

Прогнозирование выхода конечных продуктов в процессе пиролиза растительной биомассы методами машинного обучения

Роман Михайлович Бачурин[✉], Ирина Гелиевна Захарова

Тюменский государственный университет, Тюмень, Россия

Контакт для переписки: stud0000107305@study.utm.ru[✉]

Аннотация. Статья посвящена разработке методов математического моделирования выхода конечных продуктов при пиролизе растительной биомассы с использованием моделей машинного обучения.

Основная цель исследования заключается в разработке и анализе прогнозных моделей. Для обучения моделей используются данные, полученные в результате натурных экспериментов. В качестве основных признаков выступают химические и физические характеристики состава биомассы, параметры температурного режима процесса пиролиза. В качестве прогнозируемых конечных выходов продуктов пиролиза выступают жидкие продукты пиролиза, включая смолу; твердые продукты пиролиза и пиролизный газ в процентном соотношении.

На основе этих данных разработаны различные модели машинного обучения, которые проанализированы с точки зрения возможности интеграции в интеллектуальную систему поддержки научных исследований. В результате из них выбрана модель, позволяющая с достаточно высокой точностью предсказывать соотношения различных фракций выходного продукта в ходе пиролиза, что открывает новые возможности для оптимизации пиролизных установок и повышения их энергетической эффективности.

Ключевые слова: математическое моделирование, прогнозирование, пиролиз биомассы, машинное обучение, регрессионная модель, случайный лес

Цитирование: Бачурин Р. М., Захарова И. Г. 2025. Прогнозирование выхода конечных продуктов в процессе пиролиза растительной биомассы методами машинного обучения // Вестник Тюменского государственного университета. Физико-математическое моделирование. Нефть, газ, энергетика. Том 11. № 3 (43). С. 147–164. <https://doi.org/10.21684/2411-7978-2025-11-3-147-164>

Поступила 07.06.2025; одобрена 17.07.2025; принята 18.07.2025

Prediction of final product yield in the process of plant biomass pyrolysis using machine learning methods

Roman M. Bachurin[✉], Irina G. Zakharova

University of Tyumen, Tyumen, Russia

Corresponding author: stud0000107305@study.utmn.ru[✉]

Abstract. The article is devoted to the development of methods for mathematical modeling of final product yields in plant biomass pyrolysis using machine learning models. The main objective of the study is to develop and analyze predictive models. For training the models the data obtained from full-scale experiments are used. Chemical and physical characteristics of biomass composition, parameters of the pyrolysis process temperature regime act as the main attributes. The predicted final yields of pyrolysis products are liquid pyrolysis products, including tar; solid pyrolysis products and pyrolysis gas in percentage ratio.

Based on these data, various machine learning models are developed and analyzed for the possibility of integration into an intelligent research support system. As a result, a model has been selected that allows predicting the ratios of different fractions of the output product during pyrolysis with sufficiently high accuracy, which opens up new opportunities for optimizing pyrolysis plants and improving their energy efficiency.

Keywords: mathematical modeling, forecasting, biomass pyrolysis, machine learning, regression model, random forest

Citation: Bachurin, R. M., & Zakharova, I. G. (2025). Prediction of final product yield in the process of plant biomass pyrolysis using machine learning methods. *Tyumen State University Herald. Physical and Mathematical Modeling. Oil, Gas, Energy*, 11(3), 147–164. <https://doi.org/10.21684/2411-7978-2025-11-3-147-164>

Received Jun. 7, 2025; Reviewed Jul. 17, 2025; Accepted Jul. 18, 2025

Введение

Процесс пиролиза представляет собой множество физических и химических реакций, которые могут быть представлены десятками уравнений [Fatehi, Bai, 2014]. Это создает трудности в разработке адекватных математических моделей, поскольку необходимо учитывать химические взаимодействия между многочисленными компонентами топли-

ва, фазовые переходы в течение реакции, параметры реактора пиролиза [Ranzi et al., 2017]. Кинетические уравнения для различных типов биомассы также могут различаться достаточно существенно [Ranganathan, Gu, 2016]. Пиролиз часто протекает в условиях, где температура и состав сырья изменяются во времени, что требует создания нестационарных моделей.

Пиролиз представляет собой процесс термохимической переработки органических и неорганических материалов путем их разложения при недостатке кислорода.

Фазовые превращения веществ в ходе пиролиза (испарение жидкости и выделение газов) оказывают значительное влияние на тепловое состояние сырья, так как в реакторе изменяется масса вещества и вещества, покидающие реактор, уносят с собой определенное количество тепловой энергии. В ходе покидания реактора вещества участвуют в конвективном теплообмене и тоже могут переносить тепло из одних частей реактора в другие. Температурный интервал процессов выделения разных веществ в ходе пиролиза и зависимость изменения массы от температуры можно получить из результатов термогравиметрического анализа (ТГА), дифференциально-термического анализа (ДТА) и с использованием газоанализатора.

Для различных видов органического сырья для пиролиза кинетические модели разложения могут сильно различаться. Кинетические параметры уравнений найти в научных публикациях крайне сложно. Как правило, данные параметры получают из результатов ТГА [Ranzi et al., 2017].

Существуют относительно универсальные кинетические механизмы пиролиза биомассы: FLASHCHAIN [Niksa, 2020], FG-DVC [Klason, Bai, 2003], CRECK, CPD [Wooters et al., 2005; Pielsticker et al., 2024] и т. д. Так как данные механизмы содержат моделирование тепловых эффектов, химических реакций и выделение широкой номенклатуры продуктов пиролиза, они являются достаточно сложными и многоступенчатыми с точки зрения вычислительной сложности. Изначально данные механизмы были разработаны для пиролиза угля [Hasse et al., 2021], но позже были адаптированы в том числе и для растительной биомассы. При использовании данных моделей любая растительная биомасса представляется как смесь целлюлозы, гемицеллюлозы и лигнина [Parmar, 2017]. Для данных веществ составлены сложные кинетические механизмы реакций пиролиза [Sharma et al., 2015].

Основные экзотермические реакции различных видов биомассы происходят при разных температурных условиях. Температуру, необходимую для начала процесса тепловыделения, можно выяснить только экспериментальным путем с помощью ТГА и ДТА. Кроме того, из данных экспериментов можно получить температурный интервал протекания экзотермических реакций и функцию тепловыделения от температуры [Tabakaev et al., 2019]. Данная функция для разных видов растительных биомасс может существенно различаться, поэтому нужно проводить ТГА и ДТА для любого нового рассматриваемого сырья.

Для традиционного математического моделирования процесса пиролиза используются сложные вычислительные модели, которые основываются на законах сохранения массы, импульса, энергии и включают в себя начальные и граничные условия [Popescu

et al., 2020; Kaczor et al., 2020; Wijayanti et al., 2021]. В таких моделях также рассматриваются химические превращения веществ [Sharma et al., 2015; Ranganathan, Gu, 2016], обсчитываются траектории перемещения продуктов пиролиза внутри реактора методами вычислительной гидродинамики [Hasse et al., 2021], что требует дополнительных вычислительных мощностей устройства, на котором производится процесс моделирования. Таким образом, моделирование процесса пиролиза путем численного решения систем дифференциальных уравнений [Fernando et al., 2015], описывающих все вышеперечисленное, требует значительных вычислительных ресурсов.

По сравнению с классическими численными алгоритмами, методы машинного обучения применяются во многих областях и обладают рядом преимуществ, таких как более высокая скорость вычислений и возможность постоянного улучшения путем их дополнительного обучения на новых данных.

Благодаря развитию библиотек машинного обучения и удешевлению графических ускорителей, методы и технологии машинного обучения широко применяются для моделирования процесса пиролиза [Mighani et al., 2017; Ting Yan Li et al., 2021; Shen et al., 2022; Lee et al., 2024]. Исследования включают как использование классических моделей и алгоритмов [İşçen et al., 2023; Shen et al., 2022; Vikram et al., 2021; Ле и др., 2020], так и построение регрессионных моделей, основанных на искусственных нейронных сетях различных видов [Rongrong et al., 2023; Cheng et al., 2023; Balsora et al., 2023].

При этом основной сложностью является получение достаточного объема данных, требуемых для обучения модели. Открытые источники, включающие статьи в различных научных изданиях, содержат достаточно ограниченное количество данных, полученных в результате натурных экспериментов.

Целью работы является исследование прогнозирования выхода конечных продуктов пиролиза растительной биомассы в процентном соотношении методами машинного обучения.

Материалы и методы

Для обучения моделей были использованы данные результатов натурных экспериментов пиролиза растительной биомассы, полученные из открытых источников¹, которые представляют собой набор из 751 записи. Набор данных сформирован из 72 научных статей, описывающих эксперименты, методы и подходы к математическому моделированию выхода продуктов пиролиза путем их предсказания с помощью различных методов, основанных на построении регрессионных моделей. Описание данных приведено в табл. 1.

¹ Dong Z., Bai X., Xu D., Li W. 2022. Machine learning prediction of pyrolytic products of lignocellulosic biomass based on physicochemical characteristics and pyrolysis conditions // Github. URL: <https://github.com/dazhaxie666/Biomass-pyrolysis-data> (дата обращения: 03.03.2025).

Таблица 1. Исходные данные из эксперимента пиролиза биомассы**Table 1.** Data from a biomass pyrolysis experiment

| Показатель | Описание | Пример |
|-----------------|---|---------------------------|
| Biomass species | Тип биомассы | Jerusalem artichoke stick |
| M | Влажность (%) | 15.76 |
| Ash | Зольность (%) | 3.34 |
| VM | Выход летучих веществ (%) | 67.40 |
| FC | Фиксированный углерод (%) | 13.50 |
| C | Доля углерода в биомассе (%) | 45.36 |
| H | Доля водорода в сырье | 6.11 |
| O | Доля кислорода в сырье | 47.26 |
| N | Доля азота в сырье | 0.75 |
| PS | Размер частиц сырья биомассы (мм) | 0.5 |
| FT | Конечная температура процесса пиролиза | 550 |
| HR | Скорость нагревания биомассы (°С/мин) | 10.0 |
| FR | Скорость пропускания инертного газа (мл/мин) | 100.0 |
| Solid Phase | Выход твердого вещества (%) | 35 |
| Liquid Phase | Выход жидких продуктов, в том числе смолы (%) | 44.9 |
| Gas Phase | Выход газа (%) | 20.1 |

Значения части признаков пропущены в исходном наборе данных. Часть таких значений было заполнено средним значением для биомассы того же типа. Полностью отсутствовавшие значения были вычислены путем вычитания из 100% значений признаков той же группы. Например, отсутствующий признак «влажность (%)» был вычислен по формуле:

$$M = 100 - Ash - VM - FC, \quad (1)$$

где M — влажность (%); Ash — зольность (%); VM — выход летучих веществ (%); FC — фиксированный углерод (%).

Требуется обучить набор моделей машинного обучения для прогнозирования признаков: выход твердого вещества в % (Solid Phase), выход жидкого вещества в % (Liquid Phase), выход газа в % (Gas Phase) с точностью, достаточной для интеграции в интеллектуальную систему.

Solid Phase включает в себя различные твердые вещества, которые образуются в процессе пиролиза растительной биомассы.

Liquid Phase — различные жидкие вещества, включая смолу и воду, которые образуются в процессе пиролиза.

Gas Phase — газ, образованный в процессе пиролиза, который включает в себя CH_4 , CO , CO_2 , SO_2 , H_2 и т. д.

Данные были разделены на обучающую выборку и тестовую в соотношении 80 к 20 соответственно.

Для решения задачи предсказания выхода долей продуктов пиролиза в качестве классических моделей машинного обучения были выбраны регрессионные модели из библиотеки scikit-learn. Кроме того, были использованы модели LightGBM, XGBoost, CatBoost. Модели и их краткое описание приведены в табл. 2.

Таблица 2. Регрессионные модели для предсказания продуктов пиролиза

Table 2. Regression models for predicting pyrolysis products

| Модель | Описание модели |
|---------------------------|---|
| Linear Regression | Модель линейной регрессии, представляющая собой метод наименьших квадратов |
| Huber Regressor | Модифицированная версия метода наименьших квадратов. Является линейным методом, который менее чувствителен к выбросам, чем обычный метод наименьших квадратов |
| Lasso Regression | Вариация линейной регрессии, используемая для решения проблемы мультиколлинеарности и отбора наиболее информативных признаков |
| Ridge Regression | Вариация линейной регрессии, используемая для решения проблемы мультиколлинеарности, когда независимые переменные сильно коррелируют |
| SGD Regressor | Линейная регрессия с использованием стохастического градиентного спуска. Модель минимизирует регуляризованные эмпирические потери с помощью стохастического градиентного спуска. Градиент потерь вычисляется для каждой выборки, и модель обновляется итеративно с использованием убывающей скорости обучения |
| Elastic-Net Regression | Модель регуляризованной линейной регрессии, которая объединяет штрафы L1 (Lasso) и L2 (Ridge). Она полезна при работе с наборами данных, в которых наблюдается мультиколлинеарность (сильная корреляция между признаками) |
| Bayesian Ridge Regression | Модель линейной регрессии, которая использует байесовский подход для оценки своих параметров. В отличие от классической ридж-регрессии, которая добавляет штраф к функции стоимости для регуляризации, байесовская линейная регрессия рассматривает параметры как случайные величины и встраивает их неопределенность в сам процесс моделирования |
| Bayesian ARD Regression | Модель с автоматическим определением релевантности (Automatic Relevance Determination, ARD) — разновидность байесовской линейной модели, которая, в отличие от байесовской ридж-регрессии, приводит к более разреженным коэффициентам |
| Decision Tree Regression | Модель машинного обучения с учителем, которая использует структуру дерева для прогнозирования числовых значений |
| Random Forest Regression | Ансамблевая модель машинного обучения с учителем, которая использует ансамбль деревьев решений для прогнозирования числовых значений |

Продолжение табл. 2

Table 2 (continuation)

| Модель | Описание модели |
|----------------------------------|--|
| AdaBoost Regressor | Ансамблевая модель машинного обучения, основанная на принципе бустинга. Основной принцип AdaBoost заключается в обучении последовательности слабых моделей на неоднократно изменяемых версиях данных. Прогнозы всех из них затем объединяются посредством взвешенного большинства голосов для получения окончательного прогноза |
| Bagging Regressor | Ансамблевая модель машинного обучения, которая создает несколько экземпляров моделей черного ящика на случайных подмножествах исходного обучающего набора, а затем объединяет отдельные прогнозы для формирования окончательного прогноза. Эти методы используются как способ уменьшить дисперсию базовой оценки дерева решений |
| ExtraTrees Regressor | Ансамблевая модель машинного обучения, строящая множество деревьев решений, каждое из которых обучается на случайном подмножестве признаков и с использованием случайных порогов. Данная модель выбирает признаки и пороги случайным образом |
| Gradient Boosting Regressor | Ансамблевая модель машинного обучения, которая объединяет несколько слабых деревьев решений для создания более точной модели. Она добавляет базовые модели в ансамбль последовательно. Случайные деревья обучаются на остаточных ошибках, допущенных предыдущими моделями |
| Hist Gradient Boosting Regressor | Ансамблевая модель машинного обучения, которая является улучшенной версией Gradient Boosting Regressor. В ней используются методы на основе гистограмм, которые уменьшают использование памяти и ускоряют процесс обучения. Также данная модель может обрабатывать пропущенные значения |
| Stacking Regressor | Ансамблевая модель машинного обучения, которая объединяет прогнозы нескольких базовых моделей для повышения точности и стабильности итоговой модели. В отличие от простого усреднения или голосования, стекинг обучает мета-алгоритм, чтобы узнать, как лучше всего комбинировать прогнозы базовых моделей |
| KNeighbors Regressor | Модель машинного обучения, которая при прогнозировании для новой точки данных, находит k ближайших к ней точек из обучающего набора. Близость определяется с использованием метрики расстояния (например, евклидово расстояние). Т. е. значение целевой переменной для новой точки данных прогнозируется как среднее значение целевых переменных k ближайших соседей |
| SVM Regressor | Модель машинного обучения, которая использует метод опорных векторов. SVM Regressor пытается минимизировать ошибку, только если она превышает определенный порог, что делает модель менее чувствительной к выбросам |

Окончание табл. 2

Table 2 (end)

| Модель | Описание модели |
|--------------------|--|
| LightGBM Regressor | Высокопроизводительная реализация градиентного бустинга на основе деревьев решений. Модель оптимизирована для работы с большими и разреженными данными |
| XGBoost Regressor | Реализация градиентного бустинга деревьев решений, которая последовательно строит ансамбль моделей |
| CatBoost | Модель градиентного бустинга на симметричных деревьях решений |

В качестве функции потерь для измерения разницы между предсказанными значениями модели и фактическими данными, т. е. для оптимизации параметров модели в процессе обучения, была выбрана среднеквадратичная ошибка (2), так как было выяснено, что данные из натуральных экспериментов не содержат выбросов.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2, \quad (2)$$

где y_i — фактическое значение; \hat{y}_i — предсказанное значение; n — количество наблюдений.

Для оценки качества регрессионных моделей машинного обучения используются метрики, такие как средняя абсолютная ошибка (MAE) (3) и корень средней квадратичной ошибки ($RMSE$) (4) [Willmott, Matsuura, 2006].

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|, \quad (3)$$

где y_i — фактическое значение; \hat{y}_i — предсказанное значение; n — количество наблюдений.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}, \quad (4)$$

где y_i — фактическое значение; \hat{y}_i — предсказанное значение; n — количество наблюдений.

Помимо этого, для оценки моделей регрессии используется метрика коэффициент детерминации R^2 (5).

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} = 1 - \frac{SSE}{SST}, \quad (5)$$

где SSE (sum square of errors) = $\sum_{i=1}^n (y_i - \hat{y}_i)^2$ — сумма квадратов остатков (ошибок) регрессии; SST (sum square total) = $\sum_{i=1}^n (y_i - \bar{y}_i)^2$ — полная сумма квадратов; $R^2 \in [0; 1]$.

Результаты и обсуждение

Модели машинного обучения

Для предсказания выхода конечных продуктов были использованы классические регрессионные модели машинного обучения из библиотеки scikit-learn. Каждая из моделей обучена для предсказания одного признака: Solid Phase, Liquid Phase или Gas Phase.

Результаты обученных моделей и их метрики качества приведены в табл. 3.

Таблица 3. Сравнение классических регрессионных моделей машинного обучения

Table 3. Comparison of classic machine learning regression models

| Модель | Твердые в-ва | | Жидкие в-ва | | Газ | |
|-------------------------------------|---------------------|-----------|--------------------|-----------|-------------|-----------|
| | RMSE | R2 | RMSE | R2 | RMSE | R2 |
| Huber Regressor | 5.1664 | 0.5513 | 9.3632 | 0.3442 | 8.4869 | 0.0337 |
| Linear Regression | 5.1662 | 0.5514 | 9.5424 | 0.3188 | 8.4453 | 0.0431 |
| Lasso Regression | 5.7050 | 0.4529 | 9.8590 | 0.2729 | 8.1301 | 0.1132 |
| Ridge Regression | 5.1652 | 0.5515 | 9.5402 | 0.3191 | 8.4289 | 0.0468 |
| SGDRegressor Regression | 5.1556 | 0.5532 | 9.5234 | 0.3215 | 8.3614 | 0.0620 |
| ElasticNet Regression | 5.6692 | 0.4598 | 9.8192 | 0.2787 | 8.0652 | 0.1273 |
| Bayesian Ridge Regression | 5.1629 | 0.5519 | 9.5048 | 0.3242 | 8.1576 | 0.1072 |
| Bayesian ARD Regression | 5.2380 | 0.5388 | 9.5386 | 0.3194 | 8.1501 | 0.1088 |
| Decision Tree Regression | 4.1087 | 0.7162 | 3.2699 | 0.9200 | 4.1520 | 0.7687 |
| Random Forest Regression | 3.3406 | 0.8124 | 2.8885 | 0.9376 | 3.1195 | 0.8694 |
| AdaBoost Regressor | 5.5958 | 0.4737 | 7.4884 | 0.5805 | 6.2222 | 0.4806 |
| Bagging Regressor | 3.5456 | 0.7887 | 3.1001 | 0.9281 | 3.0602 | 0.8744 |
| ExtraTrees Regressor | 3.3850 | 0.8074 | 3.0055 | 0.9324 | 3.0192 | 0.8777 |
| Gradient Boosting Regressor | 3.4373 | 0.8014 | 4.1398 | 0.8718 | 3.8490 | 0.8012 |
| Hist Gradient Boosting Regressor | 3.4741 | 0.7971 | 3.4189 | 0.9126 | 3.0953 | 0.8715 |
| Stacking Regressor | 5.1646 | 0.5517 | 9.4404 | 0.3333 | 8.2151 | 0.0946 |

Окончание табл. 3

Table 3 (end)

| | Твердые в-ва | Жидкие в-ва | Газ | | |
|----------------------|--------------|-------------|--------|--------|---------------|
| KNeighbors Regressor | 4.4939 | 0.6605 | 6.2923 | 0.7038 | 5.5518 0.5865 |
| SVM Regressor | 5.7953 | 0.4355 | 9.4242 | 0.3356 | 7.1757 0.3092 |
| LightGBM Regressor | 4.2782 | 0.7156 | 3.9108 | 0.8891 | 4.3591 0.8133 |
| XGBoost Regressor | 4.0170 | 0.7493 | 3.9936 | 0.8844 | 5.3101 0.7229 |
| CatBoost | 5.6772 | 0.4993 | 7.7892 | 0.5601 | 7.7513 0.4095 |

На основе сравнения значений $RMSE$ и R^2 в табл. 3 для различных регрессионных моделей можно заключить, что ансамблевые модели и их модифицированные вариации показывают наилучшую способность предсказания значений целевых переменных $R^2 > 0.8$; $RMSE < 4.2$.

Для предсказания выхода твердых веществ наилучшие значения метрик имеют Random Forest Regression, Bagging Regressor, Extra Trees Regressor, GradientBoosting Regressor и HistGradientBoosting Regressor.

Для предсказания выхода жидкых веществ лучшие значения метрик имеют Decision Tree Regression, Random Forest Regression, Bagging Regressor, Extra Trees Regressor, GradientBoosting Regressor и HistGradientBoosting Regressor.

Для предсказания выхода газообразных веществ лучшие метрики имеют модели Random Forest Regression, Bagging Regressor, Extra Trees Regressor и HistGradientBoosting Regressor.

Далее, используя вышеперечисленные модели машинного обучения из табл. 3, были построены модели для мультитаргетной регрессии, позволяющие предсказать несколько целевых значений. Сравнительные результаты для метрик приведены в табл. 4.

Таблица 4. Сравнение мультитаргетных регрессионных моделей машинного обучения

Table 4. Comparison of multi-target machine learning regression models

| Модель | RMSE | R2 |
|--------------------------|--------|--------|
| Random Forest Regression | 3.3160 | 0.8571 |
| Bagging Regressor | 3.7126 | 0.8239 |
| Extra Trees Regressor | 3.1809 | 0.8688 |
| KNeighbors Regressor | 5.4460 | 0.6503 |
| Decision Tree Regression | 3.7636 | 0.8129 |
| LightGBM Regressor | 3.7253 | 0.7984 |
| XGBoost Regressor | 3.6829 | 0.8112 |

Сравнивая различные регрессионные модели, можно прийти к выводу, что наилучшие $RMSE$ и R^2 имеют модели Random Forest Regression и Extra Trees Regressor. Модель Extra Trees Regressor представляет собой алгоритм машинного обучения, который является

разновидностью ансамблевого метода Random Forest Regression, использующего множество деревьев решений для предсказания непрерывных значений. Он был разработан для повышения точности предсказаний и уменьшения переобучения.

Для ансамблевых моделей, показавших лучшие метрики качества, был произведен ручной подбор гиперпараметров моделей. Варьировались признаки: глубина деревьев решения; количество деревьев; количество признаков, рассматриваемых при каждом разбиении узла. Подбор не дал значительных улучшений, так как вариация значений метрик качества моделей составила: $R^2 \leq 0.01$, $RMSE \leq 0.05$.

На рис. 1 приведена круговая диаграмма, построенная на основании данных, полученных вследствие обучения регрессионной модели Extra Trees Regressor. Диаграмма показывает важность начальных признаков с точки зрения их влияния на целевые переменные, представляющие собой выход твердых, жидкых продуктов пиролиза и газа.

Существенное всего на прогноз выхода продуктов пиролиза с помощью модели Extra Trees Regressor влияют признаки, представляющие элементный состав сырья (38.2%), скорость нагревания (9.2%), конечную температуру пиролиза (14.9%) и зольность (9.2%).

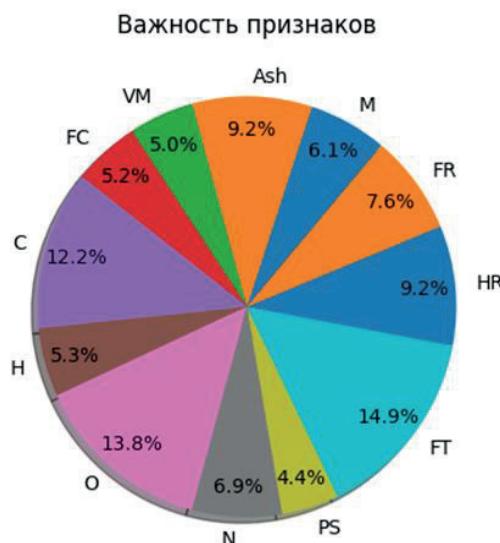


Рис. 1. Вклад признаков в прогнозную модель

Fig. 1. The features impact to prediction model

На рис. 2 приведены графики остатков для доли твердых, жидкых веществ и газа, полученных в ходе моделирования выхода продуктов в процессе пиролиза с помощью регрессионной модели Extra Trees Regressor.

В данной работе была проведена k -блочная кросс-валидация с разбиением исходного набора данных на 10 равных частей. На каждой из итераций одна из частей использовалась для проверки качества модели, а остальные 9 — для ее обучения. Та-

кой подход позволяет более объективно оценить способность модели к обобщению и уменьшить влияние случайного разбиения данных на итоговые результаты. Были получены усредненные показатели качества модели на всех блоках, что обеспечивает надежную оценку ее прогностической эффективности. Данный метод гарантирует всестороннюю проверку модели на различных подвыборках, минимизируя риск переобучения и недообучения, и является стандартной практикой при работе с ограниченным количеством данных.

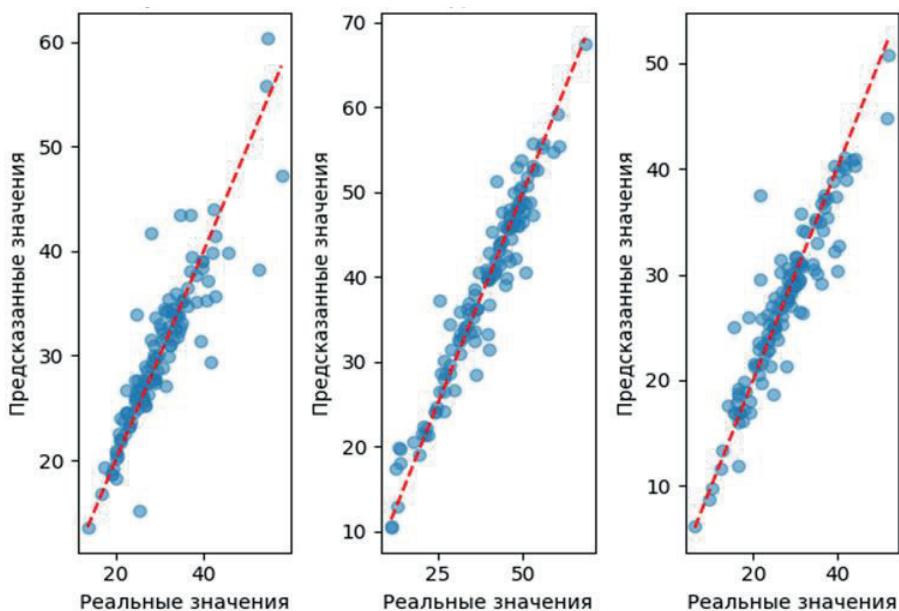


Рис. 2. График остатков предсказанных значений % выхода конечных продуктов пиролиза
Fig. 2. Residual plot of the pyrolysis predicted products yields (%)

Как показано на графиках, модель достаточно точно предсказывает выход доли жидкых веществ с $RMSE = 3.37$. Полученные значения метрик качества проверки выбранных моделей из табл. 4 согласуются со значениями метрик моделей, использующихся в задачах прогнозирования выхода продуктов пиролиза на основе данных [Shen et al., 2022]. Метод случайного леса широко применяется исследователями для прогнозирования выхода продуктов пиролиза [Zhu et al., 2019; Leng et al., 2022]. Деревья решений, входящие в случайный лес, используются для описания влияния признаков на целевые переменные и объяснения получения выходных значений [İşcen et al., 2023].

Ряд исследователей использует в качестве входных данных для обучения моделей машинного обучения процентное содержание целлюлозы, гемицеллюлозы и лигнина [Shen et al., 2022; Tsekos et al., 2021] в дополнение к признакам, описанным в табл. 1.

Заключение

Предложенный подход к математическому моделированию выхода продуктов пиролиза растительной биомассы, основанный на их предсказании с использованием алгоритмов машинного обучения, в рамках создания системы компьютерного моделирования процесса пиролиза показал свою эффективность. Это позволяет продолжить дальнейшие исследования по тематике использования алгоритмов глубокого обучения для моделирования выхода конечных продуктов на основе физико-химических характеристик исходного сырья и температурного режима пиролиза.

Ограничением для настоящего исследования являлся исходный набор данных, при появлении новых экспериментальных данных требуется дообучение модели.

Применение методов машинного обучения для моделирования процесса пиролиза осложняется необходимостью формирования репрезентативной и достаточной по объему выборки, охватывающей вариативный спектр параметров процесса и характеристик сырья, для обеспечения адекватности и высокой предиктивной способности разрабатываемых моделей.

В настоящее время разработанные регрессионные модели функционируют в виде отдельного программного приложения, однако в перспективе планируется их интеграция в состав программного комплекса, обладающего расширенным функционалом и предназначенного для моделирования процесса пиролиза растительной биомассы. Программный комплекс будет включать в себя средства для предварительного анализа данных, полученных в ходе натурных экспериментов, инструменты для изучения и визуализации корреляционных зависимостей признаков входных данных, а также функционал для прогнозирования выхода продуктов пиролиза с использованием моделей машинного обучения. Кроме того, предполагается реализация средств для осуществления расчетов теплового эффекта на основе кривых, полученных в результате ТГА.

Дальнейшие исследования предполагают изучение процесса пиролиза растительной биомассы в динамике с широким использованием методов искусственного интеллекта для прогнозирования временных рядов. Такие временные ряды, содержащие изменение температуры и состава сырья, можно получить в результате исследования ТГА и ДТА кривых. В частности, планируется выяснить, насколько точно можно прогнозировать текущий остаток сырья в различные моменты времени в ходе процесса пиролиза растительной биомассы.

Список источников

- Ле Т., Бондарев А. В., Бондарева Л. И., Монакова А. С., Баршин А. В. 2020. Применение алгоритмов машинного обучения в прогнозе результата пиролитического анализа // Proceedings of Higher Educational Establishments: Geology and Exploration. № 63(6). С. 8–19. <https://doi.org/10.32454/0016-7762-2020-63-6-8-19>
- Balsora H. K., S K., Joshi J.B., Sharma A., Chakinala A. G. 2023. Artificial neural network-based models for the prediction of biomass pyrolysis products from preliminary analysis // Industrial & Engineering Chemistry Research. No. 62(36). Pp. 14311–14319. <https://doi.org/10.1021/acs.iecr.3c01662>

- Cheng Y., Ekici E., Yildiz G., Yang Y., Coward B., Wang J. 2023. Applied machine learning for prediction of waste plastic pyrolysis towards valuable fuel and chemicals production // Journal of Analytical and Applied Pyrolysis. Vol. 169. 105857. <https://doi.org/10.1016/j.jaap.2023.105857>
- Fatehi H., Bai X. S. 2014. A comprehensive mathematical model for biomass combustion // Combustion Science and Technology. Vol. 186(4-5). Pp. 574–593. <http://dx.doi.org/10.1080/00102202.2014.883255>
- Fernando N., Amin M., Narayana M., Jayawickrama T., Asadumllah, Jayasena S. 2015. A mathematical model for pyrolysis of biomass // Proceedings of the Moratuwa Engineering Research Conference (MERCon) (Moratuwa, Sri Lanka, 2015). Pp. 6–11. IEEE. <https://doi.org/10.1109/MERCon.2015.7112311>
- Hasse Ch., Debiagi P., Wen X., Hildebrandt K., Vascellari M., Faravelli T. 2021. Advanced modeling approaches for CFD simulations of coal combustion and gasification // Progress in Energy and Combustion Science. Vol. 86. 100938. <https://doi.org/10.1016/j.pecs.2021.100938>
- İşcen A., Öznacar K., Tunç K. M. M., Günay M. E. 2023. Exploring the critical factors of biomass pyrolysis for sustainable fuel production by machine learning // Sustainability. No. 15(20). 14884. <https://doi.org/10.3390/su152014884>
- Kaczor Z., Buliński Z., Werle S. 2020. Modelling approaches to waste biomass pyrolysis: A review // Renewable Energy. Vol. 159. Pp. 427–443. <https://doi.org/10.1016/j.renene.2020.05.110>
- Klason T., Bai X.-S. 2003. FG-DVC modeling of biomass combustion in fixed bed // Proceedings of the Third Mediterranean Combustion Symposium, Marrakech, Morocco. Pp. 188–194.
- Lee H., Choi I.-H., Hwang K.-R. 2024. Machine learning prediction of bio-oil production from the pyrolysis of lignocellulosic biomass: Recent advances and future perspectives // Journal of Analytical and Applied Pyrolysis. Vol. 179. 106486. <https://doi.org/10.1016/j.jaap.2024.106486>
- Leng L., Yang L., Lei X., et al. 2022. Machine learning predicting and engineering the yield, N content, and specific surface area of biochar derived from pyrolysis of biomass // Biochar. Vol. 4. No. 63. <https://doi.org/10.1007/s42773-022-00183-w>
- Mighani M., Shahi A., Antonioni G. 2017. Catalytic pyrolysis of plastic waste products: Time series modeling using least square support vector machine and artificial neural network // 16th International Conference on Sustainable Energy Technologies — SET 2017 (17–20 July 2017, Bologna, Italy).
- Niksa S. 2020. *bio*-FLASHCHAIN® theory for rapid devolatilization of biomass 2. Predicting total yields for torrefied woods // Fuel. Vol. 263. 116645. <https://doi.org/10.1016/j.fuel.2019.116645>
- Parmar K. 2017. Biomass — An overview on composition characteristics and properties // IRAInternational Journal of Applied Sciences. Vol. 7(1). Pp. 42–51. <http://dx.doi.org/10.21013/jas.v7.n1.p4>
- Pielsticker S., Debiagi P., Cerciello F., Hasse Ch., Kneer R. 2024. Comparative analysis of pyrolysis models including SFOR, CRECK, and Bio-CPD to predict reaction kinetics and products from extracted biomass components // Fuel. Vol. 371. Part A. 131867. <https://doi.org/10.1016/j.fuel.2024.131867>

- Popescu F., Mahu R., Ion I., Rusu Eu. 2020. A mathematical model of biomass combustion physical and chemical processes // *Energies*. Vol. 13(23). 6232. <https://doi.org/10.3390/en13236232>
- Ranganathan P., Gu S. 2016. Computational fluid dynamics modelling of biomass fast pyrolysis in fluidised bed reactors, focusing different kinetic schemes // *Bioresource Technology*. Vol. 213. Pp. 333–341. <https://doi.org/10.1016/j.biortech.2016.02.042>
- Ranzi E., Debiagi P., Frassoldati A. 2017. Mathematical modeling of fast biomass pyrolysis and bio-oil formation. Note II: Secondary gas-phase reactions and bio-oil formation // *ACS Sustainable Chemistry & Engineering*. Vol. 5(4). Pp. 2882–2896. <https://doi.org/10.1021/acssuschemeng.6b03098>
- Ranzi E., Debiagi P., Frassoldati A. 2017. Mathematical modeling of fast biomass pyrolysis and bio-oil formation. Note I: Kinetic mechanism of biomass pyrolysis // *ACS Sustainable Chemistry & Engineering*. Vol. 5(4). Pp. 2867–2881. <https://doi.org/10.1021/acssuschemeng.6b03096>
- Rongrong Lu, Jing Li, Xiong Zou, Anran Wang, Hongguang Dong. 2023. Prediction of product distribution of low-medium rank coal pyrolysis using artificial neural networks model // *Journal of the Energy Institute*. Vol. 107. 101181. <https://doi.org/10.1016/j.joei.2023.101181>
- Sharma A., Pareek V., Zhang D. 2015. Biomass pyrolysis — A review of modelling, process parameters and catalytic studies // *Renewable and Sustainable Energy Reviews*. Vol. 50. Pp. 1081–1096. <https://doi.org/10.1016/j.rser.2015.04.193>
- Shen J., Yan M., Fang M., Gao X. 2022. Machine learning-based modeling approaches for estimating pyrolysis products of varied biomass and operating conditions // *Bioresource Technology Reports*. Vol. 20. 101285. <https://doi.org/10.1016/j.biteb.2022.101285>
- Tabakaev R., Astafev A., Shanenkova Yu., Dubinin Yu., Yazykov N., Yakovlev V. 2019. Thermal effects investigation during biomass slow pyrolysis in a fixed bed reactor // *Biomass and Bioenergy*. Vol. 126. Pp. 26–33. <https://doi.org/10.1016/j.biombioe.2019.05.010>
- Ting Yan Li, Huan Xiang, Yang Yang, Jiawei Wang, Güray Yıldız. 2021. Prediction of char production from slow pyrolysis of lignocellulosic biomass using multiple nonlinear regression and artificial neural network // *Journal of Analytical and Applied Pyrolysis*. Vol. 159. 105286. <https://doi.org/10.1016/j.jaap.2021.105286>
- Tsekos C., Tandurella S., de Jong W. 2021. Estimation of lignocellulosic biomass pyrolysis product yields using artificial neural networks // *Journal of Analytical and Applied Pyrolysis*. Vol. 157. 105180. <https://doi.org/10.1016/j.jaap.2021.105180>
- Vikram Sh., Rosha P., Kumar S. 2021. Recent modeling approaches to biomass pyrolysis: A review // *Energy & Fuels*. Vol. 35(9). Pp. 7406–7433. <https://doi.org/10.1021/acs.energyfuels.1c00251>
- Wijayanti W., Musyaroh, Sasongko M. N., Kusumastuti R., Sasmoko. 2021. Modelling analysis of pyrolysis process with thermal effects by using Comsol Multiphysics // *Case Studies in Thermal Engineering*. Vol. 28. 101625. <https://doi.org/10.1016/j.csite.2021.101625>
- Willmott C. J., Matsuura K. 2006. On the use of dimensioned measures of error to evaluate the performance of spatial interpolators // *International Journal of Geographical Information Science*. Vol. 20(1). Pp. 89–102. <https://doi.org/10.1080/13658810500286976>
- Wooters T., Baxter L., Fletcher T. 2005. CPD model calculations of black liquor and biomass pyrolysis // *Prepr. Symp. — Am. Chem. Soc., Div. Fuel Chem.* Vol. 50. Pp. 108–109.

Zhu X., Li Y., Wang X. 2019. Machine learning prediction of biochar yield and carbon contents in biochar based on biomass characteristics and pyrolysis conditions // *Bioresource Technology*. Vol. 288. 121527. <https://doi.org/10.1016/j.biortech.2019.121527>

References

- Le, Thi, Bondarev, A. V., Bondareva, L. I., Monakova, A. S., & Barshin, A. V. (2022). Application of machine learning algorithms in predicting pyrolytic analysis result. *Proceedings of Higher Educational Establishments. Geology and Exploration*, 63, 8–19. <https://doi.org/10.32454/0016-7762-2020-63-6-8-19> [In Russian]
- Balsora, H. K., S, K., Joshi, J.B., Sharma, A., & Chakinala, A. G. (2023). Artificial neural network-based models for the prediction of biomass pyrolysis products from preliminary analysis. *Industrial & Engineering Chemistry Research*, 62(36), 14311–14319. <https://doi.org/10.1021/acs.iecr.3c01662>
- Cheng, Y., Ekici, E., Yildiz, G., Yang, Y., Coward, B., & Wang, J. (2023). Applied machine learning for prediction of waste plastic pyrolysis towards valuable fuel and chemicals production. *Journal of Analytical and Applied Pyrolysis*, 169, 105857. <https://doi.org/10.1016/j.jaap.2023.105857>
- Fatehi, H., & Bai, X. S. (2014). A comprehensive mathematical model for biomass combustion. *Combustion Science and Technology*, 186(4-5), 574–593. <http://dx.doi.org/10.1080/0010202.2014.883255>
- Fernando, N., Amin, M., Narayana, M., Jayawickrama, T., Asadumllah, & Jayasena, S. (2015). A mathematical model for pyrolysis of biomass. In *Proceedings of the Moratuwa Engineering Research Conference (MERCon)* (pp. 6–11). IEEE. <https://doi.org/10.1109/MERCon.2015.7112311>
- Hasse, Ch., Debiagi, P., Wen, X., Hildebrandt, K., Vascellari, M., & Faravelli, T. (2021). Advanced modeling approaches for CFD simulations of coal combustion and gasification. *Progress in Energy and Combustion Science*, 86, 100938. <https://doi.org/10.1016/j.pecs.2021.100938>
- İşçen, A., Öznaçar, K., Tunç, K. M. M., & Günay, M. E. (2023). Exploring the critical factors of biomass pyrolysis for sustainable fuel production by machine learning. *Sustainability*, 15(20), 14884. <https://doi.org/10.3390/su152014884>
- Kaczor, Z., Buliński, Z., & Werle, S. (2020). Modelling approaches to waste biomass pyrolysis: A review. *Renewable Energy*, 159, 427–443. <https://doi.org/10.1016/j.renene.2020.05.110>
- Klason, T., & Bai, X.-S. (2003). FG-DVC modeling of biomass combustion in fixed bed. In *Proceedings of the Third Mediterranean Combustion Symposium* (pp. 188–194).
- Lee, H., Choi, I.-H., & Hwang, K.-R. (2024). Machine learning prediction of bio-oil production from the pyrolysis of lignocellulosic biomass: Recent advances and future perspectives. *Journal of Analytical and Applied Pyrolysis*, 179, 106486. <https://doi.org/10.1016/j.jaap.2024.106486>
- Leng, L., Yang, L., Lei, X., et al. (2022). Machine learning predicting and engineering the yield, N content, and specific surface area of biochar derived from pyrolysis of biomass. *Biochar*, 4(63). <https://doi.org/10.1007/s42773-022-00183-w>
- Mighani, M., Shahi, A., & Antonioni, G. (2017). Catalytic pyrolysis of plastic waste products: Time series modeling using least square support vector machine and artificial neural network.

In *Proceedings of the 16th International Conference on Sustainable Energy Technologies — SET 2017*.

- Niksa, S. (2020). *bio-FLASHCHAIN® theory for rapid devolatilization of biomass 2. Predicting total yields for torrefied woods*. *Fuel*, 263, 116645. <https://doi.org/10.1016/j.fuel.2019.116645>
- Parmar, K. (2017). Biomass — An overview on composition characteristics and properties. *IRAI International Journal of Applied Sciences*, 7(1), 42–51. <http://dx.doi.org/10.21013/jas.v7.n1.p4>
- Pielsticker, S., Debiagi, P., Cerciello, F., Hasse, Ch., & Kneer, R. (2024). Comparative analysis of pyrolysis models including SFOR, CRECK, and Bio-CPD to predict reaction kinetics and products from extracted biomass components. *Fuel*, 371, part A, 131867. <https://doi.org/10.1016/j.fuel.2024.131867>
- Popescu, F., Mahu, R., Ion, I., & Rusu, Eu. (2020). A mathematical model of biomass combustion physical and chemical processes. *Energies*, 13(23), 6232. <https://doi.org/10.3390/en13236232>
- Ranganathan, P., & Gu, S. (2016). Computational fluid dynamics modelling of biomass fast pyrolysis in fluidised bed reactors, focusing different kinetic schemes. *Bioresource Technology*, 213, 333–341. <https://doi.org/10.1016/j.biortech.2016.02.042>
- Ranzi, E., Debiagi, P., & Frassoldati, A. (2017). Mathematical modeling of fast biomass pyrolysis and bio-oil formation. Note II: Secondary gas-phase reactions and bio-oil formation. *ACS Sustainable Chemistry & Engineering*, 5(4), 2882–2896. <https://doi.org/10.1021/acssuschemeng.6b03098>
- Ranzi, E., Debiagi, P., & Frassoldati, A. (2017). Mathematical modeling of fast biomass pyrolysis and bio-oil formation. Note I: Kinetic mechanism of biomass pyrolysis. *ACS Sustainable Chemistry & Engineering*, 5(4), 2867–2881. <https://doi.org/10.1021/acssuschemeng.6b03096>
- Rongrong Lu, Jing Li, Xiong Zou, Anran Wang, & Hongguang Dong. (2023). Prediction of product distribution of low-medium rank coal pyrolysis using artificial neural networks model. *Journal of the Energy Institute*, 107, 101181. <https://doi.org/10.1016/j.joei.2023.101181>
- Sharma, A., Pareek, V., & Zhang, D. (2015). Biomass pyrolysis — A review of modelling, process parameters and catalytic studies. *Renewable and Sustainable Energy Reviews*, 50, 1081–1096. <https://doi.org/10.1016/j.rser.2015.04.193>
- Shen, J., Yan, M., Fang, M., & Gao, X. (2022). Machine learning-based modeling approaches for estimating pyrolysis products of varied biomass and operating conditions. *Bioresource Technology Reports*, 20, 101285. <https://doi.org/10.1016/j.biteb.2022.101285>
- Tabakaev, R., Astafev, A., Shanenkova, Yu., Dubinin, Yu., Yazykov, N., & Yakovlev, V. (2019). Thermal effects investigation during biomass slow pyrolysis in a fixed bed reactor. *Biomass and Bioenergy*, 126, 26–33. <https://doi.org/10.1016/j.biombioe.2019.05.010>
- Ting Yan Li, Huan Xiang, Yang Yang, Jiawei Wang, & Güray Yıldız. (2021). Prediction of char production from slow pyrolysis of lignocellulosic biomass using multiple nonlinear regression and artificial neural network. *Journal of Analytical and Applied Pyrolysis*, 159, 105286. <https://doi.org/10.1016/j.jaap.2021.105286>
- Tsekos, C., Tandurella, S., & de Jong, W. (2021). Estimation of lignocellulosic biomass pyrolysis product yields using artificial neural networks. *Journal of Analytical and Applied Pyrolysis*, 157, 105180. <https://doi.org/10.1016/j.jaap.2021.105180>

- Vikram, Sh., Rosha, P., & Kumar, S. (2021). Recent modeling approaches to biomass pyrolysis: A review. *Energy & Fuels*, 35(9), 7406–7433. <https://doi.org/10.1021/acs.energyfuels.1c00251>
- Wijayanti, W., Musyaroh, Sasongko, M. N., Kusumastuti, R., & Sasmoko. (2021). Modelling analysis of pyrolysis process with thermal effects by using Comsol Multiphysics. *Case Studies in Thermal Engineering*, 28, 101625. <https://doi.org/10.1016/j.csite.2021.101625>
- Willmott, C. J., & Matsuura, K. (2006). On the use of dimensioned measures of error to evaluate the performance of spatial interpolators. *International Journal of Geographical Information Science*, 20(1), 89–102. <https://doi.org/10.1080/13658810500286976>
- Wooters, T., Baxter, L., & Fletcher, T. (2005). CPD model calculations of black liquor and biomass pyrolysis. *Prepr. Symp. — Am. Chem. Soc., Div. Fuel Chem.*, 50, 108–109.
- Zhu, X., Li, Y., & Wang, X. (2019). Machine learning prediction of biochar yield and carbon contents in biochar based on biomass characteristics and pyrolysis conditions. *Bioresource Technology*, 288, 121527. <https://doi.org/10.1016/j.biortech.2019.121527>

Информация об авторах

Роман Михайлович Бачурин, аспирант кафедры программного обеспечения, Школа компьютерных наук, Тюменский государственный университет, Тюмень, Россия
stud0000107305@study.utm.ru, <https://orcid.org/0009-0001-2177-3024>

Ирина Гелиевна Захарова, кандидат физико-математических наук, профессор кафедры программного обеспечения, Школа компьютерных наук, Тюменский государственный университет, Тюмень, Россия
i.g.zakharova@utm.ru, <https://orcid.org/0000-0002-4211-7675>

Information about the authors

Roman M. Bachurin, Postgraduate Student, Department of Software, School of Computer Sciences, University of Tyumen, Tyumen, Russia
stud0000107305@study.utm.ru, <https://orcid.org/0009-0001-2177-3024>

Irina G. Zakharova, Cand. Sci. (Phys.-Math.), Professor, Department of Software, School of Computer Sciences, University of Tyumen, Tyumen, Russia
i.g.zakharova@utm.ru, <https://orcid.org/0000-0002-4211-7675>