

## Размещение экземпляра ПО, развернутого для экспертизы

Для экспертизы экземпляр ПО предоставляется в виде дистрибутива на Яндекс Диске. Он доступен по адресу: [https://disk.yandex.ru/d/scjLdVGs02Ku\\_g](https://disk.yandex.ru/d/scjLdVGs02Ku_g)

## Состав и структура дистрибутива

Программа представлена как набор папок и файлов, формирующих единый дистрибутив:

- **docker**: папка, содержащая Dockerfile - файл с описанием программного окружения и зависимостей, необходимых для работы ПО
- **DL\_template**: папка, обеспечивающая функционал основного модуля обучения нейросетевых моделей разных типов для предсказания ГФЭ.

Включает папки:

- **artifacts**, содержащую результаты работы модуля – checkpoints моделей и полногеномные предсказания моделей;
- **data**, в которую пользователь должен загрузить исходные данные для построения моделей;
- **assets**, содержащую два поясняющих изображения для шаблона Template.ipynb;
- **models**, содержащую примеры различных типов нейронных сетей;

и файлы

- **Template.ipynb**, представляющий пример работы с модулем и осуществляющий запуск работы модуля через редактирование шаблона.
- **README.mb** с документацией модуля для пользователя
- **OmicsDC**: папка, обеспечивающая функционал модуля загрузки данных. Включает папки:
  - **OmicsDC**, содержащую имплементацию функционала модуля загрузки данных

- storage для сохранения результатов по умолчанию

а также файлы:

- README.md с документацией модуля для пользователя
- Example.ipynb, представляющий пример работы с модулем и осуществляющий запуск работы модуля через редактирование шаблона
- Association: папка, обеспечивающая функционал модуля исследования признаков на ассоциацию с ГФЭ. Включает папки:
  - Association , содержащую имплементацию функционала модуля исследования признаков на ассоциацию
  - storage для сохранения результатов по умолчанию
- DomainAdaptation: папка, обеспечивающая функционал модуля трансферного обучения для переноса аннотации ГФЭ. Включает папки и файлы:
  - DomainAdaptation, содержащую имплементацию функционала модуля трансферного обучения
  - storage для сохранения результатов по умолчанию
  - conda\_requirements.yml с необходимыми библиотеками и их версиями для установки с помощью conda
  - pip.requirements с необходимыми библиотеками и их версиями для установки с помощью pip
- Interpretation: папка, обеспечивающая функционал модуля интерпретации обученных моделей. Включает файл:
  - lrp.py, содержащий имплементацию функционала модуля интерпретации обученных моделей на языке программирования Python

## Настройка программы

Для настройки программы следует:

1. Зайти на сервер с помощью протокола SSH и установить SSH-соединение с переадресацией портов, используя следующую команду:

```
ssh -L localhost:8888:localhost:8888 -p <server-port>  
<username>@<server-address>
```

Перенаправить локальный порт 8888 на удаленный порт 8888.

Заменить адрес сервера для подключения и номер порта.

2. Скачать репозиторий на сервер:

```
git clone --recurse-submodules https://github.com/hse-bioinflab/framework
```

3. Создать контейнер Docker со всеми необходимыми зависимостями:

```
cd docker  
docker build -t framework:1.0 .  
cd ..
```

4. Запустить интерфейс Jupyter Lab:

```
docker run -it --rm --runtime=nvidia --gpus all \  
  --shm-size=4GB -p 8888:8888 \  
  -v $(pwd):/workspace/ \  
  framework:1.0
```

5. Проверить свою консоль на наличие URL-адреса, начинающегося с `http://127.0.0.1:8888/lab?token =`. Скопировать URL-адрес и открыть его в своем браузере.

6. Сохранить сеанс SSH запущенным во время работы.

## Алгоритм работы программы

- Программное обеспечение начинает работу, когда пользователь запускает сервер JupyterLab и открывает пользовательскую среду в веб-браузере.
- Затем пользователь открывает шаблон, в котором, используя модуль №2, выбирает и загружает необходимые омиксные данные на сервер.
- Далее пользователь работает с Основным модулем, где ему предоставляется возможность выбора (1) омиксных данных необходимого типа из загруженных с помощью модуля №2, (2) аннотации ГФЭ для обучения, (3) полного генома анализируемого вида и (4) типа архитектуры нейронной сети для использования в предсказании.
- Пользователь запускает отредактированный им шаблон модуля №1. Программа подгружает в память ранее загруженные омиксные данные и/или данные пользователя, конструирует модель и запускает процесс обучения модели. В процессе обучения Основной модуль визуализирует прогресс и текущие метрики.
- Программа использует обученную модель для полногеномного предсказания ГФЭ и предоставляет результаты предсказания пользователю.

### **Логин и пароль для экспертов**

Логин и пароль для экспертов отсутствует