

Научная статья

УДК 004.85

DOI: 10.18577/2307-6046-2026-0-1-162-173

## РАЗРАБОТКА И ПРОГНОЗИРОВАНИЕ СВОЙСТВ ЭПОКСИДНЫХ КОМПОЗИЦИЙ МЕТОДАМИ МАШИННОГО ОБУЧЕНИЯ

К.С. Мишуров<sup>1</sup>, А.Д. Монахов<sup>1</sup>, И.А. Сарычев<sup>1</sup>

<sup>1</sup>Федеральное государственное унитарное предприятие «Всероссийский научно-исследовательский институт авиационных материалов» Национального исследовательского центра «Курчатовский институт», Москва, Россия; admin@viam.ru

**Аннотация.** Проанализировано применение методов машинного обучения для разработки эпоксидных композиций с заданными свойствами. Рассмотрены подходы к формированию обучающих наборов данных на основе ограниченного числа экспериментов. Особое внимание уделено методам активного обучения и байесовской оптимизации, позволяющим эффективно планировать эксперименты. Приведены практические примеры использования моделей машинного обучения для прогнозирования свойств разрабатываемых композиций. Показаны преимущества комбинирования различных методов машинного обучения для решения сложных задач в материаловедении.

**Ключевые слова:** эпоксидное связующее, клей, машинное обучение, активное обучение, байесовская оптимизация, прогнозирование свойств, разработка

**Для цитирования:** Мишуров К.С., Монахов А.Д., Сарычев И.А. Разработка и прогнозирование свойств эпоксидных композиций методами машинного обучения // Труды ВИАМ. 2026. № 1 (155). С. 162–173. URL: <http://www.viam-works.ru>. DOI: 10.18577/2307-6046-2026-0-1-162-173.

Scientific article

## DEVELOPMENT AND PREDICTION OF PROPERTIES OF EPOXY COMPOSITIONS USING MACHINE LEARNING METHODS

K.S. Mishurov<sup>1</sup>, A.D. Monahov<sup>1</sup>, I.A. Sarychev<sup>1</sup>

<sup>1</sup>Federal State Unitary Enterprise «All-Russian Scientific-Research Institute of Aviation Materials» of National Research Center «Kurchatov Institute», Moscow, Russia; admin@viam.ru

**Abstract.** The article analyzes the application of machine learning methods for the development of epoxy compositions with specified properties. Approaches to the formation of training data sets based on limited number of experiments are considered. Special attention is paid to the methods of active learning and Bayesian optimization methods, which allow efficient planning of experiments. Practical examples of using machine learning models for predicting the properties of developed compositions are given. The advantages of combining different machine learning methods for solving complex problems in materials science are shown.

**Keywords:** epoxy binder, adhesive, machine learning, active learning, Bayesian optimization, prediction of properties, design

**For citation:** Mishurov K.S., Monahov A.D., Sarychev I.A. Development and prediction of properties of epoxy compositions using machine learning methods. *Trudy VIAM*, 2026, no. 1 (155), pp. 162–173. Available at: <http://www.viam-works.ru>. DOI: 10.18577/2307-6046-2026-0-1-162-173.

## Введение

Эпоксидные композиции применяются в различных областях промышленности – автомобилестроении, аэрокосмической, авиационной, судостроительной, энергетической – в качестве связующих для полимерных композиционных материалов [1–5].

Разработка нового состава эпоксидной композиции с определенными свойствами – сложная комплексная задача ввиду большого выбора исходных компонентов и вариантов их сочетания. Упростить этот процесс позволяет применение статистических методов планирования эксперимента, которые позволяют снизить количество экспериментов [6]. Тематика постановки многофакторного эксперимента хорошо изучена и ее теоретические и практические аспекты описаны во многих научно-технических источниках [7, 8]. Для достижения заданных характеристик специалист, разрабатывающий новый состав, при планировании эксперимента основывается на литературных данных, в частности на анализе существующих публикаций, а также на своем профессиональном опыте и интуиции.

В настоящее время одним из направлений уменьшения трудоемкости и повышения эффективности процесса разработки новых составов эпоксидных композиций (как в количественном, так и в качественном плане) является применение методов машинного обучения (Machine Learning), которое можно кратко описать как обучение некоторой модели с прогнозом результатов поставленной задачи [9]. Суть методов и обзор возможностей применения их в различных областях материаловедения представлены в ряде работ [10–13]. Более детальное описание принципов работы методов машинного обучения представлено в специализированной научно-технической литературе [14, 15]. В самом простом случае для эпоксидных композиций задача модели – спрогнозировать температуру стеклования на основе данных о количестве отвердителя. При этом количество отвердителя – признак, а температура стеклования – прогнозируемое свойство (метка). Признак – любое свойство или характеристика данных, которые с помощью модели можно применять для прогнозирования. От точности прогноза модели будет зависеть успешность решения обратной задачи – нахождение состава с максимальной температурой стеклования путем прогнозирования с помощью модели образцов из списка виртуальных составов и дальнейшего экспериментального подтверждения свойств лучшего образца.

В данной статье разобраны подходы к созданию наборов данных на основе результатов экспериментов, необходимых для успешного применения методов машинного обучения при разработке и прогнозировании свойств эпоксидных композиций.

### Особенности создания наборов данных для прогнозирования свойств эпоксидных композиций

Получение экспериментальных данных в области полимерных материалов – это трудоемкий и дорогостоящий процесс, что является одним из основных факторов, препятствующих успешному применению машинного обучения в процессе разработки новых составов, так как от количества и качества данных зависит точность прогнозирования обученной модели. На сайте библиотеки scikit-learn в части количества данных рекомендуется начинать с 50-ти строк [16]. Количество строк данных должно превышать количество признаков как минимум в 10 раз [17]. Важное значение имеет также то, насколько в представленных данных отражены все комбинации признаков, иначе модели плохо работают, если сталкиваются с ранее не встречавшимися комбинациями. Возможным решением проблемы предоставления неполных данных может служить применение алгоритмов ресемплирования к уже существующим выборкам, что позволяет искусственно увеличить объем наблюдений – в частности, применение алгоритма на основе бутстреп-метода (bootstrap) [18]. Одним из подходов в планировании эксперимента для дальнейшего применения машинного обучения может служить создание однородной выборки из разрозненных точек данных: в данном случае могут применяться методы перебора по сетке или план «латинский квадрат» [8].

Еще одним подходом является применение методов активного обучения. В общем случае цель активного обучения – повысить точность прогнозирования модели до максимальных значений при определенных затратах на приобретение дополнительных меток, таких как затраты на изготовление дополнительных экспериментальных образцов и их испытания. Для достижения этой цели требуется проводить тщательный выбор следующего образца, а ключевым моментом в данном случае является анализ уже полученных прогнозов и метрик. Любая система активного обучения состоит из «учителя», который предоставляет результаты экспериментов, и системы запросов, которая решает, каким должен быть следующий эксперимент. Следующий эксперимент выбирают из пула спрогнозированных виртуальных экспериментов (все пространство возможных составов). Данный процесс является итерационным и может быть остановлен, если исчерпан бюджет или точность модели достигла необходимого уровня. Тема активного обучения – довольно обширна и с ее основными аспектами можно ознакомиться в следующих литературных источниках [19, 20].

Байесовская оптимизация (bayesian optimization) – метод машинного обучения, который одновременно поддерживает прогностическую модель для изучения целевой функции и принимает решения о том, как получить новые данные, чтобы уточнить имеющиеся данные о ней, используя байесовскую теорию принятия решений [21]. Для прогностической модели машинного обучения используется гауссовский процесс, а решение о том, в каком месте функции провести следующие наблюдения, принимается в области оптимизации. Можно считать байесовскую оптимизацию частным случаем активного обучения, с помощью которого решаются задачи оптимизации «черного ящика». Этот метод можно применять в самом начале разработки, когда имеется только несколько результатов экспериментов. Данный процесс также является итерационным, как показано на рис. 1.

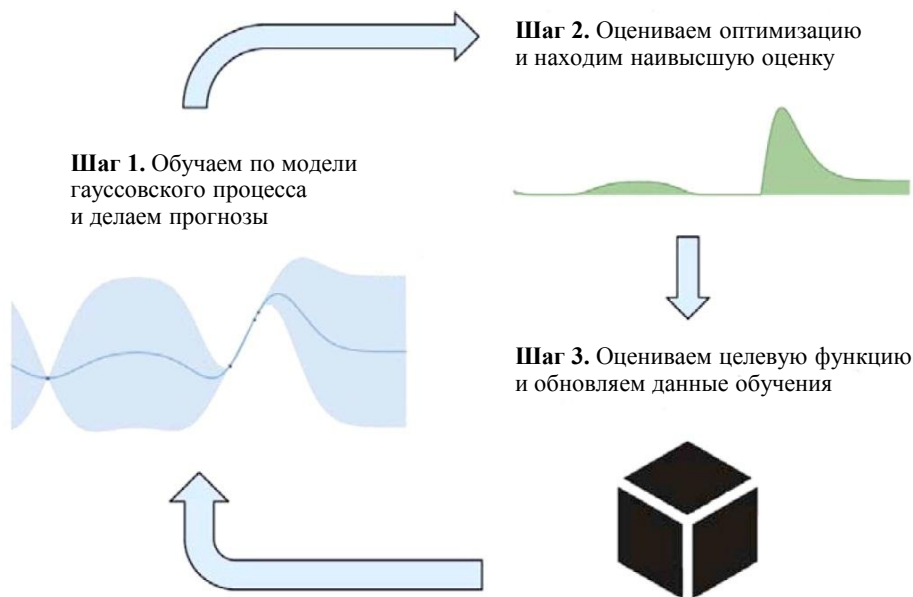


Рис. 1. Цикл байесовской оптимизации

### Прогнозирование свойств эпоксидных связующих на основе экспериментальных данных

Специалисты факультета креативной инженерии Китакусюкского национального технологического колледжа (Япония) совместно со специалистами исследовательской лаборатории химических веществ Toray Industries (Япония) применили методы машинного обучения для разработки эпоксидного связующего [22]. К разрабатываемой композиции

предъявляли следующие требования: повышенная температура стеклования  $T_g$ , низкая вязкость и короткая продолжительность отверждения. За экспериментальную основу взяли результаты испытаний более 300 составов эпоксидных связующих. В связи с отсутствием значений всех прогнозируемых характеристик для всех составов данные разделили на три части: для вязкости 88 строк, для температуры стеклования 144 строки, для продолжительности отверждения 128 строк. В целом только для 29 составов определили все три характеристики. Экспериментальные композиции содержали в составе следующие компоненты (в скобках дано количество различных компонентов): смолы (22), активные разбавители (11), спирты (8), отвердители (15), добавки (3). Обучение проводили в два этапа: сначала данные разбили на обучающую и тестовую выборки в соотношении 4:1. Для настройки гиперпараметров моделей проводили 10-кратную перекрестную проверку (Cross-Validation) с использованием фреймворка Optuna [23], в качестве основной метрики выбрали среднюю абсолютную ошибку MAE [10], чтобы избежать сильного влияния выбросов. В дополнение к ней для оценки модели также рассчитывали среднеквадратичную ошибку (RMSE) в качестве вспомогательной метрики. В рамках исследования обучены следующие модели машинного обучения: регрессия частично наименьших квадратов (Partial Least Squares Regression – PLS), метод опорных векторов (Support Vector Regression – SVR), регрессия на основе случайного леса (Random Forest Regression – RF), ридж-регрессия ядра (Kernel Ridge Regression – KRR), а также полносвязная искусственная нейронная сеть (Artificial Neural Network – ANN). Лучшие результаты для вязкости (на тестовых данных было только одно существенное отклонение) и температуры стеклования показала модель KRR с гауссовым ядром, для продолжительности отверждения – модель RF. Данные по метрикам представлены в табл. 1.

Таблица 1

Данные метрик лучших моделей

Параметр (модель)	Значения параметров для метрик			
	MAE	MAE-тест	RMSE	RMSE-тест
Вязкость (KRR)	0,056	0,551	0,084	1,66
Температура стеклования (KRR)	0,018	0,028	0,025	0,038
Продолжительность отверждения (RF)	0,055	0,136	0,076	0,218

Второй этап обучения проводили на лучших моделях с подобранными ранее гиперпараметрами: использовали все данные для повышения точности прогнозирования (результаты представлены на диаграммах рассеяния (Scatter plot) на рис. 2) – чем ближе точка к прямой линии, тем лучше прогноз модели. Использование всех данных увеличило качество прогнозирования моделей.

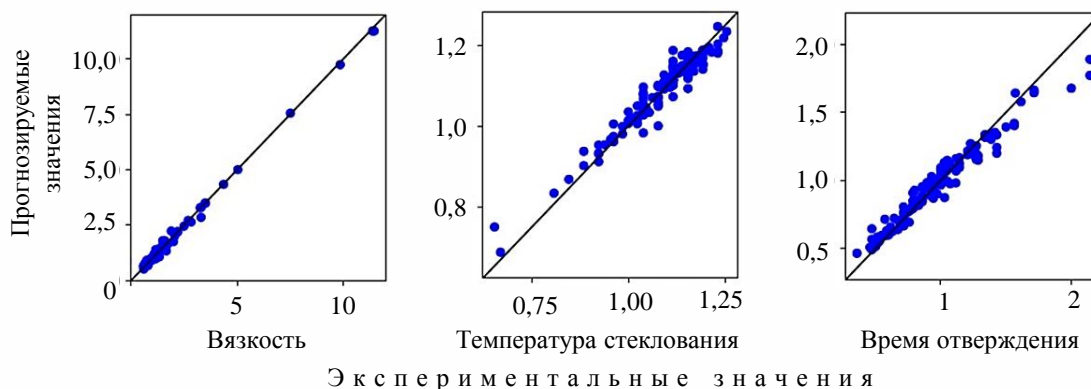


Рис. 2. Диаграммы рассеяния прогнозируемых и экспериментальных данных

Далее определяли пространство всех возможных составов, среди которых осуществляли поиск лучшего состава. Количество проверяемых составов ограничено, во-первых, доступностью компонентов, а во-вторых, выбором ограниченных пределов стехиометрического соотношения смолы и отвердителя: можно предположить, что данные ограничения сделаны исходя из стехиометрии и опыта разработчиков. Количество композиций составило 1 044 288, что позволяет воспользоваться простейшим методом перебора по сетке. Авторы работы [22] полагают, что в случае увеличения количества компонентов метод перебора по сетке будет слишком требователен к вычислительным ресурсам, и необходимо будет прибегнуть к другим методам – например, таким как генетический алгоритм [24]. Из всех прогнозируемых композиций только 42 удовлетворяли требованиям к свойствам, из них выбрали пять для экспериментального подтверждения – верификации. Следует отметить, что соотношения компонентов для выбранных композиций выходят за рамки соотношений для исходных экспериментальных композиций, использованных для обучения моделей. Это указывает на экстраполяционный характер полученных решений, что может снижать достоверность прогнозов и требует последующего экспериментального подтверждения полученных результатов. Однако в результате экспериментальной проверки один состав удовлетворил всем требованиям к связующему с большим отрывом от остальных – результаты исследования представлены в табл. 2.

Таблица 2

**Результаты генерации и проверки наиболее удачного состава связующего**

Параметр	Относительное значение характеристики		
	прогнозируемое	фактическое	требуемое
Вязкость	0,921	0,64	$\leq 1$
Температура стеклования	1,130	1,15	$\geq 1$
Продолжительность отверждения	0,870	0,90	$\leq 1$

В данной работе все компоненты составов закодированы, а полученные характеристики нормализованы, что исключает возможность непосредственного изучения полученных составов связующего. Исследование наглядно демонстрирует потенциал применения методов машинного обучения для поиска наиболее эффективных сочетаний компонентов, обеспечивающих выполнение заданных требований. В дальнейшем разработчики планируют расширить набор применяемых моделей машинного обучения, а также осуществить подбор гиперпараметров при помощи инструментов AutoML (Automated Machine Learning).

Разработчики из Японии применили машинное обучение для разработки клеевого состава, чтобы достичь максимальной прочности клеевого соединения  $\sigma_{адгез}$  [25]. За основу клея взяты эпоксидные смолы на основе бисфенола А и отвердителя полипропиленгликоля с концевыми аминогруппами (джеффамин). В данной работе использованы смолы с молекулярной массой  $M_{wE}$ : 370, 1650, 2900, 3800 г/моль и отвердители с  $M_{wC}$ : 230, 400, 2000, 4000 г/моль; четыре стехиометрических соотношения экспериментальных композиций с компонентами  $r$  – от 0,75 до 1,50 (шаг 0,25); четыре температуры отверждения  $T_{отв}$  – от 90 до 210 °С (шаг 40 °С). Получено всего 256 возможных комбинаций. Далее составлены два «греко-латинских» квадрата 4×4 и получены 32 комбинации составов экспериментальных образцов, которые испытаны на прочность клеевого соединения. Около 35 % результатов составили композиции со значениями  $\sigma_{адгез}$  до 4 МПа, среди которых имелись композиции, не обладающие достаточной технологичностью для изготовления образцов. Максимальное значение составило 31,9 МПа ( $M_{wE} = 370$  г/моль;  $M_{wC} = 230$  г/моль;  $r = 1,50$ ;  $T_{отв} = 130$  °С), среднее значение 10,2 МПа.

В качестве моделей машинного обучения выбраны эластичная сеть (Elastic Net), случайный лес и градиентный бустинг. Обучение проводилось с 32-кратной перекрестной проверкой, использовались метрики MAE, RMSE и  $R^2$  [10]. Лучший результат показала модель градиентного бустинга, по которой получены следующие значения по метрикам MAE: 3,7 МПа, RMSE: 5,4 МПа и  $R^2 = 0,68$ . Данная модель использована для прогнозирования оставшихся 224 композиций, полученные значения ранжированы по возрастанию. Далее выбраны пять наилучших составов для дальнейшего экспериментального подтверждения и дообучения модели. Эти этапы являются первым циклом активного обучения из трех. После каждого этапа точность модели улучшалась и достигла – MAE: 3,0 МПа, RMSE: 4,0 МПа и  $R^2 = 0,85$ . На третьем этапе наилучший результат, полученный экспериментально, составил 29,0 МПа (прогнозируемый 25,4 МПа) с составом –  $Mw_E = 370$  г/моль,  $Mw_C = 400$  г/моль;  $r = 1,00$ ;  $T_{отв} = 210$  °С. Третий этап активного обучения стал последним в связи с тем, что среднее абсолютное отклонение модели MAE составило 3,0 МПа, что сопоставимо с максимальным отклонением в результатах эксперимента, для которого получено значение 3,5 МПа. Далее реализован процесс байесовской оптимизации с уточнением экспериментальных условий: параметр  $r$  может изменяться с шагом 0,01;  $T_{отв}$  – с шагом в 1 °С, параметры  $Mw_E$  и  $Mw_C$  останутся дискретными значениями. При байесовской оптимизации выбрано четыре композиции с присвоением рейтинга, данные композиции проверены экспериментально. Все четыре композиции имели одинаковые значения  $Mw_E = 370$  г/моль,  $Mw_C = 400$  г/моль и отличались значениями  $r$  и  $T_{отв}$ . Композиция с самым низким рейтингом показала экстремально высокий результат  $\sigma_{адгез} = 35,8$  МПа и значения  $r = 0,89$  и  $T_{отв} = 209$  °С. Следует отметить, что значение  $r$  для данной композиции меньше, чем у первоначальных лучших составов с показателями 1,00–1,50, а также меньше, чем у композиций с более высоким рейтингом: 1,11–1,30. Авторы статьи [25] сделали вывод, что столь высокий показатель  $\sigma_{адгез}$  связан с нахождением оптимального баланса между прочностью и эластичностью клеевого состава, который, по их мнению, напрямую зависит от значения стехиометрического соотношения компонентов. В статье также довольно подробно проанализированы закономерности, выявленные с помощью модели машинного обучения, и установлено, что они соответствуют закономерностям из научно-технических литературных данных о клеевых соединениях.

Коллектив разработчиков из Германии применил машинное обучение для нахождения составов на основе диглицидилового эфира бисфенола А, отвержденного семью различными аминокислотами в присутствии ускорителя отверждения – замещенной мочевины, с максимально и минимально возможной температурой стеклования  $T_g$ , и обучил модель, способную прогнозировать составы с заданной температурой  $T_g$  [26]. Первоначальная цель исследования – проведение минимально необходимого количества экспериментов в процессе разработки. Изготовили образцы, в которых каждую аминокислоту взяли в стехиометрическом соотношении со смолой, равном 1:1 (из расчета на активный водород). Для экспериментальных составов случайным образом выбирали процентное соотношение в каждом образце – например, экспериментальный состав мог быть таким: 50 % L-аргинин, 40 % L-цитруллин, 10 % гаммааминомасляная кислота (ГАМК), что в сумме составляло 100 %. Образцы также могли содержать составы со всеми аминокислотами. В каждый экспериментальный состав добавляли ускоритель DYNARD UR400. Первоначально случайным образом выбрано пять процентных соотношений, изготовлены экспериментальные образцы и определена их температура  $T_g$ . Количество каждого компонента являлось признаком, а температура  $T_g$  – прогнозируемой характеристикой (меткой). Для выбора следующих экспериментальных составов применяли байесовскую оптимизацию и активное обучение. В качестве прогностической модели для байесовской оптимизации использовали гауссовский процесс с ядром

Матерна. Модель обучали на первоначальных пяти образцах, затем выполняли прогноз по всем возможным виртуальным экспериментальным составам ( $10^6$  вариантов). Полученные средние значения и стандартное отклонение использовали в качестве основы для байесовской оптимизации, предлагающей варианты составов, которые проверены экспериментально и показали максимальную и минимальную температуры стеклования (область БО (II) на рис. 3), которая соответствует первой фазе байесовской оптимизации.

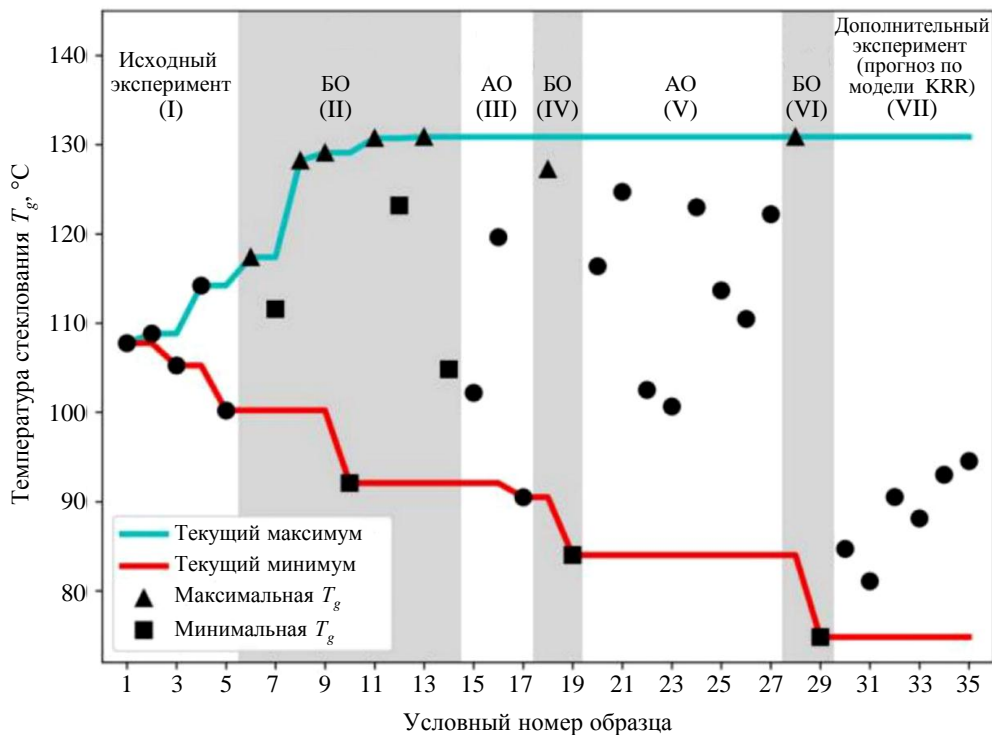


Рис. 3. Этапы проведения байесовской оптимизации (БО), совмещенные с этапами активного обучения (АО)

Авторы статьи [26] обращают внимание на разное поведение стратегии байесовской оптимизации – только эксплуатация для поиска максимальной температуры  $T_g$  и смешанный тип с этапами исследования для поиска минимальной температуры  $T_g$ , благодаря которым улучшается прогнозирующая способность модели. Уже в рамках первой фазы байесовской оптимизации с помощью модели смогли найти температуру  $T_g$  (131 °C) выше, чем у образца на основе L-цитруллина, который показал самую высокую температуру  $T_g$  (129 °C) среди образцов на основе только одной аминокислоты, что может свидетельствовать о наличии надмолекулярных эффектов, возникающих при использовании смеси аминокислот. Однако нельзя сделать вывод, что найден глобальный максимум функции, так как количество комбинаций составов огромно, а истинная функция неизвестна. Далее следовала фаза активного обучения (область АО (III) на рис. 3), в которой использовали метод выборки неопределенности (uncertainty sampling) [18] – в результате выбраны образцы, показывающие самую высокую неопределенность прогнозов. На каждом этапе активного обучения из обучающего набора данных случайным образом формировали десять подвыборок объемом 70 % от общей выборки (бутстреп-метод). Для каждой из подвыборок активно обучали отдельную модель KRR. Далее модели использовали для прогнозирования температуры  $T_g$  по всем  $10^6$  виртуальным экспериментам. Применение активного обучения оправдало ожидание разработчиков – на следующем этапе байесовской оптимизации (область БО (IV) на рис. 3)

найден состав с более низкой температурой  $T_g$ , состав с более высокой температурой  $T_g$  не обнаружен. Далее проводился этап активного обучения (область АО (V) на рис. 3), включающий экспериментальную проверку восьми составов. На следующем этапе байесовской оптимизации (область БО (VI) на рис. 3) обнаружены составы с высокой (первоначальное значение 131 °C так и не было превышено) и низкой (76 °C) температурой  $T_g$ , что ниже, чем у образца на основе ГАМК, который показал самую низкую температуру  $T_g$  (81 °C) среди образцов на основе только одной аминокислоты. На этом этапе авторы статьи [26] приняли решение, что задача поиска составов выполнена, однако обучение моделей на таком небольшом (29 образцов) и неоднородном наборе данных не привело бы к созданию модели, с помощью которой можно точно прогнозировать составы с необходимой температурой  $T_g$ , поэтому на следующем этапе добавили еще несколько точек данных. Последняя стадия (область (VII) на рис. 3) включала создание образцов, которые по прогнозам попадут в диапазон температур от 80 до 100 °C (так как проведение предыдущих этапов привело к тому, что большинство точек расположены выше 100 °C). Полученные экспериментальные данные подтвердили прогнозы, что можно считать первой проверкой модели.

Далее протестированы следующие регрессионные модели: регрессия на основе гауссовских процессов (Gaussian process regression – GPR) ридж-регрессия ядра (Kernel ridge regression – KRR), K-ближайших соседей (K-nearest neighbors – KNN), градиентный бустинг (Gradient boosting regression – GBR), метод опорных векторов (Support vector regression – SVR), метод наименьших квадратов (Least squares – LS), линейная регрессия с L1-регуляризацией (LASSO), случайный лес (Random forest – RF). Первоначально модели тестировали с гиперпараметрами по умолчанию. Прогнозирующую способность моделей проверяли при помощи  $k$ -кратной перекрестной проверки, число  $k$  ( $k$  – количество частей, на которое разбиваются исходные данные) генерировалось случайным образом и для каждой модели формировалось 200 наборов данных (с различным соотношением обучающей части к тестовой). Параметры  $k$ , показавшие лучшие значения метрик, указаны рядом с названиями моделей на рис. 4.

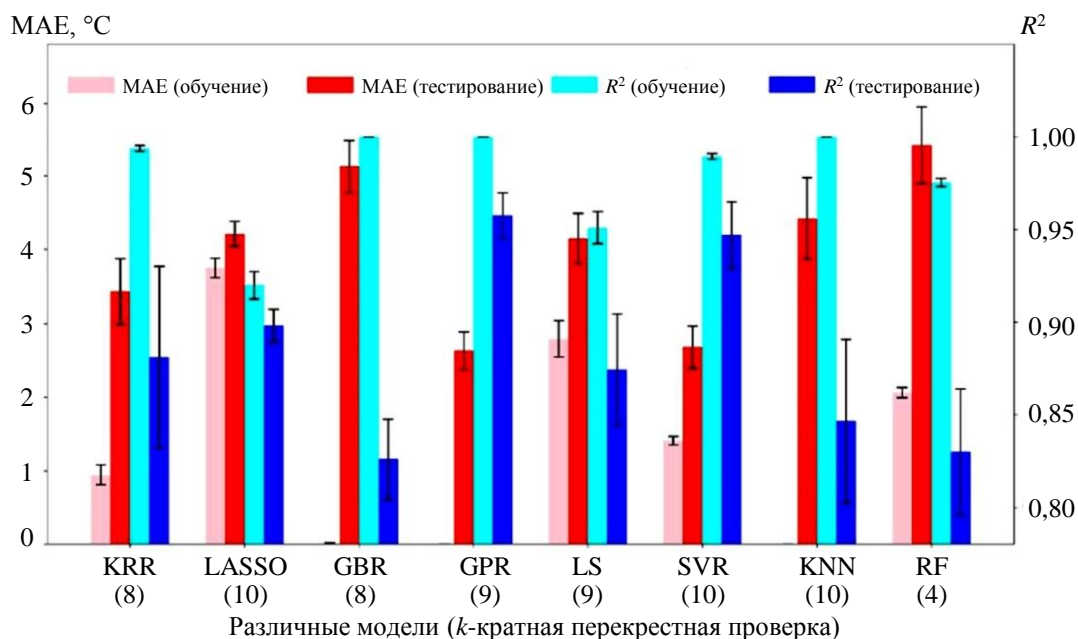


Рис. 4. Сравнение различных моделей обучения по основным метрикам – MAE и коэффициенту детерминации  $R^2$

Все модели показали схожие характеристики по средней абсолютной ошибке (MAE) на тестовом наборе для температуры  $T_g = 2,6-5,4$  °C. Среди них наилучшие результаты показали непараметрические модели – GPR и SVR, которые не предполагают конкретной формы базовой функции: они обеспечили наименьшее значение MAE и наивысший коэффициент детерминации ( $R^2$ ) на тестовом наборе данных. Модели GPR, GBR и KNN продемонстрировали переобучение, так как имеют самую большую разницу между значениями MAE на обучающем и тестовом наборах. Исходя из тех же принципов, модель LASSO демонстрирует наименьшее переобучение. Авторами статьи [26] предпринята попытка к уменьшению переобучения с помощью оптимизации гиперпараметров на примере модели KRR: все данные разделили на обучающий, контрольный (Validation Set) и тестовый наборы, что с учетом небольшого количества исходных данных, хотя и показало уменьшение переобучения, однако также продемонстрировало существенную зависимость показателей метрик от используемого обучающего/контрольного набора после проведения 10 различных оптимизаций. По мнению авторов статьи [26], более простым способом уменьшения переобучения может быть увеличение исходного набора данных.

Лучшей моделью для понимания взаимосвязей между компонентным составом и температурой  $T_g$  является модель LASSO с использованием регуляризации L1, которая может обращать весовые коэффициенты в ноль, что упрощает уравнение и тем самым позволяет облегчить интерпретацию результатов. Авторы статьи [26] отмечают как интересный тот факт, что состав с самой высокой температурой  $T_g$  не содержит L-цитруллин, который показал самую высокую температуру  $T_g$  (129 °C). По мнению авторов статьи [26], это весомый аргумент в пользу того, что при проведении экспериментов с многомерными параметрами не следует полагаться исключительно на интуицию.

Одной из возможных причин применения полученных результатов может быть снижение стоимости разрабатываемых композиций, так как прогнозируемые составы демонстрируют схожую температуру  $T_g$  и часть составов содержит относительно более дешевые компоненты. Сравнение стоимости L-цитруллина с комбинацией аминокислот из состава, показавшего более высокую температуру  $T_g$ , демонстрирует разницу в 33,6 % в пользу применения комбинации аминокислот.

### **Заключения**

Проанализировав представленные в обзоре статьи, можно выделить два подхода разработчиков к применению машинного обучения в экспериментальной работе: первый – на основе существующих данных, второй, подразумевающий создание нового набора данных для обучения моделей. Первый подход может быть реализован в том числе, когда разработчики используют опубликованные данные других исследователей в качестве обучающего набора. Понимание того, что все данные текущих экспериментов могут быть использованы в дальнейшем и что к будущим разрабатываемым составам могут быть предъявлены требования по многим характеристикам, приводит к тому, что исследователям-разработчикам необходимо проводить широкий спектр испытаний с заделом на будущее, особенно если основные временные и стоимостные затраты идут на изготовление экспериментального состава. С другой стороны, к временным затратам на изготовление образцов для испытаний множества характеристик надо подходить рационально. В контексте обоих подходов особую актуальность приобретает применение методов активного обучения, в частности байесовской оптимизации. Эти подходы позволяют более рационально формировать обучающую выборку путем приоритетного отбора наиболее информативных экспериментов. Такой подход не только снижает общее количество необходимых испытаний, что позволяет существенно сократить временные и экономические затраты, но и повысить качество и обобщающую способность итоговой модели.

Таким образом, понимание возможностей машинного обучения позволяет изначально выстроить экспериментальную работу так, чтобы свести к минимуму количество образцов, но при этом добиться максимального уровня свойств, а также получить работающую модель, с помощью которой в будущем можно прогнозировать составы с другими целевыми свойствами, если таковые понадобятся.

Следует отметить, что в рассмотренных в данном обзоре статьях довольно подробно описан каждый этап проделанной работы, что позволит другим исследователям применить полученные знания, что особенно актуально с учетом отсутствия стандартов и методик для проведения подобной работы и общую направленность обучающей литературы в области машинного обучения на работу с большими базами данных.

#### Список источников

1. Каблов Е.Н., Чурсова Л.В., Бабин А.Н., Мухаметов Р.Р., Панина Н.Н. Разработки ФГУП «ВИАМ» в области расплавных связующих для полимерных композиционных материалов // Полимерные материалы и технологии. 2016. Т. 2. № 2. С. 37–42.
2. Каблов Е.Н. Материалы нового поколения и цифровые технологии их переработки // Вестник Российской академии наук. 2020. Т. 90. № 4. С. 331–334. DOI: 10.31857/S0869587320040052.
3. Малышева Г.В., Мараховский П.С., Баринов Д.Я., Николаев Е.В. Оптимизация режимов отверждения стеклопластиков на основе эпоксидного связующего // Авиационные материалы и технологии. 2023. № 2 (71). С. 94–103. URL: <http://www.journal.viam.ru> (дата обращения: 17.04.2025). DOI: 10.18577/2713-0193-2023-0-2-94-103.
4. Ткачук А.И., Донецкий К.И., Терехов И.В., Караваев Р.Ю. Применение терморективных связующих для изготовления полимерных композиционных материалов методами безавтоклавного формования // Авиационные материалы и технологии. 2021. № 1 (62). С. 22–33. URL: <http://www.journal.viam.ru> (дата обращения: 17.04.2025). DOI: 10.18577/2713-0193-2021-0-1-22-33.
5. Каблов Е.Н., Чурсова Л.В., Лукина Н.Ф., Куцевич К.Е., Рубцова Е.В., Петрова А.П. Исследование эпоксидно-полисульфоновых полимерных систем как основы высокопрочных клеев авиационного назначения // Клеи. Герметики. Технологии. 2017. № 3. С. 7–12.
6. Мошинский Л. Эпоксидные смолы и отвердители. Тель-Авив: Аркадия пресс Лтд., 1995. 370 с.
7. Адлер Ю.П., Маркова Е.В., Грановский Ю.В. Планирование эксперимента при поиске оптимальных условий. М.: Наука, 1976. 140 с.
8. Монтомери Д.К. Планирование эксперимента и анализ данных. Пер. с англ. Л.: Судостроение, 1980. 384 с.
9. Орешко Е.И., Ерасов В.С., Сibaев И.Г., Луценко А.Н., Шершак П.В. Алгоритмы машинного обучения (обзор). Часть 1. Задачи классификации и регрессии. Линейные алгоритмы в машинном обучении. Применение алгоритмов машинного обучения для расчета прочностных характеристик материалов // Авиационные материалы и технологии. 2022. № 3 (68). С. 130–146. URL: <http://www.journal.viam.ru> (дата обращения: 20.04.2025). DOI: 10.18577/2713-0193-2022-0-3-130-146.
10. Орешко Е.И., Ерасов В.С., Сibaев И.Г., Луценко А.Н., Шершак П.В. Алгоритмы машинного обучения (обзор). Часть 2. Метрики машинного обучения. Решающие деревья и ансамбли решающих деревьев. Алгоритм нейронной сети по прогнозированию свойств ферритномартенситной стали // Авиационные материалы и технологии. 2022. № 4 (69). С. 132–146. URL: <http://www.journal.viam.ru> (дата обращения: 20.04.2025). DOI: 10.18577/2713-0193-2022-0-4-132-146.
11. Шоничев И.Д., Тынченко В.С., Бородулин А.С., Музыка С.С. Применение технологий машинного обучения при проектировании и производстве конструкций из композиционных материалов // Все материалы. Энциклопедический справочник. 2024. № 6. С. 18–27. DOI: 10.31044/1994-6260-2024-0-6-18-27.
12. Монахов А.Д., Яковлев Н.О. Применение метода глубокого обучения при исследовании характеристик трещиностойкости // Труды ВИАМ. 2024. № 6 (136). С. 80–91. URL: <http://www.viam-works.ru> (дата обращения: 20.04.2025). DOI: 10.18577/2307-6046-2024-0-6-80-91.
13. Xu P., Ji X., Li M. et al. Small data machine learning in materials science // npj Computational Materials. 2023. Vol. 9. Art. 42. DOI: 10.1038/s41524-023-01000-z.

14. Серрано Л. Грокаем машинное обучение. СПб.: Питер, 2022. 320 с.
15. Шолле Ф. Глубокое обучение на Python. СПб.: Питер, 2021. 400 с.
16. Choosing the right estimator // Scikit-learn. Machine learning in Python. URL: [http://scikit-learn.org/stable/tutorial/machine\\_learning\\_map](http://scikit-learn.org/stable/tutorial/machine_learning_map) (дата обращения: 16.04.2025).
17. Теобальд О. Машинное обучение для абсолютных новичков. М.: Бомбора, 2024. 208 с.
18. Сутубалов А.И., Подживотов Н.Ю., Шершак П.В., Яковлев Н.О. Оценка адгезионной прочности системы «волокно–полимерная матрица» с помощью непараметрического бутстреп-метода // Труды ВИАМ. 2024. № 11 (141). С. 100–112. URL: <http://www.viam-works.ru> (дата обращения: 17.04.2025). DOI: 10.18577/2307-6046-2024-0-11-100-112.
19. Aggarwal C.C. Data Mining: The Textbook. N.-Y.: Springer, 2015. 734 p.
20. Монарх Р. Машинное обучение с участием человека. М.: ДМК Пресс, 2022. 498 с.
21. Нгуен К. Байесовская оптимизация на практике. Астана: АЛИСТ, 2024. 416 с.
22. Taniguchi S., Uemura K., Tamaki S. et al. Multi-objective Optimization of the Epoxy Matrix System Using Machine Learning // Results in Materials. 2023. Vol. 17 (1). Art. 100376. DOI: 10.1016/j.rinma.2023.100376.
23. Akiba T.T., Sano S., Yanase T. Optuna: a next-generation hyperparameter optimization framework // Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD'19). N.-Y., 2019. P. 2623–2631. DOI: 10.1145/3292500.3330701.
24. Ришал Х. Грокаем алгоритмы искусственного интеллекта. СПб.: Питер, 2023. 368 с.
25. Pruksawan S., Lambard G., Samitsu S. et al. Prediction and optimization of epoxy adhesive strength from a small dataset through active learning // Science and Technology of Advanced Materials. 2019. Vol. 20 (1). P. 1010–1021. DOI: 10.1080/14686996.2019.1673670.
26. Albuquerque R.Q., Rothenhäusler F., Ruckdäschel H. Designing formulations of bio-based, multi-component epoxy resin systems via machine learning // MRS Bulletin. 2024. No. 49. P. 59–70. DOI: 10.1557/s43577-023-00504-9.

#### References

1. Kablov E.N., Chursova L.V., Babin A.N., Mukhametov R.R., Panina N.N. Developments of FSUE VIAM in the field of melt binders for polymer composite materials. *Polimernye materialy i tekhnologii*, 2016, vol. 2, no. 2, pp. 37–42.
2. Kablov E.N. New generation materials and digital technologies for their processing. *Vestnik Rossiyskoy akademii nauk*, 2020, vol. 90, no. 4, pp. 331–334. DOI: 10.31857/S0869587320040052.
3. Malysheva G.V., Marakhovskiy P.S., Barinov D.Ya., Nikolaev E.V. Optimization of the curing modes of fiber-glass based on epoxy binder. *Aviation materials and technologies*, 2023, no. 2 (71), pp. 94–103. Available at: <http://www.journal.viam.ru> (accessed: April 17, 2025). DOI: 10.18577/2713-0193-2023-0-2-94-103.
4. Tkachuk A.I., Donetskyy K.I., Terekhov I.V., Karavaev R.Yu. The use of thermosetting matrices for the manufacture of polymer composite materials by the non-autoclave molding methods. *Aviation materials and technology*, 2021, no. 1 (62), pp. 22–33. Available at: <https://www.journal.viam.ru> (accessed: April 17, 2025). DOI: 10.18577/2713-0193-2021-0-1-22-33.
5. Kablov E.N., Chursova L.V., Lukina N.F., Kutsevich K.E., Rubtsova E.V., Petrova A.P. Study of epoxy-polysulfone polymer systems as the basis for high-strength adhesives for aviation purposes. *Klei. Germetiki. Tekhnologii*, 2017, no. 3, pp. 7–12.
6. Moshinsky L. *Epoxy resins and hardeners*. Tel Aviv: Arcadia press Ltd., 1995, 370 p.
7. Adler Yu.P., Markova E.V., Granovsky Yu.V. *Experimental design in the search for optimal conditions*. Moscow: Nauka, 1976, 140 p.
8. Montgomery D.K. *Experimental design and data analysis*. Trans. from Engl. Leningrad: Shipbuilding, 1980, 384 p.
9. Oreshko E.I., Erasov V.S., Sibayev I.G., Lutsenko A.N., Shershak P.V. Machine learning algorithms (review). Part 1. Classification and regression tasks. Linear algorithms in machine learning. Application of machine learning algorithms for calculating the strength characteristics of materials. *Aviation materials and technologies*, 2022, no. 3 (68), pp. 130–146. Available at: <http://www.journal.viam.ru> (accessed: April 20, 2025). DOI: 10.18577/2713-0193-2022-0-3-130-146.

10. Oreshko E.I., Erasov V.S., Sibayev I.G., Lutsenko A.N., Shershak P.V. Machine learning algorithms (review). Part 2. Machine learning metrics. Decision trees and ensembles of decision trees. An algorithm for neural network to forecast properties of ferrite-martensite steel. *Aviation materials and technologies*, 2022, no. 4 (69), pp. 132–146. Available at: <http://www.journal.viam.ru> (accessed: April 20, 2025). DOI: 10.18577/2713-0193-2022-0-4-132-146.
11. Shonichev I.D., Tynchenko V.S., Borodulin A.S., Muzyka S.S. Application of machine learning technologies in the design and production of structures from composite materials. *Vse materialy. Entsiklopedicheskiy spravochnik*, 2024, no. 6, pp. 18–27. DOI: 10.31044/1994-6260-2024-0-6-18-27.
12. Monakhov A.D., Yakovlev N.O. Application of the deep learning method in studying crack resistance characteristics. *Trudy VIAM*, 2024, no. 6 (136), pp. 80–91. Available at: <http://www.viam-works.ru> (accessed: April 20, 2025). DOI: 10.18577/2307-6046-2024-0-6-80-91.
13. Xu P., Ji X., Li M. et al. Small data machine learning in materials science. *npj Computational Materials*, 2023, vol. 9, art. 42. DOI: 10.1038/s41524-023-01000-z.
14. Serrano L. *Grokking Machine Learning*. St. Petersburg: Piter, 2022, 320 p.
15. Scholle F. *Deep Learning in Python*. St. Petersburg: Piter, 2021, 400 p.
16. Choosing the right estimator. *Scikit-learn. Machine learning in Python*. Available at: [http://scikit-learn.org/stable/tutorial/machine\\_learning\\_map](http://scikit-learn.org/stable/tutorial/machine_learning_map) (accessed: April 16, 2025).
17. Theobald O. *Machine Learning for Absolute Beginners*. Moscow: Bombora, 2024, 208 p.
18. Sutubalov A.I., Podzhivotov N.Yu., Shershak P.V., Yakovlev N.O. Assessment of the adhesive strength of the fiber–polymer matrix system using the nonparametric bootstrap method. *Trudy VIAM*, 2024, no. 11 (141), pp. 100–112. Available at: <http://www.viam-works.ru> (accessed: April 17, 2025). DOI: 10.18577/2307-6046-2024-0-11-100-112.
19. Aggarwal C.C. *Data Mining: The Textbook*. N.-Y.: Springer, 2015, 734 p.
20. Monarch R. *Machine Learning with Human Interaction*. Moscow: DMK Press, 2022, 498 p.
21. Nguyen K. *Bayesian Optimization in Practice*. Astana: ALIST, 2024, 416 p.
22. Taniguchi S., Uemura K., Tamaki S. et al. Multi-objective Optimization of the Epoxy Matrix System Using Machine Learning. *Results in Materials*, 2023, vol. 17 (1), art. 100376. DOI: 10.1016/j.rinma.2023.100376.
23. Akiba T.T., Sano S., Yanase T. Optuna: a next-generation hyperparameter optimization framework. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD'19)*. N.-Y., 2019, pp. 2623–2631. DOI: 10.1145/3292500.3330701.
24. Rishal H. *Grokking Artificial Intelligence Algorithms*. St. Petersburg: Piter, 2023, 368 p.
25. Pruksawan S., Lambard G., Samitsu S. et al. Prediction and optimization of epoxy adhesive strength from a small dataset through active learning. *Science and Technology of Advanced Materials*, 2019, vol. 20 (1), pp. 1010–1021. DOI: 10.1080/14686996.2019.1673670.
26. Albuquerque R.Q., Rothenhäusler F., Ruckdäschel H. Designing formulations of bio-based, multicomponent epoxy resin systems via machine learning. *MRS Bulletin*, 2024, no. 49, pp. 59–70. DOI: 10.1557/s43577-023-00504-9.

#### Информация об авторах

**Мишуров Константин Сергеевич**, ведущий инженер, НИЦ «Курчатовский институт» – ВИАМ, [admin@viam.ru](mailto:admin@viam.ru)

**Монахов Антон Дмитриевич**, инженер 2 категории, НИЦ «Курчатовский институт» – ВИАМ, [admin@viam.ru](mailto:admin@viam.ru)

**Сарычев Игорь Анатольевич**, научный сотрудник, НИЦ «Курчатовский институт» – ВИАМ, [admin@viam.ru](mailto:admin@viam.ru)

#### Information about the authors

**Konstantin S. Mishurov**, Leading Engineer, NRC «Kurchatov Institute» – VIAM, [admin@viam.ru](mailto:admin@viam.ru)

**Anton D. Monahov**, Second Category Engineer, NRC «Kurchatov Institute» – VIAM, [admin@viam.ru](mailto:admin@viam.ru)

**Igor A. Sarychev**, Researcher, NRC «Kurchatov Institute» – VIAM, [admin@viam.ru](mailto:admin@viam.ru)

Статья поступила в редакцию 02.06.2025; одобрена и принята к публикации после рецензирования 19.06.2025.  
The article was submitted 02.06.2025; approved and accepted for publication after reviewing 19.06.2025.