

doi 10.18699/vjgb-25-49

# Геномное прогнозирование признаков растений популярными методами машинного обучения

К.Н. Козлов <sup>1</sup>, М.П. Банкин <sup>1</sup>, Е.А. Семенова<sup>2</sup>, М.Г. Самсонова <sup>1</sup> Санкт-Петербургский политехнический университет Петра Великого, Санкт-Петербург, Россия<sup>2</sup> Дальневосточный государственный аграрный университет, Благовещенск, Амурская область, Россия m.g.samsonova@gmail.com

**Аннотация.** Быстро накапливающийся массив геномных данных – секвенированных геномов сельскохозяйственных растений – позволил получить обширные результаты по геномному прогнозированию и выявлению ассоциаций однонуклеотидных полиморфизмов с фенотипическими признаками. Во многих случаях для обнаружения новых связей фенотипов с генотипами предпочтительно использовать методы машинного обучения, глубокого обучения и искусственного интеллекта, в особенности объяснимого, способные распознавать сложные закономерности. Вручную было отобрано 80 источников, при этом ограничения по дате выхода не ставилось, основной интерес представляла оригинальность предлагаемого подхода или модификации для применения в задаче геномного прогнозирования. В статье рассмотрены модели для геномного прогнозирования, сверточные нейронные сети, объяснимый искусственный интеллект и большие языковые модели. Уделено внимание подходам к дополнению данных, переносу знаний, методам снижения размерности и гибридным методам. Приведен пример современного способа кодирования больших геномных данных в искусственные изображения, преимуществом которых являются наглядная визуализация и возможность использования известных моделей для извлечения признаков. Исследования в области модельно-специфичных и модельно-независимых методов интерпретации решения моделей представлены тремя основными категориями: зондирование, возмущение и суррогатная модель. В рассмотренных примерах отражены основные современные тренды в изучаемой области. Отмечены растущая роль больших языковых моделей, в том числе основанных на трансформерах, для обработки генетического кода, а также разрабатываемые методы аугментации данных. Дополнительным преимуществом применения языковой модели может стать возможность формулировать запросы на близком к естественному языку и получать ответы за относительно короткое время. Среди гибридных подходов выделена перспективность сочетания моделей машинного обучения и моделей развития растений на основе биофизических и биохимических процессов. Поскольку методы машинного обучения и искусственного интеллекта находятся в фокусе внимания как специалистов в различных прикладных областях, так и фундаментальных ученых, а кроме того, вызывают общественный резонанс, количество посвященных этим темам работ имеет взрывной рост.

**Ключевые слова:** геномное прогнозирование; фенотип растений; машинное обучение; глубокое обучение; искусственный интеллект

**Для цитирования:** Козлов К.Н., Банкин М.П., Семенова Е.А., Самсонова М.Г. Геномное прогнозирование признаков растений популярными методами машинного обучения. *Вавиловский журнал генетики и селекции*. 2025;29(3):458-466. doi 10.18699/vjgb-25-49

**Финансирование.** Исследование финансируется Министерством науки и высшего образования Российской Федерации в рамках программы Исследовательского центра мирового уровня «Передовые цифровые технологии» (соглашение № 075-15-2020-311 от 20.04.2022).

## Genomic prediction of plant traits by popular machine learning methods

К.Н. Kozlov <sup>1</sup>, М.П. Bankin <sup>1</sup>, Е.А. Semenova<sup>2</sup>, М.Г. Samsonova <sup>1</sup> Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia<sup>2</sup> Far Eastern State Agrarian University, Blagoveshchensk, Amur region, Russia m.g.samsonova@gmail.com

**Abstract.** A rapid growth of the available body of genomic data has made it possible to obtain extensive results in genomic prediction and identification of associations of SNPs with phenotypic traits. In many cases, to identify new relationships between phenotypes and genotypes, it is preferable to use machine learning, deep learning and artificial intelligence, especially explainable artificial intelligence, capable of recognizing complex patterns. 80 sources were manually selected; while there were no restrictions on the release date, the main attention was paid to the originality of the proposed approach for use in genomic prediction. The article considers models for genomic prediction, convolutional neural networks, explainable artificial intelligence and large language models. Attention is

paid to Data Augmentation, Transfer Learning, Dimensionality Reduction methods and hybrid methods. Research in the field of model-specific and model-independent methods for interpretation of model solutions is represented by three main categories: sensing, perturbation, and surrogate model. The considered examples reflect the main modern trends in this area of research. The growing role of large language models, including those based on transformers, for genetic code processing, as well as the development of data augmentation methods, are noted. Among hybrid approaches, the prospect of combining machine learning models and models of plant development based on biophysical and biochemical processes is emphasized. Since the methods of machine learning and artificial intelligence are the focus of attention of both specialists in various applied fields and fundamental scientists, and also cause public resonance, the number of works devoted to these topics is growing explosively.

**Key words:** genomic prediction; plant phenotype; machine learning; deep learning; artificial intelligence

**For citation:** Kozlov K.N., Bankin M.P., Semenova E.A., Samsonova M.G. Genomic prediction of plant traits by popular machine learning methods. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov J Genet Breed.* 2025;29(3):458-466. doi 10.18699/vjgb-25-49

## Введение

К настоящему времени накоплен и продолжает быстро увеличиваться огромный массив геномных данных – секвенированных геномов сельскохозяйственных растений, таких как нут, маш, соя, пшеница, рожь, лен и т. д. (Брагина и др., 2019; Ichihara et al., 2023; Chamorro-Padial et al., 2024; Tang et al., 2024). Получены многочисленные аннотации, нашли свое применение классические методы геномного прогнозирования и полногеномного поиска ассоциаций, обнаружены ОНП, связанные с различными важными фенотипическими признаками (Hayes, 2013).

Во многих случаях представляющие интерес для селекции растений фенотипические признаки являются скоррелированными, что требует использования многопризнаковых моделей для получения статистически значимых прогнозов. Для решения подобных задач могут быть применены методы машинного обучения, глубокого обучения и искусственного интеллекта (ИИ), в особенности объяснимого, способные распознавать сложные закономерности во входных данных и обобщать извлеченные знания.

При отборе статей основное внимание уделялось оригинальности предлагаемого подхода или модификации для применения в задаче геномного прогнозирования. В базе научных публикаций PubMed (<https://pubmed.ncbi.nlm.nih.gov/>, дата доступа 7 ноября 2024 г.) при поиске по терминам «растения геномное прогнозирование машинное обучение» (plants genomic prediction machine learning) за период с 2010 г. по настоящее время наблюдается экспоненциальный рост числа публикаций с небольшим снижением темпа роста после 2021 г. (рис. 1).

Вручную было отобрано 80 источников, при этом ограничения по дате выхода не ставилось. Самая давняя работа относилась к 1988 г., большая часть (60 %) датирована 2020 г. и новее, а 20 % рассмотренных статей было издано за последние два года (рис. 2).

## Геномное прогнозирование

Геномное прогнозирование направлено на прогнозирование фенотипа организма с учетом данных об однонуклеотидных полиморфизмах (ОНП) (Meuwissen et al., 2001). Все множество методов ГП можно разделить на две группы: линейные и непараметрические. Линейные методы, такие как BLUP, хорошо работают для аддитивных признаков. Они моделируют фенотип как функцию вкладов различных факторов, например индивидуальных

### Список сокращений

- ОНП (SNP, single nucleotide polymorphism) – однонуклеотидный полиморфизм
- ГП (GP, genomic prediction) – геномное прогнозирование
- GBLUP (genomic best linear unbiased prediction) – геномное наилучшее линейное несмещенное прогнозирование
- МО (ML, machine learning) – машинное обучение
- RRBLUP (ridge regression best linear unbiased prediction) – гребневая регрессия с наилучшим линейным несмещенным прогнозированием
- CNN (convolutional neural network) – сверточная нейронная сеть
- AIO (artificial image object) – искусственное изображение
- PCA (principal component analysis) – анализ главных компонент
- Объяснимый ИИ (XAI, explainable artificial intelligence) – объяснимый искусственный интеллект
- DT (decision trees) – деревья решений
- RF (random forest) – случайный лес
- БЯМ – большая языковая модель
- GPT (generative pretrained transformer) – генеративная предобученная модель

маркеров, погодных параметров, полевых условий и т. д. С другой стороны, непараметрические методы машинного обучения, такие как машины опорных векторов, случайный лес и градиентный бустинг, могут моделировать нелинейные признаки, обеспечивая огромную гибкость для адаптации к сложным ассоциациям между генотипом и фенотипом (Montesinos-López et al., 2021).

Инструменты геномного прогнозирования на основе статистических методов, таких как геномное наилучшее линейное несмещенное прогнозирование (GBLUP), широко используются в селекции сельскохозяйственных культур. Однако эти инструменты не предназначены для учета нелинейных связей в многомерных наборах данных или для работы с наборами данных высокой размерности: например, изображения, собранные беспилотными летательными аппаратами. Алгоритмы машинного обучения имеют потенциал превзойти точность прогнозирования текущих инструментов, используемых для прогнози-

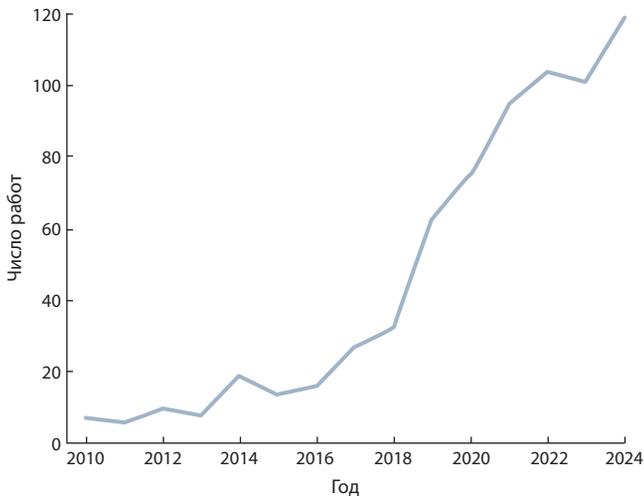


Рис. 1. Рост числа публикаций в PubMed.

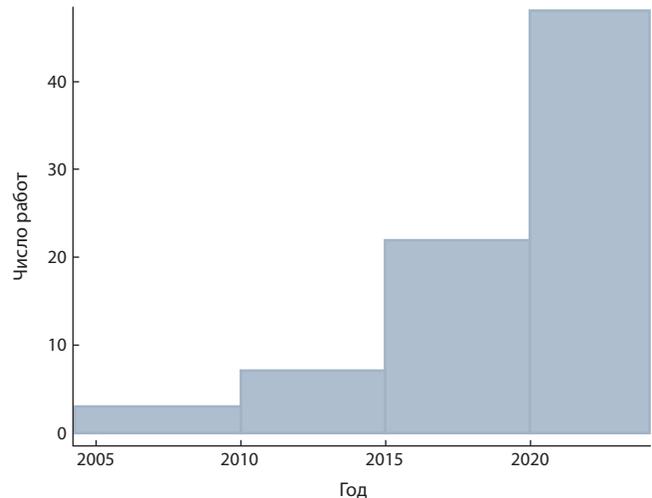


Рис. 2. Распределение рассмотренных работ по годам.

вания фенотипических признаков исходя из геномных данных, благодаря их способности автономно извлекать признаки и представлять их связи на нескольких уровнях абстракции (Danilevicz et al., 2022).

Точность прогнозирования зависит от качества и предварительной обработки фенотипических данных, платформы, используемой для получения геномной информации, схемы скрещивания популяции, внутренней генетической архитектуры признака, генетической структуры популяции, того, как рассматривается взаимодействие генотипа и среды, и метода прогнозирования (de Los Campos et al., 2013).

В работе (Sandhu et al., 2021) показано, что модели глубокого обучения превосходят традиционные модели гребневой регрессии с наилучшим линейным несмещенным прогнозированием (RRBLUP) и байесовские модели при всех сценариях прогнозирования. Методы машинного обучения использовались для увеличения статистической мощности моделей. Для применения многоэтапного машинного обучения был предложен новый пакет BioM2 (Zhang S. et al., 2024) для системы статистических расчетов R, который имеет возможности использовать стратификацию и агрегацию многомерных данных на основе биологической информации для повышения эффективности обучения моделей. В данном случае стратификация позволяет строить подмножества данных, например обучающие и тестовые выборки, управляя соотношением количества объектов из разных групп, в том числе ОНП в генах, участвующих в различных процессах.

В то же время агрегация многомерных данных позволяет использовать более простые и легче интерпретируемые модели, которые могут быть уточнены в ходе многоэтапного обучения. К. Tong с коллегами (2024) предложили инновационную вычислительную систему PlantMine, которая объединяет методы отбора признаков и машинного обучения для эффективного определения основных SNP на примере критически важных факторов для улучшения признаков риса. С помощью набора данных из проекта “3000 Rice Genomes Project” были применены различные алгоритмы обработки данных. Результаты подчеркивают

эффективность объединения отбора признаков с машинным обучением для точного определения основных SNP, предлагают многообещающий путь для ускорения селекции новых сортов растений с повышенной урожайностью и устойчивостью к стрессу. Эффективность модели в целом больше зависела от алгоритма прогнозирования, чем от метода отбора предикторов. Среди всех моделей наилучшим образом отличились методы машинного обучения на основе деревьев решений (случайные леса и градиентный бустинг), тогда как классические байесовские методы проявляли склонность к переобучению (Sirsat et al., 2022).

### Сверточные нейронные сети и искусственные изображения

Среди методов машинного обучения CNN обеспечивают наилучшую способность находить скрытые закономерности или признаки из данных и лучше всего подходят для анализа изображений (Pook et al., 2020; Montesinos-López et al., 2021). Искусственные изображения (АИО) – это новая концепция представления геномных данных, которую можно использовать для кодирования больших геномных данных путем рассмотрения отдельных генетических вариантов как пикселей (Galli et al., 2022). Преимуществами АИО являются удобная простая визуализация, компактность и возможность применения огромного количества методов, разработанных для анализа и классификации изображений (Chen X. et al., 2021b), в частности CNN (Chen X. et al., 2021a). Следовательно, АИО может использоваться CNN для задач регрессии и классификации (Bavykina et al., 2022).

В статье (Bazgir et al., 2020) показан алгоритм оптимизации упаковки данных в АИО. Пакет DeepFeature, предложенный в работах (Sharma et al., 2019, 2021), был создан для преобразования экспериментальных данных большого объема, например геномных или транскриптомных, в форму, оптимальную для обучения модели CNN. Входной вектор преобразуется в матрицу с использованием t-SNE, kernel PCA, PHATE или UMAP, а наименьший прямоугольник, содержащий все элементы, находится с

помощью алгоритма выпуклой оболочки. Выполняется вращение для выравнивания изображения, при этом декартовы координаты преобразуются в индексы пикселей.

Применение CNN к обработке АЮ позволяет вычислять и визуализировать влияние различных факторов на окончательное решение модели. Работа (Liu et al., 2019) считается первым исследованием, в котором применена карта значимости для определения наиболее важных предикторов у сои. В этом исследовании пропуски в данных рассматривали как новый генотип, в результате каждый SNP был закодирован четырьмя двоичными значениями. Значение значимости каждого генотипа было рассчитано как максимальное абсолютное значение градиентов среди этих четырех каналов кодирования, медианное значение популяции использовалось в качестве измерения вклада однонуклеотидных полиморфизмов.

В работе (Xie et al., 2024) архитектура ResNet, широко используемая в методах глубокого обучения, представлена для применения в моделях геномной селекции. Поскольку каждый локус делает разный вклад в итоговый фенотип, для модели геномной селекции больше подходят последовательные свертки, чем объединение слоев. Таким образом, был предложен алгоритм глубокого обучения ResGS, который значительно нивелирует проблему деградации, т. е. снижения производительности с увеличением глубины модели, что позволяет повысить точность прогноза по сравнению с традиционными методами (Wu H. et al., 2024).

В последнее время все больше внимания уделяется внутренним механизмам сверточных нейронных сетей и причине, по которой сеть принимает определенные решения (Wang et al., 2020). Предложено несколько методов, которые включают подходы, основанные на перестановках данных и обратном распространении ошибок (Zhang X., Gao, 2020), алгоритмы на основе градиента (Selvaraju et al., 2020) и карты активации классов (Wang et al., 2020). Карта значимости представляет пространственные области, связанные с определенным классом на данном изображении (Simonyan et al., 2014). Карты активации классов обеспечивают визуальное объяснение для одного входного изображения (Chattopadhyay et al., 2018; Selvaraju et al., 2020), но чувствительны к архитектуре модели. Картирование активации классов с градиентным весом (Grad-CAM) использует градиенты любого целевого понятия, поступающего в конечный сверточный слой, для создания грубой карты локализации, выделяющей важные области на изображении для прогнозирования класса (Selvaraju et al., 2017).

Score-CAM, в отличие от предыдущих подходов, основанных на картировании активации классов, избавляется от зависимости от градиентов, получая вес каждой карты активации в процессе прямого расчета сети для экземпляров целевого класса, конечный результат получается с помощью линейной комбинации весов и карт активации (Wang et al., 2020). Grad-CAM++ (Chattopadhyay et al., 2018), модификация Grad-CAM (Selvaraju et al., 2020), обобщает CAM для моделей без глобальных слоев объединения. LayerCAM (Jiang et al., 2021) может генерировать надежные карты активации классов из комбинации карт активации классов из разных слоев CNN.

## Объяснимый искусственный интеллект

Объяснимый ИИ направлен на преодоление проблемы черного ящика и обеспечение понимания того, как системы ИИ принимают решения. Интерпретируемые модели машинного обучения могут объяснить, как они делают прогнозы, и определить факторы, влияющие на их результаты. Однако большинство современных интерпретируемых методов МО были разработаны для таких областей, как, например, компьютерное зрение, что делает прямое применение к проблемам биоинформатики сложным без настройки и адаптации к доменной области.

Интерпретируемая модель МО может выявить факторы, которые влияют (например, статистически значимые признаки) на ее результаты, и объяснить взаимодействия между ними (Molnar, 2022). В зависимости от уровня абстракции методы можно разделить на методы локальной и глобальной интерпретируемости. В то время как локальные методы сосредоточены на интерпретации отдельных прогнозов, глобальные пытаются объяснить поведение всей модели в форме диаграмм или списков. Разрабатываются различные варианты специфичных для модели и независимых от модели интерпретируемых методов МО, на основе которых может быть построена система ХАИ для повышения ее локальной и глобальной интерпретируемости (Wachter et al., 2018), однако чаще всего эти методы используются для улучшения визуализации (Weber et al., 2023). Линейные модели, деревья решений (DT) и системы на основе правил менее сложны и по своей сути интерпретируемы. Однако они менее точны по сравнению с ансамблями на основе деревьев, такими как случайные леса (RF) и глубокие нейронные сети, что приводит к компромиссу между точностью и интерпретируемостью.

Предложено и разработано множество специфичных и независимых от структуры модели интерпретируемых методов МО (Azodi et al., 2020). Все эти методы в основном делятся на три основные категории: зондирование, возмущение и суррогатная модель. Примеры методов зондирования – методы на основе градиента, такие как градиентно-взвешенное отображение активации класса (Grad-CAM++) и распространение послойной релевантности (LRP) (Guidotti et al., 2018). Широко используемым методом, основанным на возмущениях, являются аддитивные объяснения Шепли (SHAP). Метод SHAP основан на теории коалиционных игр, т. е. на среднем предельном вкладе признака и способе распределения выигрышей между его игроками (Cubitt, 1991).

Поскольку интерпретируемость достигается за счет компромисса между точностью и сложностью, в исследованиях предложено обучение простой интерпретируемой модели для имитации сложной модели (Molnar, 2022). Суррогатная или простая прокси-модель – это стратегия интерпретации модели, включающая обучение изначально интерпретируемой модели путем аппроксимации локальных прогнозов черного ящика (Stiglic et al., 2020; Molnar, 2022).

Большинство инструментов и библиотек построения суррогатных моделей разработано с целью улучшения интерпретируемости и объяснимости моделей МО как черного ящика, охватывающих общие проблемы в области

компьютерного зрения, интеллектуального анализа текста или структурированных данных, и основано на известных интерпретируемых методах машинного обучения, таких как LIME (Ribeiro et al., 2016), понимание модели через объяснение подпространства (MUSE) (Lakkaraju et al., 2019), SHAP (Lundberg, Lee, 2017) и его варианты: ядро SHAP и дерево SHAP, график частичной зависимости (PDP), индивидуальное условное ожидание (ICE), важность признаков перестановки (PFI) и контрфактуальные объяснения (CE) (Wachter et al., 2018).

### Большие языковые модели

В последнее время использование больших языковых моделей (БЯМ) в различных областях науки получило широкое распространение, в том числе для расшифровки генетического текста с целью прогнозирования проявления полезных признаков на примере растений. БЯМ, например GPT-4, покорили мир, демонстрируя удивительные возможности во владении естественным языком, что незамедлительно побудило исследователей к адаптации БЯМ к языку другого типа – геному, для решения сложных задач на базе больших объемов накопленных данных. Во многом успех БЯМ обусловлен использованием в архитектуре блоков внимания на основе трансформеров. Применение таких архитектурных решений позволило широко известной нейросети AlphaFold2 (Google DeepMind, 2021 г.) предсказывать трехмерные структуры белка с невиданной до сегодняшнего дня точностью. AlphaFold3 (2024 г.), по заявлениям разработчиков, впервые превосходит физические методы в своих предсказаниях трехмерной структуры белка, а также взаимодействие белков друг с другом и с другими веществами. БЯМ компании Profluent позволила создать искусственный белок для редактирования генома, сопоставимый по эффективности с природным, но обладающий гораздо большей специфичностью.

Переход к широкому практическому применению результатов этих достижений требует глубокого понимания лежащих в основе механизмов с учетом сложных взаимодействий, и ускорения поиска ответов на встающие на практике вопросы. В частности, существует необходимость с большей степенью надежности переходить от выявления ассоциированных с признаком ОНП к определению генов, влияющих на признак. Кроме того, требуется учитывать взаимодействие генов между собой, а также рассматривать не только один признак, но и пары связанных признаков. Решение описанной проблемы будет невозможно без привлечения новейших достижений информатики, нашедших выражение в виде искусственного интеллекта на основе БЯМ. Дополнительным плюсом применения такого подхода будет возможность формулировать запросы на близком к естественному языку и получать ответы за относительно короткий срок.

В последнее время исследования в этом направлении заметно активизировались. Так, например, в обзоре (Consens et al., 2023), посвященном применению трансформероподобных моделей к генетическим данным, отмечено бурное развитие в этой области при рассмотрении более 100 недавних работ. Наблюдается использование больших языковых моделей, не только основанных на трансформерах, но и использующих так называемый слой Нуена (Poli

et al., 2023) для обработки геномных данных (Nguyen et al., 2023). Один из интересных подходов – возможность предобучения таких моделей на последовательностях геномов без применения фенотипов.

На сегодняшний день максимальная длина входной последовательности среди общедоступных БЯМ на основе трансформера для ДНК ограничена всего  $3 \times 10^4$  нуклеотидов для архитектуры GENA-LM. Для смягчения этого ограничения в работе (Kuratov et al., 2024) изучена эффективность использования модифицированной архитектуры трансформера с рекуррентной памятью (RMT) в модели GENA-LM для множественных задач геномного анализа, требующих обработки длинных последовательностей ДНК. Полученные в работе (Kuratov et al., 2024) результаты показали, что дополнение GENA-LM с помощью RMT приводит к существенному повышению производительности.

S. Jubair с коллегами (2021) предложили метод на основе трансформероподобной нейронной сети для предсказания тяжести фузариоза и связанного с ним накопления опасного микотоксина дезоксиниваленола, с применением геномных и фенотипических данных о популяции ячменя. Показано превосходство частотного кодирования маркеров и отмечены проблемы высоких требований модели к объему памяти при задействовании большого количества маркеров, число которых снижалось за счет использования отбора по информационному критерию.

В работе (Wu C. et al., 2023) представлена модель геномной селекции на основе глубокой нейронной сети с применением трансформеров, сверточных слоев и модуля предвзвешенной информации. Архитектура модели использует кодирование позиций маркеров тригонометрическими функциями, быстрое преобразование Фурье, линейную функцию активации с гауссовой функцией ошибок (GELU) и включает блок сверточной сети, блок трансформеров и блок регрессора. Модель была применена к пяти наборам данных, на которых превзошла четыре метода, использованные для сравнения.

Важный источник снижения точности моделей предсказания фенотипа по геномным данным – отсутствие учета межгенных взаимодействий. В работе (Cui et al., 2022) предложен подход к выявлению взаимодействий между генами и их учету в модели глубокого обучения для предсказания фенотипа. В архитектуру глубокой нейронной сети добавлен слой, представляющий гены как скрытые узлы разреженной сети. Важную часть подхода также составляет использование значения Шапли для скрытых узлов геномного слоя для определения важности влияния взаимодействий на решение модели.

### Дополнение данных

Обучение больших языковых моделей требует большого объема данных, потому что количество неизвестных параметров велико. В статьях (Jubair et al., 2021; Wu C. et al., 2023) рассматриваются модели на основе трансформероподобной нейронной сети для прогнозирования фенотипа по генотипу. GPTransformer в работе (Jubair et al., 2021) содержит два кодирующих блока Transformer, использует два узла для каждого слоя внимания; каждый блок Transformer содержит 256 скрытых нейронов. Выходные

данные представляют собой вектор, являющийся входом сети прямого распространения, которая содержит один выходной нейрон. Применяется функция потерь среднеквадратической ошибки (MSE). Для обучения и анализа использован набор данных из 400 генотипов, фенотипированных в трех географических областях и двух годах, т. е. 2400 записей, при этом коэффициент корреляции Пирсона между предсказанием модели и данными составил 0.6 и позволил получить значимые результаты.

Основанную на трансформероподобной нейронной сети модель GPformer для прогнозирования фенотипа по генотипу в работе (Wu C. et al., 2023) отдельно обучали и тестировали на наборах данных Soybean999, Maize282, Rice469, Wheat599 и Wheat2403, имеющих записи 999, 282, 469, 599 и 2403 соответственно. Полученный в результате коэффициент корреляции Пирсона составил 0.4–0.8 для различных вариантов.

Дополнительным средством, как и в случае моделей глубокого обучения для обработки изображений, может послужить аугментация данных, работы по которой для моделей глубокого обучения в области биоинформатики появились недавно. Так, например, L. Ji с коллегами (2024) предложили новый подход к аугментации данных о биологических последовательностях, в котором изменяется порядок хромосом. Такой метод генерации дополнительных данных может быть применен для обучения, потому что модели не могут использовать номер хромосомы как предиктор. В статье (Montesinos-López et al., 2024) рассмотрен метод смешивания, который предлагает независимый от предметной области подход к аугментации, основанный на предположении, что линейная комбинация векторов признаков должна примерно соответствовать линейной комбинации соответствующих им целевых значений. В работе (Vilov, Heinig, 2022) аугментация данных успешно использована для обучения классификатора геномных вариантов. В. Yelmen с коллегами (2021) представили подходы на основе генеративной сети (GAN) и машины Больцмана (RBM) для составления синтетических геномов. В описанных работах авторам удалось улучшить точность и обобщающую способность моделей, поэтому аугментация данных может быть задействована для расширения имеющегося набора данных для обучения больших языковых моделей.

T.-H. Zhang с коллегами (2021) предложили метод прогнозирования классификации энхансеров на сильные и слабые с использованием дополненных данных и сверточной нейронной сети ES-ARCNN. Для обучения ES-ARCNN были применены два приема дополнения данных: обратное дополнение и сдвиг для ранее идентифицированных энхансеров.

### Перенос знаний

Перенос знаний (transfer learning) позволяет создавать эффективные модели для целевой области, используя знания из другой, но связанной с ней исходной области. В медицинских исследованиях перенос знаний может значительно повысить точность прогнозирования заболеваний для групп населения, не имеющих достаточных данных, при несбалансированных данных (Gao, Cui,

2022). Этот подход обладает также большим потенциалом для улучшения прогнозирования сложных фенотипических признаков, таких, например, как «урожайность растений», хотя работает не во всех случаях (Kovalev et al., 2018). Наиболее широко этот подход применяется при использовании моделей извлечения характерных черт из изображений, предобученных на универсальных наборах данных, а затем подвергающихся тонкой настройке на относительно ограниченном специализированном наборе данных (Kirchler et al., 2022).

Для упрощения применения подхода с переносом знаний для построения моделей прогнозирования фенотипа по генотипу J. Li с коллегами (2024) предложили эффективную реализацию TrG2P. Разработанная структура сначала использует сверточные нейронные сети для обучения моделей с применением геномных данных и фенотипических признаков, имеющих более простые зависимости, чем сложный целевой признак, например «урожайность». Затем параметры сверточных слоев этих предварительно обученных моделей переносятся в задачу прогнозирования целевого признака, а полносвязные слои повторно обучаются, таким образом приводя к повышению точности прогнозирования результирующей модели (Li et al., 2024).

### Методы снижения размерности

Взрывной рост объемов данных не только обуславливает беспрецедентный прогресс в биоинформатике и возможности для выполнения предиктивного моделирования (Han, Liu, 2022), но и создает проблемы для существующих методов и инструментов ИИ, такие как неоднородность данных, высокая размерность и объем (Karim et al., 2021). Анализ главных компонент (PCA) и изометрическое отображение признаков (Isomap) широко применяются в качестве методов снижения размерности (Fournier, Aloise, 2019). Однако представления, полученные с помощью этих методов, часто теряют основные свойства (Aggarwal, Reddy, 2014), что делает их менее эффективными против известного явления, называемого проклятием размерности, особенно для высокоразмерных наборов данных (Fournier, Aloise, 2019).

### Гибридные методы

С ростом вычислительных мощностей все чаще модели машинного обучения комбинируются в сложные гибридные методы. Например, в работе (Chen C. et al., 2024) были рассмотрены алгоритмы, на первом этапе использующие BayesR/GWAS для идентификации подмножества из 1000 маркеров с умеренными и большими предельными аддитивными эффектами, а затем задействованы сети внимания для получения прогнозов на основе этих эффектов и их взаимодействий. Гибридные методы с сетями внимания дали наименьшую дисперсию точности прогнозирования по всем проверочным наборам данных и наименьшую среднеквадратичную ошибку, что может быть критерием, заслуживающим рассмотрения в практических программах селекции. В статье (Ramzan et al., 2020) проблему обнаружения большого количества локусов с малым эффектом на фенотип предлагается решать с использованием

процедуры из двух этапов. На первом этапе кубическими сплайнами аппроксимируются значения тестовой статистики Уалда (Wald test), а геномные регионы с пиками сплайнов, которые оказываются выше ожидаемых, рассматриваются как локусы количественных признаков (QTL). Затем SNP в этих QTL ранжируются по степени их связи с фенотипом с применением подхода случайного леса. M. Nascimento с коллегами (2024) предложили модель Stacking Ensemble Learning (SEL), объединяющую несколько моделей для потенциально еще более точного прогнозирования важных признаков, на примере селекции кофе *Coffea arabica*.

Перспективным направлением является комбинирование моделей машинного обучения и моделей развития растений на основе биофизических и биохимических процессов (CGM). Было выдвинуто предположение, что такой подход может улучшить прогнозы интегративных признаков путем разложения их на более простые промежуточные признаки с лучше выраженной наследуемостью (Lague et al., 2024). В проведенном исследовании предсказательная способность комбинированного подхода CGM-GP превзошла предсказательную способность моделей геномной селекции без интеграции CGM, независимо от используемого метода регрессии. CGM имитирует нелинейные (причинные) реакции растений на окружающую среду через параметры модели, представляющие генотипическую чувствительность к этим реакциям (G×E). Таким образом, калиброванные CGM для генотипа могут быть полезны для прогнозирования его производительности в неизвестных условиях, но они не могут предсказывать неизвестные генотипы (Lague et al., 2019).

## Заключение

Большое разнообразие методов машинного обучения и искусственного интеллекта находит применение для решения задач биоинформатики растений, таких как геномное прогнозирование важных фенотипических признаков. Поскольку МО и ИИ находятся в фокусе внимания как специалистов в различных прикладных областях, так и фундаментальных ученых, а кроме того, вызывают общественный резонанс, количество посвященных этим темам работ имеет взрывной рост.

В рассмотренных примерах отражены основные современные тренды в исследуемой области, такие как использование общих моделей машинного обучения и их специализированных для биоинформатики вариантов, самостоятельно и в сочетании с методами снижения размерности и отбора признаков, построения систем объяснимого ИИ и гибридных методов. Отдельно следует упомянуть возрастающую роль больших языковых моделей, в том числе основанных на трансформерах, и связанные с ними методы аугментации данных, необходимые для обучения моделей с огромным числом параметров. В связи с проблемой недостаточности или несбалансированности данных могут быть задействованы методы с переносом знаний.

Важную роль для применения МО и ИИ играет кодирование исходных данных, например, упомянутые искусственные изображения позволяют использовать мощный и хорошо себя зарекомендовавший в различных задачах аппарат сверточных нейронных сетей для выделения ха-

рактерных паттернов в данных, а также методы на основе карт значимости для выявления важных предикторов.

В связи с развитием интернета вещей, мобильных устройств, автономных роботов стало развиваться направление упрощения моделей и алгоритмов для переноса вычислений на эти устройства. Однако данная тема достаточно широка, требует отдельного рассмотрения и не была затронута в этой работе.

## Список литературы / References

- Брагина М.К., Афонников Д.А., Салина Е.А. Прогресс в секвенировании геномов растений – направления исследований. *Вавиловский журнал генетики и селекции*. 2019;23(1):38-48. doi 10.18699/VJ19.459
- [Bragina M.K., Afonnikov D.A., Salina E.A. Progress in plant genome sequencing: research directions. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov J Genet Breed*. 2019;23(1):38-48. doi 10.18699/VJ19.459 (in Russian)]
- Aggarwal C.C., Reddy C.K. (Eds) *Data Clustering: Algorithms and Applications*. New York: Chapman and Hall/CRC, 2014. doi 10.1201/9781315373515
- Azodi C.B., Tang J., Shiu S.-H. Opening the black box: interpretable machine learning for geneticists. *Trends Genet*. 2020;36(6):442-455. doi 10.1016/j.tig.2020.03.005
- Bavykina M., Kostina N., Lee C.-R., Schaffleitner R., Bishop-von Wettberg E., Nuzhdin S.V., Samsonova M., Gursky V., Kozlov K. Modeling of flowering time in *Vigna radiata* with artificial image objects, convolutional neural network and random forest. *Plants*. 2022; 11(23):3327. doi 10.3390/plants11233327
- Bazgir O., Zhang R., Dhruva S.R., Rahman R., Ghosh S., Pal R. Representation of features as images with neighborhood dependencies for compatibility with convolutional neural networks. *Nat Commun*. 2020;11(1):4391. doi 10.1038/s41467-020-18197-y
- Chamorro-Padial J., Garcia R., Gil R. A systematic review of open data in agriculture. *Comput Electron Agric*. 2024;219:108775. doi 10.1016/j.compag.2024.108775
- Chattopadhyay A., Sarkar A., Howlader P., Balasubramanian V.N. Grad-CAM++: generalized gradient-based visual explanations for deep convolutional networks. In: 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Tahoe, NV, USA. IEEE, 2018;839-847. doi 10.1109/WACV.2018.00097
- Chen C., Bhuiyan S.A., Ross E., Powell O., Dinglasan E., Wei X., Atkin F., Deomano E., Hayes B. Genomic prediction for sugarcane diseases including hybrid Bayesian-machine learning approaches. *Front Plant Sci*. 2024;15:1398903. doi 10.3389/fpls.2024.1398903
- Chen X., Chen D.G., Zhao Z., Balko J.M., Chen J. Artificial image objects for classification of breast cancer biomarkers with transcriptome sequencing data and convolutional neural network algorithms. *Breast Cancer Res*. 2021a;23(1):96. doi 10.1186/s13058-021-01474-z
- Chen X., Chen D.G., Zhao Z., Zhan J., Ji C., Chen J. Artificial image objects for classification of schizophrenia with GWAS-selected SNVs and convolutional neural network. *Patterns*. 2021b;2(8):100303. doi 10.1016/j.patter.2021.100303
- Consens M.E., Dufault C., Wainberg M., Forster D., Karimzadeh M., Goodarzi H., Theis F.J., Moses A., Wang B. To transformers and beyond: large language models for the genome. *arXiv*. 2023. doi 10.48550/arXiv.2311.07621
- Cubitt R. The Shapley value: essays in honor of Lloyd S. Shapley. *Econ J*. 1991;101(406):644-646. doi 10.2307/2233574
- Cui T., El Mekkaoui K., Reinvald J., Havulinna A.S., Marttinen P., Kasiki S. Gene-gene interaction detection with deep learning. *Commun Biol*. 2022;5(1):1238. doi 10.1038/s42003-022-04186-y
- Danilevicius M.F., Gill M., Anderson R., Batley J., Bennamoun M., Bayer P.E., Edwards D. Plant genotype to phenotype prediction using machine learning. *Front Genet*. 2022;13:822173. doi 10.3389/fgene.2022.822173

- de Los Campos G., Hickey J.M., Pong-Wong R., Daetwyler H.D., Calus M.P.L. Whole-genome regression and prediction methods applied to plant and animal breeding. *Genetics*. 2013;193(2):327-345. doi 10.1534/genetics.112.143313
- Fournier Q., Aloise D. Empirical comparison between autoencoders and traditional dimensionality reduction methods. In: 2019 IEEE Second International Conference on Artificial Intelligence and Knowledge Engineering (AIKE), Sardinia, Italy. IEEE, 2019;211-214. doi 10.1109/AIKE.2019.00044
- Galli G., Sabadin F., Yassue R.M., Galves C., Carvalho H.F., Crossa J., Montesinos-López O.A., Fritsche-Neto R. Automated machine learning: a case study of genomic “image-based” prediction in maize hybrids. *Front Plant Sci*. 2022;13:845524. doi 10.3389/fpls.2022.845524
- Gao Y., Cui Y. Deep transfer learning provides a *Pareto* improvement for multi-ancestral clinico-genomic prediction of diseases. *bioRxiv*. 2022. doi 10.1101/2022.09.22.509055
- Guidotti R., Monreale A., Ruggieri S., Pedreschi D., Turini F., Gianotti F. Local rule-based explanations of black box decision systems. *arXiv*. 2018. doi 10.48550/arXiv.1805.10820
- Han H., Liu X. The challenges of explainable AI in biomedical data science. *BMC Bioinformatics*. 2022;22(Suppl. 12):443. doi 10.1186/s12859-021-04368-1
- Hayes B. Overview of statistical methods for genome-wide association studies (GWAS). In: *Genome-Wide Association Studies and Genomic Prediction. Methods in Molecular Biology*. Vol. 1019. Totowa, NJ: Humana Press, 2013;149-169. doi 10.1007/978-1-62703-447-0\_6
- Ichihara H., Yamada M., Kohara M., Hirakawa H., Ghelfi A., Tamura T., Nakaya A., ... Komaki A., Fawcett J.A., Sugihara E., Tabata S., Isobe S.N. Plant GARDEN: a portal website for cross-searching between different types of genomic and genetic resources in a wide variety of plant species. *BMC Plant Biol*. 2023;23(1):391. doi 10.1186/s12870-023-04392-8
- Ji L., Hou W., Xiong L., Zhou H., Liu C., Li L., Yuan Z. GSCNN: a genomic selection convolutional neural network model based on SNP genotype and physical distance features and data augmentation strategy. *Res Square*. 2024. doi 10.21203/rs.3.rs-3991262/v1
- Jiang P.-T., Zhang C.-B., Hou Q., Cheng M.-M., Wei Y. LayerCAM: exploring hierarchical class activation maps for localization. *IEEE Trans Image Process*. 2021;30:5875-5888. doi 10.1109/TIP.2021.3089943
- Jubair S., Tucker J.R., Henderson N., Hiebert C.W., Badea A., Domaratzki M., Fernando W.G.D. GPTransformer: a transformer-based deep learning method for predicting Fusarium related traits in barley. *Front Plant Sci*. 2021;12:761402. doi 10.3389/fpls.2021.761402
- Karim M.R., Beyan O., Zappa A., Costa I.G., Rebholz-Schuhmann D., Cochez M., Decker S. Deep learning-based clustering approaches for bioinformatics. *Brief Bioinform*. 2021;22(1):393-415. doi 10.1093/bib/bbz170
- Kirchler M., Konigorski S., Norden M., Meltendorf C., Kloft M., Schurmann C., Lippert C. transferGWAS: GWAS of images using deep transfer learning. *Bioinformatics*. 2022;38(14):3621-3628. doi 10.1093/bioinformatics/btac369
- Kovalev M.S., Igolkina A.A., Samsonova M.G., Nuzhdin S.V. A pipeline for classifying deleterious coding mutations in agricultural plants. *Front Plant Sci*. 2018;9:1734. doi 10.3389/fpls.2018.01734
- Kuratov Y., Shmelev A., Fishman V., Kardymon O., Burtsev M. Recurrent memory augmentation of GENA-LM improves performance on long DNA sequence tasks. In: Workshop Machine Learning for Genomics Explorations (MLGenX). 2024. Available: <https://openreview.net/pdf?id=K6711CX90x>
- Lakkaraju H., Kamar E., Caruana R., Leskovec J. Faithful and Customizable Explanations of Black Box Models. In: Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society (AIES '19). New York, NY, USA: Association for Computing Machinery, 2019;131-138. doi 10.1145/3306618.3314229
- Larue F., Fumey D., Rouan L., Soulié J.-C., Roques S., Beurrier G., Luquet D. Modelling tiller growth and mortality as a sink-driven process using *Ecomeristem*: implications for biomass sorghum ideotyping. *Ann Bot*. 2019;124(4):675-690. doi 10.1093/aob/mcz038
- Larue F., Rouan L., Pot D., Rami J.-F., Luquet D., Beurrier G. Linking genetic markers and crop model parameters using neural networks to enhance genomic prediction of integrative traits. *Front Plant Sci*. 2024;15:1393965. doi 10.3389/fpls.2024.1393965
- Li J., Zhang D., Yang F., Zhang Q., Pan S., Zhao X., Zhang Qi., Han Y., Yang J., Wang K., Zhao C. TrG2P: a transfer-learning-based tool integrating multi-trait data for accurate prediction of crop yield. *Plant Commun*. 2024;5(7):100975. doi 10.1016/j.xplc.2024.100975
- Liu Y., Wang D., He F., Wang J., Joshi T., Xu D. Phenotype prediction and genome-wide association study using deep convolutional neural network of soybean. *Front Genet*. 2019;10:1091. doi 10.3389/fgene.2019.01091
- Lundberg S., Lee S.-I. A unified approach to interpreting model predictions. In: Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17). Red Hook, NY, USA: Curran Associates Inc., 2017;4768-4777. doi 10.48550/arXiv.1705.07874
- Meuwissen T.H., Hayes B.J., Goddard M.E. Prediction of total genetic value using genome-wide dense marker maps. *Genetics*. 2001;157(4):1819-1829. doi 10.1093/genetics/157.4.1819
- Molnar C. Interpretable Machine Learning. A Guide for Making Black Box Models Explainable. Independently published, 2022
- Montesinos-López O.A., Montesinos-López A., Mosqueda-Gonzalez B.A., Montesinos-López J.C., Crossa J., Ramirez N.L., Singh P., Valladares-Anguiano F.A. A zero altered Poisson random forest model for genomic-enabled prediction. *G3 (Bethesda)*. 2021;11(2):jkaa057. doi 10.1093/g3journal/jkaa057
- Montesinos-López O.A., Solis-Camacho M.A., Crespo-Herrera L., Saint Pierre C., Huerta Prado G.I., Ramos-Pulido S., Al-Nowibet K., Fritsche-Neto R., Gerard G., Montesinos-López A., Crossa J. Data augmentation enhances plant-genomic-enabled predictions. *Genes*. 2024;15(3):286. doi 10.3390/genes15030286
- Nascimento M., Nascimento A.C.C., Azevedo C.F., de Oliveira A.C.B., Caixeta E.T., Jarquin D. Enhancing genomic prediction with Stacking Ensemble Learning in Arabica Coffee. *Front Plant Sci*. 2024;15:1373318. doi 10.3389/fpls.2024.1373318
- Nguyen E., Poli M., Faizi M., Thomas A., Birch-Sykes C., Wornow M., Patel A., Rabideau C., Massaroli S., Bengio Y., Ermon S., Baccus S.A., Ré C. HyenaDNA: long-range genomic sequence modeling at single nucleotide resolution. In: Proceedings of the 37th International Conference on Neural Information Processing Systems (NIPS '23). Red Hook, NY, USA: Curran Associates Inc., 2023; 43177-43201. doi 10.48550/arXiv.2306.15794
- Poli M., Massaroli S., Nguyen E., Fu D.Y., Dao T., Baccus S., Bengio Y., Ermon S., Ré C. Hyena hierarchy: towards larger convolutional language models. In: Proceedings of the 40th International Conference on Machine Learning (ICML '23). Vol. 202. JMLR.org, 2023;28043-28078. doi 10.48550/arXiv.2302.10866
- Pook T., Freudenthal J., Korte A., Simianer H. Using local convolutional neural networks for genomic prediction. *Front Genet*. 2020; 11:561497. doi 10.3389/fgene.2020.561497
- Ramzan F., Gültas M., Bertram H., Cavero D., Schmitt A.O. Combining random forests and a signal detection method leads to the robust detection of genotype-phenotype associations. *Genes (Basel)*. 2020;11(8):892. doi 10.3390/genes11080892
- Ribeiro M.T., Singh S., Guestrin C. “Why should i trust you?”: explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16). New York, NY, USA: Association for Computing Machinery, 2016;1135-1144. doi 10.1145/2939672.2939778
- Sandhu K., Patil S.S., Pumphrey M., Carter A. Multitrait machine- and deep-learning models for genomic selection using spectral informa-

- tion in a wheat breeding program. *Plant Genome*. 2021;14(3):e20119. doi 10.1002/tpg2.20119
- Selvaraju R.R., Cogswell M., Das A., Vedantam R., Parikh D., Batra D. Grad-CAM: visual explanations from deep networks via gradient-based localization. In: 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy. IEEE, 2017;618-626. doi 10.1109/ICCV.2017.74
- Selvaraju R.R., Cogswell M., Das A., Vedantam R., Parikh D., Batra D. Grad-CAM: visual explanations from deep networks via gradient-based localization. *Int J Comput Vis*. 2020;128(2):336-359. doi 10.1007/s11263-019-01228-7
- Sharma A., Vans E., Shigemizu D., Boroevich K.A., Tsunoda T. DeepInsight: a methodology to transform a non-image data to an image for convolution neural network architecture. *Sci Rep*. 2019; 9(1): 11399. doi 10.1038/s41598-019-47765-6
- Sharma A., Lysenko A., Boroevich K.A., Vans E., Tsunoda T. DeepFeature: feature selection in nonimage data using convolutional neural network. *Brief Bioinform*. 2021;22(6):bbab297. doi 10.1093/bib/bbab297
- Simonyan K., Vedaldi A., Zisserman A. Deep inside convolutional networks: visualising image classification models and saliency maps. *arXiv*. 2014. doi 10.48550/arXiv.1312.6034
- Sirsat M.S., Obléssuc P.R., Ramiro R.S. Genomic prediction of wheat grain yield using machine learning. *Agriculture*. 2022;12(9):1406. doi 10.3390/agriculture12091406
- Stiglic G., Kocbek P., Fijacko N., Zitnik M., Verbert K., Cilar L. Interpretability of machine learning based prediction models in healthcare. *WIREs Data Min Knowl Discovery*. 2020;10(5):e1379. doi 10.1002/widm.1379
- Tang F.H.M., Nguyen T.H., Conchedda G., Casse L., Tubiello F.N., Maggi F. CROPGRIDS: a global geo-referenced dataset of 173 crops. *Sci Data*. 2024;11:413. doi 10.1038/s41597-024-03247-7
- Tong K., Chen X., Yan S., Dai L., Liao Y., Li Z., Wang T. PlantMine: a machine-learning framework to detect core SNPs in rice genomics. *Genes*. 2024;15(5):603. doi 10.3390/genes15050603
- Vilov S., Heinig M. Neural network approach to somatic SNP calling in WGS samples without a matched control. *bioRxiv*. 2022. doi 10.1101/2022.04.14.488223
- Wachter S., Mittelstadt B., Russell C. Counterfactual explanations without opening the black box: automated decisions and the GDPR. *Harvard J Law Technol*. 2018;31(2):841-887
- Wang H., Wang Z., Du M., Yang F., Zhang Z., Ding S., Mardziel P., Hu X. Score-CAM: score-weighted visual explanations for convolutional neural networks. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Seattle, WA, USA. IEEE, 2020;111-119. doi 10.1109/CVPRW50498.2020.00020
- Weber L., Lapuschkin S., Binder A., Samek W. Beyond explaining: opportunities and challenges of XAI-based model improvement. *Inf Fusion*. 2023;92:154-176. doi 10.1016/j.inffus.2022.11.013
- Wu C., Zhang Y., Ying Z., Li L., Wang J., Yu H., Zhang M., Feng X., Wei X., Xu X. A transformer-based genomic prediction method fused with knowledge-guided module. *Brief Bioinform*. 2023;25(1):bbad438. doi 10.1093/bib/bbad438
- Wu H., Gao B., Zhang R., Huang Z., Yin Z., Hu X., Yang C.-X., Du Z.-Q. Residual network improves the prediction accuracy of genomic selection. *Anim Genet*. 2024;55(4):599-611. doi 10.1111/age.13445
- Xie Z., Xu X., Li L., Wu C., Ma Y., He J., Wei S., Wang J., Feng X. Residual networks without pooling layers improve the accuracy of genomic predictions. *Theor Appl Genet*. 2024;137(6):138. doi 10.1007/s00122-024-04649-2
- Yelmen B., Decelle A., Ongaro L., Marnetto D., Tallec C., Montinaro F., Furtlehner C., Pagani L., Jay F. Creating artificial human genomes using generative neural networks. *PLoS Genet*. 2021; 17(2):e1009303. doi 10.1371/journal.pgen.1009303
- Zhang S., Li P., Wang S., Zhu J., Huang Z., Cai F., Freidel S., Ling F., Schwarz E., Chen J. BioM2: biologically informed multi-stage machine learning for phenotype prediction using omics data. *Brief Bioinform*. 2024;25(5):bbae384. doi 10.1093/bib/bbae384
- Zhang T.-H., Flores M., Huang Y. ES-ARCNN: predicting enhancer strength by using data augmentation and residual convolutional neural network. *Anal Biochem*. 2021;618:114120. doi 10.1016/j.ab.2021.114120
- Zhang X., Gao J. Measuring feature importance of convolutional neural networks. *IEEE Access*. 2020;8:196062-196074. doi 10.1109/ACCESS.2020.3034625

**Конфликт интересов.** Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию 26.11.2024. После доработки 23.01.2025. Принята к публикации 23.01.2025.