

УДК 004.048: 631.46

<https://doi.org/10.20914/2304-4691-2024-4-21-22>

**ПРИМЕНЕНИЕ МОДЕЛЕЙ ГЛУБОКОГО ОБУЧЕНИЯ ПРИ РЕАЛИЗАЦИИ
ВЫСОКОПРОИЗВОДИТЕЛЬНОГО ФЕНОТИПИРОВАНИЯ В СЕЛЕКЦИИ
СЕЛЬСКОХОЗЯЙСТВЕННЫХ КУЛЬТУР**

А.В. Линкина, В.И. Елсуков

Воронежский институт высоких технологий, г. Воронеж, Россия

Одним из стратегических направлений в отечественном агропромышленном комплексе в условиях санкционных рисков в настоящее время является развитие селекционной науки для обеспечения продовольственной безопасности. Принята Федеральная программа научно-технологического развития сельского хозяйства до 2025 г. (утв. Постановлением Правительства Российской Федерации от 25 августа 2017 г. № 996), к основным приоритетам которой отнесены формирование условий и развитие научно-технологической базы отечественной селекции и семеноводства. Поэтому значительный импульс развития получает проведение высокопроизводительного фенотипирования, позволяющее проводить автоматизированный цифровой анализ больших выборок растений в рамках крупномасштабных селекционно-генетических исследований. Поскольку такой метод предполагает оценку большого количества признаков, отражающих закономерности взаимодействия морфологических, физиологических и иных характеристик биоценозов и условий внешних факторов среды, то весьма перспективным является применение алгоритмов машинного обучения. Кроме того, стоит отметить, что происходит накопление больших объемов данных в произвольном формате. Эта информация включает неструктурированные данные, изображения, отчеты на естественном языке. При высокопроизводительном фенотипировании (High-throughput phenotyping, НТР) используются данные дистанционного зондирования, что приводит к накоплению большого количества пространственных и

временных данных. Объединение данных, полученных с помощью технологии НТР на основе анализа изображений и информации от полевых датчиков, с геномными и другими омиксными данными, отражающими реакцию растений на условия окружающей среды, открывает возможности для оптимизации селекционных процессов и выявления новых фенотипов сельскохозяйственных культур, которые могут более эффективно использовать доступные ресурсы и проявлять устойчивость к изменяющимся климатическим условиям.

Большая часть данных, отвечающих за хозяйственные признаки растений, например, устойчивость к болезням растений и климатическим факторам, объем биомассы, урожайность и т. п., учитываются полигенно. Основная идея нового подхода при НТР состоит в том, что для создания объема выборки можно применять в том числе отдельные геномные локусы. Это позволяет при анализе выборок сократить ошибку идентификации QTL, а также увеличить размер выборки для методов поиска полногеномных ассоциаций.

В ходе реализации данного научного подхода учитываются не только малоструктурированные данные, в том числе технологии NLP, но также необходимо анализировать картографические изображения, полученные при проведении дистанционного зондирования. Спутниковые изображения или сведения, полученные при спектральной съемке, должны быть обработаны алгоритмами компьютерного зрения. Авторами были проведены исследования, посвященные обработке мультиспектральных изображений (индексы NDVI и CVI). Обработка производилась в среде Jupyter Notebook, использовались стандартные библиотеки OpenCV, Scikit-Image, Matplotlib. Также использовался модуль arcsu.ia в ГИС-платформе ArcGis. Было произведено формирование обучающей и тестовой выборки в соотношении 70:30, использовано 849 уникальных строк. Для формирования обучающей выборки по стандартной методике произведена предварительная бинарная сегментация на основе нормализованного вегетационного индекса, затем произведена доработка сегментации классов, произведено обучение классификатора. Архитектура сверточной сети состояла из 11 слоев и одного полносвязного слоя с глубиной каскада равной 4. Изображения, используемые для обучения, представляли собой участки площадью от 4,00 до 8,40 га, на которых возделывались сорта озимой мягкой пшеницы (*Triticum aestivum* L.). Установлено, что при оценке генотип-средового воздействия по индексу продуктивности растений на полученных данных наиболее высокие результаты показал сорт Воронежская-12 и Крестьянка.

Таким образом, технология классификации спектральных изображений при глубоком нейросетевом анализе является весьма эффективным инструментом в оценке генотип-средового воздействия и в дальнейшем может способствовать установлению высокоценных хозяйственных признаков селекционируемых сельскохозяйственных растений. С учетом развития феномики как науки применение машинного обучения является одним из перспективных направлений при реализации высокопроизводительного фенотипирования.

Работа выполнена при грантовой поддержке Федерального агентства по делам молодёжи (Росмолодёжь) Соглашение № 091-10-2023-069 от 23.05.2023 г. проект «Наука рядом».

Литература

1. Новаковский АБ Существующие программные решения для обработки геоботанических данных. Молодежь и наука на Севере. Материалы докладов II Всероссийской (XVII) молодежной научной конференции (с элементами научной школы), 2013:101–104.
2. Susan L Hoops, Dan Knights, LMdist: Local Manifold distance accurately measures beta diversity in ecological gradients. *Bioinformatics*.2023; № 39(12):72
3. Линкина АВ Применение методов искусственного интеллекта при работе с геопространственными данными. Научно-технический прогресс в сельскохозяйственном производстве. материалы XVII Международной научно-практической конференции, 2022:196–201.