



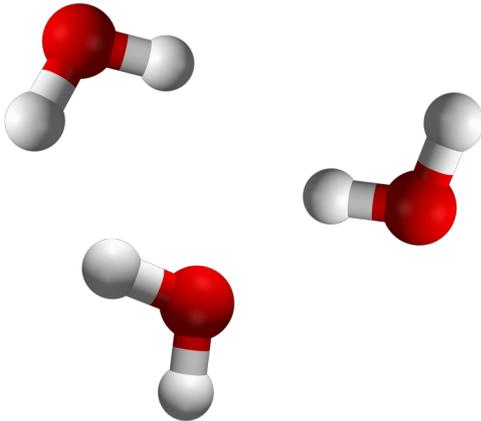
# Дескрипторы для определения локальной структуры атомистических моделей аморфных льдов

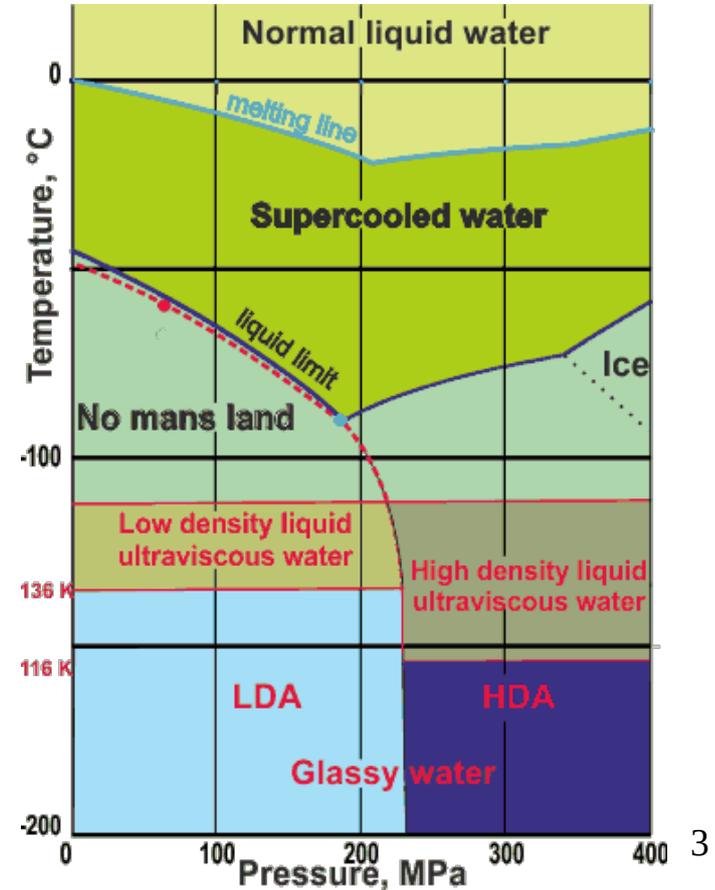
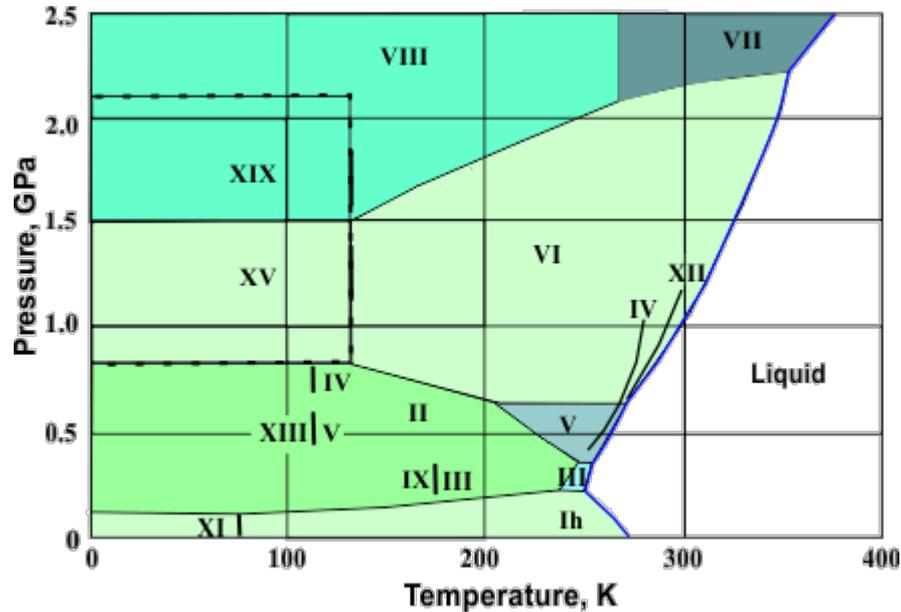
Выполнили:  
Яшин П. А.  
Баганцова Е. Р.  
Щуплецова А. А.  
Стегайлов В. В.



# Содержание

- Актуальность исследований
- Постановка задачи
- Данные и инструменты
- Результаты
- Выводы





Chaplin M. F. Structure and properties of water in its various states

//Encyclopedia of water: science, technology, and society. – 2019. – С. 1-19.



# Актуальность



Аморфный лёд распространён во вселенной



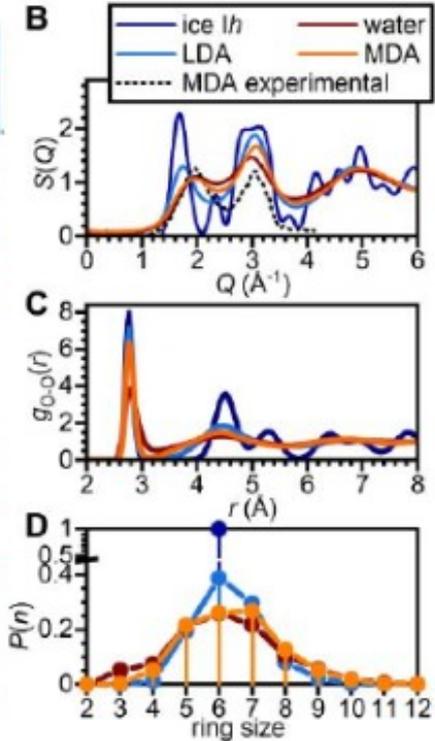
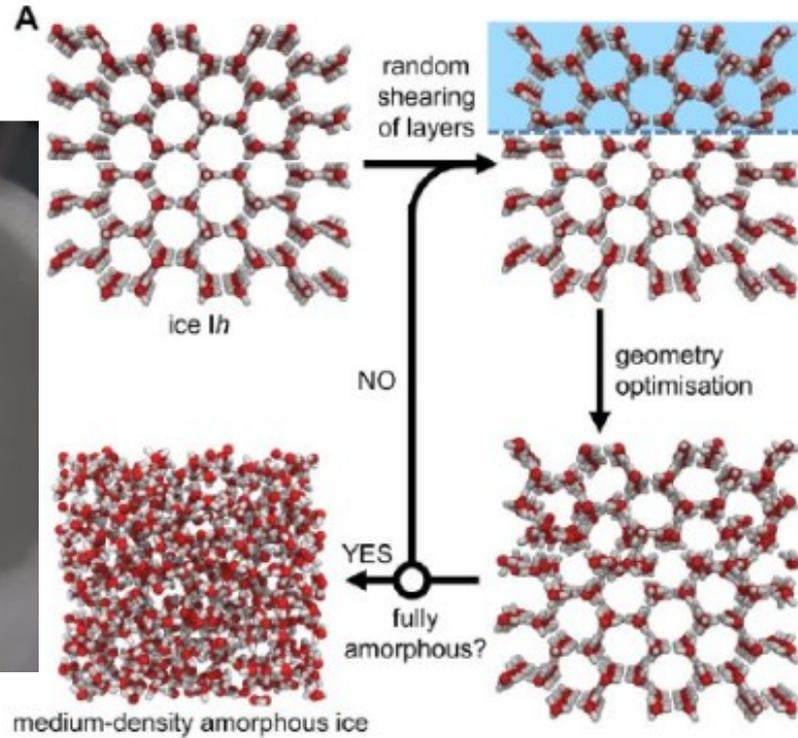
Аморфный лёд используется в криоэлектронной микроскопии



Аморфный лёд может быть полезен в крионике

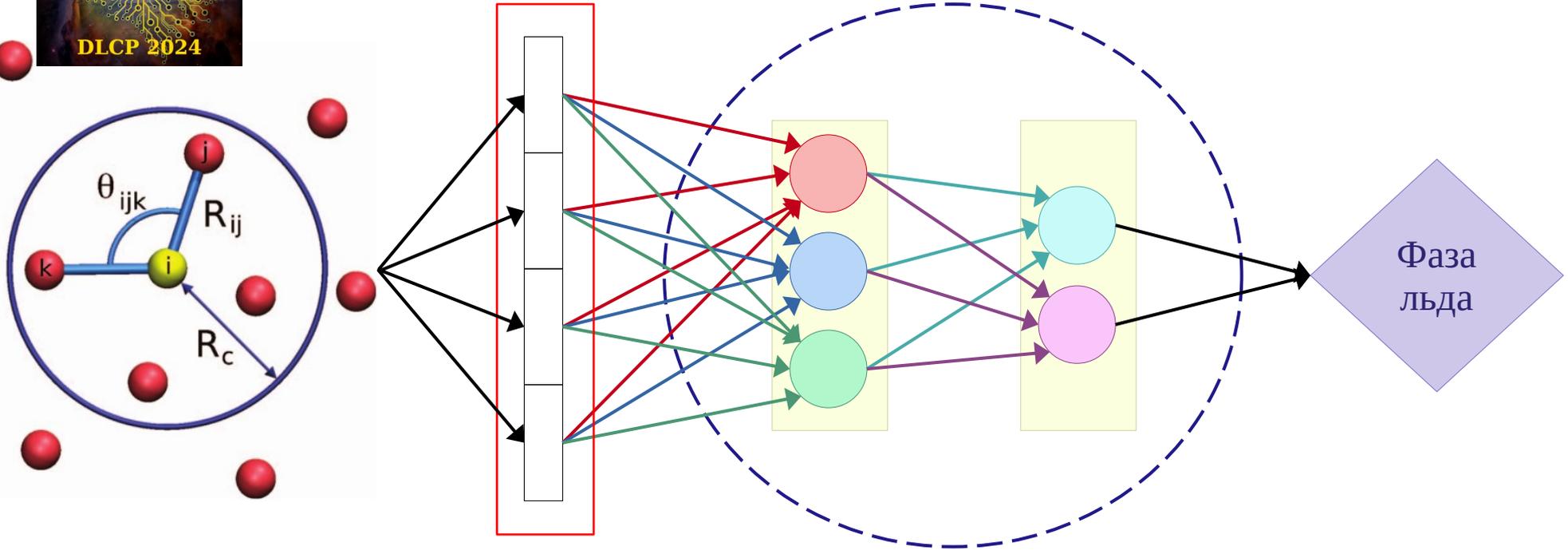


# Аморфный лёд средней плотности





# Решение - разработка модели



**Для классической модели ищем оптимальный способ описания системы - дескриптор**

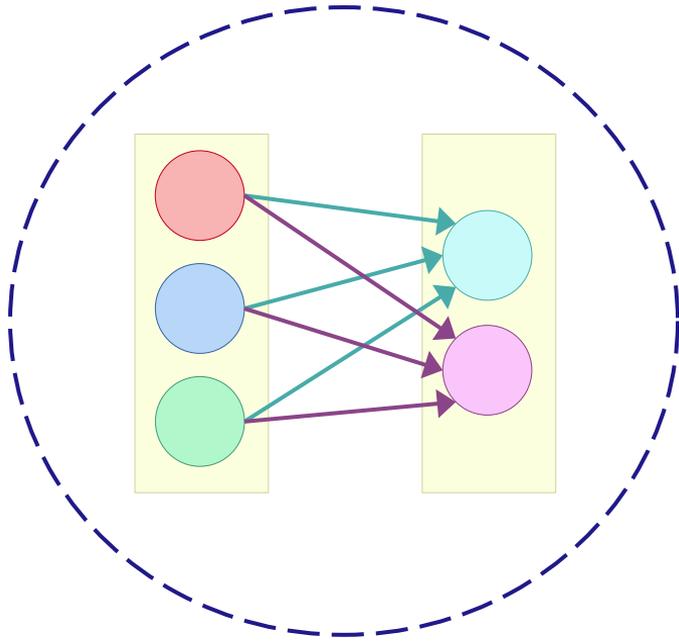


## Цели и задачи

- Изучить теоретические основы работы современных дескрипторов
- Рассмотреть возможность применения известных дескрипторов для решения задачи о распознавания различных фаз воды
- На основе полученных данных разработать простую модель, способную решать задачу о распознавании локальных атомистических структур водных систем.



# Нейронная сеть



Входной слой:

$$N \times 100$$

Скрытый слой:

$$100 \times 100$$

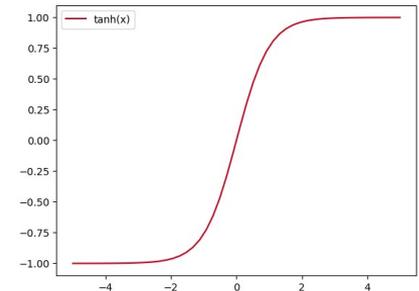
$N$  – размерность входного вектора

Функция активации:  $\tanh(x)$

Выходной слой:

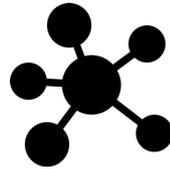
$$100 \times 9$$

Оптимизатор: ADAM





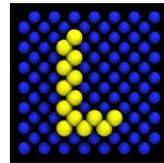
# Данные



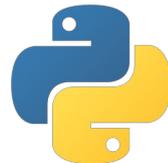
8 льдов (включая LDA, HDA, MDA) и вода



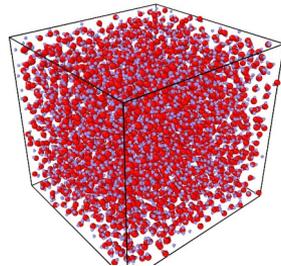
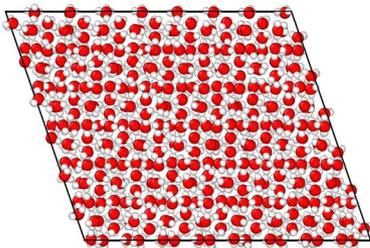
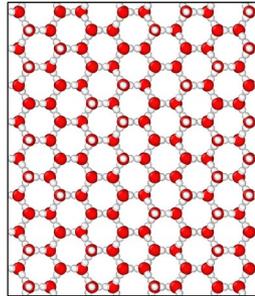
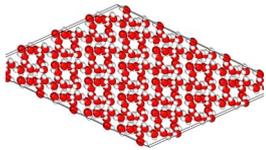
GenIce2 — получение исходных структур



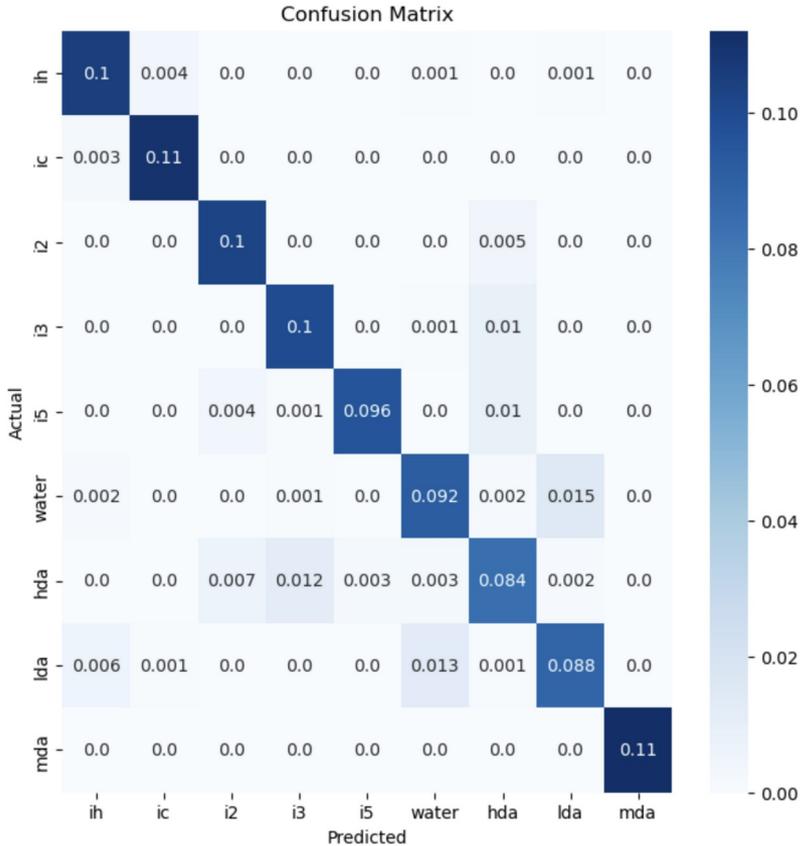
180 MD – траекторий, полученных при помощи LAMMPS



Составление выборки структур и дескрипторы на Python



# Результаты: функции симметрии



Функции обрезки:

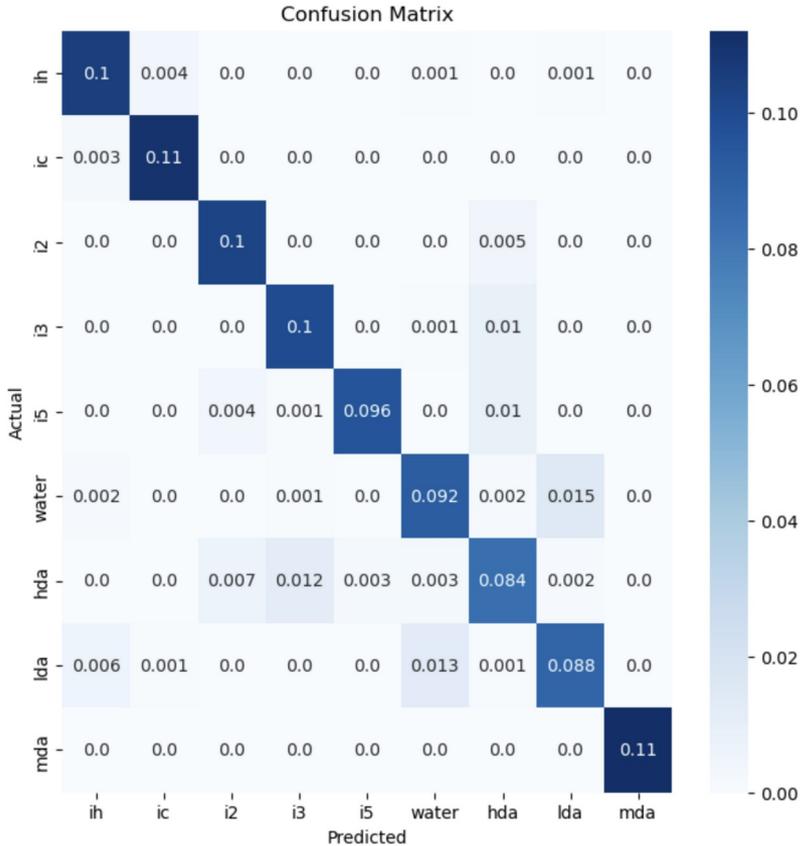
$$f_c(R) = \begin{cases} [1 + \exp\{\alpha(R - R_c + \varepsilon_c)\}]^{(-1)} & \text{if } R < R_c, \\ 0 & \text{else} \end{cases}$$

$$f_b(R) = \begin{cases} \cos^2[\nu(R - a_1)] & \text{if } a_1 - \frac{\pi}{2\nu} < R < a_1, \\ 1 & \text{if } a_1 < R < a_r, \\ \cos^2[\nu(R - a_r)] & \text{if } a_r < R < a_r + \frac{\pi}{2\nu}, \\ 1 & \text{else} \end{cases}$$



**Результат: Accuracy  $\approx 88,92\%$**

# Результаты: функции симметрии



Используемые функции

$$G_2^{[i]}(r) = \sum_{j \neq i} e^{-\eta(R_{ij} - R_s)^2} f_c(R_{ij})$$

$$G_3^{[i]}(r) = \sum_{j \neq i} \cos(\kappa R_{ij}) f_c(R_{ij})$$

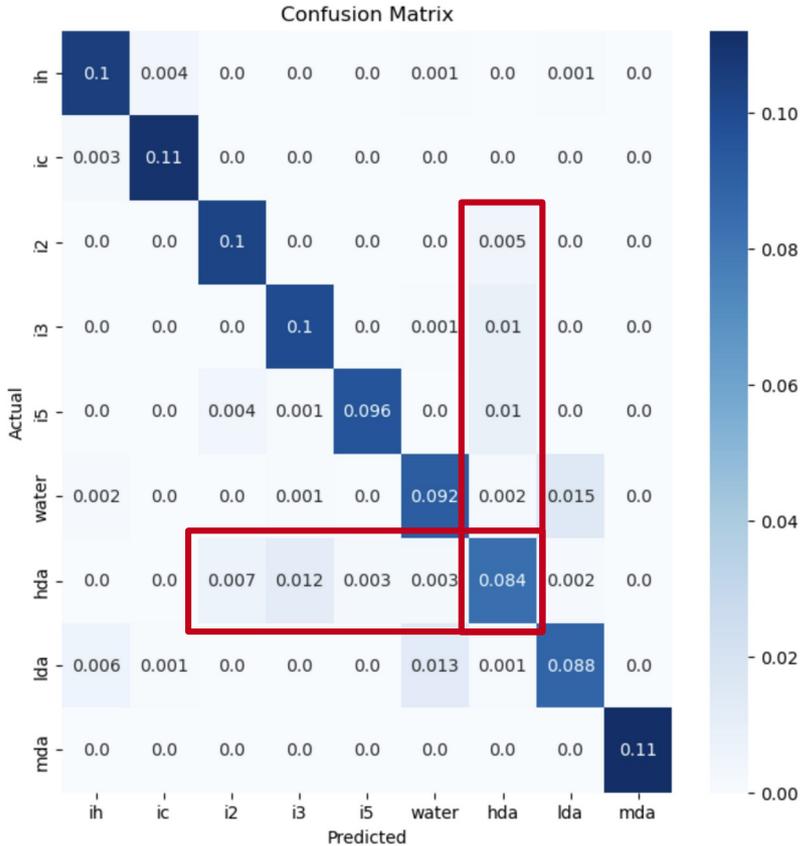
$$G_4^{[i]}(r) = \frac{1}{2^\xi} \sum_{j, k \neq i} ([1 + \lambda \cos(\theta_{ijk})]^\xi \times e^{-\eta(R_{ij}^2 + R_{ik}^2 + R_{jk}^2)} f_c(R_{ij}) f_c(R_{ik}) f_c(R_{jk}))$$

$$G_8^{[i]}(r) = \frac{1}{2} \sum_{j, k \neq i} \sin[\eta(\theta_{ijk} - \alpha)] f_b(R_{ij}) f_b(R_{ik})$$



**Результат: Accuracy ≈ 88,92 %**

# Результаты: функции симметрии



Используемые функции

$$G_2^{[i]}(r) = \sum_{j \neq i} e^{-\eta(R_{ij} - R_s)^2} f_c(R_{ij})$$

$$G_3^{[i]}(r) = \sum_{j \neq i} \cos(\kappa R_{ij}) f_c(R_{ij})$$

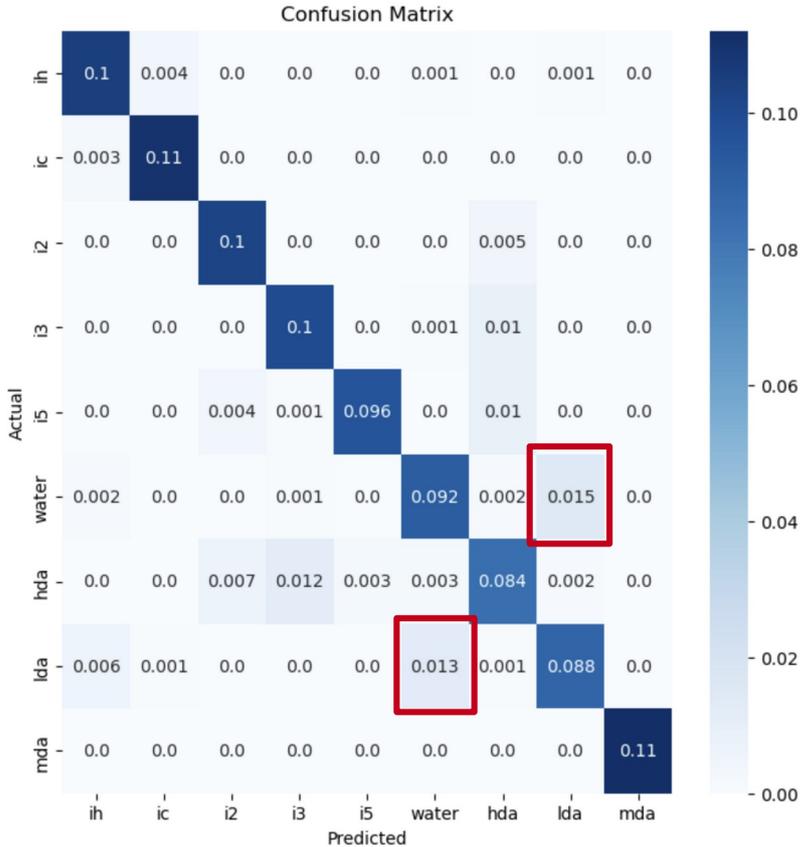
$$G_4^{[i]}(r) = \frac{1}{2^\xi} \sum_{j, k \neq i} ([1 + \lambda \cos(\theta_{ijk})]^\xi \times e^{-\eta(R_{ij}^2 + R_{ik}^2 + R_{jk}^2)} f_c(R_{ij}) f_c(R_{ik}) f_c(R_{jk}))$$

$$G_8^{[i]}(r) = \frac{1}{2} \sum_{j, k \neq i} \sin[\eta(\theta_{ijk} - \alpha)] f_b(R_{ij}) f_b(R_{ik})$$



**Результат: Accuracy ≈ 88,92 %**

# Результаты: функции симметрии



Используемые функции

$$G_2^{[i]}(r) = \sum_{j \neq i} e^{-\eta(R_{ij} - R_s)^2} f_c(R_{ij})$$

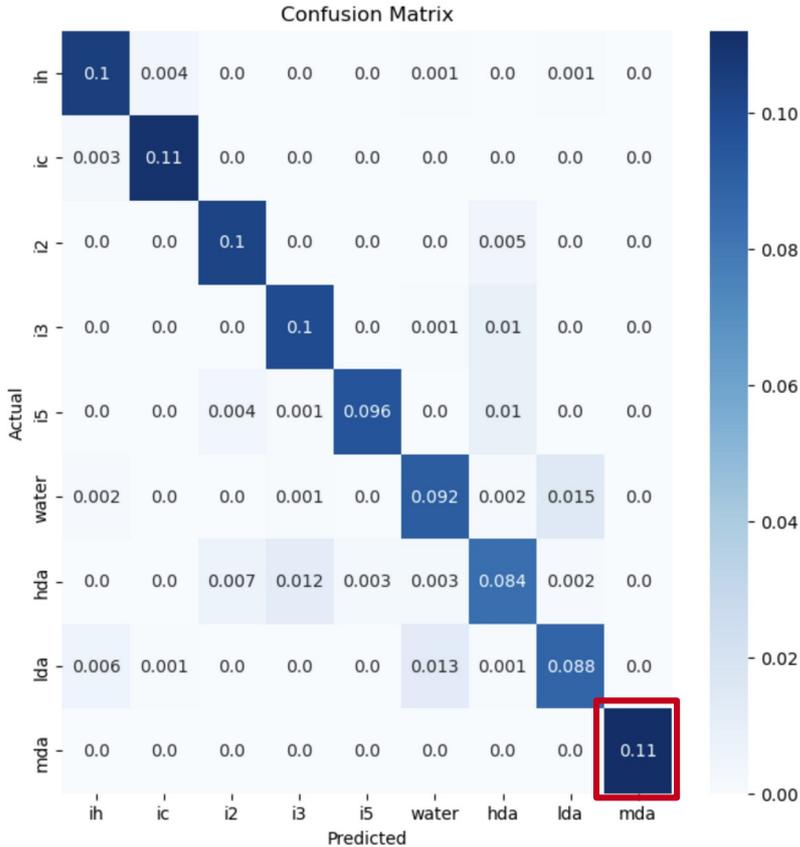
$$G_3^{[i]}(r) = \sum_{j \neq i} \cos(\kappa R_{ij}) f_c(R_{ij})$$

$$G_4^{[i]}(r) = \frac{1}{2^\xi} \sum_{j, k \neq i} ([1 + \lambda \cos(\theta_{ijk})]^\xi) \times e^{-\eta(R_{ij}^2 + R_{ik}^2 + R_{jk}^2)} f_c(R_{ij}) f_c(R_{ik}) f_c(R_{jk})$$

$$G_8^{[i]}(r) = \frac{1}{2} \sum_{j, k \neq i} \sin[\eta(\theta_{ijk} - \alpha)] f_b(R_{ij}) f_b(R_{ik})$$

**Результат: Accuracy  $\approx 88,92\%$**

# Результаты: функции симметрии



**Результат: Accuracy  $\approx 88,92\%$**

Используемые функции

$$G_2^{[i]}(r) = \sum_{j \neq i} e^{-\eta(R_{ij} - R_s)^2} f_c(R_{ij})$$

$$G_3^{[i]}(r) = \sum_{j \neq i} \cos(\kappa R_{ij}) f_c(R_{ij})$$

$$G_4^{[i]}(r) = \frac{1}{2^\xi} \sum_{j, k \neq i} ([1 + \lambda \cos(\theta_{ijk})]^\xi \times e^{-\eta(R_{ij}^2 + R_{ik}^2 + R_{jk}^2)} f_c(R_{ij}) f_c(R_{ik}) f_c(R_{jk}))$$

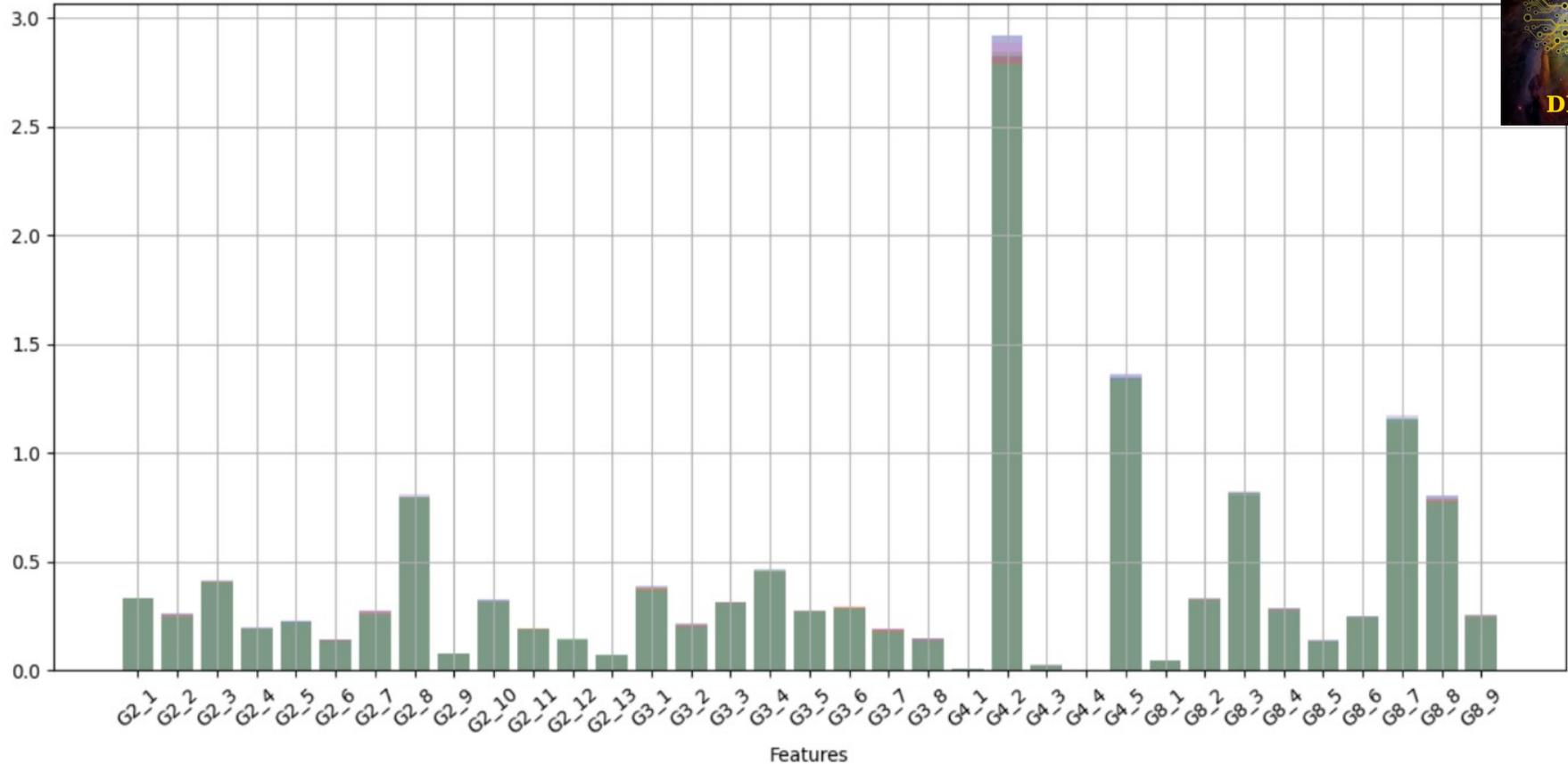
$$G_8^{[i]}(r) = \frac{1}{2} \sum_{j, k \neq i} \sin[\eta(\theta_{ijk} - \alpha)] f_b(R_{ij}) f_b(R_{ik})$$



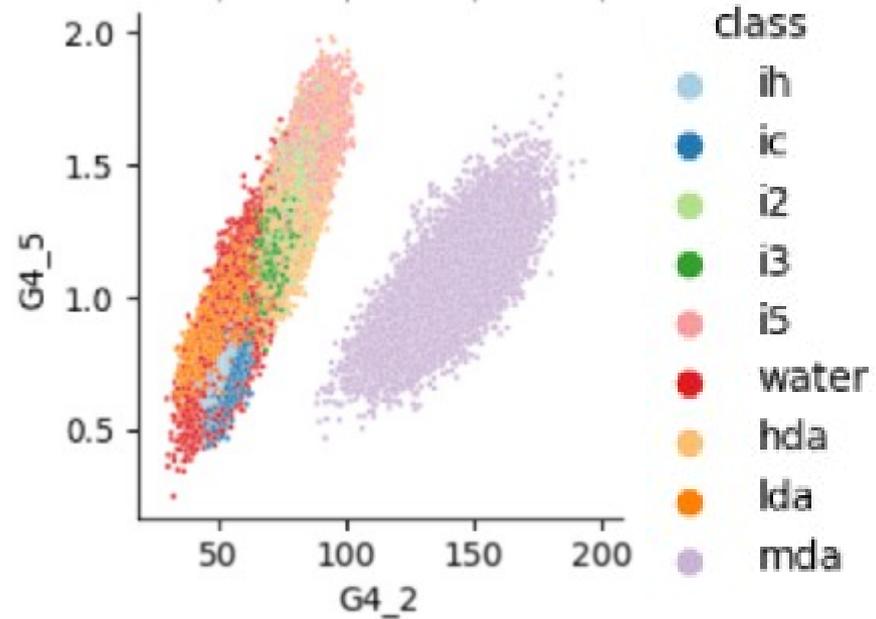
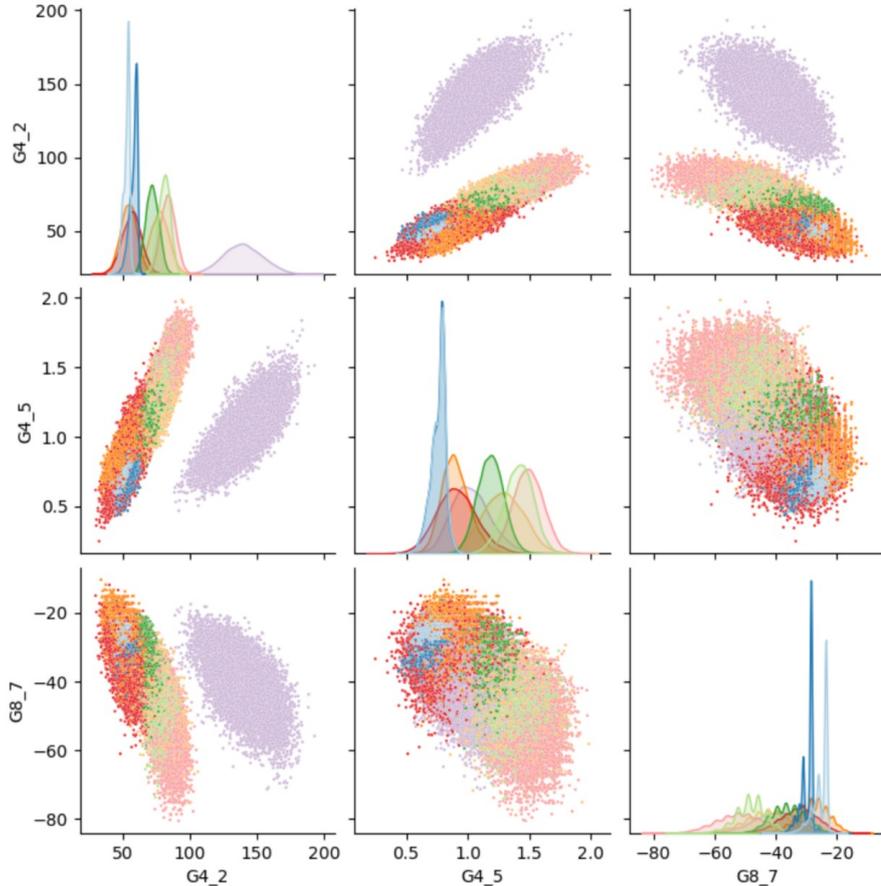
# Результаты: функции симметрии



Average Feature Importances

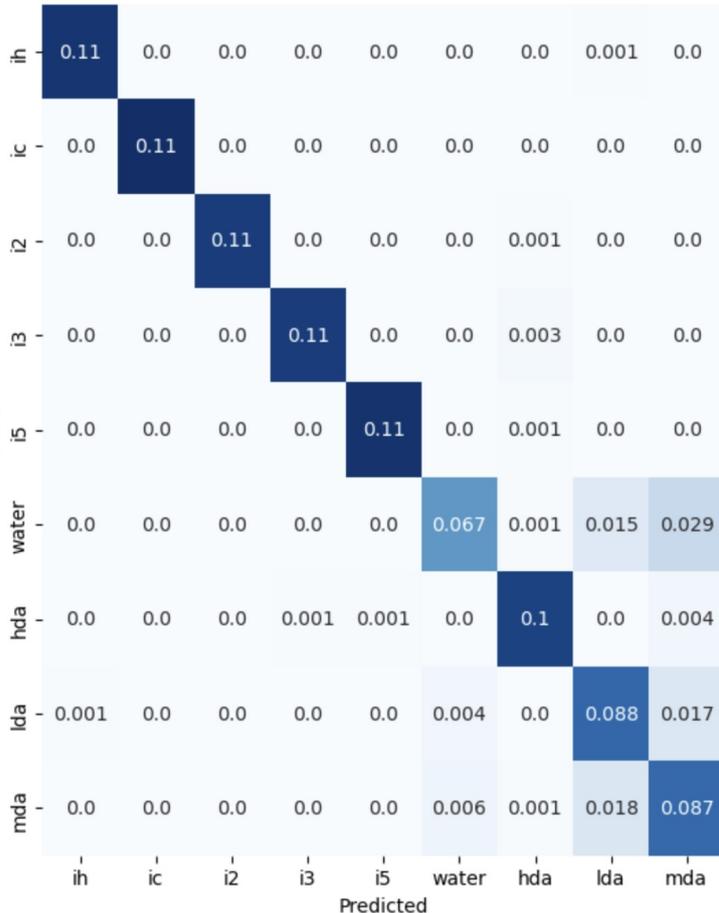


# Результаты: функции симметрии





Confusion Matrix



$$B_{j_1, j_2, j} = \sum_{m'_1, m_1 = -j_1}^{j_1} \sum_{m'_2, m_2 = -j_2}^{j_2} \sum_{m', m = -j}^j (C_{m', m}^j)^* \times C_{j_1 m_1 j_2 m_2}^{j m} C_{j_1 m'_1 j_2 m'_2}^{j m'} C_{m'_1 m_1}^{j_1} C_{m'_2 m_2}^{j_2}$$

$C_{m', m}^j$  - коэффициенты разложения атомарной плотности по Вигнеровским матрицам

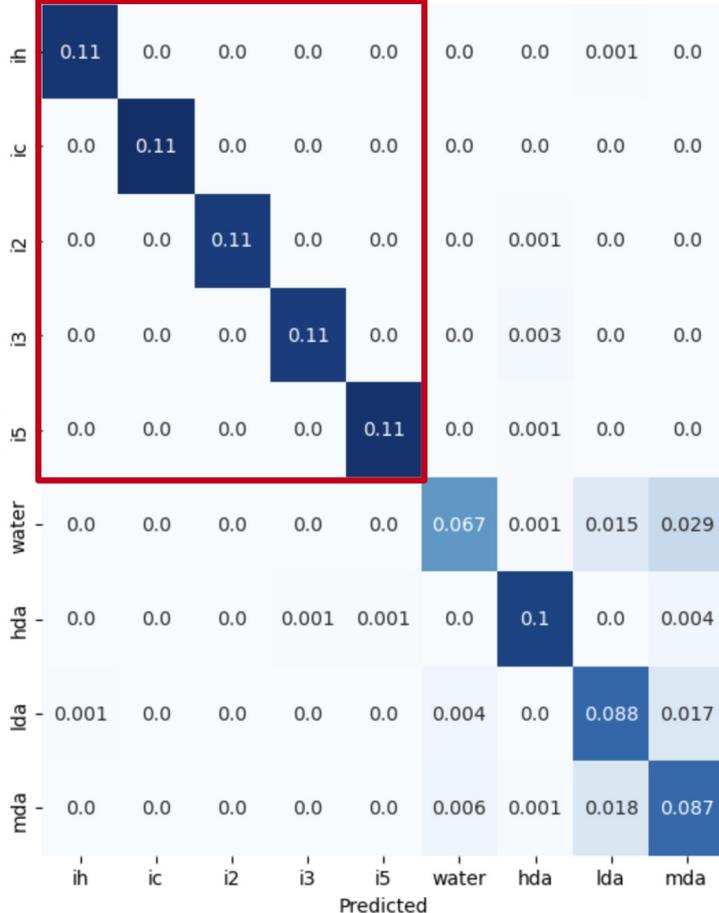
$C_{j_1 m_1 j_2 m_2}^{j m}$  - коэффициенты Клебша-Гордана

**Результат: Accuracy ≈ 89,41 %**

# Результаты: биспектральные компоненты



Confusion Matrix



$$B_{j_1, j_2, j} = \sum_{m'_1, m_1 = -j_1}^{j_1} \sum_{m'_2, m_2 = -j_2}^{j_2} \sum_{m', m = -j}^j (C_{m', m}^j)^* \times C_{j_1 m_1 j_2 m_2}^{j m} C_{j_1 m'_1 j_2 m'_2}^{j m'} C_{m'_1 m_1}^{j_1} C_{m'_2 m_2}^{j_2}$$

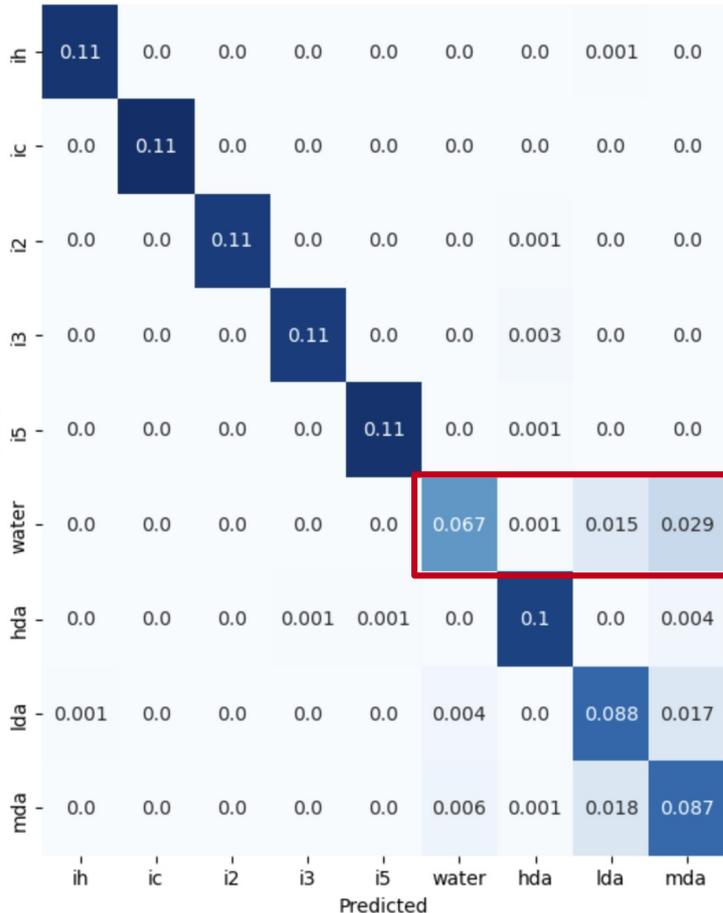
$C_{m', m}^j$  - коэффициенты разложения атомарной плотности по Вигнеровским матрицам

$C_{j_1 m_1 j_2 m_2}^{j m}$  - коэффициенты Клебша-Гордана

**Результат: Accuracy ≈ 89,41 %**



Confusion Matrix



$$B_{j_1, j_2, j} = \sum_{m'_1, m_1 = -j_1}^{j_1} \sum_{m'_2, m_2 = -j_2}^{j_2} \sum_{m', m = -j}^j (C_{m', m}^j)^* \times C_{j_1 m_1 j_2 m_2}^{j m} C_{j_1 m'_1 j_2 m'_2}^{j m'} C_{m'_1 m_1}^{j_1} C_{m'_2 m_2}^{j_2}$$

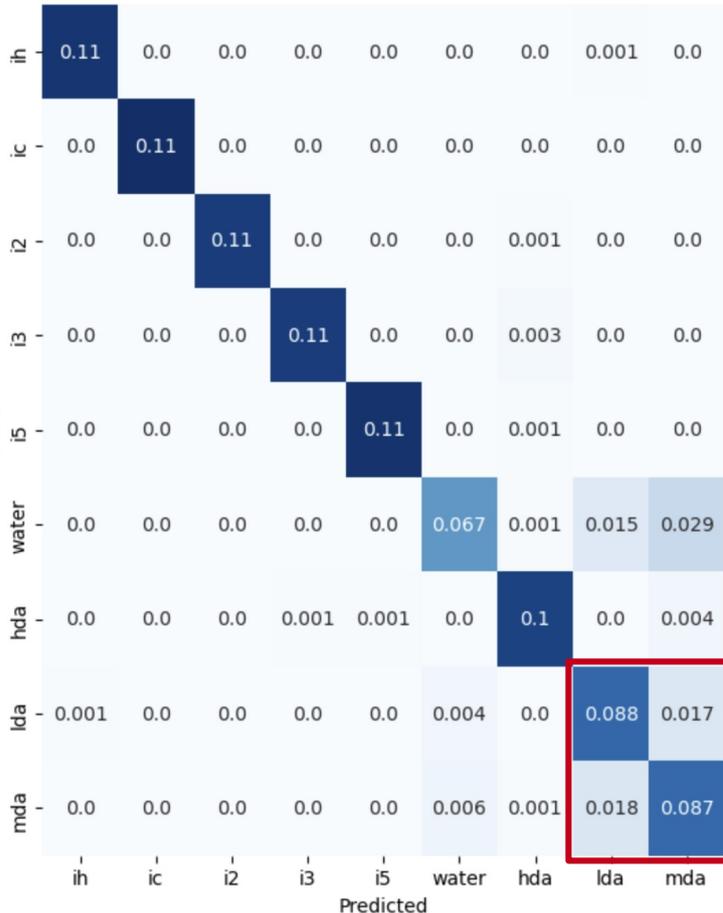
$C_{m', m}^j$  - коэффициенты разложения атомарной плотности по Вигнеровским матрицам

$C_{j_1 m_1 j_2 m_2}^{j m}$  - коэффициенты Клебша-Гордана

**Результат: Accuracy ≈ 89,41 %**



Confusion Matrix

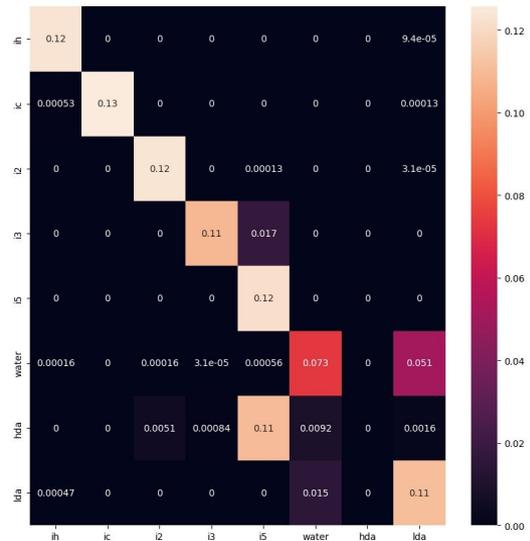
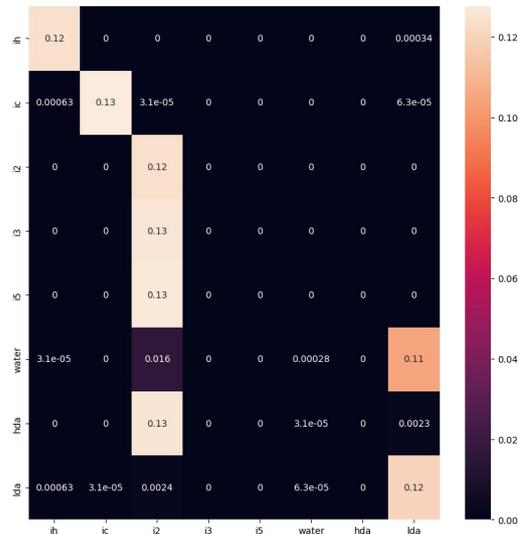
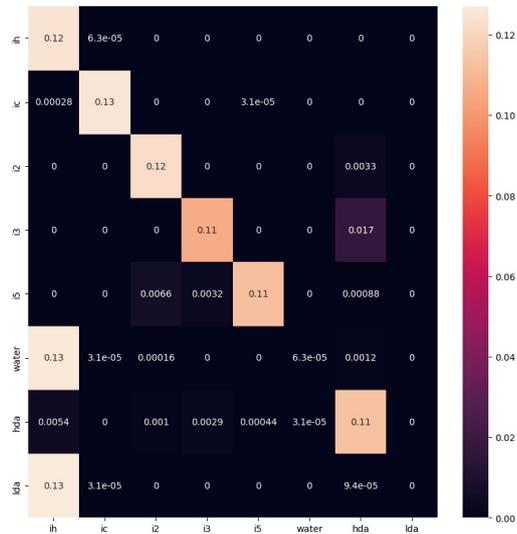


$$B_{j_1, j_2, j} = \sum_{m'_1, m_1 = -j_1}^{j_1} \sum_{m'_2, m_2 = -j_2}^{j_2} \sum_{m', m = -j}^j (c_{m', m}^j)^* \times C_{j_1 m'_1 j_2 m'_2}^{j m} C_{j_1 m'_1 j_2 m'_2}^{j m'} C_{m'_1 m_1}^{j_1} C_{m'_2 m_2}^{j_2}$$

$c_{m', m}^j$  - коэффициенты разложения атомарной плотности по Вигнеровским матрицам

$C_{j_1 m'_1 j_2 m'_2}^{j m}$  - коэффициенты Клебша-Гордана

**Результат: Accuracy ≈ 89,41 %**

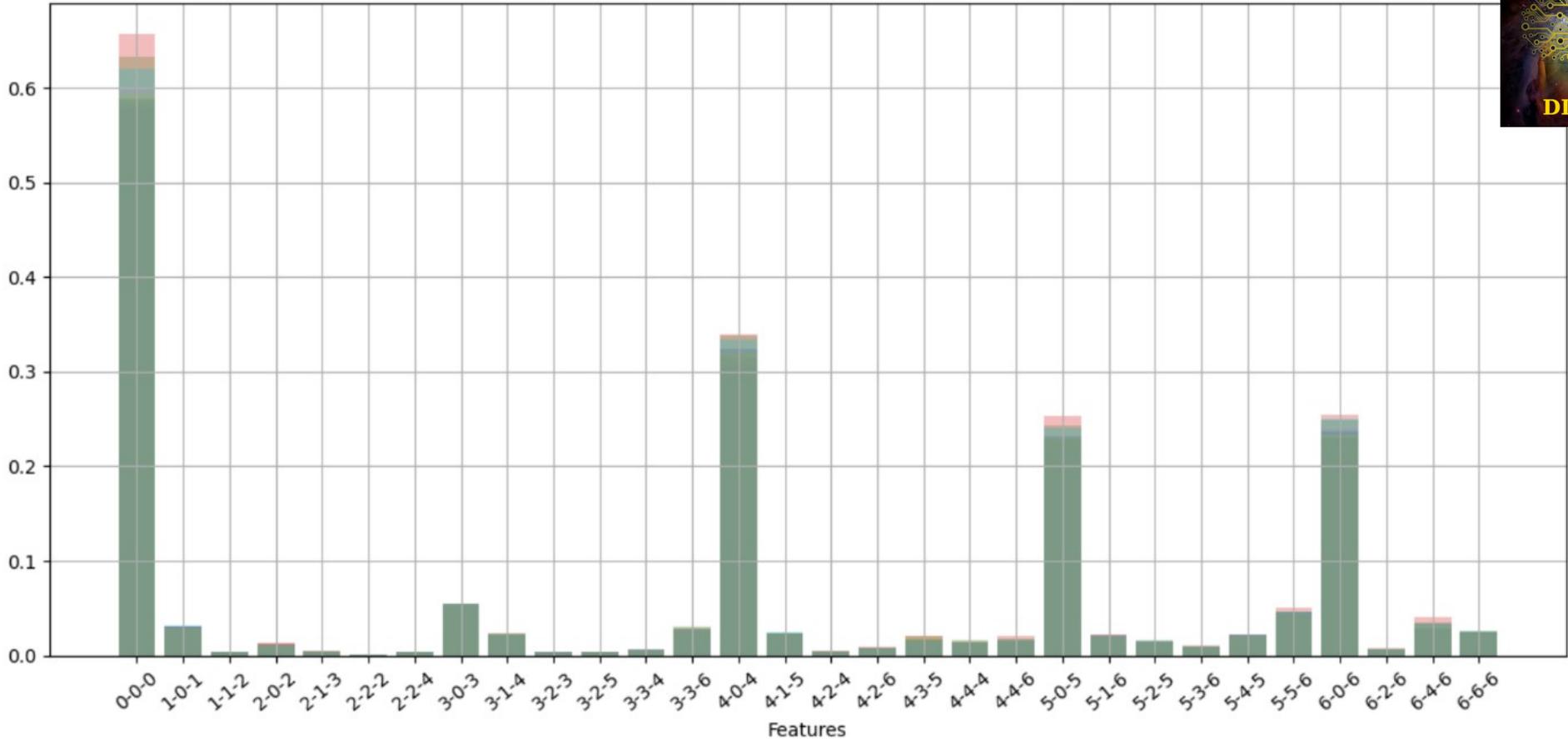


**Проблемы с обучением вероятно связаны со сложной формой поверхности, на которой происходит оптимизация**

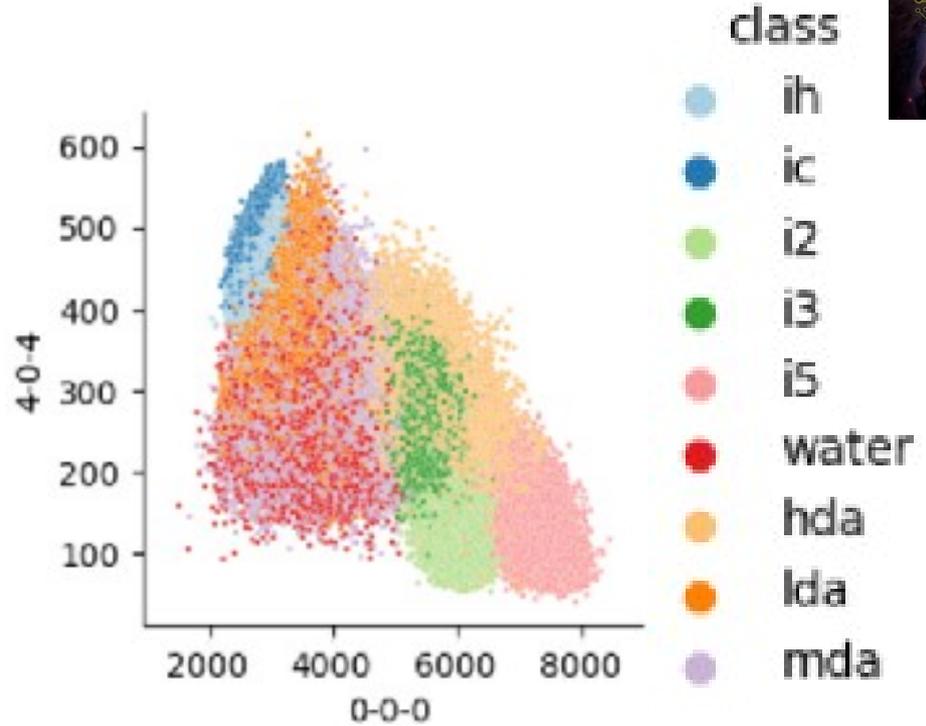
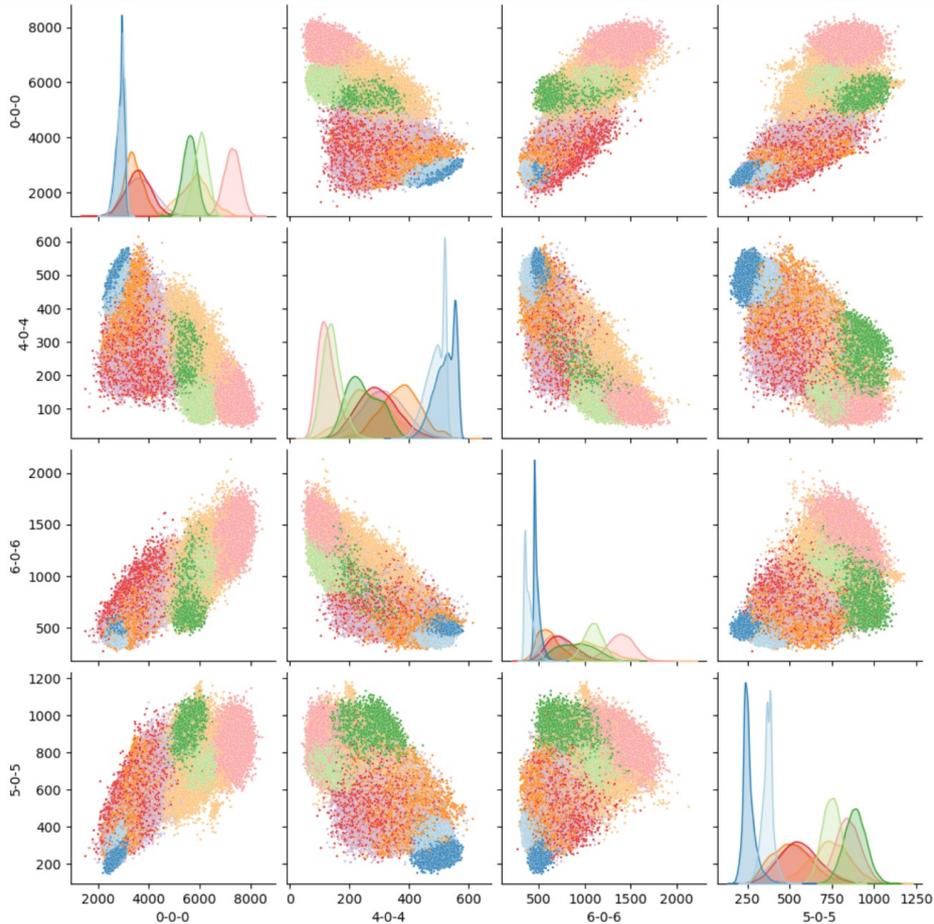
**Кроме того, вычисление биспектральных компонент — существенно более ресурсозатратная задача**

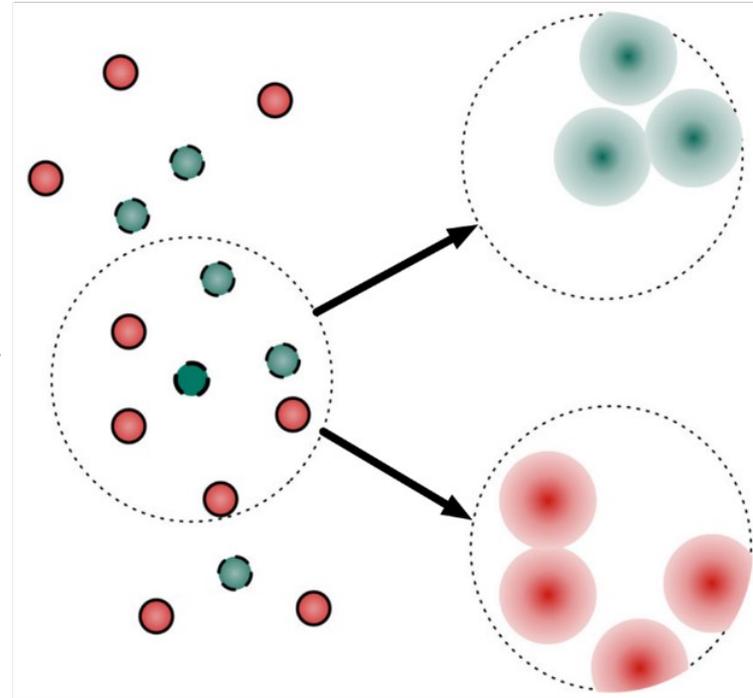
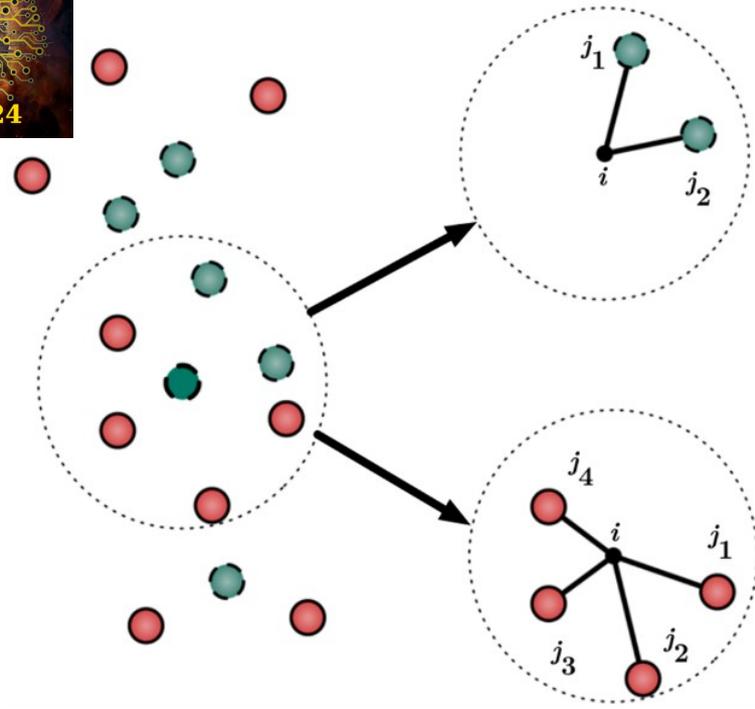
# Результаты: биспектральные компоненты

Average Feature Importances



# Результаты: биспектральные компоненты





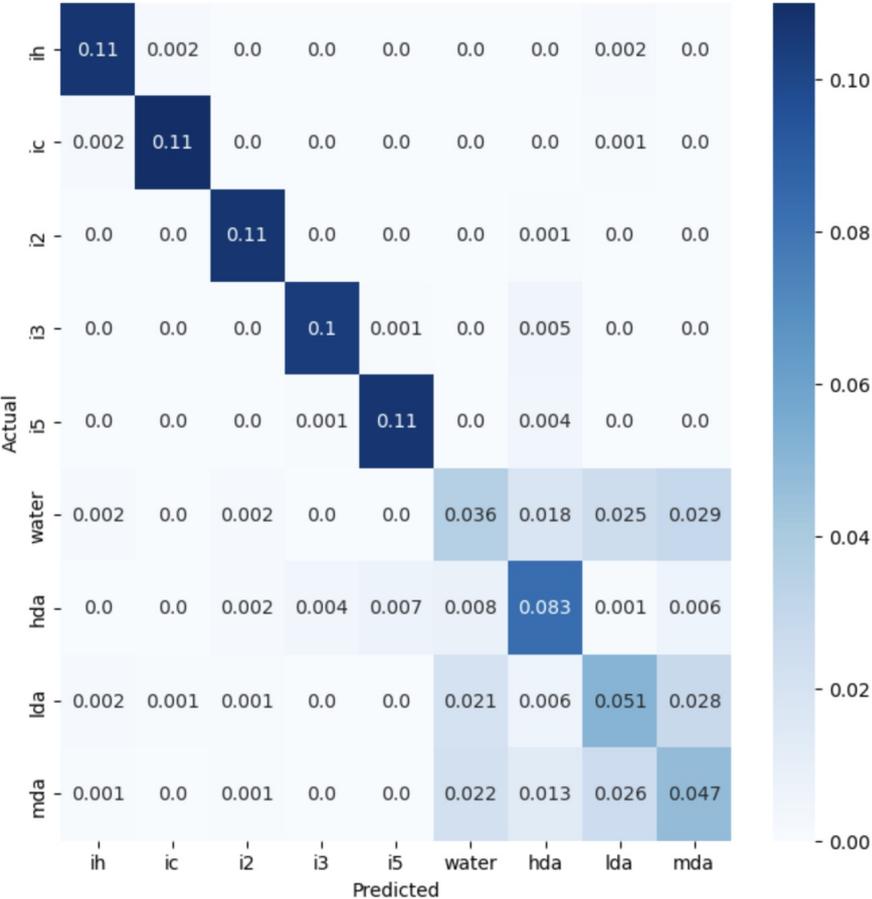
$$\rho_i(r) = \delta(r) + \sum_{i' \neq i} f_c(R_{ii'}) w_{i'} \delta(r - R_{ii'})$$

$$\rho(r) = \sum_i e^{-\alpha|r-r_i|^2} = \sum_{nlm} c_{nlm} g_n(r) Y_{lm}(\hat{r})$$

# Результаты: SOAR (37 компонент)



Confusion Matrix



$$P_{nn'l} \equiv \pi \sqrt{\frac{8}{2l+1}} \sum_m C_{nlm} (C_{n'lm})^*$$

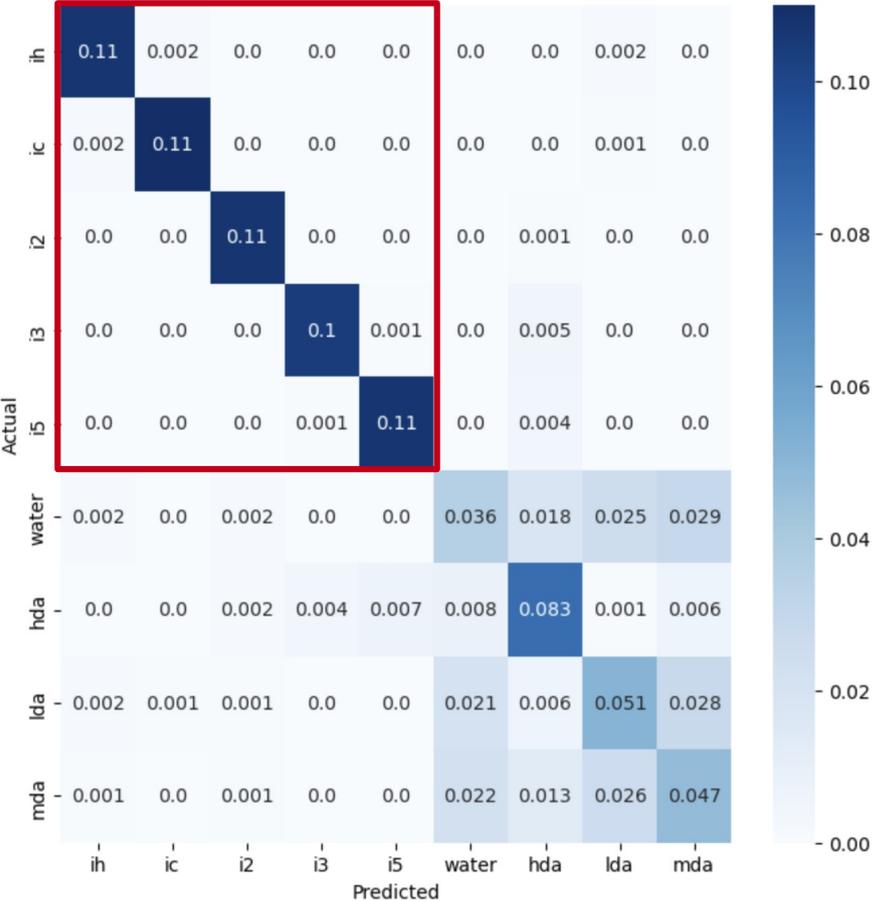
$C_{nlm} = \int g_n(r) Y_{lm}(\theta, \phi) \rho(r) dr$  - коэффициенты разложения атомарной плотности по радиально базисным и сферическим функциям

**Результат: Accuracy ≈ 75,23 %**

# Результаты: SOAR (37 компонент)



Confusion Matrix



$$P_{nn'l} \equiv \pi \sqrt{\frac{8}{2l+1}} \sum_m C_{nlm} (C_{n'lm})^*$$

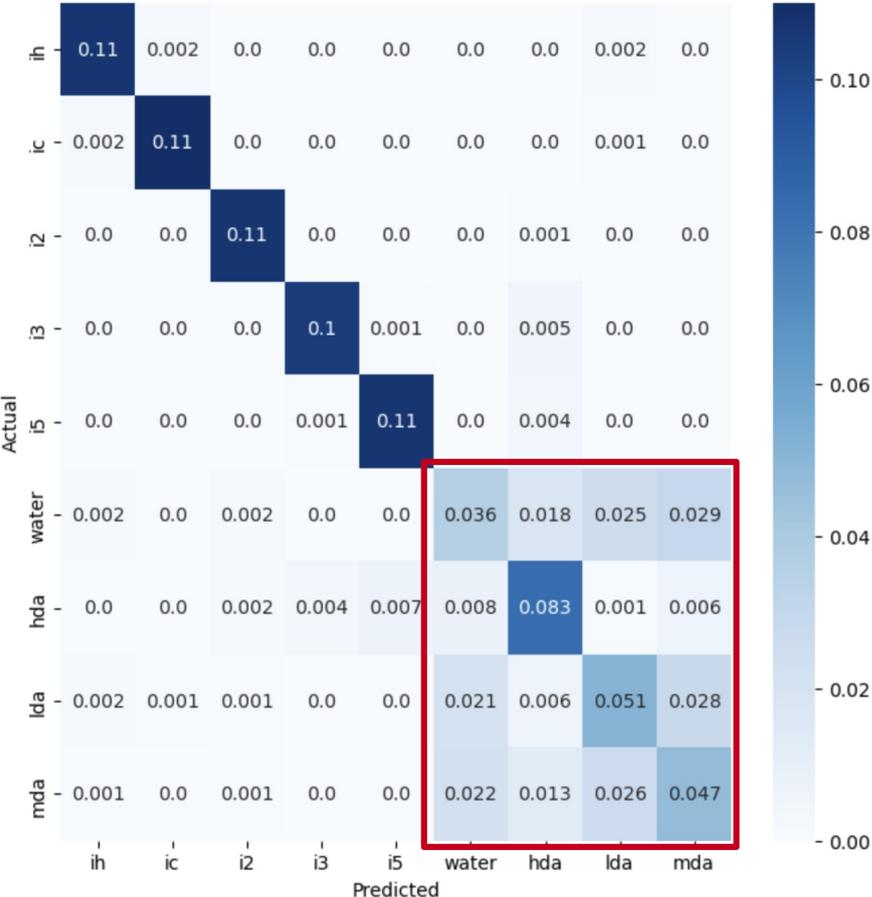
$C_{nlm} = \int g_n(r) Y_{lm}(\theta, \phi) \rho(r) dr$  - коэффициенты разложения атомарной плотности по радиально базисным и сферическим функциям

**Результат: Accuracy ≈ 75,23 %**

# Результаты: SOAR (37 компонент)



Confusion Matrix



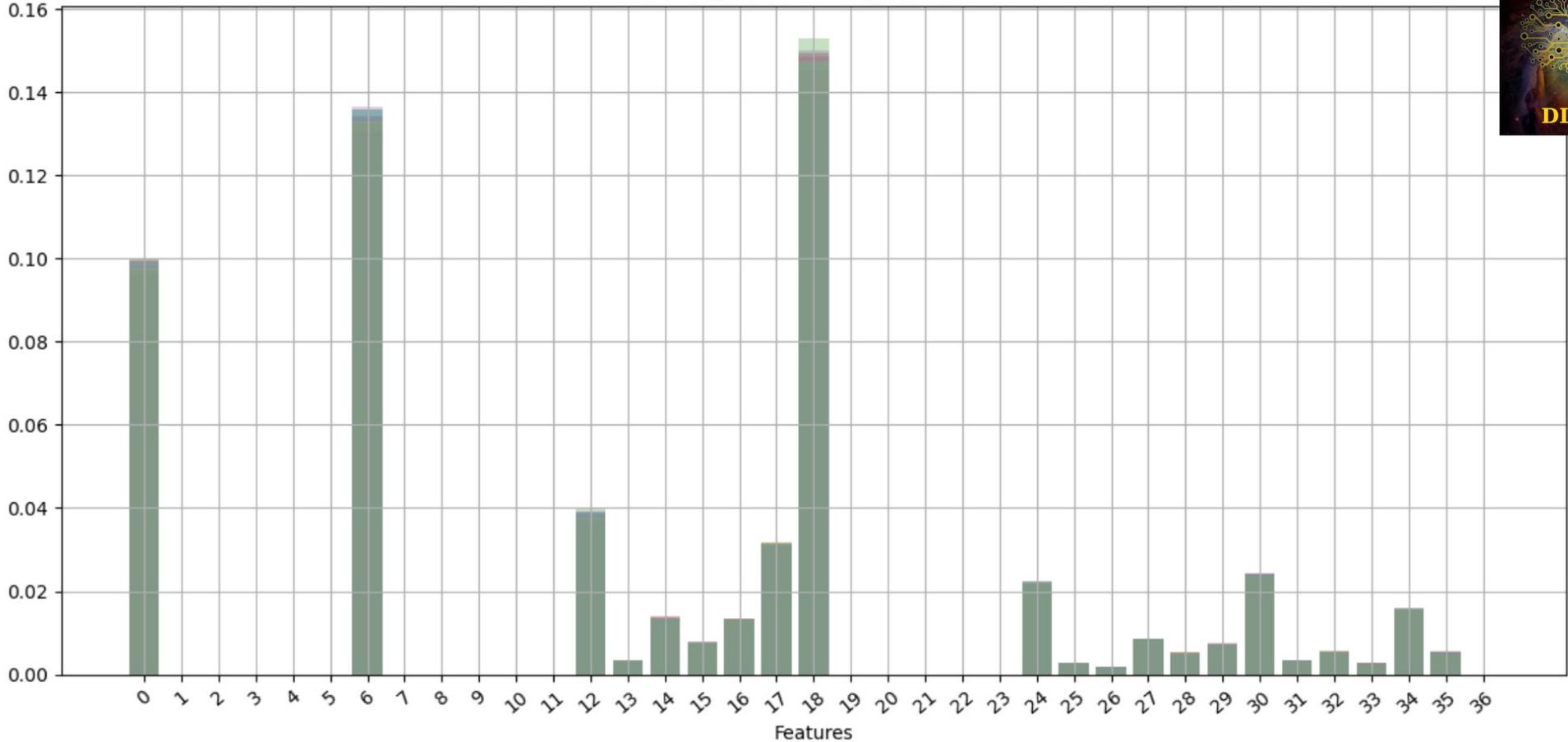
$$P_{nn'l} \equiv \pi \sqrt{\frac{8}{2l+1}} \sum_m C_{nlm} (C_{n'lm})^*$$

$C_{nlm} = \int g_n(r) Y_{lm}(\theta, \phi) \rho(r) dr$  - коэффициенты разложения атомарной плотности по радиально базисным и сферическим функциям

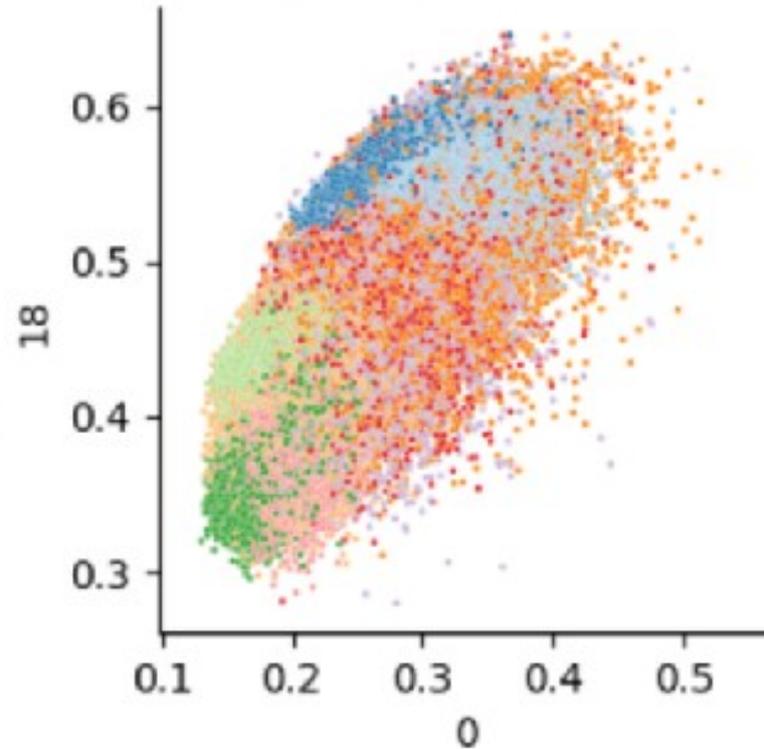
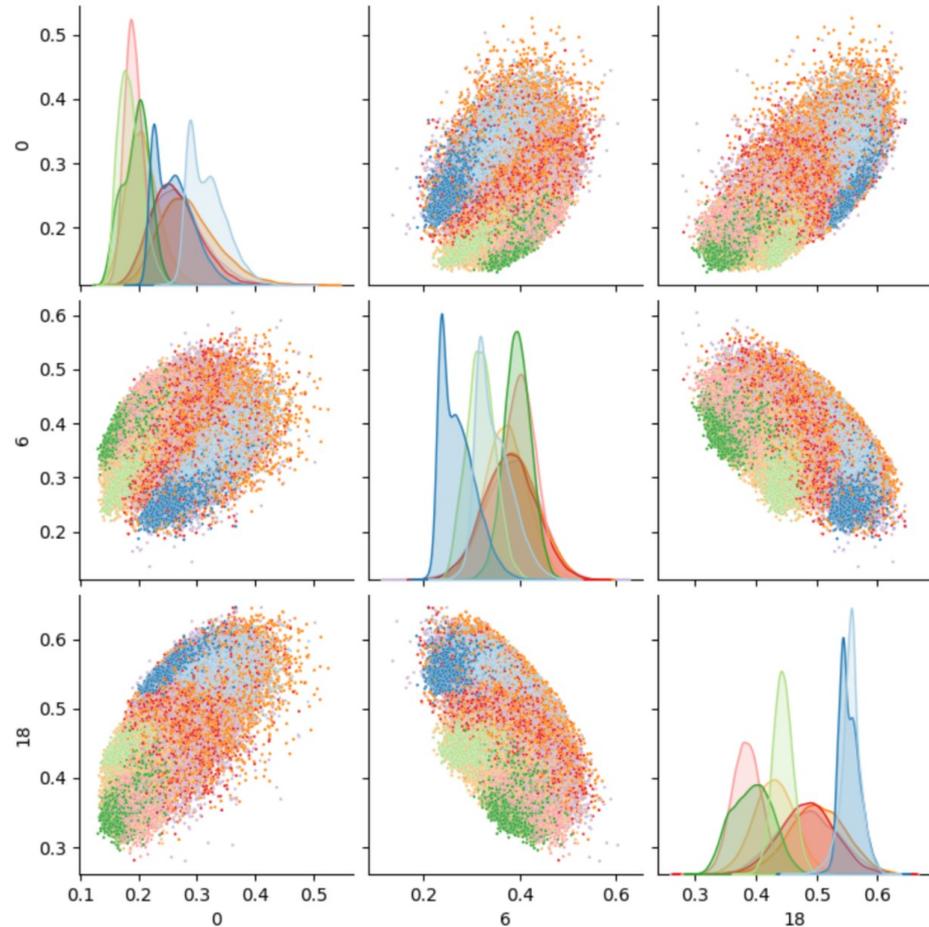
**Результат: Accuracy ≈ 75,23 %**

# Результаты: SOAR (37 компонент)

Average Feature Importances



# Результаты: SOAR (37 компонент)

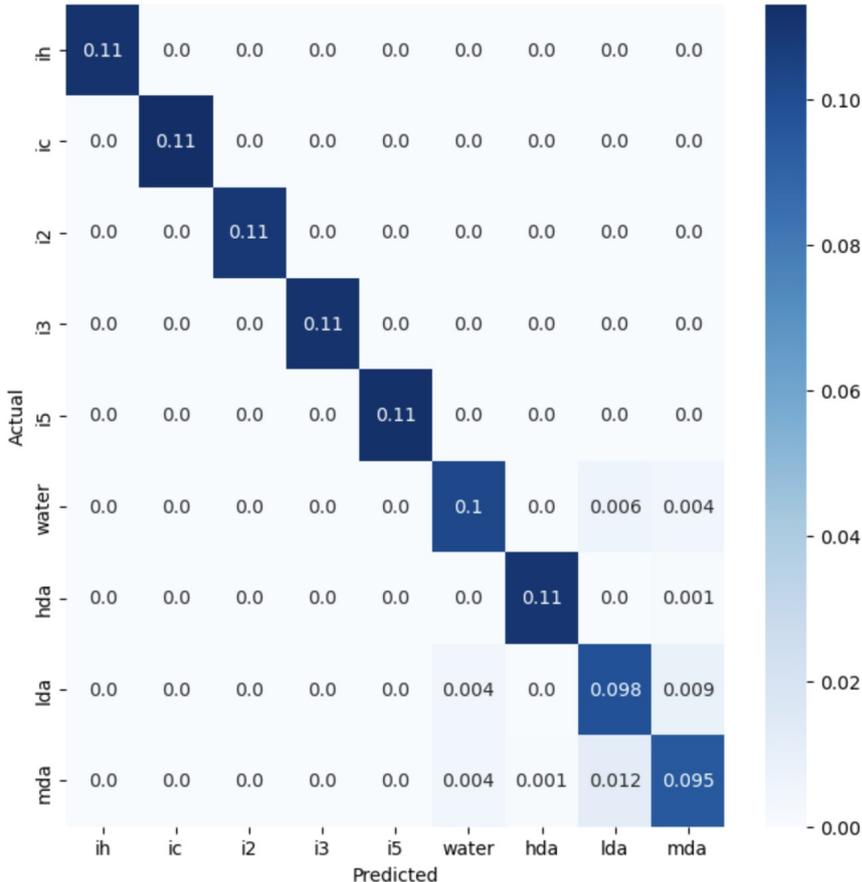


- class
- ih
  - ic
  - i2
  - i3
  - i5
  - water
  - hda
  - lda
  - mda

# Результаты: SOAP (325 компонент)



Confusion Matrix



$$P_{nn'l} \equiv \pi \sqrt{\frac{8}{2l+1}} \sum_m C_{nlm} (C_{n'lm})^*$$

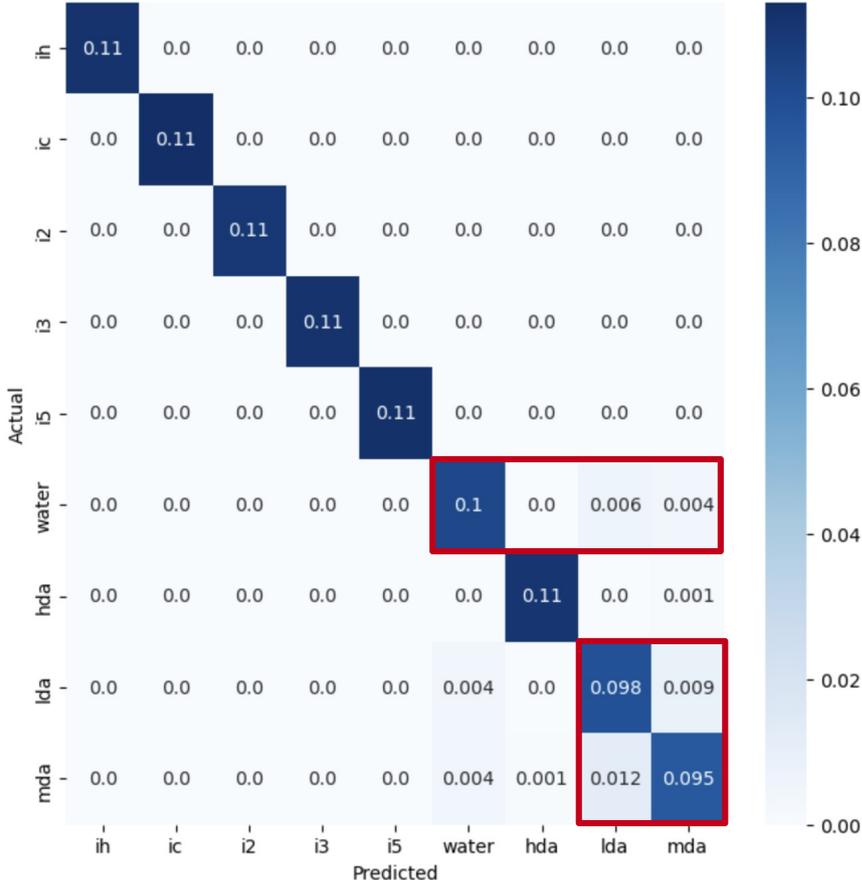
$C_{nlm} = \int g_n(r) Y_{lm}(\theta, \phi) \rho(r) dr$  - коэффициенты разложения атомарной плотности по радиально базисным и сферическим функциям

**Результат: Accuracy ≈ 95,88 %**

# Результаты: SOAR (325 компонент)



Confusion Matrix



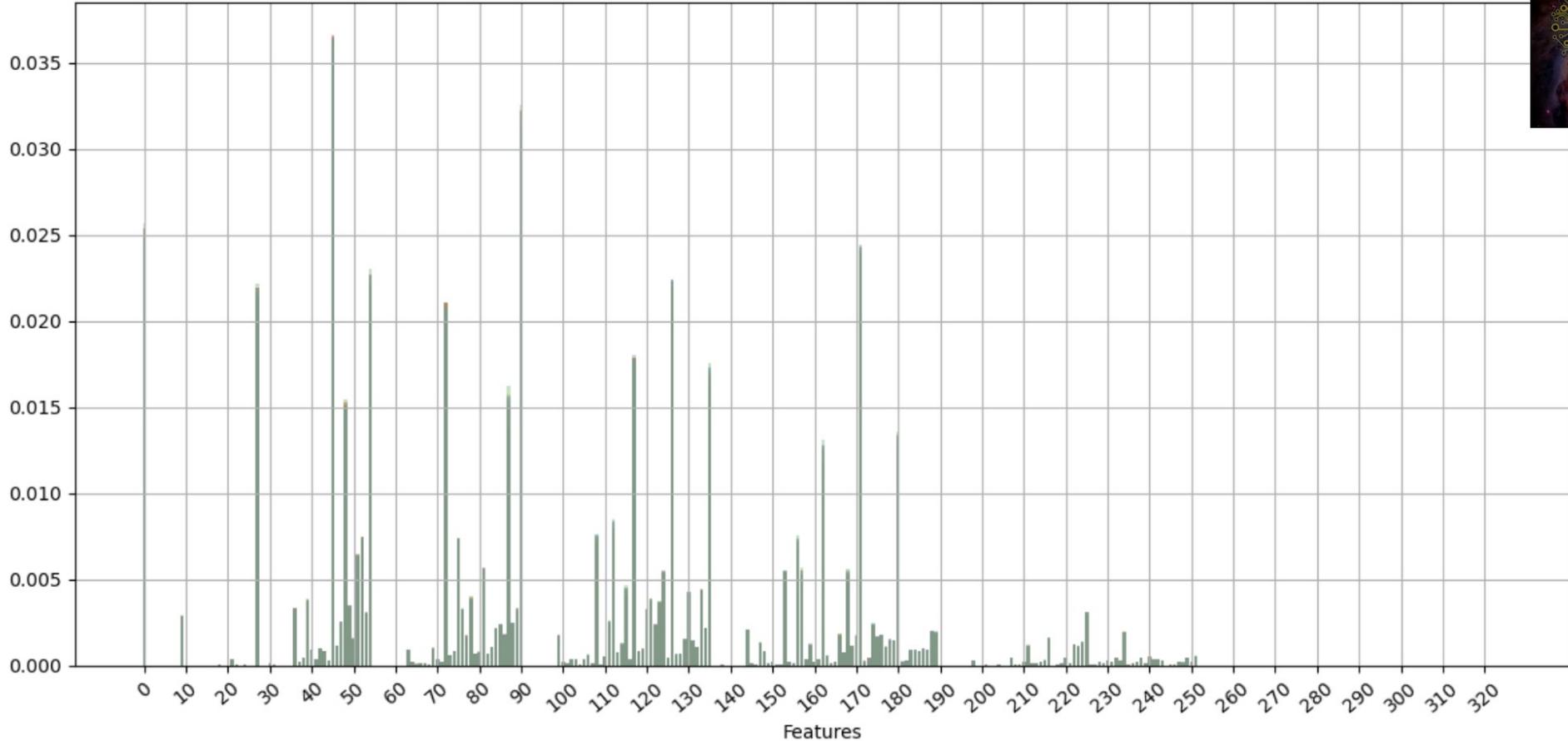
$$P_{nn'l} \equiv \pi \sqrt{\frac{8}{2l+1}} \sum_m C_{nlm} (C_{n'lm})^*$$

$C_{nlm} = \int g_n(r) Y_{lm}(\theta, \phi) \rho(r) dr$  - коэффициенты разложения атомарной плотности по радиально базисным и сферическим функциям

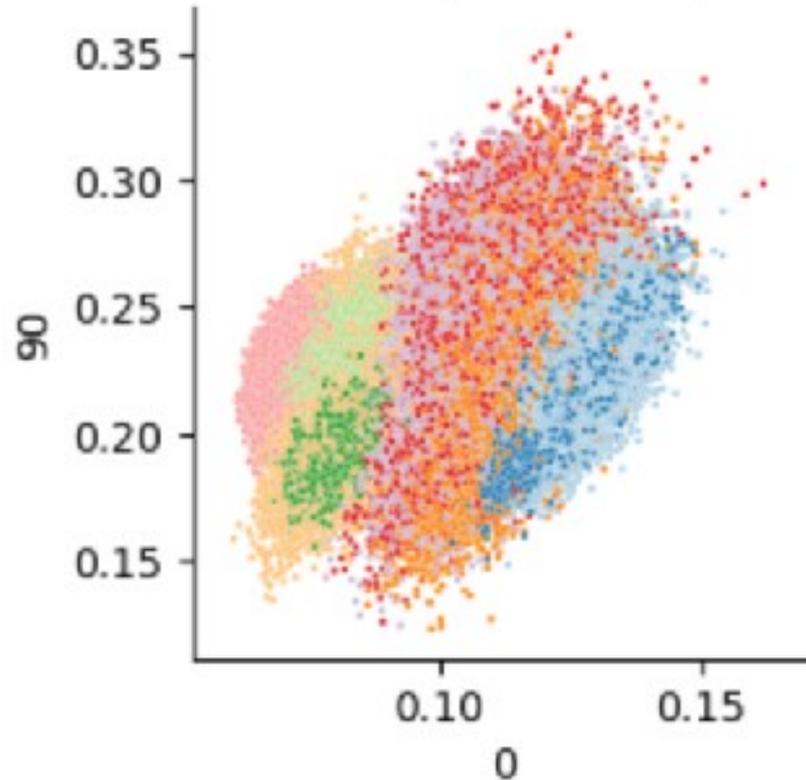
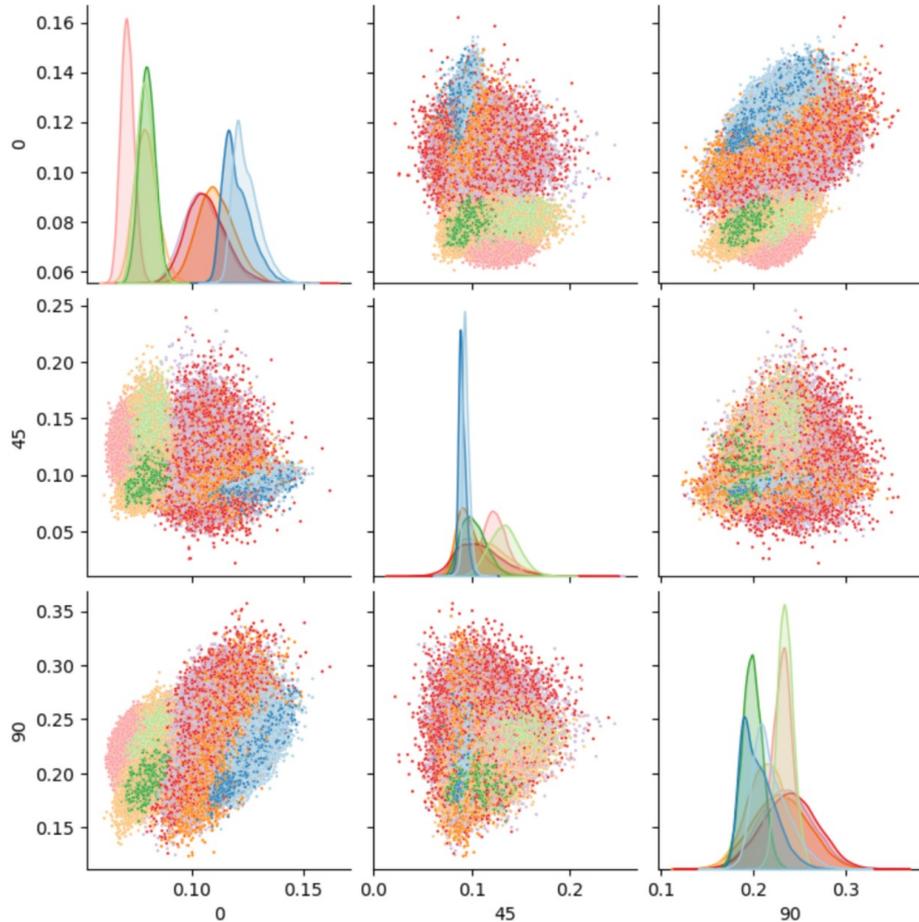
**Результат: Accuracy ≈ 95,88 %**

# Результаты: SOAR (325 компонент)

Average Feature Importances



# Результаты: SOAR (325 компонент)



- class
- ih
  - ic
  - i2
  - i3
  - i5
  - water
  - hda
  - lda
  - mda

# Результаты: время вычислений



Дескриптор	Время вычисления / число признаков
Bispectrum	15 мин/30
SOAP <sub>1</sub>	12 мин/37
SOAP <sub>2</sub>	16 мин/325

Вычисления производились на 16-ти ядрах процессора Intel Core i7 12700K

Функции симметрии не принимали участия в эксперименте, так как условия вычислений были различны



- Полученные при помощи трёх типов дескрипторов результаты позволяют осуществить оптимальный с точки зрения времени вычислений и необходимой точности результата выбор метода ОПИСАНИЯ Faure Beaulieu Z., Deringer V. L., Martelli F. //J. Chem. Phys. 160, 081101 (2024)
- Наличие признака (функции симметрии), явным образом выделяющего MDA в задаче о распознавании водных фаз, может служить очередным подтверждением уникальности структуры и её отличие от других аморфных льдов.
- Компоненты дескриптора SOAP могут быть не информативны с точки зрения структурного описания системы.

# Дополнительные слайды: определение важности (Integrated Gradients)



$$\text{IntegratedGradients}_i(x) ::= (x_i - x'_i) \times \int_{\alpha=0}^1 \frac{\partial F(x' + \alpha \times (x - x'))}{\partial x_i} d\alpha$$

Здесь

- $i$  – номер признака
- $x$  – некоторый вход
- $x'$  – базисный вход
- $\alpha$  – параметр, задающий прямую