



ИССЛЕДОВАНИЕ ПЕРСПЕКТИВНОСТИ ИСПОЛЬЗОВАНИЯ НЕЙРОННЫХ СЕТЕЙ ДЛЯ ПОИСКА НОВЫХ ВЫСОКОМОЛЕКУЛЯРНЫХ ВЫСОКОЭНЕРГЕТИЧЕСКИХ ВЕЩЕСТВ

Амосова Елена С., Волохов В.М., Парахин В.В., Лемперт Д.Б., Акостелов И.И.,
Воеводин В.В.

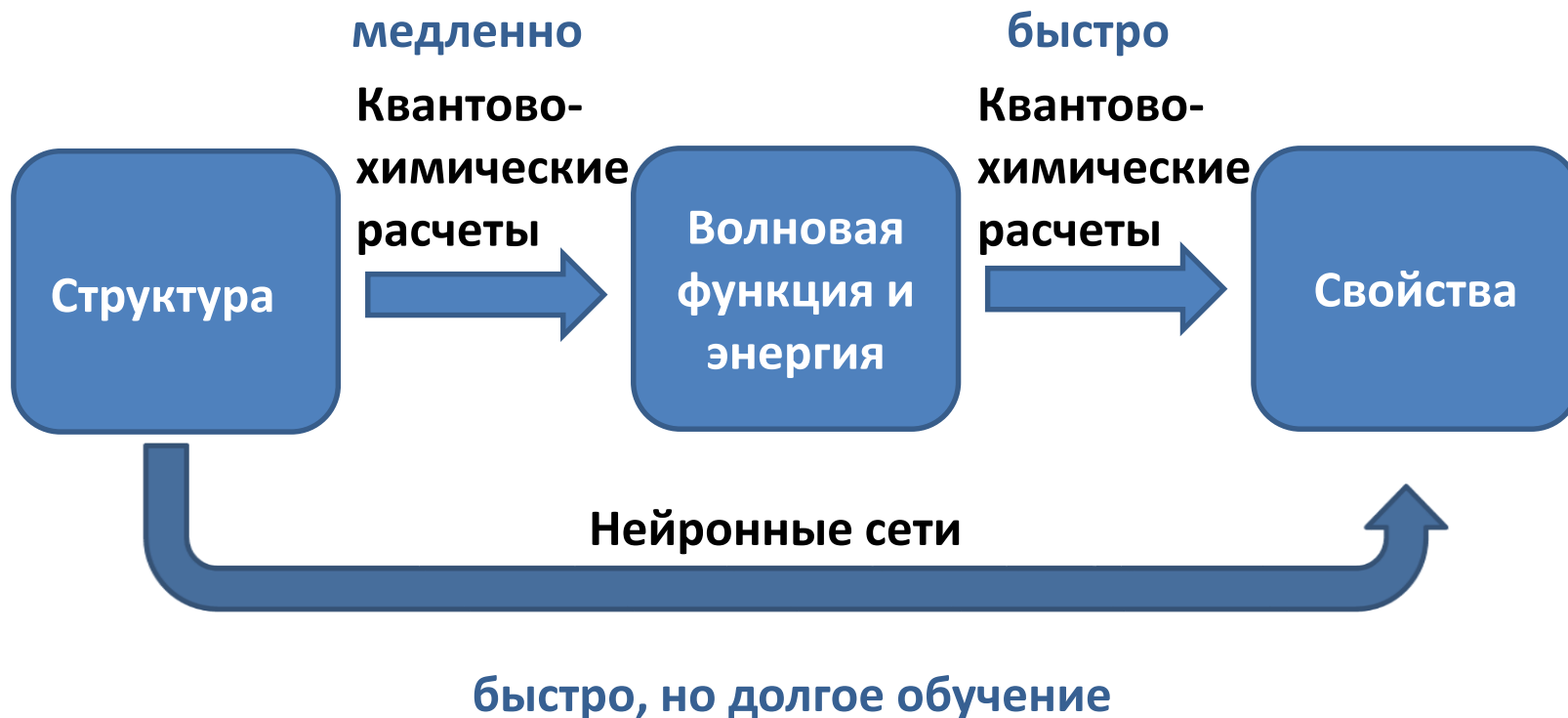
*Федеральный исследовательский центр проблем химической физики и
медицинской химии РАН, Черноголовка, Россия, aes@icpr.ac.ru*

Институт органической химии им. Н.Д. Зелинского РАН, Москва, Россия

Московский государственный университет им. М.В. Ломоносова, Москва, Россия

*Научно-исследовательский вычислительный центр Московского
государственного университета им. М.В. Ломоносова, Москва, Россия*

Исследование высокоэнергетических соединений



Вычислительные ресурсы

Volta2 (Lomonosov-2):

Intel Xeon Gold 6240 processors (18 cores, 2.60GHz - 1497.6 GFlop/s) with Nvidia Tesla V100 graphics accelerators (900-2G500-0010-000, 1246 MHz, 7 TFlop/s)

Нейронные сети

1. Создают эффективные векторные представления молекул, учитывая их трехмерную геометрию и химическую структуру.
2. Способны моделировать сложные химические взаимодействия между атомами и функциональными группами в молекулах.
3. Учитывают трехмерную структуру молекул, включая расстояния и углы между атомами, а также их пространственное расположение.
4. Обучаются на основе квантово-химических данных, что обеспечивает высокую точность предсказаний химических свойств и взаимодействий.
5. Могут быть использованы для предсказания различных химических свойств молекул, анализа молекулярных взаимодействий, дизайна новых химических соединений и других приложений в химическом моделировании.

Использованные нейронные сети

Allegro¹

DimeNet++²

PaiNN³

TorchMD-NET ET⁴

1. Musaelian, A., Batzner, S., Johansson, A. et al. Learning local equivariant representations for large-scale atomistic dynamics // Nat. Commun. 2023. Vol. 14. 579. DOI: [10.1038/s41467-023-36329-y](https://doi.org/10.1038/s41467-023-36329-y)
2. Gasteiger, J., Giri, S., Margraf, J.T., et al. Fast and Uncertainty-Aware Directional Message Passing for Non-Equilibrium Molecules // Machine Learning for Molecules Workshop at NeurIPS. 2020. DOI: [10.48550/arXiv.2011.14115](https://doi.org/10.48550/arXiv.2011.14115)
3. Schütt, K., Unke, O., and Gastegger, M. Equivariant message passing for the prediction of tensorial properties and molecular spectra // Proceedings of the 38th International Conference on Machine Learning, in Proceedings of Machine Learning Research. 2021. Vol. 139. P. 9377–9388. URL: <https://proceedings.mlr.press/v139/schutt21a.html>
4. Thölke, P., De Fabritiis, G. TorchMD-NET: Equivariant Transformers for Neural Network based Molecular Potentials // Proceedings of the 10th International Conference on Learning Representations. 2022. DOI: [10.48550/arXiv.2202.02541](https://doi.org/10.48550/arXiv.2202.02541)

Наборы данных

GDB-17



- Набор сгенерированных химических структур
- До 17 атомов C, H, N, O, F
- Изменяется и дополняется
- Молекула – граф: атомы – узлы, химические связи – рёбра
- Активно используется в различных областях науки

QM9



- 134 тысячи молекул в газовой фазе из GDB-17
- До 9 тяжёлых атомов C, N, O, F
- Геометрия, число атомов, связей, углов, длины связей
- Расчётные данные: энергия основного состояния, энтальпия образования, теплоёмкость и т.д.
- Популярный выбор в машинном обучении

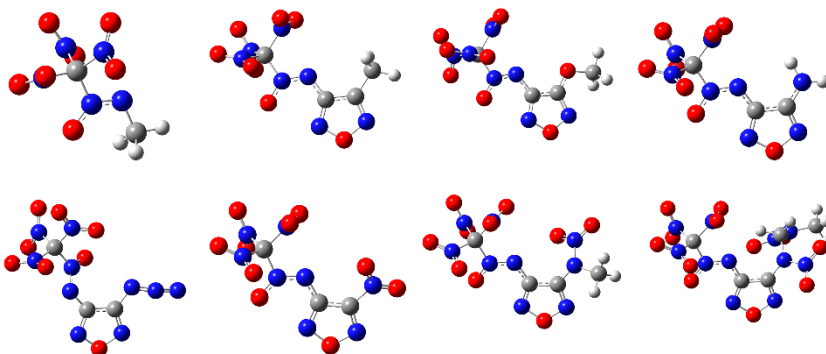
Обучение и валидация

4 архитектуры + 1 набор данных

Обучение 80%

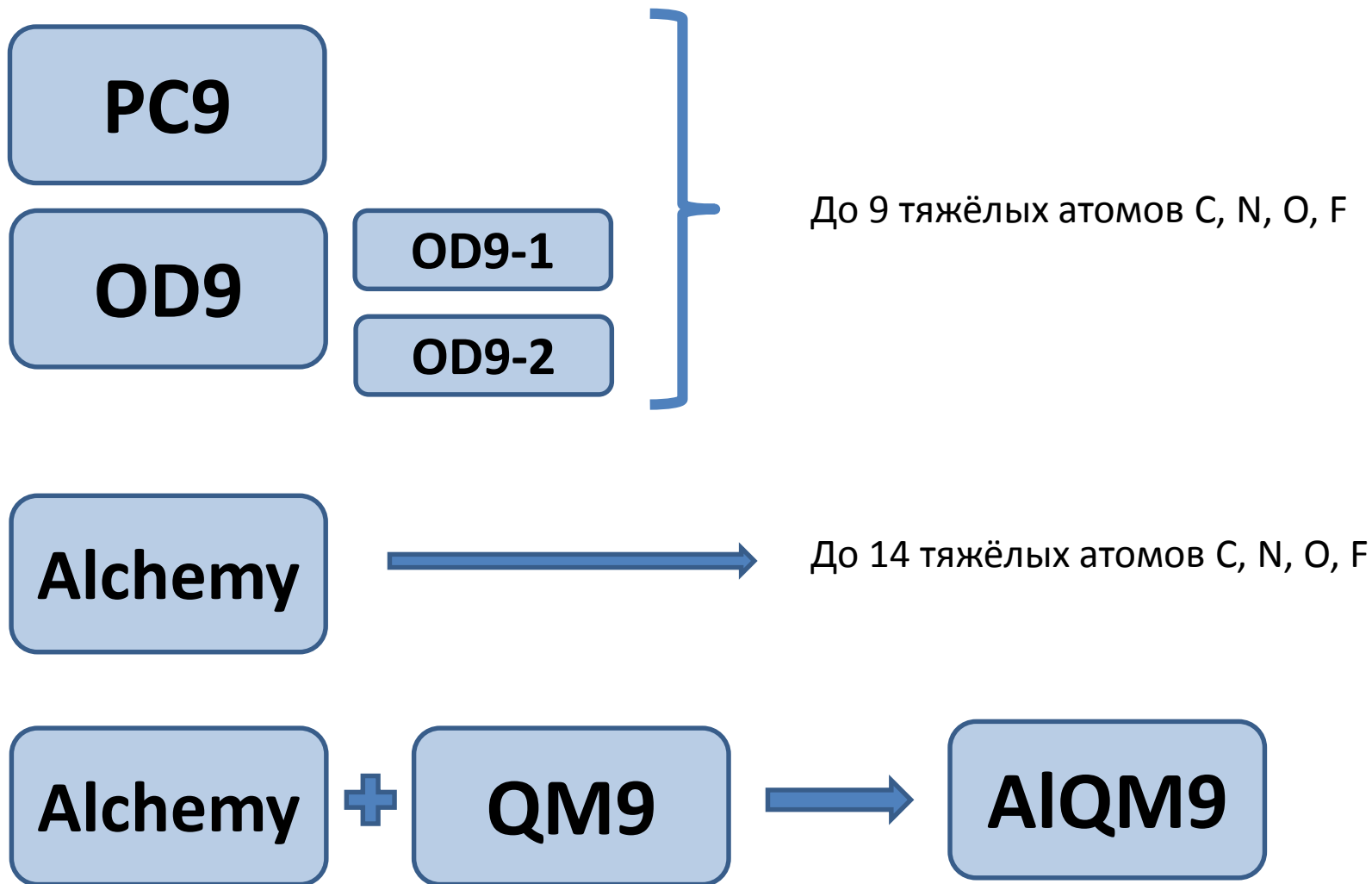
Валидация 20%

Среднее абсолютное отклонение 0.2 ккал/моль



Ошибка сильно варьируется
До 24 ккал/моль

Дополнительные наборы данных

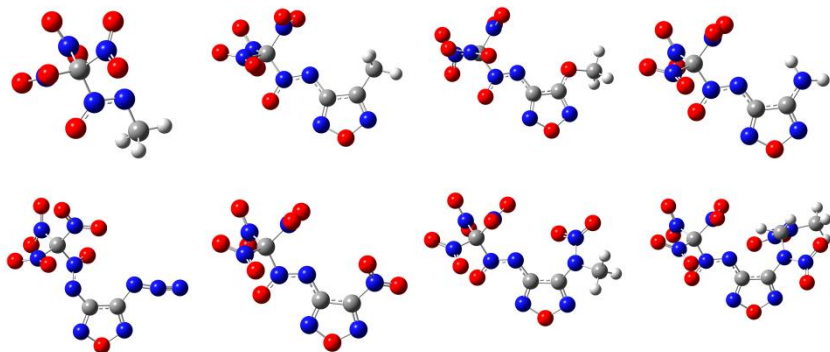


Нейросетевые модели

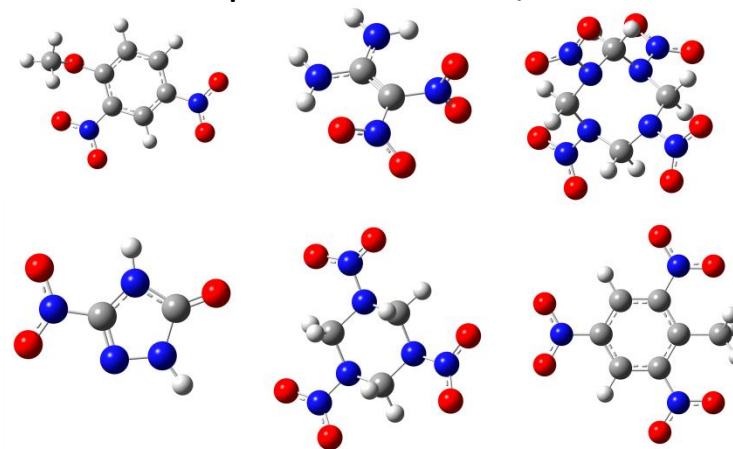
	QM9	PC9	OD9	OD9-1	OD9-2	Alchemy	AIQM9
Allegro	Allegro QM9	Allegro PC9	Allegro OD9	Allegro OD9-1	Allegro OD9-2	Allegro Alchemy	Allegro AIQM9
DimeNet++	DimeNet++ QM9	DimeNet++ PC9	DimeNet++ OD9	DimeNet++ OD9-1	DimeNet++ OD9-2	DimeNet++ Alchemy	DimeNet++ AIQM9
PaiNN	PaiNN QM9	PaiNN PC9	PaiNN OD9	PaiNN OD9-1	PaiNN OD9-2	PaiNN Alchemy	PaiNN AIQM9
TorchMD-NET ET	TorchMD-NET ET QM9	TorchMD-NET ET PC9	TorchMD-NET ET OD9	TorchMD-NET ET OD9-1	TorchMD-NET ET OD9-2	TorchMD-NET ET Alchemy	TorchMD-NET ET AIQM9

Тестовые группы веществ

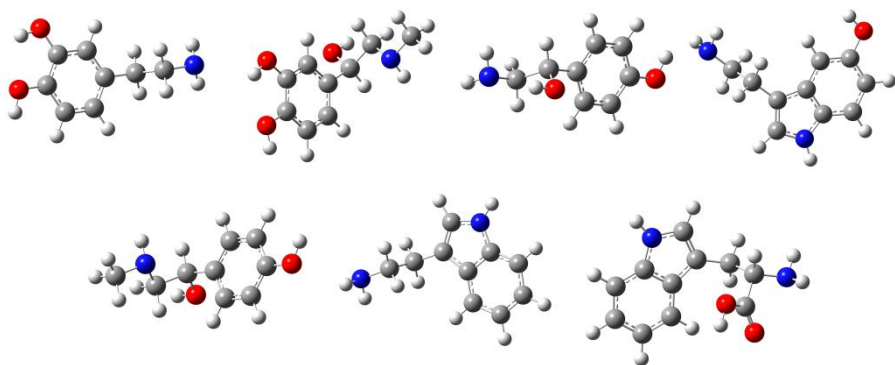
Азокси-соединения



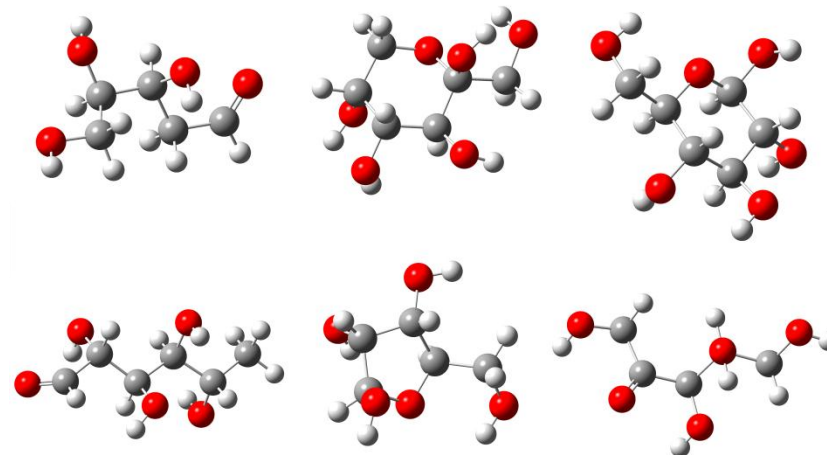
Взрывчатые вещества



Нейромодуляторы и нейротрансмиттеры

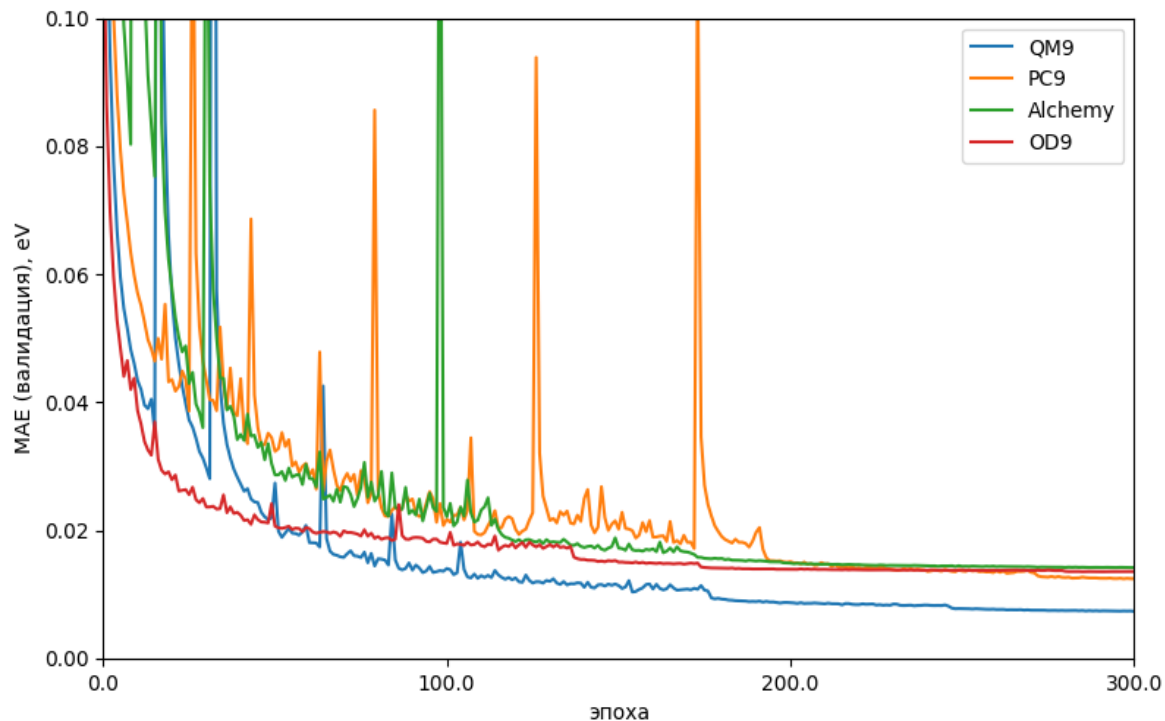


Моносахариды



Обучение PaiNN

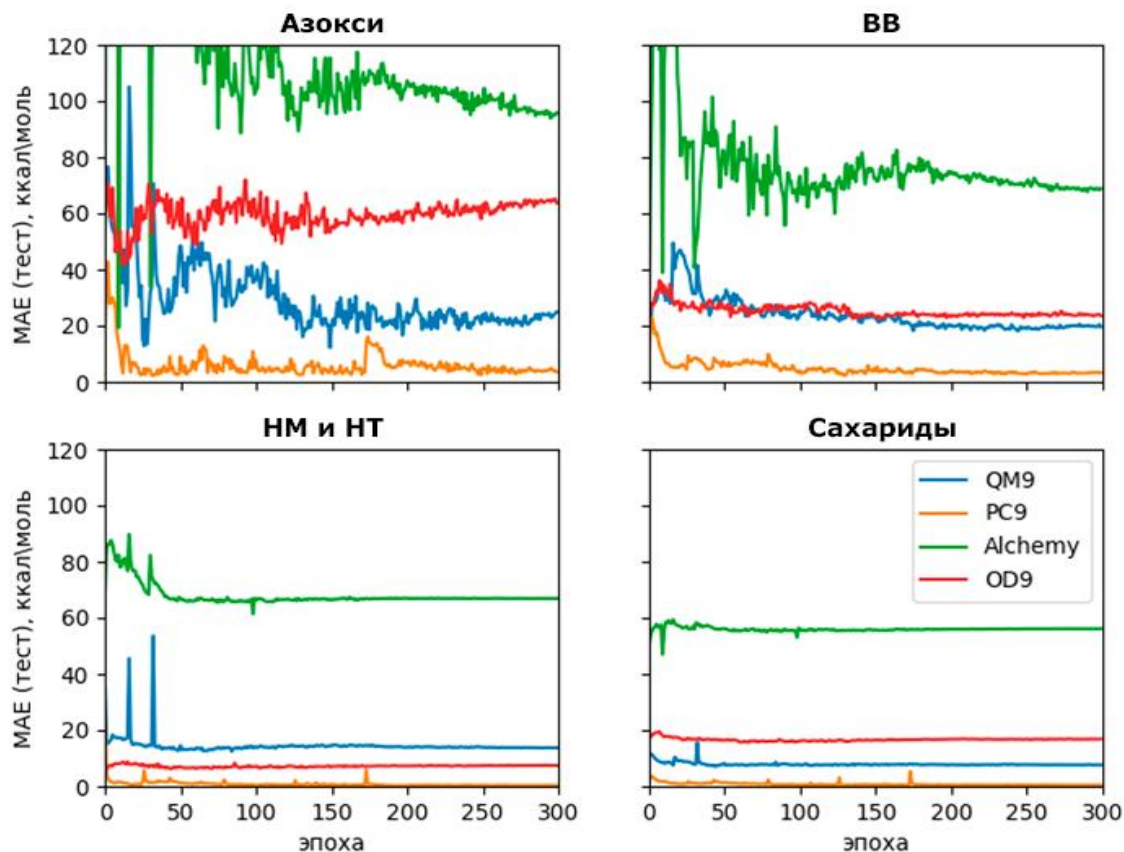
Кривые обучения в зависимости от тренировочного набора данных



При обучении самая высокая точность – QM9

Тестирование PaiNN

Кривые точности предсказания энергии атомизации различных групп веществ в зависимости от тренировочного набора данных.



При тестировании самая высокая точность – PC9

Точность предсказания (PaINN)

Точность предсказания энергии атомизации различных групп веществ (средняя абсолютная ошибка в ккал/моль) в зависимости от тренировочного набора данных

	Азокси	ВВ	НМ и НТ	Сахариды
PC9	2,86	3,03	0,40	0,29
QM9	27,61	20,8	13,43	7,61
Alchemy	94,93	68,78	66,97	56,18
ALQM9	190,74	64,93	66,18	56,26
OD9	64,89	24,23	7,50	16,79
OD9 (0)	51,14	26,24	6,49	13,33
OD9 (1)	181,0	78,63	1,75	18,84

Результаты

Точность модели – среднее абсолютное от точности для каждой из тестовых групп.

Обобщаемость – стандартное отклонение точности для каждой из тестовых групп.

Наиболее точная и обобщаемая модель – **TorchMD-NET ET**, обученная на **PC9**. CAO – **0.8** ккал/моль. CO – **0.2** ккал/моль

Среди других архитектур лучшие показатели у обученных на **PC9**.

Наименьшая точность – тестовая группа соединений с **азокси-** группой (до 28 тяжёлых атомов). CAO – до **190** ккал/моль.

Выводы

- Было проведено исследование перспективности использования нейронных сетей для поиска новых высокоэнергетических веществ и предсказания их физико-химических свойств на примере нейросетей Allegro (2023), DimeNet++ (2020), PaiNN (2021), TorchMD-NET ET (2022) и наборов данных QM9, PC9, OD9 (и его двух подмножества), Alchemy, а также объединения QM9 и Alchemy (AIQM9).
- Обученные модели были применены для предсказания энергии атомизации 4 тестовых групп веществ, ранжированных по структурной сложности. Наиболее точной и обобщаемой оказалась архитектура TorchMD-NET ET, обученная на PC9 (среднее абсолютное отклонение от расчетных данных составило 0.8 ккал/моль).
- Все рассмотренные нейросетевые модели демонстрируют наименьшую точность при тестировании на группе веществ, содержащих до 28 тяжёлых атомов.

Работа выполнена с использованием оборудования Центра коллективного пользования сверхвысокопроизводительными вычислительными ресурсами МГУ имени М.В. Ломоносова.

Работа выполнена при поддержке Российского научного фонда (проект **23-71-00005**), а также по темам государственных заданий, №№ госрегистрации **124013100856-9** and **124020100045-5**.

СПАСИБО ЗА ВНИМАНИЕ!