Прогнозирование межвидовых переходов на основе анализа геномов РНК-вирусов, составляющих виром животных и птиц.

Цель: Создание искусственной нейросети для определения пандемического потенциала и вероятности межвидового перехода зоонозных РНК-вирусов.

Задачи:

1. Формирование коллекций геномов зоонозных РНК вирусов, имеющих опасность межвидового перехода

2. Выбор целевых участков геномов зоонозных РНК вирусов, задействованных в межвидовом переходе

3. Разработка алгоритма создания базы данных на основе целевых участков генома вирусов

4. Создание обучающей базы данных на основе известных зоонозных РНК-вирусов, совершивших межвидовой переход с последующим пандемическим распространением (*Coronaviridae, Orthomyxoviridae, Flaviviridae*).

5. Разработка искусственной нейросети для решения задач классификации РНК-вирусов по пандемическому потенциалу

6. Разработка искусственной нейросети для решения задач прогнозирования межвидового перехода зоонозных РНК-вирусов

7. Обучение искусственных нейросетей, оптимизация алгоритмов подготовки данных, анализа, контроля качества результатов

Искусственный интеллект (ИИ) и алгоритмы глубокого обучения являются потенциальными методами прогнозирования возможности распространения РНК-вирусов и обеспечения пандемической безопасности. Они стали неотъемлемой частью современной научной методологии, предлагая автоматизированные процедуры прогнозирования явления на основе прошлых наблюдений, разгадывая лежащие в основе закономерности в данных и создавая представление о проблеме на основе больших объёмов данных [1].

Так, используя комбинацию нескольких нейронных сетей, определенных с использованием ряда различных стратегий кодирования последовательностей, включая скрытое кодирование марковской модели, можно обеспечить точное предсказание аффинности связывания пептида с рецепторами главного комплекса гистосовместимости [2].

Анализ основ модельных предсказаний позволил предположить существование обобщаемых характеристик вирусных геномов, которые не зависят от таксономических связей вирусов и которые могут преадаптировать вирусы для заражения людей. Оценка зоонозного риска на основе генома обеспечивает быстрый и малозатратный подход, позволяющий осуществлять основанный на фактических данных эпиднадзор за вирусами, и повышает осуществимость последующей биологической и экологической характеристики вирусов [3].

Сохраняющаяся угроза появления и рецидива флавивирусов с пандемическим потенциалом подчеркивает необходимость детального фундаментального понимания биологии этих вирусов, иммунных реакций, которые могут сдерживать их [4].

РНК-вирусы отличаются от ДНК-вирусов тем, что у них много мутаций, и, следовательно, они обладают большей универсальностью. Эти мутация обеспечивают их непрерывное развитие, которое приводит к лучшей адаптации вируса [5].

Со временем появилось несколько вирусов с пандемическим потенциалом. Рост заболеваемости коронавирусом чрезвычайно интенсивных респираторных заболеваний в 2002 году (SARS-CoV), широкое распространение гриппа H1N1 в 2009 году, циркуляция штаммов гриппа H5N1 и H5N7, и более поздний рост заболеваемости коронавирусом ближневосточных респираторных заболеваний (MERS-CoV) сохраняют опасность пандемического распространения и сейчас. Несмотря на существенные различия в их морфологии, эти широко распространенные вирусы обладают рядом общих свойств. Это зоонозные РНК-респираторные вирусы, которые лишь иногда передаются между людьми в своей первоначальной форме, но после трансформации начинают более продуктивно распространятся от человека к человеку.

Машинное обучение - один из инструментов, используемых для анализа данных о мутациях. Методы машинного обучения предлагают помощь, предсказывая влияние несинонимичных однонуклеотидных полиморфизмов на стабильность белка, его работу и лекарственную устойчивость [6].

Изучение правил отображения мутаций, влияющих на поведение белка, и их использование для сбора неиспользуемых важных мутаций, которые будут устойчивы к определенным лекарствам, является одной из целей, для которых используются методы машинного обучения [7]. Другой целью является прогнозирование потенциальной вторичной структуры на основе основных структурных последовательностей [8–10].

Другой тенденцией является предсказание вариантов отдельных нуклеотидов последовательностей РНК. Последовательность РНК считается серией из четырех отдельных состояний и, таким образом, следует за заменами нуклеотидов в ходе эволюции последовательности. В работе исследователей из Египта рекуррентная нейронная сеть продемонстрировала замечательную реализацию во многих областях исследований. В этом исследовании удалось манипулировать нуклеотидами A, C, T и G для обработки данных ДНК. Используя однократный вектор для представления последовательностей ДНК и применяя модель повторяющихся нейронных сетей LSTM, эта работа прокладывает путь к новым горизонтам, где возможно предсказание мутаций, таких как эволюция вируса [6].

**Список литературы**

1. Mohamed T., Sayed S., Salah A., и др. Deep Neural Network for Virus Mutation Prediction: A Comprehensive Review // Integrating Meta-Heuristics and Machine Learning for Real-World Optimization Problems. – Springer, 2022. – С. 225–255.

2. Nielsen M., Lundegaard C., Worning P., et al. Reliable prediction of T-cell epitopes using neural networks with novel sequence representations // Protein Science. – 2003. – Vol. 12, № 5. – P. 1007–1017.

3. Mollentze N., Babayan S.A., Streicker D.G. Identifying and prioritizing potential human-infecting viruses from their genome sequences // PLoS Biology. – 2021. – Т. 19, № 9.– С. e3001390.

4. Pierson T.C., Diamond M.S. The continued threat of emerging flaviviruses: 6 // Nature Microbiology. – Nature Publishing Group, 2020. – Vol. 5, № 6. – P. 796–812.

5. Wilson B.A., Garud N.R., Feder A.F., и др. The population genetics of drug resistance evolution in natural populations of viral, bacterial and eukaryotic pathogens // Molecular ecology. – Wiley Online Library, 2016. – Т. 25, № 1.– С. 42–66.

6. Mohamed T., Sayed S., Salah A., et al. Long Short-Term Memory Neural Networks for RNA Viruses Mutations Prediction // Mathematical Problems in Engineering. – Hindawi, 2021. – Vol. 2021. – P. e9980347.