
УТВЕРЖДЕНА

Решением Ученого совета
АНО ВО «Центральный университет»
«25» мая 2026 г.
Протокол № 2

**Рабочая программа дисциплины (модуля)
«Глубокое обучение в биоинформатике»**

Направление подготовки: 02.04.01 Математика и компьютерные науки

Направленность (профиль) подготовки: III в биотехе

Квалификация (степень) выпускника: магистр

Форма обучения: очная

Срок освоения программы: 2 года

Год набора: 2026

**Москва
2026**

Содержание

1. Краткая характеристика дисциплины (модуля)	3
2. Перечень планируемых результатов обучения.....	5
3. Тематический план.....	9
4. Содержание дисциплины (модуля).....	9
5. Учебно-методическое обеспечение	11
6. Материально-техническое обеспечение	11
7. Методические и оценочные материалы	13

1. Краткая характеристика дисциплины (модуля)

Рабочая программа дисциплины (модуля) «Глубокое обучение в биоинформатике» составлена в соответствии с федеральным государственным образовательным стандартом высшего образования – магистратура по направлению 02.04.01 Математика и компьютерные науки, профиль ИИ в биотехе, утвержденный приказом Министерства науки и высшего образования Российской Федерации № 810 от 23.08.2017 года.

Изучение дисциплины (модуля) «Глубокое обучение в биоинформатике» позволяет подготовить специалистов, способных эффективно применять передовые методы искусственного интеллекта для анализа сложных биологических данных, что критически важно для развития персонализированной медицины, дизайна лекарств и фундаментальных исследований в области молекулярной биологии.

Место дисциплины (модуля) в структуре образовательной программы

Настоящая дисциплина (модуль) включена в учебный план по программе подготовки магистратуры по направлению 02.04.01 Математика и компьютерные науки, профиль ИИ в биотехе и входит в обязательную часть Блока 1 как дисциплина по выбору.

Дисциплина (модуль) изучается на 1 курсе во 2 семестре.

Цель изучения дисциплины (модуля): формирование у обучающихся теоретических знаний и практических навыков в области применения методов глубокого обучения для анализа биологических данных, включая последовательности, структуры белков и одноклеточные данные, с целью решения актуальных задач биоинформатики и биомедицины.

Задачи изучения дисциплины (модуля):

- изучить архитектуры глубоких нейронных сетей и их применение для анализа биологических данных;
- освоить принципы обучения нейросетей, включая прямое и обратное распространение ошибки, выбор оптимизаторов и функций потерь;
- научиться подготавливать и кодировать биологические данные для обучения моделей с учётом их специфики;
- овладеть навыками реализации, валидации и интерпретации нейросетевых моделей в контексте биологических гипотез;
- развить компетенции в использовании современных фреймворков, облачных платформ и практик воспроизводимого анализа при решении междисциплинарных задач.

В результате освоения дисциплины (модуля) обучающийся должен:

знать:

- основные архитектуры глубоких нейронных сетей: полносвязные, свёрточные (CNN), рекуррентные (RNN, LSTM, GRU), трансформеры;
- принципы работы алгоритмов обучения нейронных сетей: прямое распространение, обратное распространение ошибки, оптимизаторы (SGD, Adam и др.), функции потерь;
- особенности биологических данных: последовательности ДНК/РНК/белков, структура белков, экспрессия генов, метагеномные данные;
- основные задачи биоинформатики, решаемые с помощью глубокого обучения:

- предсказание структуры белков, классификация последовательностей, аннотация геномов, прогнозирование взаимодействий белков, анализ single-cell данных;
- методы предобработки биологических данных: кодирование последовательностей (one-hot, эмбединги), нормализация, балансировка выборок;
 - принципы валидации моделей в биоинформатике: кросс-валидация, метрики качества (точность, полнота, F1, AUC и др.), работа с несбалансированными данными;
 - современные биоинформатические инструменты и фреймворки: TensorFlow, PyTorch, Biopython, Hugging Face (для биологических трансформеров, например, DNABERT, ProtBERT);
 - этические и вычислительные аспекты использования ИИ в биомедицине: интерпретируемость моделей, воспроизводимость, защита персональных данных.

уметь:

- формулировать биологические задачи как задачи машинного обучения (классификация, регрессия, прогнозирование);
- подготавливать и обрабатывать биологические данные для обучения нейронных сетей;
- реализовывать и обучать нейронные сети для анализа биологических последовательностей и структур;
- использовать предобученные модели (transfer learning) в биоинформатике (например, ProtT5, ESM);
- оценивать качество моделей с учётом специфики биологических данных;
- интерпретировать результаты моделирования в контексте биологических гипотез;
- работать с биологическими базами данных (UniProt, PDB, NCBI, GTEx и др.) для получения обучающих данных.

владеть:

- навыками программирования на Python в контексте глубокого обучения (библиотеки: NumPy, Pandas, Scikit-learn, PyTorch/TensorFlow);
- навыками работы с GPU-ускорением и облачными платформами (Google Colab, AWS, Google Cloud) для обучения моделей;
- навыками визуализации биологических данных и результатов моделирования (Matplotlib, Seaborn, Plotly, PyMOL, Chimera);
- навыками воспроизводимого анализа: использование Jupyter-ноутбуков, Git, Docker;
- навыками критической оценки научных публикаций в области ИИ в биоинформатике;
- навыками командной работы при решении междисциплинарных задач (биология + ИИ).

2. Перечень планируемых результатов обучения

Компетенции, формируемые в результате освоения дисциплины (модуля) при проведении учебных занятий в форме контактной работы обучающихся с педагогическими работниками Университета и в форме самостоятельной работы обучающихся:

Компетенция	Содержание компетенции	Индикатор компетенции	Перечень планируемых результатов обучения по дисциплине (модулю)
УК-1.	Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий	УК-1.1.	Знает основные принципы системного подхода и методы критического анализа, а также теоретические основы стратегического планирования и принятия решений
		УК-1.2.	Умеет применять методы системного анализа для выявления ключевых проблемных ситуаций, формулировать гипотезы и разрабатывать альтернативные стратегии действий на основе полученных данных
		УК-1.3.	Имеет практический опыт в проведении анализа реальных проблемных ситуаций в рамках проектов, способен вырабатывать и обосновывать стратегии действий, учитывая различные аспекты и последствия
УК-2.	Способен управлять проектом на всех этапах его жизненного цикла	УК-2.1.	Знает основные методологии управления проектами, ключевые этапы жизненного цикла проекта и инструменты для планирования и контроля
		УК-2.2.	Умеет разрабатывать проектную документацию, устанавливать цели и задачи проекта, а также эффективно распределять ресурсы и управлять рисками на всех этапах его реализации
		УК-2.3.	Имеет практический опыт в управлении реальными проектами, включая планирование, исполнение и завершение, а также в оценке результатов и проведении анализа успешности проекта
ОПК-1.	Способен находить, формулировать и решать актуальные и значимые проблемы прикладной и	ОПК-1.1.	Знает основные методы и подходы к решению задач в области биологии, медицины и биотехнологий, включая

	компьютерной математики		алгоритмы, математическое моделирование и теорию оптимизации, а также современные инструменты и технологии, используемые в этой области
		ОПК-1.2.	Умеет анализировать и формулировать математические задачи в области биологии, медицины, биотехнологий, применять соответствующие методы и алгоритмы для их решения, а также интерпретировать и представлять результаты в понятной и доступной форме
		ОПК-1.3.	Имеет практический опыт работы над проектами или исследованиями в области прикладной и компьютерной математики, включая анализ биологических данных, моделирование процессов и обработку экспериментальных результатов.
ОПК-3.	Способен самостоятельно создавать прикладные программные средства на основе современных информационных технологий и сетевых ресурсов, в том числе отечественного производства	ОПК-3.1.	Знает основные принципы программирования, архитектуры программного обеспечения и современные языки программирования, а также особенности отечественных информационных технологий и сетевых ресурсов
		ОПК-3.2.	Умеет разрабатывать прикладные программные средства, используя современные инструменты и технологии, а также интегрировать их с сетевыми ресурсами для решения конкретных задач
		ОПК-3.3.	Имеет практический опыт разработки программных средств, используемых при построении математических моделей в области биологии, медицины, биотехнологий
ПК-2.	Способен математически корректно ставить и решать естественнонаучные и прикладные задачи в области в области	ПК-2.1.	Знает основные методы математического моделирования и теоретические основы современной биологии и биоинформатики, необходимые для корректной постановки задач

	биологии, медицины или биотехнологий	ПК-2.2.	Умеет формулировать математические модели биологических процессов, выбирать адекватные методы решения и интерпретировать результаты в профессиональном контексте
		ПК-2.3.	Имеет практический опыт постановки и решения прикладных задач с использованием методов ИИ в области биологии, медицины или биотехнологий
ПК-3.	Способен решать задачи профессиональной деятельности и применять методы искусственного интеллекта в прикладных и фундаментальных исследованиях в области биологии, медицины или биотехнологий	ПК-3.1.	Знает принципы и методы решения профессиональных задач, включая применение профильно-специализированных знаний в области ИИ для фундаментальных и прикладных исследований в области биологии, медицины или биотехнологий
		ПК-3.2.	Умеет применять математические и компьютерные методы, формулировать результаты и анализировать их последствия для научных и практических решений
		ПК-3.3.	Имеет опыт участия в проектах, где методы ИИ применялись для решения исследовательских и прикладных задач в области биологии, медицины или биотехнологий
ПК-5.	Способен передавать результаты прикладных задач в виде профессиональных рекомендаций в области искусственного интеллекта, биоинформатики и биотехнологии	ПК-5.1.	Знает методы формулирования рекомендаций и терминологию области искусственного интеллекта и биоинформатики
		ПК-5.2.	Умеет интерпретировать результаты анализа и моделирования и формулировать рекомендации для практического применения в области биологии, медицины или биотехнологий
		ПК-5.3.	Имеет опыт разработки рекомендаций по результатам проектов в области искусственного интеллекта и биотехнологии
ПК-6.	Способен разрабатывать программное обеспечение для решения прикладных задач в области искусственного	ПК-6.1.	Знает основные языки программирования, методы разработки программного обеспечения, а также принципы проектирования и архитектуры

	интеллекта, биоинформатики и биотехнологии		программных систем, применяемых в области искусственного интеллекта, биоинформатики и биотехнологии
		ПК-6.2.	Умеет анализировать прикладные задачи, разрабатывать алгоритмы и реализовывать их в виде программного обеспечения, используя современные инструменты и технологии, а также проводить тестирование и отладку созданных решений
		ПК-6.3.	Имеет практический опыт разработки программного обеспечения в рамках реальных проектов, включая участие в командах, где были успешно реализованы решения для конкретных прикладных задач в области искусственного интеллекта, биоинформатики и биотехнологии

3. Тематический план

№ п/п	Наименование раздела дисциплины (модуля)	Трудоемкость, академические часы				ТКУ (текущий контроль успеваемости)
		Очная форма				
		Аудиторная работа		Контроль	Самостояте льная работа	
		Лекции	Семинары (практичес кие занятия)			
1	Теоретические основы глубокого обучения и инструменты разработки	6	12		14	Домашнее задание Творческое задание Тест
2	Архитектуры глубокого обучения для анализа изображений	6	12		14	Домашнее задание Тест
3	Нейронные модели обработки естественного языка	6	12		14	Домашнее задание Творческое задание
4	Рекомендательные системы	3	6		12	Домашнее задание Тест
5	Обработка звука	3	6		12	Домашнее задание Тест
6	Генеративные модели и графовые нейронные сети	5	10		14	Домашнее задание Тест
7	Мультимодальные и прикладные модели глубокого обучения	6	12		14	Домашнее задание Тест
	Экзамен			6		
Итого:		30	60	6	94	
Объем дисциплины (модуля) (в ак. ч.)		190				
Объем дисциплины (модуля) (в зач. ед.)		5				

4. Содержание дисциплины (модуля)

№п/п	Наименование раздела дисциплины (модуля)	Содержание дисциплины (модуля) по темам
1	Теоретические основы глубокого обучения и инструменты разработки	Введение в глубокое обучение: архитектуры нейронных сетей, регуляризация, методы оптимизации и алгоритм обратного распространения ошибки Операция свёртки и практические основы разработки нейросетевых моделей в PyTorch
2	Архитектуры глубокого обучения для анализа изображений	Сверточные нейронные сети для компьютерного зрения Методы сегментации изображений, дистилляция моделей и практическая работа с задачами компьютерного зрения Методы детекции объектов

3	Нейронные модели обработки естественного языка	Векторные представления слов, Word2vec и рекуррентные нейронные сети Токенизация текста и архитектура Transformer: BERT, GPT, Mixture-of-Experts и параметрически эффективная адаптация моделей Дообучение и применение больших языковых моделей: Hugging Face и методы LoRA
4	Рекомендательные системы	Введение в рекомендательные системы
5	Обработка звука	Нейросетевые модели обработки звука и речи
6	Генеративные модели и графовые нейронные сети	Vision Transformer, состязательные генеративные сети и вариационные автоэнкодеры; генеративное моделирование Графовые нейронные сети, PyTorch Geometric
7	Мультимодальные и прикладные модели глубокого обучения	Мультимодальные модели: CLIP и современные подходы к анализу визуальных данных Диффузионные модели и прикладные сценарии глубокого обучения; практикум по диффузионному моделированию Интеграция прикладных моделей глубокого обучения

5. Учебно-методическое обеспечение

Университет располагает полным набором лицензионного и свободно распространяемого программного обеспечения, включая продукты отечественного производства.

Каждый студент в течение всего периода обучения получает индивидуальный неограниченный доступ к электронно-библиотечной системе и электронной информационно-образовательной среде университета. Эти системы предоставляют возможность доступа к ресурсам из любой точки, где есть подключение к сети Интернет, как на территории университета, так и за его пределами.

Студентам обеспечен удаленный доступ к современным профессиональным базам данных и информационным справочным системам.

Основная литература:

1. Шолле, Ф. Глубокое обучение на Python : практическое руководство / Ф. Шолле. - Санкт-Петербург : Питер, 2023. - 576 с. - (Библиотека программиста). - ISBN 978-5-4461-1909-7. - Текст : электронный. - URL: <https://znanium.com/catalog/product/2123389>.

2. Флах, П. Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных : учебник / П. Флах. - 2-е изд. - Москва.:ДМК Пресс, 2023. - 401 с. - ISBN 978-5-89818-300-4. - Текст : электронный. - URL: <https://znanium.ru/catalog/product/2085038>.

Дополнительная литература:

1. Платонов, А. В. Машинное обучение : учебник для вузов / А. В. Платонов. — 2-е изд. — Москва : Издательство Юрайт, 2026. — 78 с. — (Высшее образование). — ISBN 978-5-534-22200-5. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/600893>.

6. Материально-техническое обеспечение

Университет располагает материально-технической базой, соответствующей действующим противопожарным правилам и нормам и обеспечивающей проведение всех видов дисциплинарной и междисциплинарной подготовки, практической и научно-исследовательской работ обучающихся, предусмотренных учебным планом.

Помещения, которые представляют собой учебные аудитории для проведения занятий лекционного типа, занятий семинарского (практического) типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, а также помещения для самостоятельной работы и помещения для хранения и профилактического обслуживания учебного оборудования. Помещения укомплектованы специализированной мебелью и техническими средствами обучения, служащими для представления учебной информации большой аудитории.

Изучение дисциплины (модуля) обеспечивается в учебных аудиториях, оснащенных:

- столами и стульями;
- компьютерной техникой;
- механическими калькуляторами;
- специализированным оборудованием, включая демонстрационное оборудование.

Помещения для самостоятельной работы обучающихся, в том числе приспособленные для использования инвалидами и лицами с ограниченными возможностями здоровья, оснащены компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду Университета.

Обучающимся предоставляется доступ (в том числе удаленный) к ресурсам информационно-телекоммуникационной сети «Интернет», электронным ресурсам (в том числе электронным библиотечным системам, современным профессиональным базам данных и информационным справочным системам):

№	Наименование портала (издания, курса, документа)	Ссылка
1.	Научная электронная библиотека elibrary.ru библиотека	https://elibrary.ru/defaultx.asp
2.	База данных для IT-специалистов	https://habr.com
3.	База данных ScienceDirect	https://www.sciencedirect.com
4.	Официальный сайт Министерства науки и высшего образования Российской Федерации	https://minobrnauki.gov.ru/
5.	Федеральный портал «Российское образование»	https://www.edu.ru/
6.	Информационная система "Единое окно доступа к образовательным ресурсам"	http://window.edu.ru/
7.	Единая коллекция цифровых образовательных ресурсов	http://school-collection.edu.ru/
8.	Федеральный центр информационно - образовательных ресурсов	http://fcior.edu.ru/

Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), в том числе комплект лицензионного программного обеспечения, современные профессиональные базы данных и информационные справочные системы:

Наименование ПО	Производство	Лицензионное / свободно распространяемое
Операционные системы:		
Microsoft Imagine (Windows Client, Server)	зарубежное	лицензионное
Браузеры:		
Яндекс.Браузер	отечественное	свободно распространяемое
Google Chrome	зарубежное	свободно распространяемое
Офисные приложения:		
Microsoft Imagine (Visio, OneNote)	зарубежное	лицензионное
TeXstudio	зарубежное	свободно распространяемое
Adobe Acrobat Reader	зарубежное	свободно распространяемое
Программное обеспечение для планирования и учета времени:		
Toggle app	зарубежное	свободно распространяемое
Системы управления проектами:		
Microsoft Imagine (Project)	зарубежное	лицензионное
Системы управления базами данных:		

Microsoft Imagine (SQL Server)	зарубежное	лицензионное
Системы резервного копирования (backup):		
Acronis Backup Advanced for HyperV	зарубежное	лицензионное
Справочно-правовые системы:		
КонсультантПлюс: справочно-правовая система	отечественное	лицензионное
Средства антивирусной защиты:		
Kaspersky Endpoint Security для бизнеса Стандартный Russian Edition	отечественное	лицензионное
Среды разработки:		
Visual Studio Code	зарубежное	свободно распространяемое
Bash (Unix shell)	зарубежное	свободно распространяемое
Anaconda	зарубежное	свободно распространяемое
Robotic Operating System	зарубежное	свободно распространяемое
CopelliaSim	зарубежное	свободно распространяемое
Google Colaboratory	зарубежное	свободно распространяемое
Пакеты программных средств и библиотек:		
AutoPsy	зарубежное	свободно распространяемое
Interactive Disassembler (IDA)	зарубежное	свободно распространяемое
Системы управления библиографической информацией:		
Zotero	зарубежное	свободно распространяемое
Сервисы и службы:		
Bind	зарубежное	свободно распространяемое
Docker	зарубежное	свободно распространяемое

7. Методические и оценочные материалы

Методические указания для обучающихся по освоению дисциплины (модуля)

В процессе изучения дисциплины (модуля) «Глубокое обучение в биоинформатике» в рамках текущего контроля успеваемости используются такие виды учебной работы, как лекции, семинары, домашние задания, творческие задания, тесты, а также различные виды самостоятельной работы обучающихся по заданию преподавателя, направленные на развитие навыков профессиональной лексики, закрепление практических профессиональных компетенций, поощрение инициатив.

Лекция – систематическое, последовательное, монологическое изложение преподавателем учебного материала, как правило, теоретического характера.

В процессе лекций рекомендуется вести конспект лекций: кратко и схематично фиксировать основные идеи, выводы и обобщения лекции; выделять важные мысли, ключевые слова и термины. Необходимо отметить вопросы или материалы, которые вызывают затруднения, и попытаться найти ответы в рекомендованной литературе. Если разобраться в материале не удастся, следует сформулировать вопрос и задать его преподавателю на консультации или во время семинарского (практического) занятия.

Семинар – это форма учебной деятельности, проводимая в учебном заведении под руководством преподавателя, где студенты активно участвуют в обсуждениях, Электронный документ

практических заданиях и других формах взаимодействия.

Для успешной подготовки к семинару рекомендуется заранее ознакомиться с темой занятия и основными материалами, чтобы иметь возможность активно участвовать в обсуждении. Также полезно подготовить вопросы и идеи для обсуждения, что поможет глубже понять материал и продемонстрировать заинтересованность.

Домашнее задание – набор задач по темам недели.

При работе над домашними заданиями важно внимательно ознакомиться с требованиями и сроками выполнения. Рекомендуется разбивать задания на этапы, чтобы избежать перегрузки и лучше усвоить материал, использовать различные источники информации, включая учебники и онлайн-ресурсы, для более глубокого понимания темы.

Творческое задание — это форма учебной деятельности, где студенты самостоятельно разрабатывают оригинальные проекты, интегрируя математические методы и компьютерные технологии для решения нестандартных задач и демонстрации инновационного мышления.

Для успешного выполнения творческого задания рекомендуется глубоко исследовать тему, экспериментировать с различными подходами и документировать процесс, чтобы не только решить задачу, но и представить результаты в виде отчета или презентации, подчеркивающей креативность и практическую применимость.

Тест – особая форма проверки знаний. Проводится после освоения одной или нескольких тем и свидетельствует о качестве понимания основных понятий изучаемого материала. Тестовые задания составлены к ключевым понятиям, основным разделам, важным терминологическим категориям изучаемой дисциплины.

Для подготовки к тесту необходимо знать терминологический аппарат дисциплины, понимать смысл научных категорий и уметь их использовать в профессиональной лексике. Владение понятийным аппаратом, включённым в тестовые задания, позволяет преподавателю быстро проверить уровень понимания студентами важных методологических категорий.

Бонусные баллы — это оценки, которые студенты могут получить за выполнение дополнительных заданий. Формат бонусных баллов позволяет студентам улучшить общую оценку по дисциплине (модулю) и стимулирует углубленное изучение материала.

Самостоятельная работа – работа студентов, направленная на углубленное изучение отдельных тем и вопросов учебной дисциплины (модуля).

В процессе самостоятельной работы студенты взаимодействуют с рекомендованными материалами при минимальном участии преподавателя. Задачи студента включают работу с конспектами лекций (обработка текста), повторное изучение учебных материалов планов и тезисов ответов, изучение дополнительных тем, выполнение учебно-исследовательских заданий и другое.

Система оценивания результатов обучения по дисциплине (модулю)

Критерии получения уровня и оценивания сформированности компетенций по дисциплине (модулю) «Глубокое обучение в биоинформатике»

Оценивание уровня учебных достижений обучающихся по дисциплине (модулю) осуществляется в виде текущего контроля успеваемости и промежуточной аттестации.

Промежуточная аттестация по дисциплине (модулю) осуществляется в форме *экзамена*, при этом проводится оценка компетенций, сформированных по дисциплине.

Для оценивания текущего контроля успеваемости и промежуточной аттестации используется десятибалльная шкала оценивания, которая соотносится с традиционной пятибалльной шкалой следующим образом, если иное не указано в силлабусе дисциплины:

Десятибалльная оценка	Пятибалльная оценка	Общая характеристика результата обучения по дисциплине (модулю)
10	Отлично	Студент полностью владеет знаниями, изложенными в рабочей программе, и глубоко осмысляет дисциплину (модуль). Он самостоятельно и логически последовательно отвечает на все вопросы, акцентируя внимание на наиболее важном. Умеет анализировать, сравнивать, классифицировать, обобщать, конкретизировать и систематизировать изученный материал, выделяя ключевые моменты и устанавливая причинно-следственные связи. Четко формулирует ответы, уверенно интерпретирует результаты анализов и других исследований, а также решает сложные задачи. Студент хорошо знаком с методами исследования, необходимыми для практической деятельности, и умеет связывать теоретические аспекты дисциплины (модуля) с практическими задачами.
9	Отлично	
8	Отлично	
7	Хорошо	Студент обладает знаниями предмета почти в полном объеме рабочей программы и самостоятельно, логически последовательно и всесторонне отвечает на все вопросы, акцентируя внимание на наиболее значимых моментах. Он умеет анализировать, сравнивать, классифицировать, обобщать, конкретизировать и систематизировать изученный материал, выделяя его ключевые аспекты и устанавливая причинно-следственные связи. Формулирует свои ответы, уверенно интерпретирует результаты анализов и других исследований, а также решает сложные ситуационные задачи. Студент хорошо знаком с методами исследования, необходимыми для практической деятельности, и умеет связывать теоретические аспекты предмета с практическими задачами.
6	Хорошо	
5	Удовлетворительно	Студент обладает базовыми знаниями по дисциплине (модулю), но испытывает трудности при самостоятельных ответах и использует неточные формулировки. В ходе ответов он допускает ошибки, касающиеся сути вопросов. Студент способен решать только самые простые задачи и владеет лишь минимальным набором методов исследования.
4	Удовлетворительно	
3	Не сдан	Студент не овладел обязательным минимумом знаний по предмету и не может ответить на вопросы, даже если
2	Не сдан	

Десятибалльная оценка	Пятибалльная оценка	Общая характеристика результата обучения по дисциплине (модулю)
1	Не сдан	преподаватель задает дополнительные наводящие вопросы.

Дисциплина (модуль) «Глубокое обучение в биоинформатике» оценивается следующим образом:

Активность	Вес	Количество	Описание
Домашние задания	40%	10	Устные ответы на вопросы, список которых известен студенту заранее
Творческие задания	30%	2	Форма учебной деятельности, где студентам предлагается разработать оригинальный проект или решить нестандартную задачу
Тесты	10%	5	Форма проверки знаний студентов через ответы на вопросы с выбором из нескольких вариантов ответов
Экзамен	20%	1	Письменная или устная работа над заданием, направленным на проверку полученных знаний и навыков по дисциплине (модулю)

Формула расчёта итоговой оценки по дисциплине (модулю) «Глубокое обучение в биоинформатике»: $\langle 0,4 \times \text{Среднее за домашние задания} + 0,3 \times \text{Среднее за творческие задания} + 0,1 \times \text{Среднее за тесты} + 0,2 \times \text{Экзамен} \rangle$.

В рамках изучения дисциплины (модуля) возможно получение бонусных баллов.

Текущий контроль успеваемости обучающихся по дисциплине (модулю)

Примерные домашние задания

Домашнее задание 1. Классификация белковых последовательностей с помощью RNN

1. Используйте набор данных белков, разделённых по функциональным классам (например, ферменты, транспортные белки).
2. Закодируйте последовательности (one-hot или BLOSUM).
3. Постройте модель на основе LSTM для классификации.
4. Оцените качество (ассурасу, F1), используя перекрёстную проверку.
5. Проанализируйте, как длина последовательности влияет на качество модели.

Домашнее задание 2. Предсказание вторичной структуры белка с помощью CNN

1. Возьмите данные из PDB с аннотированной вторичной структурой.
2. Преобразуйте последовательности в окна с контекстом, закодируйте аминокислоты.
3. Постройте свёрточную сеть для предсказания α -спиралей, β -слоёв и нерегулярных участков.
4. Сравните с baseline-моделью (логистическая регрессия).
5. Визуализируйте активации фильтров и интерпретируйте их биологический смысл.

Домашнее задание 3. Анализ экспрессии генов с помощью автоэнкодера

1. Используйте данные single-cell RNA-seq (например, из базы 10x Genomics).
2. Нормализуйте данные и отфильтруйте шум.
3. Постройте вариационный автоэнкодер (VAE) для снижения размерности.
4. Визуализируйте клетки в латентном пространстве (t-SNE/UMAP).
5. Сопоставьте кластеры с известными типами клеток.

Примерное описание задания и критерии оценивания к творческому заданию

«Разработка и интерпретация нейросетевой модели для предсказания патогенности мутаций в генах»

Описание задания:

Студент самостоятельно формулирует биологическую задачу, связанную с анализом последовательностей ДНК или белков, и разрабатывает нейросетевую модель для её решения. Задание включает сбор данных, предобработку, выбор архитектуры, обучение, валидацию и интерпретацию результатов. Допускается использование предобученных моделей (ProtT5, ESM, DNABERT).

Примеры направлений:

- предсказание патогенности SNP;
- классификация промоторных последовательностей;
- предсказание связывания транскрипционных факторов;
- обнаружение генов в геноме с помощью CNN.

Критерии оценивания:

- Актуальность и корректность постановки биологической задачи.
- Качество и обоснованность предобработки данных, отсутствие утечек.
- Обоснованный выбор архитектуры и методов обучения (включая оптимизаторы, регуляризацию).
- Глубина анализа и интерпретация результатов в биологическом контексте.
- Воспроизводимость, оформление кода (ноутбук, README), визуализация результатов.

Примерные задания для теста

Вопрос 1.

Какая архитектура наиболее подходит для анализа последовательностей белков?

1. Свёрточная сеть (CNN)
2. Полносвязная сеть
3. Рекуррентная сеть (LSTM)
4. Vision Transformer

Правильный ответ: 3

Вопрос 2.

Электронный документ

Что делает операция свёртки в CNN при анализе изображений?

1. Уменьшает размерность за счёт PCA
2. Выделяет локальные признаки (края, текстуры)
3. Заменяет пропущенные значения
4. Нормализует данные

Правильный ответ: 2

Вопрос 3.

Какой метод позволяет дообучить большую модель с минимальным числом обновляемых параметров?

1. Полное переобучение
2. LoRA (Low-Rank Adaptation)
3. Увеличение размера выборки
4. Изменение функции потерь

Правильный ответ: 2

Вопрос 4.

Какой фреймворк используется для работы с графовыми нейронными сетями в биоинформатике?

1. Biopython
2. PyTorch Geometric
3. Matplotlib
4. Seaborn

Правильный ответ: 2

Вопрос 5.

Что такое векторное представление (эмбеддинг) аминокислоты в модели ProtBERT?

1. Случайное число
2. Физико-химическое свойство
3. Многомерный числовой вектор, отражающий контекст
4. Номер в алфавите

Правильный ответ: 3

Вопрос 6.

Какой метрикой следует оценивать модель при несбалансированных классах в задаче классификации?

1. Accuracy
2. F1-мера
3. MSE
4. R^2

Правильный ответ: 2

Задания для промежуточной аттестации по дисциплине (модулю)

№ п/п	Задание	Ответ	Компетенция
1.	Какая сеть подходит для анализа изображений белковых кристаллов?	Свёрточная (CNN)	ПК-2
2.	Как называется процесс передачи ошибки от выхода к входу в нейросети?	Обратное распространение	УК-1

		ие	
3.	Какой оптимизатор часто используется в глубоком обучении благодаря адаптивной скорости обучения?	Adam	ПК-3
4.	Какой метод позволяет снизить переобучение за счёт случайного отключения нейронов?	Dropout	ОПК-1
5.	Какой тип сети используется для анализа временных последовательностей?	RNN	ПК-2
6.	Как называется векторное представление слова или аминокислоты?	Эмбединг	УК-2
7.	Какой фреймворк используется для дообучения биологических трансформеров?	Hugging Face	ПК-5
8.	Какой метод позволяет визуализировать активации в свёрточной сети?	Grad-CAM	ПК-6
9.	Какой тип данных используется в single-cell анализе?	Экспрессия генов на уровне клеток	ОПК-3
10.	Какой алгоритм используется для предсказания структуры белка?	Трансформер (например, AlphaFold)	ПК-3
11.	Какой слой в CNN отвечает за уменьшение размерности?	Пулинг	ОПК-1
12.	Какой метод позволяет генерировать новые данные?	Вариационный автоэнкодер	ПК-5
13.	Какой графовый метод используется для предсказания взаимодействий?	GNN	ПК-6
14.	Какой показатель учитывает и точность, и полноту?	F1-мера	УК-1
15.	Какой подход позволяет использовать предобученную модель на новой задаче?	Перенос обучения	ПК-2
16.	Какой инструмент используется для визуализации 3д-структуры белка?	PyMOL	ПК-6
17.	Какой метод применяется для балансировки несбалансированных выборок?	Апсемплинг / взвешивание	ОПК-3
18.	Какой формат данных часто используется для хранения структур белков?	PDB	УК-2
19.	Какой оптимизатор использует момент?	SGD с моментом	ПК-3
20.	Какой архитектурный элемент лежит в основе BERT?	Трансформер	ПК-5