркивают необходимость дальнейшего развития и совершенствования алгоритмов классификации для дальнейшего снижения субъективности в экологическом картировании.

Таблица 2

**Топ-10 наиболее важных и репрезентативных признаков для классификации лесных сообществ**

Предиктор	Важность	F-Value	P-Value
TriVI_max	1.319 %	1093.94	< 0.0000
TriVI_stdDev	1.175 %	1046.95	< 0.0000
TGI_mean	0.934 %	435.02	< 0.0000
SEVI_median	0.919 %	319.74	< 0.0000
TGI_median	0.871 %	541.51	< 0.0000
FCVI_max	0.870 %	1075.45	< 0.0000
SR_median	0.828 %	580.48	< 0.0000
GRVI_median	0.792 %	607.80	< 0.0000
TriVI_mean	0.748 %	910.28	< 0.0000
NDYI_median	0.732 %	202.37	< 0.0000

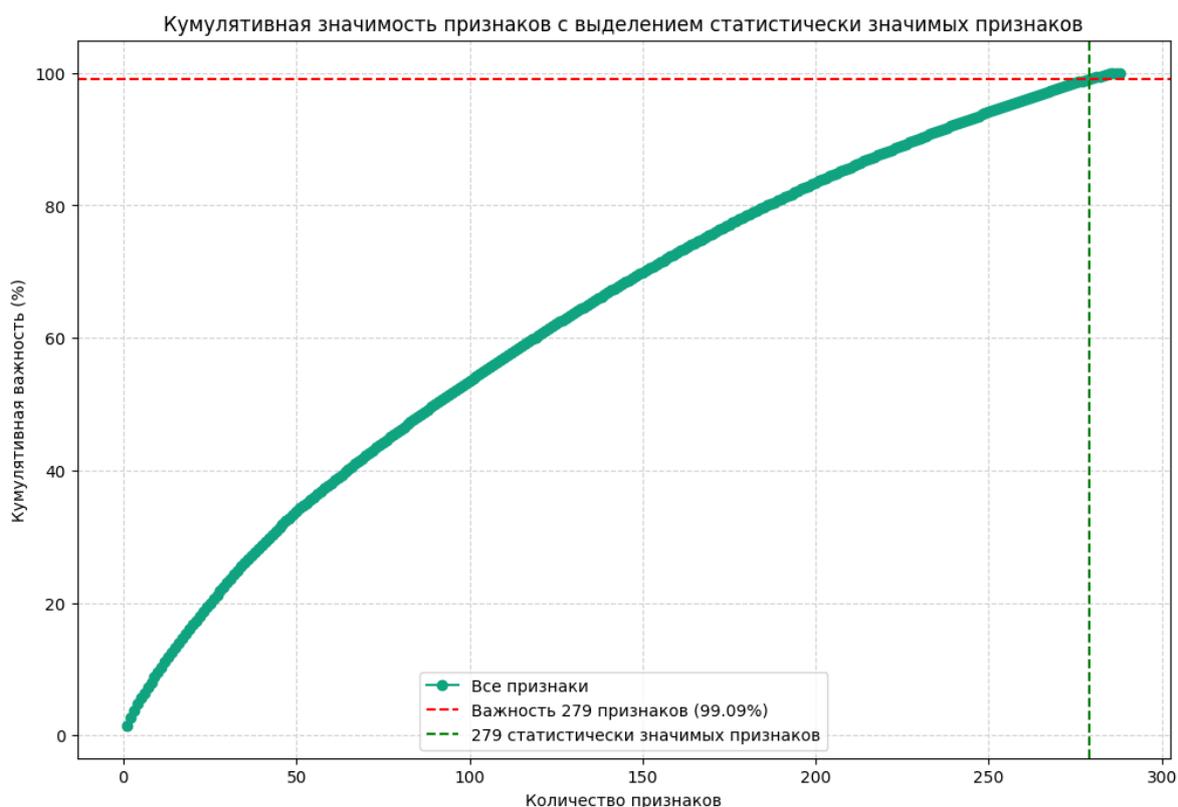


Рис. 2. Кумулятивная важность предикторов модели лесных сообществ РТ

Будущие исследования могут изучить интеграцию дополнительных источников данных дистанционного зондирования и передовых алгоритмов машинного обучения для повышения точности классификации [8]. Кроме того, исследование подчеркивает важность создания стандартизированных, воспроизводимых методов классификации растительности, которые могут применяться во всем мире, способствуя более эффективному управлению и сохранению природных ресурсов.

Работа выполнена за счет средств субсидии, выделенной Казанскому федеральному университету для выполнения государственного задания в сфере научной деятельности, проект № FZSM-2024-0004.

### Библиографические ссылки

1. *Kozhevnikova M., Prokhorov V.* Syntaxonomy of the xero-mesophytic oak forests in the Republic of Tatarstan (Eastern Europe) // *Vegetation Classification and Survey*. 2021. Т. 2. Р.47–58.

2. *Гафуров А. М.* Использование глубоких нейронных сетей для картографирования сельскохозяйственных культур Республики Татарстан Казань // *Актуальные вопросы геодезии и геоинформационных систем: Тезисы и доклады*. 2019. С. 84–90.

3. *Geography and Geocology of Russia in the Mosaic of River Basins / O. P. Yermolaev [et al.] // Geography and Natural Resources*. 2023. Т. 44. № 3. Р. 208–214.

4. *Pelleg D.* X-means: Extending k-means with efficient estimation of the number of clusters // *International Conference on Machine Learning*. 2000. Р. 727–734.

5. *Global land use extent and dispersion within natural land cover using Landsat data / M. C. Hansen [et al.] // Environmental Research Letters*. 2022. Т. 17. № 3. С.034050.

6. *Prokhorov V., Rogova T., Kozhevnikova M.* Vegetation Database of Tatarstan // *Phytocoenologia*. 2017. Т. 47. № 3. Р. 309–313.

7. Использование архивных пространственных данных при анализе истории землепользования территории болгарских городищ Предволжья Республики Татарстан / *М. А. Иванов [и др.] // Региональные Геосистемы*. 2020. Т. 44. № 4. С. 388–403.

8. *Geospatial modelling of changes in land use/land cover dynamics using Multi-layer Perceptron Markov chain model in Rajshahi City, Bangladesh / N. N. Dey [et al.] // Environmental Challenges*. 2021. Т. 4. Р.100148.