

## **ПРИМЕНЕНИЕ ДОЛГОЙ КРАТКОСРОЧНОЙ ПАМЯТИ В ЗАДАЧАХ КЛАССИФИКАЦИИ ГЕНЕТИЧЕСКИХ МУТАЦИЙ**

**Канищев И.С.**

Вятский государственный университет,  
Кафедра электронных вычислительных машин  
Россия, 610000, г.Киров, ул.Московская, д.36,  
E-mail: kanishchev.ilya@gmail.com

В течение последних нескольких лет было много сказано о том, как точная медицина и, более конкретно, как генетическое тестирование будет способствовать методам лечения заболеваний, таких как рак. Однако продвижение в этой области происходит мелкими шагами из-за огромного объема ручной работы.

После секвенирования раковая опухоль может иметь тысячи генетических мутаций. Основная задача состоит в том, чтобы отличить мутации, которые способствуют росту опухоли от нейтральных. В настоящее время эта интерпретация выполняется вручную. Это очень трудоемкая задача, когда клинический патологоанатом должен вручную анализировать и классифицировать каждую генетическую мутацию на основе данных из текстовой литературы.

Набор данных, представленный компанией Memorial Sloan Kettering Cancer Center (MSKCC), необходимо использовать для обучения моделей машинного обучения, для того чтобы помочь специалистам значительно ускорить их исследования.

Данные, основанные на нескольких тысячах аннотаций онкологов по всему миру, представлены двумя различными наборами. Один содержит в себе информацию о генетических мутациях, в то время как другой предоставляет клинические данные, которые эксперты используют для классификации.

Существует девять разных классов, на которые можно классифицировать генетическую мутацию. В обучающем наборе имеется 3321 различных идентификаторов, содержащих 264 различных экспрессий генов и 2996 различных вариаций мутаций. Тестирующая выборка содержит на 70% больший объем данных.

В основе обучающей модели лежит рекуррентная нейронная сеть LSTM. Данная структура сети хорошо подходит для изучения задач классификации, обработки и анализа текстовой информации в случаях, когда важные события разделены временными задержками с неопределенной продолжительностью и границы. Модель реализована с помощью библиотеки глубокого обучения Keras. Обучение сети проводилось с применением технологии CUDA.