

**ОПТИЧЕСКОЕ РАСПОЗНАВАНИЕ ХИМИЧЕСКИХ СТРУКТУР: ОБЗОР И  
ТЕСТИРОВАНИЕ**

**Головинский Р.П. (ИТМО), Чистяков Д.А. (ИТМО), Тонкий И.С. (ИТМО)  
Научный руководитель – доктор химических наук, профессор Скорб Е.В. (ИТМО)**

**Введение.** Цель распознавания изображений молекул обусловлена необходимостью оцифровки химических данных, которая в свою очередь требует инструменты, позволяющие декодировать и понимать сложные молекулярные структуры на основе визуальных данных. Этот технологический сдвиг обусловлен важностью преобразования химической информации в цифровой формат, что позволяет проводить более сложный анализ и принимать обоснованные решения [1].

**Основная часть.** Существует несколько подходов к распознаванию изображений молекул, наиболее перспективный из которых основан на обнаружении примитивных частей молекулы и их интерпретации в виде графика [2]. Этот подход включает в себя различные механизмы, основанные на методе оптического распознавания химической структуры (OCSR) и реализованные в виде программных пакетов с открытым исходным кодом, таких как OSRA [3], Imago [4], и Molvec [5]. В связи с расширением области использования нейронных сетей также в рамках вышеуказанного подходы особое внимание уделяется механизмам, основанным на методах машинного обучения. В данной работе были рассмотрены такие программы, как ChemGrapher [6], DECIMER [7], DECIMER 1.0 [8], DECIMER.ai [9], Img2Mol [10], Micer [11], Image2SMILES [12], MolScribe [13], SwinOCSR [14], MolMiner [15], и AutoChemplete [16], которые используют различные алгоритмы машинного обучения и модели для распознавания молекул; выявлены основные преимущества и недостатки данных программ, а также описан ряд нерешенных проблем при реализации механизмом распознавания изображений молекул. Данные программы были протестированы на реальных изображениях из статей.

**Выводы.** В данной работе было выявлено необходимость развития программных пакетов, качество распознавания молекул которыми не зависит от количества атомов и разрешения изображения молекулы.

**Список использованных источников:**

1. Segler M. H. S., Preuss M., Waller M. P. Planning chemical syntheses with deep neural networks and symbolic AI //Nature. – 2018. – Т. 555. – №. 7698. – С. 604–610.
2. Aspuru-Guzik A., Lindh R., Reiher M. The matter simulation (r) evolution //ACS central science. – 2018. – Т. 4. – №. 2. – С. 144–152.
3. Filippov I. V., Nicklaus M. C. Optical structure recognition software to recover chemical information: OSRA, an open source solution//J. Chem. Inf. Model. – 2009. – Т. 49. – №. 3. – С. 740-743.

Автор \_\_\_\_\_ Головинский Р. П.

Автор \_\_\_\_\_ Чистяков Д. А.

Автор \_\_\_\_\_ Тонкий И. С.

Научный руководитель \_\_\_\_\_ Скорб Е. В.