

**КУЛИКОВ**  
**Дмитрий Сергеевич**

магистратура, Петрозаводский государственный университет (Петрозаводск, Россия),  
[dima23030202@mail.ru](mailto:dima23030202@mail.ru)

## **АВТОМАТИЗАЦИЯ ОБРАБОТКИ РЕНТГЕНОГРАММ И МОДЕЛИРОВАНИЕ СТРУКТУРЫ МЕТОДАМИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА**

**Научный руководитель:**  
Прусский Андрей Иванович

**Рецензент:**  
Рогов Александр  
Александрович

Статья поступила: 05.05.2026;  
Принята к публикации: 27.06.2026;  
Размещена в сети: 27.06.2026.

**Аннотация.** В работе рассматривается задача автоматизации рентгеноструктурного анализа целлюлозы, полученной из вторичного бумажного сырья, что важно для оценки качества переработанной макулатуры. Цель исследования — разработка Python-комплекса для предобработки дифрактограмм, расчёта степени кристалличности, идентификации примесей и моделирования структуры образца. Используются фильтр Савицкого–Голея, метод AsLS, аппроксимация пиков профилем псевдо-Войта и нейросетевой анализ. Показана высокая согласованность автоматического и ручного расчёта и существенное сокращение времени обработки.

**Ключевые слова:** полимеры, целлюлоза, искусственный интеллект, рентгеноструктурный анализ

**Для цитирования:** Куликов Д. С. Автоматизация обработки рентгенограмм и моделирование структуры методами искусственного интеллекта // StudArctic Forum. 2026. Т. 11, № 2. С. 78–84.

Переработка бумажных и картонных отходов — ключевой приоритет для устойчивого развития экономики. Ежегодно по всему миру образуются миллионы тонн такого мусора, который вполне способен вернуться в производственный цикл. При этом вторичная макулатура накапливает загрязнения от предыдущих циклов переработки и отличается сниженной кристалличностью по сравнению с первозданной целлюлозой [Прусский: 308].

Рентгеноструктурный анализ по-прежнему выступает ключевым методом исследования надмолекулярной структуры переработанных целлюлозных полимеров. В литературе [Park: 3] выполнено сопоставление различных алгоритмов расчёта степени кристалличности (СК), при этом доказано, что итоговые значения существенно варьируются в зависимости от выбора аморфной части и вида функций, аппроксимирующих кривую. При ручной обработке дифрактограмм одного и того же образца целлюлозного гидрогеля погрешность определения СК может достигать 5–7 % [Saurov: 890]. Для снижения субъективности и повышения воспроизводимости целесообразно внедрять автоматизированные процессы, сочетающие асимметричное вычитание фона (например, алгоритмы AsLS, airPLS или SNIP) с подбором кристаллических пиков профилями псевдо-Войта.

Развитие технологий искусственного интеллекта существенно расширило аналитические возможности: сегодня исследователи способны не только автоматизировать работу с экспериментальными данными, но и выполнять количественный фазовый анализ с

помощью нейросетевых алгоритмов. Свою роль здесь сыграли свёрточные нейронные сети [Salgado: 8] для определения крупных наборов рентгеновских спектров, однако предложенная ими методика не позволяла точно определять концентрацию посторонних фаз. Из литературы известно, что целлюлоза после вторичной обработки, включает в себя химические элементы процессов переработки.

В литературе моделирование наноцеллюлозы обычно сводится к анализу либо упорядоченных кристаллитов, либо чистых аморфных областей. При этом практические образцы фибрилл после механического воздействия и кислотного гидролиза характеризуются выраженной кривизной и скрученностью, а их вклад в искажение дифракционных картин до сих пор не систематизирован.

В данной работе предполагается, что внедрение автоматизированной предобработки рентгенограмм, с применением нейросетевой калибровки примесей многократно ускорит процедуру изучения надмолекулярной структуры. Цель исследования — создание автоматизированного Python-комплекса для анализа рентгеновских порошковых дифрактограмм наноцеллюлозы, полученной из вторичного сырья. Комплекс должен рассчитывать степень кристалличности, распознавать и количественно оценивать примеси с помощью нейронной сети, а также позволять моделировать структуру для анализа надмолекулярной составляющей.

Для достижения цели были поставлены следующие задачи:

- 1) Разработать алгоритм автоматической предобработки дифрактограмм, обеспечивающий подавление шума, удаление фона и выделение кристаллических рефлексов с корректным разделением вкладов кристаллической и аморфной фаз.
- 2) Рассчитать степень кристалличности по модифицированному методу Руланда и сравнить с ручной обработкой.
- 3) Спроектировать алгоритм, основанный на предсказании нейросетевой модели, обученную на синтетических дифрактограммах, для количественной оценки массовой доли примеси.
- 4) Выполнить моделирование ячейки образца и реализовать алгоритмы изгиба и кручения фибриллы с использованием библиотеки `rumatgen`.

Работа основана на сочетании экспериментальных методов рентгеноструктурного анализа, вычислительных алгоритмов обработки сигналов (библиотеки `SciPy`, `NumPy`) и методов машинного обучения с использованием библиотек `TensorFlow/Keras`. Для атомного моделирования использована библиотека `rumatgen` и собственные скрипты трансформации координат. Все расчёты выполнены в среде `Jupyter Notebook` на языке `Python 3.9`.

\* \* \* \* \*

Объекты исследования были спрессованы в таблетки диаметром 5 мм, толщиной порядка  $\sim 1$  мм. Данные образцы были получены путем обработки бумажной макулатуры соляной кислотой при  $T=40^{\circ}\text{C}$ . Материал для выявления примеси – пульперкартон, плотный упаковочный материал, полностью изготовленный из вторсырья.

Первым этапом обработки является сглаживание экспериментальных данных фильтром Савицкого–Голея [Герасимов: 22]. Этот фильтр сохраняет форму пиков и их ширину. Использовано 11 точек сглаживания и полином 2-ой степени. Фильтр реализован функцией `scipy.signal.savgol_filter`.

Коррекция фона выполнена методом асимметричных наименьших квадратов (AsLS). Метод основан на минимизации взвешенной суммы квадратов разностей между сглаженной кривой и исходными данными с асимметричным штрафом за отрицательные значения. Параметры алгоритма:  $\lambda=10^5$  (коэффициент гладкости),  $p=0,01$  (асимметрия). Количество итераций равняется 15. Полученная фоновая кривая вычиталась из сглаженной картины рассеяния.

Для поиска локальных максимумов использован алгоритм `find_peaks` из `scipy.signal` с заданной фильтрацией:

- расстояние между пиками не менее  $0,3^\circ$  ( $2\theta$ ),
- ширина на полувысоте от  $0,2^\circ$  до  $2,5^\circ$ .

Автоматически найденные пики проверялись на достоверность: пики с шириной более  $3^\circ$  считались недостоверными и удалялись.

Аппроксимация каждого пика выполнялась функцией псевдо-Войта (смесь Гаусса и Лоренца):

$$pV(x_i) = \eta \cdot L(x_i) + (1-\eta) \cdot G(x_i)$$

где  $\eta$  выражает соотношение между функциями и является уточняемым параметром. Параметр принимает значения от 0 до 1.

Аморфный максимум описывался одним широким гауссовым контуром с фиксированным центром в области  $18\text{--}22^\circ$  ( $2\theta$ ) для целлюлозы I $\beta$ . Подбор параметров осуществлялся методом нелинейных наименьших квадратов [Marquardt: 4]. На рисунке 1 показано выделение «аморфного» пика образца.

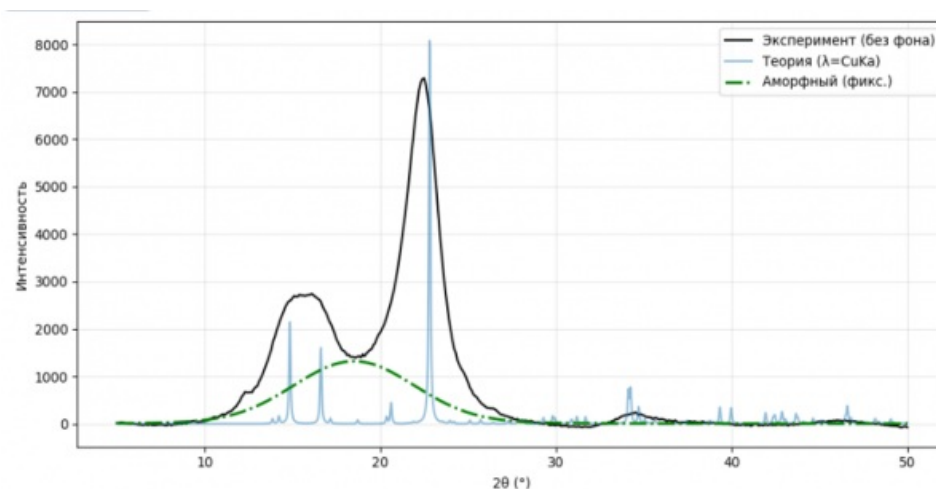


Рис. 1. Аморфный пик исследуемого образца

Для расчёта СК использовался метод Руланда [Прусский: 175]. Идея заключается в этом методе в том, что мы раскладываем сумму интегральных интенсивностей: кристаллических максимумов ( $I_{кр.}$ ), которая прямо пропорционально соотносится с кристаллической фазой и аморфного ( $I_{аморф.}$ ) неструктурного максимума. Формула для расчёта СК:

$$СК = I_{кр} / (I_{кр} + I_{аморф}) \times 100\%$$

Для исследуемого образца автоматический расчёт дал значение СК=63,5 %. Ручная аппроксимация дала результат СК=64,6 %. Расхождение составило 1,1 %, При этом время обработки сократилось с 20–30 минут до 3–5 секунд на один образец.

Качественный фазовый анализ выполнялся путём сравнения находящихся в области  $2\theta$  от  $10^\circ$  до  $60^\circ$  на медном излучении ( $\lambda=1.5406 \text{ \AA}$ ) экспериментальных пиков с эталонными значениями из локальной базы данных. Критерии поиска:

- 1) точность по углу  $2\theta$  ( $\Delta(2\theta)=0.3^\circ$ ),
- 2) точность по межплоскостному расстоянию  $d$  ( $\Delta(d)=0.15 \text{ \AA}$ ),
- 3) точность по  $I/I_{max}$  ( $\Delta(I)=25\%$ ),
- 4) совпадение 2 из 3 пиков(минимум).

Для образца пульперкартона локальная база надёжно идентифицировала кальцит. Дополнительно был проведён поиск по открытой базе Materials Project [Ong: 317] с

использованием официального интерфейса прикладного программирования (API).

Для предсказания количества фазы в образце была обучена полносвязная нейронная сеть.

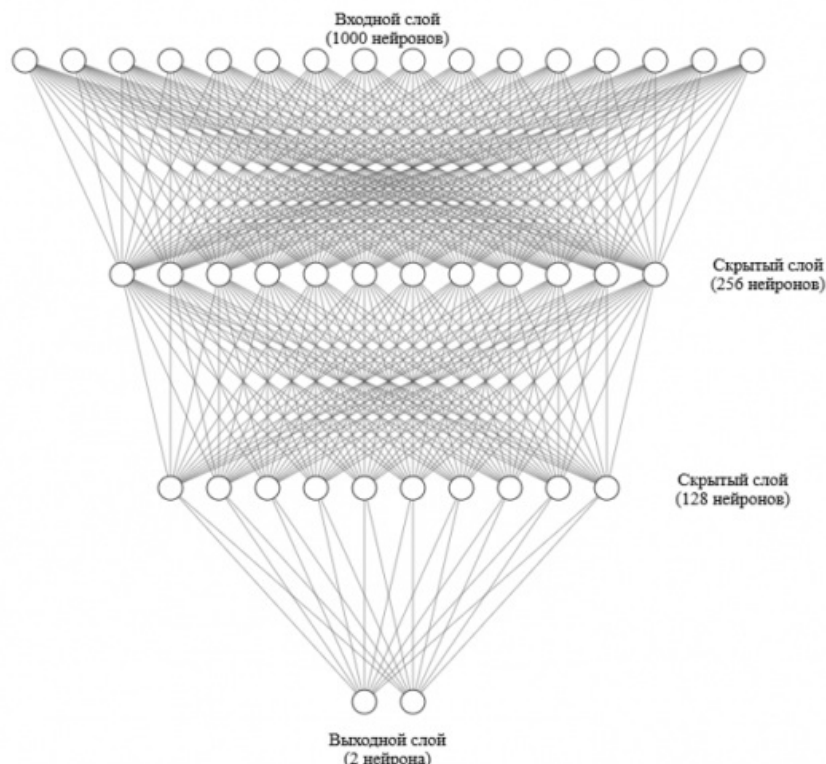


Рис. 2. Архитектура нейронной сети для предсказания массовой доли в образце

Архитектура нейронной сети:

- Входной слой: 1000 нейронов.
- Скрытый слой 1: 256 нейронов, активация ReLU, dropout 0,3.
- Скрытый слой 2: 128 нейронов, активация ReLU, dropout 0,3.
- Выходной слой: 2 нейрона с активацией softmax.
- Функция потерь – кросс-энтропия, оптимизатор – Adam
- Метрика – средняя абсолютная ошибка (MAE) между предсказанной и истинной массовой.

Так как как качественная съёмка одного образца может занимать время от трёх часов до трёх суток, было принято решение использовать аугментацию данных. Была сгенерирована синтетическая выборка из 15000 рентгенограмм. Создание максимально достоверной выборки проводилось с использованием таких методов:

- 1) сдвиг пиков,
- 2) гауссов шум,
- 3) аморфный фон,
- 4) уширение пиков.

Выборка разделена на обучающую (80 %) и тренировочную (20 %). Обучение проводилось в течение 25 эпох. Для экспериментальной дифрактограммы образца пульперкартона сеть предсказала: целлюлоза – 73,1 %, CaCO<sub>3</sub> – 21,9 % с MAE=2,3 % Полученное значение хорошо согласуется с данными технических условий ГОСТ 1760-2014<sup>1</sup>.

Для атомного моделирования за основу взята кристаллическая структура целлюлозы Iβ. С помощью библиотеки rumatgen [Ong] произведена трансляция ячейки по векторам **a**, **b**, **c** с коэффициентами размножения ячейки 5, 4, 11. Такая атомный кластер включает в себя

около 18000 атомов.

Изгиб фибриллы реализован через разбиение кластера на слои вдоль оси Y. Смещение по x:

$$\Delta x = R \cdot (1 - \cos(\theta)),$$

где:

$$\theta = y / R,$$

L – длина суперячейки по Y,

Кручение фибриллы моделировалось путём поворота каждого атома вокруг оси Y на угол

$$\varphi = (y / L) \cdot 120^\circ$$

Поворот выполнялся в плоскости XZ. Полученная структура (рисунок 3) демонстрирует спиралевидную форму, характерную для целлюлосодержащих продуктов.

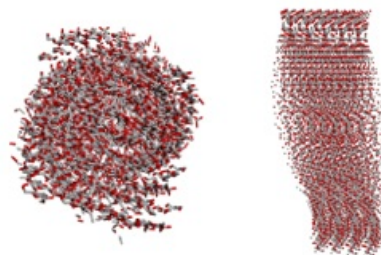


Рис. 3. Кластер целлюлозы (5a×4b×11c) после применения изгиба (R=500 Å) и кручения (120°)

В результате выполнения работы был разработан полностью автоматизированный программный комплекс на Python для обработки рентгенограмм наноцеллюлозы, полученной из макулатурной массы (газетная бумага, картон). Основные научные и практические результаты:

1. Реализованы и оптимизированы процедуры сглаживания фильтром Савицкого – Голея и очистки фона асимметричным методом наименьших квадратов (AsLS). Разработан алгоритм автоматического поиска и аппроксимации кристаллических пиков профилем псевдо-Войта с выделением аморфного гауссова контура.

2. По модифицированному методу Руланда автоматически рассчитана степень кристалличности для образца целлюлозы из пульперкартона – 63,5 %. Расхождение с ручной обработкой составило менее 1,1 %, что подтверждает высокую точность автоматизации. Время обработки сокращено с 30 минут до 2–3 минут.

3. Проведён качественный фазовый анализ с использованием локальной базы данных и открытой базы Materials Project. Для образцов из вторичного сырья выполнено количественное определение массовой доли CaCO<sub>3</sub> с помощью полносвязной нейронной сети, обученной на синтетической выборке. Предсказанная доля CaCO<sub>3</sub> составила 21,9 %, что соответствует требованиям ГОСТ 1760-2014 для макулатуры. Средняя абсолютная ошибка модели на тестовой выборке – 2,5 %.

4. Атомное моделирование. Создан метод моделирования кластеров ячеек, добавлены алгоритмы изгиба и кручения, позволяющие создавать реалистичные модели деформированных фибрилл.

Разработанные алгоритмы обеспечивают высокую воспроизводимость и скорость анализа рентгенограмм целлюлозосодержащих материалов. Значения, предсказанные с помощью моделей ИИ, совпадают с полученными традиционной обработкой, что говорит о корректности метода. Автоматизация расчёта степени кристалличности и фазового состава ускоряет обработку данных и закладывает основу для моделирования дифракции и прогноза структурно-зависимых свойств целлюлозных гидрогелей и композитов.

## Примечания

<sup>1</sup> ГОСТ 10700-97. Макулатура бумажная и картонная. Технические условия: межгосуд. стандарт Российской Федерации: дата введения 2003-01-01. Москва: Стандартинформ, 2002. С. 1.

## Список литературы

*Герасимов Р.Д.* Модуль обработки данных дифрактограмм для электронного журнала. Нахождение дифракционных максимумов / Р.Д. Герасимов, А.Д. Стариков // Известия Томского политехнического университета. Промышленная кибернетика. 2024. Т. 2, № 3. С. 20-27. DOI: 10.18799/29495407/2024/3/63

*Прусский А.И.* Использование модифицированного метода Руланда для расчёта степени кристалличности образцов целлюлозы / А.И. Прусский, Н.Е. Котельникова // BIOAsia-Altai: материалы IV Международного биотехнологического форума. Барнаул, 2024. Т. 4, № 1. С. 173-177.

*Marquardt D.W.* An algorithm for least-squares estimation of nonlinear parameters // Journal of the Society for Industrial and Applied Mathematics. 1963. Vol. 11, No. 2. P. 431-441. DOI: 10.1137/0111030

*Ong S.P.* Python materials genomics (pymatgen): A robust, open-source Python library for materials analysis / S.P. Ong, W.D. Richards, A. Jain, G. Hautier, M. Kocher, S. Cholia, D. Gunter, V.L. Chevrier, K.A. Persson, G. Ceder // Computational Materials Science. 2013. Vol. 68. P. 314-319. DOI: 10.1016/j.commatsci.2012.10.028

*Park S.* Cellulose crystallinity index: measurement techniques and their impact on interpreting cellulase performance / S. Park, J.O. Baker, M.E. Himmel, P.A. Parilla, D.K. Johnson // Biotechnology for Biofuels. 2010. Vol. 3. Article 10. DOI: 10.1186/1754-6834-3-10

*Salgado J.E.* Automated classification of big X-ray diffraction data using deep learning models / J.E. Salgado, S. Lerman, Zh. Du, Ch. Xu, N. Abdolrahim // npj Computational Materials. 2023. Vol. 9. Article 214. DOI: 10.1038/s41524-023-01164-8

*Saurov S.K.* Comparative study of powder celluloses and cellulose hydrogels by WAXS method. Impact of measurement technique and computation on variability of results / S.K. Saurov, K. Svedström, A. Mikhailidi, N. Kotelnikova // Cellulose Chemistry and Technology. 2019. Vol. 53, No. 9-10. P. 885-896. DOI: 10.35812/CelluloseChemTechnol.2019.53.86

Dmitry S. KULIKOV

master's degree, Petrozavodsk State University  
(Petrozavodsk, Russia),  
[dima23030202@mail.ru](mailto:dima23030202@mail.ru)

### AUTOMATED X-RAY IMAGE PROCESSING AND STRUCTURAL MODELLING USING AI METHODS

**Scientific adviser:**

Andrey I. Prusskiy

**Reviewer:**

Aleksandr Rogov

Paper submitted on: 05/05/2026;

Accepted on: 06/27/2026;

Published online on: 06/27/2026.

**Abstract.** This paper addresses automating X-ray diffraction analysis of cellulose from recycled paper, an important step in assessing recycled paper quality. The aim was to develop a Python-based toolkit for pre-processing diffractograms, calculating the degree of crystallinity, identifying impurities, and modelling sample structure. The authors employed the Savitzky–Golay filter, the AsLS method, peak approximation using a pseudo-Voigt profile, and neural network analysis. The results show a high degree of agreement between automatic and manual calculations, along with a significant reduction in processing time.

**Keywords:** polymers, cellulose, artificial intelligence, X-ray diffraction analysis

**For citation:** Kulikov, D. S. Automated X-ray Image Processing and Structural Modelling Using AI Methods. *StudArctic Forum*. 2026, 11 (2): 78–84.

#### References

Gerasimov R.D., Starikov A.D. Diffractogram data processing module for electronic journal. Finding diffraction maximum. *Bulletin of the Tomsk Polytechnic University. Industrial Cybernetics*, 2024, Vol. 2, No. 3, pp. 20-27. DOI: 10.18799/29495407/2024/3/63 (In Russ.)

Prusskiy A.I., Kotelnikova N.E. The use of a modified Ruland method for calculating the degree of crystallinity of cellulose samples. *BIOAsia-Altai: proceedings of the IV International Biotechnological Forum*, Barnaul, 2024, Vol. 4, No. 1, pp. 173-177. (In Russ.)

Marquardt D.W. An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics*, 1963, Vol. 11, No. 2, pp. 431-441. DOI: 10.1137/0111030

Ong S.P., Richards W.D., et al. Python materials genomics (pymatgen): A robust, open-source Python library for materials analysis. *Computational Materials Science*, 2013, Vol. 68, pp. 314-319. DOI: 10.1016/j.commatsci.2012.10.028

Park S., Baker J.O., et al. Cellulose crystallinity index: measurement techniques and their impact on interpreting cellulase performance. *Biotechnology for Biofuels*, 2010, Vol. 3, Article 10. DOI: 10.1186/1754-6834-3-10

Salgado J.E., Lerman S., et al. Automated classification of big X-ray diffraction data using deep learning models. *npj Computational Materials*, 2023, Vol. 9, Article 214. DOI: 10.1038/s41524-023-01164-8

Saurov S.K., Svedström K., et al. Comparative study of powder celluloses and cellulose hydrogels by WAXS method. Impact of measurement technique and computation on variability of results. *Cellulose Chemistry and Technology*, 2019, Vol. 53, No. 9–10, pp. 885–896. DOI: 10.35812/CelluloseChemTechnol.2019.53.86