

М.А. Гендельманның 110 жылдығына арналған «Сейфуллин оқулары – 19» халықаралық ғылыми - практикалық конференциясының материалдары = Материалы международной научно-практической конференции «Сейфуллинские чтения – 19, посвященной 110-летию М.А. Гендельмана». - 2023.- Т. I, Ч. IV. – С. 20-22.

**УДК 631.1(528.88)**

## **ОЦЕНКА ТОЧНОСТИ КЛАССИФИКАЦИИ ЗЕМЛЕПОЛЬЗОВАНИЯ И ЗЕМНОГО ПОКРОВА С ИСПОЛЬЗОВАНИЕМ АЛГОРИТМА RANDOM FOREST**

*Мусайф Г., докторант*

*Казахский агротехнический исследовательский университет им. С.*

*Сейфуллина, г.Астана*

Формирование и развитие инфраструктуры пространственно-временных данных (ИПД) предполагает автоматическую классификацию землепользования и земного покрова (ЗЗП) как одной из её функций [1]. Однако, классификация ЗЗП достаточно сложный процесс, которая нуждается в оценке точности используемых алгоритмов [2,3]. Прежде всего их делят на контролируемые и неконтролируемые. Неконтролируемая классификация используется, когда данных очень мало для классификации изображения. Контролируемые виды классификации, имеют определенные преимущества перед неконтролируемыми и в свою очередь состоят из параметрических и непараметрических методов. Особенность параметрических алгоритмов заключается в том, что для каждого класса в каждой полосе имеется нормальное распределение.

На данном этапе развития дистанционного зондирования Земли (ДЗЗ) более широко используются непараметрические методы классификации ЗЗП, основанные на машинном обучении. Чтобы выбрать метод классификации ЗЗП, перед началом этого процесса мы сравнили Random forest (RF), CART и SVM. RF использует построение большого количества решений и применение голосования к результатам. Основным примером используется для восстановления  $N$  учебных комплектов из аутентичного набора данных. Далее для каждой обучающей выборки строится дерево решений («лес»). Каждое дерево решений автономно и не связано с другим [4]. CART основан на построении дерева решений, аналогичного в принципе методу RF. В отличие от RF, CART представляет собой единое дерево решений и не объединяет большое количество деревьев решений [5]. SVM использует построение правила классификации с помощью опорных векторов и имеет более низкую точность классификации вблизи границы, разделяющей классы [6]. Поэтому мы использовали RF в качестве основного алгоритма классификации LULC по изображениям Landsat 7/8 с 1991 по 2021 годы.

Исходя из этого целью наших исследований была проверка применимости алгоритма Random Forest для контролируемой классификации

мультивременного ряда ЗЗП Бурабайского района Акмолинской области. Задачи работы сводились к определению общей точности классификации, коэффициента Каппа и ошибок пользователя и потребителя. Для данного исследования были выбраны пять классов ЗЗП: пашня, пастбища, водные объекты, лес и застроенные территории.

Оценка точности классификации по алгоритму Random Forest была определена с помощью матрицы ошибок. В матрице ошибок общий показатель достоверности классификации определяется как количество правильно классифицированных точек, расположенных по диагонали таблицы. На самом деле это число может быть случайным. Чтобы учесть этот факт, часто при обобщении результатов используют так называемый коэффициент или индекс К - Каппа, вносящий поправку на случайность [2].

Вычисление общей точности и коэффициента Каппа показали (таблица 1), что результаты классификации являются достаточно надежными. Так общая точность классификации изменялись в пределах 0,85-0,98, а коэффициент Каппа для классифицированных изображений от 0,85, до 0,97. Средняя общая точность этого набора данных составила  $0,92 \pm 0,044$ , а для Каппа  $0,89 \pm 0,05$ ; для точности пользователя  $0,94 \pm 0,03$ ; точность производителя  $0,94 \pm 0,03$ .

Таблица 1 - Оценка точности ЗЗП с 1999 по 2021 годы

| Годы    | Точность  |           |              |               |
|---------|-----------|-----------|--------------|---------------|
|         | Общая     | Карра     | Пользователя | Производителя |
| 1999    | 0,97      | 0,96      | 0,96         | 0,96          |
| 2000    | 0,92      | 0,89      | 0,93         | 0,93          |
| 2001    | 0,90      | 0,84      | 0,94         | 0,94          |
| 2002    | 0,91      | 0,88      | 0,93         | 0,93          |
| 2003    | 0,90      | 0,86      | 0,92         | 0,92          |
| 2004    | 0,93      | 0,91      | 0,96         | 0,96          |
| 2005    | 0,95      | 0,94      | 0,96         | 0,96          |
| 2006    | 0,89      | 0,85      | 0,90         | 0,90          |
| 2007    | 0,93      | 0,91      | 0,95         | 0,95          |
| 2008    | 0,89      | 0,85      | 0,93         | 0,93          |
| 2009    | 0,85      | 0,79      | 0,91         | 0,91          |
| 2010    | 0,93      | 0,91      | 0,94         | 0,94          |
| 2011    | 0,88      | 0,84      | 0,90         | 0,90          |
| 2012    | 0,86      | 0,80      | 0,91         | 0,91          |
| 2013    | 0,93      | 0,90      | 0,91         | 0,91          |
| 2014    | 0,87      | 0,82      | 0,90         | 0,90          |
| 2015    | 0,94      | 0,93      | 0,96         | 0,96          |
| 2016    | 0,98      | 0,97      | 0,97         | 0,97          |
| 2017    | 0,95      | 0,94      | 0,97         | 0,97          |
| 2018    | 0,94      | 0,92      | 0,95         | 0,95          |
| 2019    | 0,94      | 0,92      | 0,95         | 0,95          |
| 2020    | 0,98      | 0,97      | 0,98         | 0,98          |
| 2021    | 0,97      | 0,97      | 0,97         | 0,97          |
| Average | 0,92±0,04 | 0,89±0,05 | 0,94±0,03    | 0,94±0,03     |

Таблица 2 - Оценка точности различных типов ЗЗП

| Классы ЗЗП             | Точность     |               |
|------------------------|--------------|---------------|
|                        | Пользователя | Производителя |
| Посевы                 | 0,87±0,09    | 0,88±0,09     |
| Пастбища               | 0,86±0,06    | 0,87±0,08     |
| Лес                    | 0,97±0,05    | 0,99±0,01     |
| Водные объекты         | 0,99±0,01    | 0,94±0,08     |
| Застроенные территории | 0,99±0,02    | 0,99±0,03     |

По разным видам землепользования (таблица 2) за период 1999–2021 гг. точность пользователей и точность производителей для лесов (0,97±0,005; 0,99±0,01), водоемов (0,99±0,01; 0,94±0,08) и застроенных территории (0,99±0,02; 0,99±0,03) были относительно высокими, но эти показатели для пашни (0,87±0,09; 0,88±0,09) и пастбищ были относительно низкими (0,86±0,06; 0,87±0,08).

Таким образом, приведенные результаты оценки точности классификации и их ошибки убедительно показывают надежность выполненной нами работы. Это открывает путь к автоматизации процессов тематического картографирования результатов классификации землепользования и созданию аграрной инфраструктуры пространственно-временных данных.

Исследование выполнено в рамках программного-целевого финансирования по линии МСХ РК согласно научно-технической программе «Исследование влияния государственной политики в отрасли сельского хозяйства на развитие кооперационных процессов в АПК, устойчивого развития сельских территорий и обеспечения продовольственной безопасности» ИРН BR10764919 под руководством д.б.н., профессора О.Ә.Әліпбеки.

### Список литературы

- 1 Алипбеки О.А., Алипбекова Ч.А. Разработка пространственных данных: создание и формирование [Текст] / Нур-Султан, 2020, Издательство: КазАТУ им. С.Сейфуллина. - С.340. ISBN 978-601-257-284-1.
- 2 Лурье И.К., Косиков А.Г. Теория и практика цифровой обработки изображений [Текст]: Под ред. А.М. Берлянта. - М.: Научный мир, 2003. - 168 с. ISBN 5-89176-231-5. - Текст: непосредственный.
- 3 Jensen, J.R. Introductory Digital Image Processing: A Remote Sensing Perspective [Text]: 3th edition. 2005.- 543p.
- 4 Matarira, D.; Mutanga, O.; Naidu, M. Google Earth Engine for Informal Settlement Mapping: A Random Forest Classification Using Spectral and Textural Information [Text] / *Remote Sens.* - 2022. - №14. - P.5130. <https://doi.org/10.3390/rs14205130>
- 5 Wei-Yin, L. Fifty Years of Classification and Regression Trees [Text] / *International Statistical Review*, - 201. - № 82. - P. 329–348. <https://doi.org/10.1111/insr.12016>
- 6 Support Vector Machines in Machine Learning (SVM): 2023 Guide <https://www.knowledgehut.com/blog/data-science/support-vector-machines-in-machine-learning>

Научный руководитель: д.б.н., профессор Әліпбеки О.Ә.