

УДК 004.942+624.131.37

DOI: [10.26102/2310-6018/2026.56.5.002](https://doi.org/10.26102/2310-6018/2026.56.5.002)

Реконструкция c , φ и E_{50} по лабораторным данным: интерпретируемый ансамбль и сравнение моделей

Н.Р. Тишин^{1,2}✉

¹Московский государственный технический университет им. Н.Э. Баумана, Москва, Российская Федерация

²АО «МОСТДОРГЕОТРЕСТ», Москва, Российская Федерация

Резюме. Статья посвящена задаче восстановления прочностных и деформационных характеристик грунтов: удельного сцепления c , угла внутреннего трения φ и секущего модуля деформации E_{50} по физическим и классификационным признакам, доступным в массовых лабораторных протоколах. Актуальность работы обусловлена тем, что в инженерно-геологической практике механические параметры определяются не для всех образцов, хотя именно они необходимы при расчетах оснований и параметризации геотехнических моделей. В работе использован архив лабораторных данных по грунтам, для которого выполнены контроль качества, фильтрация, формирование информативного признакового описания и независимая внешняя проверка. Для решения задачи проведено сравнение моделей машинного обучения для табличных данных, включая CatBoost, FT-Transformer и многозадачную нейросеть, а также рассмотрен интерпретируемый ансамбль моделей. Дополнительно выполнен анализ важности признаков, позволяющий оценить физическую согласованность получаемых прогнозов. Показано, что наилучшее качество достигается при использовании ансамбля с доминирующим вкладом CatBoost (FT-Transformer (0,10) + CatBoost (0,90)) с $WAPE = 13,16\%$, $R^2 = 0,877$ и $Acc_{\pm 20\%} = 76,36\%$. На тестовой выборке лучшие решения обеспечивают высокое качество восстановления целевых параметров, а внешняя валидация на независимом объекте подтверждает устойчивость подхода. Установлено, что наиболее надежно восстанавливаются параметры c и φ , тогда как прогнозирование E_{50} является более сложной задачей из-за повышенной чувствительности этого показателя к условиям испытаний и структурным особенностям грунта. Практическая значимость работы состоит в том, что предложенный подход позволяет обоснованно восстанавливать недостающие механические параметры грунтов по данным стандартных лабораторных определений и может использоваться в цифровых системах инженерно-геологического моделирования, обработке лабораторных данных и подготовке расчетных параметров для инженерной практики.

Ключевые слова: инженерная геология, механика грунтов, восстановление параметров, табличные данные, CatBoost, FT-Transformer, многозадачное обучение, ансамблирование, SHAP.

Для цитирования: Тишин Н.Р. Реконструкция c , φ и E_{50} по лабораторным данным: интерпретируемый ансамбль и сравнение моделей. *Моделирование, оптимизация и информационные технологии*. 2026;14(5). URL: <https://moitvvt.ru/ru/journal/article?id=2249> DOI: 10.26102/2310-6018/2026.56.5.002

Reconstruction c , φ and E_{50} from laboratory data: interpretable ensemble and model comparison

N.R. Tishin^{1,2}✉

¹Bauman Moscow State Technical University, Moscow, the Russian Federation

²JSC MOSTDORGEOTREST, Moscow, the Russian Federation

Abstract. The article addresses the problem of reconstructing the strength and deformation characteristics of soils, namely cohesion c , internal friction angle φ , and secant deformation modulus E_{50} , from physical and classification features available in routine laboratory reports. The relevance of the study is due to the fact that, in engineering geological practice, mechanical parameters are not determined for all samples, although these parameters are essential for foundation design calculations and for the parameterization of geotechnical models. The study is based on an archive of laboratory soil data, for which quality control, filtering, informative feature engineering, and independent external validation were performed. To solve the problem, a comparative analysis of machine learning models for tabular data was carried out, including CatBoost, FT-Transformer, and a multitask neural network, and an interpretable model ensemble was also considered. In addition, feature importance analysis was performed to assess the physical consistency of the obtained predictions. It is shown that the best performance is achieved by an ensemble with a dominant contribution from CatBoost, namely FT-Transformer (0.10) + CatBoost (0.90), yielding mean $WAPE = 13.16\%$, mean $R^2 = 0.877$ and mean $Acc_{\pm 20\%} = 76.36\%$. On the test set, the best solutions provide high-quality reconstruction of the target parameters, while external validation on an independent site confirms the robustness of the approach. It was found that the parameters c and φ are reconstructed most reliably, whereas predicting E_{50} is a more challenging task due to the greater sensitivity of this parameter to testing conditions and the structural features of the soil. The practical significance of the study lies in the fact that the proposed approach enables a justified reconstruction of missing mechanical soil parameters from standard laboratory test data and can be used in digital systems for engineering geological modeling, laboratory data processing, and preparation of design parameters for engineering practice.

Keywords: engineering geology, soil mechanics, parameter reconstruction, tabular data, CatBoost, FT-Transformer, multi-task learning, ensembling, SHAP.

For citation: Tishin N.R. Reconstruction c , φ and E_{50} from laboratory data: interpretable ensemble and model comparison. *Modeling, Optimization and Information Technology*. 2026;14(5). (In Russ.). URL: <https://moitvvt.ru/ru/journal/article?id=2249> DOI: 10.26102/2310-6018/2026.56.5.002

Введение

Современная инженерная геология и механика грунтов переживают фундаментальную трансформацию, обусловленную экспоненциальным ростом объемов данных и развитием вычислительных мощностей. Определение прочностных и деформационных характеристик грунтов остается одной из ключевых, но одновременно самых «дорогих» по времени задач инженерной геологии: такие параметры, как угол внутреннего трения φ , удельное сцепление c и модуль деформации E_{50} непосредственно входят в расчеты устойчивости, осадок и деформаций оснований и сооружений, а также в параметризацию моделей типа Mohr-Coulomb и Hardening Soil¹ [1]. На практике получение этих параметров требует репрезентативных серий лабораторных испытаний, строгого контроля качества проб и согласованной обработки результатов; при больших программах изысканий это приводит к неизбежным пробелам в данных: часть образцов имеет только физические показатели (плотности, пористости, пластичности), либо часть – прочностные параметры без деформационных или наоборот. В результате в инженерных проектах возникает системная потребность «достаивать» отсутствующие характеристики по доступным измерениям и классификациям, сохраняя физический смысл и обеспечивая сопоставимость параметров между объектами и лабораториями² [2]. За последние годы машинное обучение стало одним из наиболее практичных инструментов для такой задачи: модели способны аппроксимировать нелинейные связи между индексными характеристиками, грансоставом и механическими параметрами, а

¹ Papavasileiou S. *Manuals Latest – PLAXIS*. Bentley Communities. URL: https://bentleysystems.service-now.com/community?id=kb_article&sysparm_article=KB0107989 (дата обращения: 10.03.2026).

² ГОСТ 25100–2020. Грунты. Классификация. Москва: Стандартинформ; 2020. 37 с.

также учитывать взаимодействия факторов, которые трудно устойчиво описать простой регрессией [2].

В период с 2021 по 2026 год в научной литературе наблюдается консолидация усилий по внедрению методов машинного обучения (Machine Learning, ML) для создания суррогатных моделей, способных предсказывать механические параметры на основе более доступных физических индексов и данных зондирования. Этот переход к «дата-центричной» геотехнике (data-centric geotechnics) постулирует, что точность прогнозных моделей зависит не столько от сложности алгоритма, сколько от качества, объема и репрезентативности обучающих данных [3, 4]. Этот вывод подтверждается и последующими исследованиями в области геотехнического машинного обучения [5].

При этом в геотехнике сохраняются две критичные методические проблемы. Первая – обобщаемость: многие работы по прогнозу параметров прочности используют выборки в пределах сотен-нескольких тысяч наблюдений и часто не демонстрируют внешнюю проверку на независимом объекте/площадке, что повышает риск переобучения на «локальных» лабораторных закономерностях и снижает доверие к переносу результата [2, 6]. Аналогичные ограничения отмечаются и в других исследованиях геотехнических ML-моделей [7, 8]. Вторая – интерпретируемость: инженерная практика требует не только точности, но и понимания, почему модель дала то или иное значение, какие признаки определяют прогноз. Именно поэтому направления Explainable AI (XAI) и методы атрибуции вклада признаков (в частности, SHAP) быстро становятся стандартом для геотехнических приложений [9, 10].

Для решения этих проблем в работе используется массив лабораторных данных, кратно превосходящий типичные объемы в публикациях по оценке φ , и c : после фильтрации формируется выборка порядка 12600 строк при исходном массиве 16600 строк, что существенно повышает статистическую устойчивость и позволяет более жестко отделять обучение, валидацию и тестирование. Также показана независимая валидация на реальном объекте (54 пробы), не входящем в изначальную выборку – это принципиально усиливает доказательность вывода об обобщаемости модели. Задача решается сразу для трех связанных целевых параметров φ , c и E_{50} , что ближе к инженерному сценарию применения, чем одиночные модели [6, 7]. Сходные подходы обсуждаются и в ряде других исследований [8]. Наконец, выбрана современная и при этом прагматичная стратегия моделирования табличных данных: сильный бустинг по деревьям (CatBoost) [11], табличный трансформер FT-Transformer [12] и ансамблирование с подбором весов, что соответствует общим выводам литературы по табличному DL: «универсально лучшего» семейства моделей нет, а ансамбль часто дает выигрывает в устойчивости и профиле ошибок [11, 12]. Сходные выводы приводятся и в других работах по табличным моделям глубокого обучения [13, 14]. Дополнительные подтверждения получены в более поздних исследованиях [15].

Материалы и методы

Эмпирической базой исследования послужил масштабный архив данных лабораторных испытаний грунтов, собранный в ходе инженерно-геологических изысканий на различных объектах строительства. В рамках воспроизводимого пайплайна сначала выполняется фильтрация по полноте и адекватности ключевых полей и формируются независимые разбиения train/val/test в долях 70/15/15 при фиксированном seed, что необходимо для честного сравнения моделей и для последующего ансамблирования без утечки информации между подвыборками.

Признаковое описание в статье намеренно «инженерное»: используется компактный набор, который физически интерпретируем и реально доступен в массовых

протоколах. В итоговой конфигурации применяются 22 признака: 8 числовых и 14 категориальных. Числовой блок включает грансостав, плотность грунта r , плотность твердых частиц r_s , сухую плотность r_d , коэффициент пористости e , индекс пластичности I_p и индекс текучести I_l . Именно эти величины в литературе являются важными предикторами прочности и деформативности связных грунтов, поскольку косвенно описывают минералогический состав, структуру и состояние грунта [2, 6]. Аналогичные зависимости между индексными характеристиками и механическими параметрами отмечены и в других исследованиях [7, 8].

Ключевым элементом метода является перевод «сырых» гранулометрических долей в физически осмысленные агрегированные параметры и классы. По значениям содержания гранулометрических фракций формируется кривая процентного прохода $P(D)$. Затем значения D_{10} , D_{50} и D_{60} определяются методом линейной интерполяции в логарифмической шкале диаметров, после чего рассчитывается коэффициент неоднородности $C_u = D_{60}/D_{10}$. Такой способ согласуется с принятой практикой представления гранулометрических кривых, поскольку размер частиц обычно анализируется в логарифмической шкале. Далее вводится гранулометрическая классификация PLAXIS¹ по D_{50} с пятью классами: very coarse $D_{50} > 10$, coarse $2 < D_{50} < 10$, medium ($0,25 < D_{50} < 2$), fine ($0,075 < D_{50} < 0,25$), very fine $D_{50} < 0,075$. Эти категории затем кодируются one-hot энкодингом и добавляются к признакам, что позволяет модели по-разному «читать» одинаковые индексные показатели в разных гранулометрических режимах.

Вторая линия типизации – «тип грунта» по нормативной классификации. В представленной реализации тип грунта задается правилами, согласованными с ГОСТ 25100–2020 «Грунты. Классификация»², который используется как основной норматив. Выделяются 9 типов грунтов: торф, пески (гравелистый/крупный/средней крупности/мелкий/пылеватый), связные грунты (супесь, суглинок и глина), которые также кодируются с помощью one-hot энкодинга.

Выбор моделей и схема ансамбля опираются на специфику геотехнических признаков. В качестве базового семейства выбран градиентный бустинг по деревьям: он остается одним из самых сильных и устойчивых подходов на табличных данных, способный моделировать нелинейности и взаимодействия признаков без ручного задания базиса [11, 15]. Дополнительные примеры успешного применения подобных методов приводятся в работах [16, 17]. В частности, CatBoost удобен при наличии категориальных признаков и дает конкурентный уровень качества в задачах регрессии [11]. В проекте CatBoost реализован в каскадной постановке: сначала обучаются модели для φ и c , затем их предсказания добавляются как высокоуровневые признаки при прогнозе E_{50} . Эта идея согласуется с общей логикой метамоделирования и stacking-подходов: использовать выходы более простых/близких задач как информативные представления для более сложной цели [11, 18], а также отражает инженерную связанность параметров прочности и деформативности в реальных грунтах¹ [1].

Как альтернативное семейство используется FT-Transformer, адаптированный к табличным данным через токенизацию признаков в эмбединги и обработку self-attention слоями [12]. Важно, что литература по tabular DL подчеркивает: глубокие модели не всегда лучше бустинга, но дают иной профиль обобщения и ошибок, поэтому грамотный ансамбль может быть практичнее, чем ставка на одну архитектуру [12, 13]. Похожие выводы приводятся и в других исследованиях табличных DL-моделей [14].

Третья модель в проекте – многозадачная нейросеть (MTL) с общим backbone и тремя головами под (φ , c и E_{50}), что методически оправдано: многозадачное обучение

позволяет извлекать общие представления из связанных задач и часто стабилизирует обучение при шумных измерениях и частичной неполноте данных [19].

Гиперпараметры моделей подбирались с помощью перебора всех комбинаций в заданной сетке:

- {'depth': [6; 8; 10; 12; 14], 'learning_rate': [0,05; 0,01; 0,03; 0,05], 'iterations': [3000; 10000; 20000; 30000]} для CatBoost;
- {'d_token': [96; 128; 192], 'n_blocks': [2, 3, 4], 'n_heads': [8], 'lr': [1e⁻⁴; 5e⁻⁴], 'dropout': [0,1; 0,2], 'batch_size': [32, 64, 128]} для FT-Transformer;
- {'hidden_layers': [[128; 64], [256; 128; 64], [64; 32]], 'lr': [0,0005; 0,001; 0,005; 0,01], 'batch_size': [32, 64, 128, 256], 'dropout': [0,0; 0,1; 0,2]} для MTL.

Итоговое решение в работе эмпирически оптимизированное: финальный прогноз строится как взвешенный ансамбль предсказаний моделей в трансформированном пространстве с последующим обратным преобразованием в исходную шкалу.

$$\hat{y}_{ens} = \sum_{k=1}^K w_k \hat{y}_k, \quad \sum_{k=1}^K w_k = 1, \quad w_k > 0, \quad (1)$$

где y – истинный (наблюдаемый) вектор таргетов. В задаче это $[E_{50}^{(100)}, c, \varphi]$. \hat{y}_{ens} – итоговое (ансамблевое) предсказание для одного объекта/пробы после объединения моделей. Это то, что используется как финальный прогноз в статье и в приложении. $\hat{y}^{(k)}$ – предсказание k -й модели (модели № k) для того же объекта/пробы. Пример: $\hat{y}^{(1)}$ – CatBoost, $\hat{y}^{(2)}$ – FT-Transformer, $\hat{y}^{(3)}$ – MTL-сеть и т. д. K – число моделей в ансамбле. Например, если объединяете CatBoost и FT-Transformer, то $K = 2$. w_k – вес k -й модели в ансамбле (сколько «доверия» отдаем ее прогнозу). Чем больше w_k , тем сильнее влияние модели k на итог \hat{y}_{ens} . $\sum_{k=1}^K w_k = 1$ – условие нормировки весов: веса образуют выпуклую комбинацию. Это гарантирует, что итоговое предсказание – усреднение в масштабе прогнозов. $w_k > 0$ – условие положительности весов (иногда допускают $w_k \geq 0$). Положительные веса означают, что ни одна модель не «вычитается» из другой, а итог – интерпретируемое средневзвешенное (Рисунок 1).

Веса ансамбля подбирались полным перебором весов на симплексе для трех моделей. Перебирались все тройки весов (2):

$$w_{mtl} + w_{ft} + w_{cat} = 1, \quad w_i \in \{0; 0,05; 0,10; \dots; 1,00\}, \quad (2)$$

что дает 231 вариант (число решений уравнения $i + j + k = 20$). Для каждого варианта строилось предсказание ансамбля, после чего считались метрики в исходном масштабе таргетов.

Качество моделей и ансамбля оценивается в исходной шкале на тестовой выборке и на внешнем объекте. В проекте используется WAPE (3), как основная метрика оптимизации (средняя по таргетам), что уместно при широком диапазоне значений таргетов:

$$WAPE = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{\sum_{i=1}^n |y_i|} \cdot 100\%, \quad (3)$$

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

$$Acc_{\pm 20\%} = \frac{1}{n} \sum_{i=1}^n \mathbb{1} \left(\frac{|y_i - \hat{y}_i|}{\max(|y_i|, \varepsilon)} \leq 0,20 \right) \cdot 100\%, \quad (5)$$

где $Acc_{\pm 20\%}$ – процент предсказаний, попавших в допуск $\pm 20\%$ от истинного значения. n – число наблюдений, y_i – истинное значение таргета для i -й пробы, \hat{y}_i – предсказанное значение таргета для i -й пробы, \bar{y} – среднее истинных значений по выборке, $\mathbb{1}(\cdot)$ –

индикаторная функция: равна 1, если условие внутри скобок истинно, равна 0, если условие ложно.

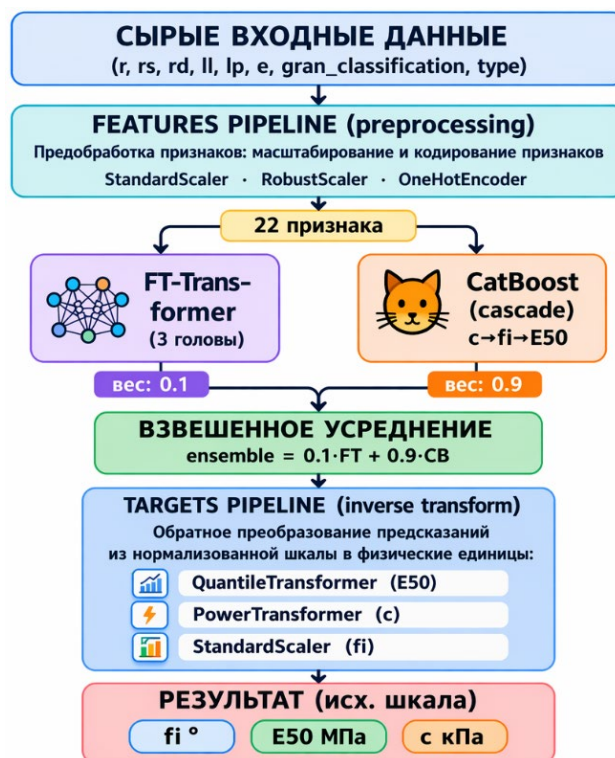


Рисунок 1 – Схема пайплайна обработки данных
 Figure 1 – Schematic of the data processing pipeline

Дополнительно приводятся R^2 (4) и точность попадания в допуск $Acc_{\pm 20\%}$ (5).

Блок совместимости построен следующим образом: для CatBoost используются точные SHAP-значения, для FT-Transformer – permutation feature importance, а для ансамбля – взвешенное суммирование вкладов пропорционально весам моделей. Выбор SHAP методически обоснован тем, что это один из наиболее принятых и теоретически formalизованных подходов к локальной и глобальной интерпретации предсказаний сложных моделей [20], и он активно применяется в прикладной геотехнике (например, в задачах разжижения, прочности и устойчивости) [10, 20].

В сравнении с близкими работами по прогнозу параметров прочности, включая современные публикации 2023–2025 гг. [6, 7], а также ряд других исследований в этой области [8, 21], предлагаемая работа отличается не использованием экзотической архитектуры, а совокупностью признаков научной добросовестности и инженерной воспроизводимости [22, 23]. К таким признакам относятся: использование большого массива данных, строгий и воспроизводимый препроцессинг с сохранением пайплайнов обработки, а также систематическое сравнение семейств моделей, наиболее подходящих для табличных данных [11, 12]. Подобный подход дополнительно опирается на результаты исследований, показывающих эффективность градиентного бустинга и связанных методов для табличных задач [13, 14]. В рамках работы также выполнена оптимизация ансамбля по фиксированному протоколу и проведена независимая внешняя валидация на отдельном объекте, дополненная ХАИ-контролем интерпретируемости моделей [15, 16]. Теоретические основания используемых ансамблевых методов и подходов машинного обучения широко обсуждаются в классической литературе по анализу данных [24].

Результаты

После загрузки исходного массива лабораторных данных (16660 строк) выполнены контроль качества и фильтрация, в результате чего сформирован очищенный датасет из 12596 наблюдений. На этом этапе дополнительно выполнено приведение деформационного параметра к единому уровню напряжений: рассчитан новый столбец E_{50} (пересчет на 100 кПа), причем корректное значение удалось получить для 9029 из 12596 записей, что отражает неизбежную неоднородность исходной лабораторной отчетности и различия в доступности параметров для пересчета. Для дальнейшего моделирования использовалась строгая постановка задачи восстановления параметров E_{50} , c и φ на подмножестве с полными таргетами ($N=9029$), с разбиением на Train/Validation/Test = 6319/1355/1355, что обеспечивает честное сравнение моделей и воспроизводимость оценки качества.

На тестовой выборке проведено сравнение трех базовых подходов: CatBoost, FT-Transformer и MTL-нейросеть, а также их комбинаций путем усреднения предсказаний. Сравнение выполнялось в общей постановке трехтаргетной регрессии с единым допуском $\pm 20\%$ для метрики $Acc_{\pm 20\%}$, при этом метрики рассчитывались в исходном масштабе целевых переменных после обратного преобразования. Результаты приведены в Таблице 1 и содержат агрегированные метрики по трем таргетам и/или таргет-специфичные показатели. R^2 используется как мера объясненной дисперсии, а $Acc_{\pm 20\%}$ отражает долю прогнозов, попавших в инженерный допуск $\pm 20\%$ в относительной шкале.

Таблица 1 – Одиночные модели на тестовой выборке
Table 1 – Performance of individual models on the test set

Модель	$\overline{R^2}$	$\overline{Acc_{\pm 20\%}}$, %	$\overline{R^2_{E_{50}}}$	$\overline{Acc_{\pm 20\%_{E_{50}}}}$, %	$\overline{R^2_c}$	$\overline{Acc_{\pm 20\%_c}}$, %	$\overline{R^2_\varphi}$	$\overline{Acc_{\pm 20\%_\varphi}}$, %
MTL	0,859	72,5	0,69	56,8	0,934	61,8	0,951	98,9
FT-Transformer	0,846	72,1	0,652	55,4	0,935	62,0	0,950	98,9
CatBoost	0,878	76,3	0,738	60,7	0,936	69,3	0,957	98,9

По средним метрикам на тесте CatBoost достигает $\overline{R^2} = 0,878$ и $\overline{Acc_{\pm 20\%}} = 76,3\%$, Для сравнения: MTL дает $\overline{R^2} = 0,859$, FT-Transformer – $\overline{R^2} = 0,846$; простые ансамбли улучшают качество не во всех случаях и не для всех таргетов, что указывает на различие «сильных сторон» моделей и отсутствие универсального выигрыша от простого усреднения.

Для поиска лучшего ансамбля в качестве основного критерия оптимизации использовался \overline{WAPE} (средняя взвешенная абсолютная процентная ошибка) по трем таргетам (как устойчивый к масштабу и удобный для инженерной интерпретации показатель), а R^2 и $\overline{Acc_{\pm 20\%}}$ использовались как контрольные. Результаты поиска оптимальных весов ансамбля на симплексе при дискретизации шага 0,05 приведены в Таблице 2. Таблица демонстрирует устойчивость выбора и величину выигрыша от оптимизации весов относительно одиночных моделей.

По итогам weight-search лучшая конфигурация по \overline{WAPE} – ансамбль ft_transformer+catboost с весами: $w_{ft} = 0,10$, $w_{cat} = 0,90$, $w_{mtl} = 0,00$. Ее итоговые метрики на тестовых данных: $\overline{WAPE} = 13,2\%$, $\overline{R^2} = 0,877$, $\overline{Acc_{\pm 20\%}} = 76,4\%$ (при этом CatBoost один дает $\overline{R^2} = 0,877$, $\overline{WAPE} = 13,2\%$, то есть весовой ансамбль дает небольшой, но воспроизводимый выигрыш именно по \overline{WAPE}).

Таблица 2 – Оптимизация весов ансамбля (Тор-10 по минимальному \overline{WAPE})
Table 2 – Optimization of ensemble weights (Top 10 configurations with the lowest mean \overline{WAPE})

Rank	Комбинация	Весы	\overline{WAPE} , %	$\overline{R^2}$	$\overline{Acc}_{\pm 20\%}$, %	$WAPE_{E_{50}}$, %	$WAPE_c$, %	$WAPE_\varphi$, %
1	ft_transformer+catboost	{ft: 0,10, cat: 0,90}	13,16	0,877	76,36	19,66	14,87	4,97
2	ft_transformer+catboost	{ft: 0,05, cat: 0,95}	13,17	0,877	76,31	19,62	14,90	4,96
3	mtl+ft_transformer+catboost	{mtl: 0,05, ft: 0,05, cat: 0,90}	13,17	0,877	76,29	19,65	14,88	4,97
4	ft_transformer+catboost	{ft: 0,15, cat: 0,85}	13,17	0,877	76,43	19,70	14,83	4,98
5	mtl+ft_transformer+catboost	{mtl: 0,05, ft: 0,10, cat: 0,85}	13,17	0,877	76,38	19,69	14,85	4,98
6	mtl+catboost	{mtl: 0,05, cat: 0,95}	13,17	0,877	76,41	19,62	14,93	4,97
7	catboost	{cat: 1,00}	13,17	0,877	76,33	19,60	14,96	4,97
8	mtl+ft_transformer+catboost	{mtl: 0,10, ft: 0,05, cat: 0,85}	13,18	0,878	76,21	19,69	14,88	4,98
9	mtl+catboost	{mtl: 0,10, cat: 0,90}	13,18	0,878	76,29	19,66	14,92	4,98
10	mtl+ft_transformer+catboost	{mtl: 0,05, ft: 0,15, cat: 0,80}	13,19	0,877	76,33	19,75	14,82	4,99

В Таблице 3 приведено сводное сопоставление трех одиночных моделей (MTL, FT-Transformer, CatBoost) и выбранного по результатам weight-search лучшего ансамбля (веса $w_{cat} = 0,90$, $w_{ft} = 0,10$, $w_{mtl} = 0,00$). Таблица является основанием для фиксации финальной конфигурации модели, используемой далее во внешней валидации и интерпретации.

Таблица 3 – Сравнение одиночных моделей и лучшего ансамбля моделей
Table 3 – Comparison of individual models and the best-performing ensemble

Модель	Весы	$\overline{R^2}$	\overline{WAPE} , %	$\overline{Acc}_{\pm 20\%}$, %
MTL	–	0,859	14,34	72,50
FT-Transformer	–	0,846	14,57	72,08
CatBoost	–	0,877	13,18	76,33
Best ensemble	{cat: 0,90, ft: 0,10}	0,877	13,16	76,36

Агрегированные метрики скрывают важную инженерную деталь: сложность восстановления параметров неодинакова. Для φ CatBoost показывает практически предельное качество на тесте: $R^2 = 0,957$ при $Acc_{\pm 20\%} = 98,97\%$. Такая устойчивость соответствует физическому смыслу задачи: при наличии индексов состояния и классификационных признаков угол внутреннего трения восстанавливается наиболее однозначно и с меньшим влиянием лабораторной методики по сравнению с деформационными модулями.

При прогнозировании удельного сцепления с различия между моделями выражены слабее, чем для E_{50} , а наиболее высокое значение R^2 достигается у более широких комбинаций, однако CatBoost остается конкурентоспособным: на тесте он обеспечивает $R^2 = 0,936$ и $Acc_{\pm 20\%} = 69,30\%$. При этом добавление нейросетевых

компонент может улучшать R^2 точно (в отдельных областях пространства признаков), но не приводит к гарантированному росту «инженерной» точности в допуске $\pm 20\%$.

Наиболее трудным таргетом является E_{50} : даже лучшая модель (CatBoost) дает $R^2 = 0,738$ и $Acc_{\pm 20\%} = 60,74\%$. Это ожидаемо для деформационных модулей: E_{50} значительно чувствительнее к протоколу испытания, состоянию напряжений, структуре и неоднородности образцов, а также к артефактам пересчета к единому уровню напряжений. Поэтому в интерпретации результатов именно по E_{50} следует рассматривать не только R^2 , но и метрики относительной ошибки и обязательно подтверждать переносимость на независимых данных.

Ключевым подтверждением практической применимости методики является внешняя проверка на независимом объекте, не участвовавшем в обучении и подборе весов ансамбля ($N = 54$ пробы). Для внешнего теста использован взвешенный ансамбль FT-Transformer (0,1) + CatBoost (0,9), выбранный как прагматичный компромисс между устойчивостью бустинга на табличных данных и способностью трансформера учитывать сложные взаимодействия признаков.

Валидационный объект характеризуется преобладанием связных грунтов при малых и средних глубинах отбора 0,3–11,55 м (среднее $\sim 4,59$ м). По нормативной типизации доминируют суглинки ($\sim 55,6\%$) и супеси ($\sim 22,2\%$), доля песчаных грунтов составляет $\sim 22,2\%$ (пески средней крупности $\sim 11,1\%$ и мелкие $\sim 11,1\%$). В инженерно-литологическом описании это согласуется с протоколами: наиболее частые разности – «суглинок тугопластичный/твердый/полутвердый/мягкопластичный» и «супесь твердая/пластичная», а пески представлены неоднородными средней плотности. Гранулометрически по классификации PLAXIS преобладают классы very fine ($\sim 77,8\%$), при меньших долях fine ($\sim 11,1\%$) и medium ($\sim 11,1\%$), что задает мелкодисперсный режим интерпретации модели (существенная роль индексов состояния и плотностных характеристик). Физическое состояние в среднем соответствует относительно плотным/умеренно деформируемым грунтам: коэффициент пористости $e \sim 0,47$ (0,31–0,64), индекс текучести по доступным значениям $I_l \sim 0,20$ ($-0,24 \dots 0,67$; доля $I_l > 0$ среди ненулевых измерений $\sim 71\%$), индекс пластичности $I_p \sim 8,36\%$ (5,5–11,4), Прочностные параметры имеют «инженерно устойчивый» уровень: $c \sim 39$ кПа (1–90 кПа) и $\varphi \sim 28,25^\circ$ ($17,6 \text{--} 36,7^\circ$). E_{50} лежит в диапазоне $\sim 10,9 \text{--} 26,6$ МПа (в среднем $\sim 18,49$ МПа, медиана $\sim 17,61$ МПа), то есть распределение E_{50} на объекте сравнительно собранное (без экстремально тяжелых хвостов), что принципиально влияет на профиль ошибок во внешней валидации. Распределение реальных и предсказанных значений E_{50} , c и φ представлено на Рисунках 2, 3.

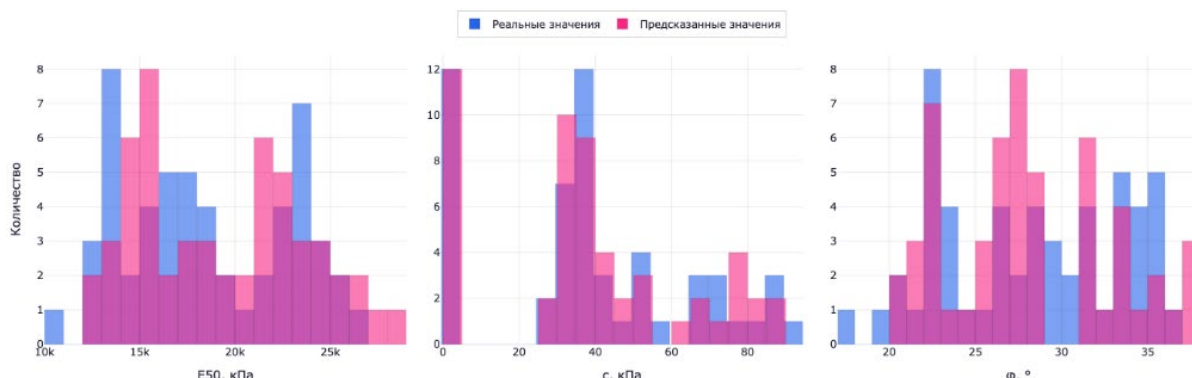


Рисунок 2 – Распределение предсказанных и реальных значений значений в валидационном объекте

Figure 2 – Distribution of predicted and observed values in the validation site

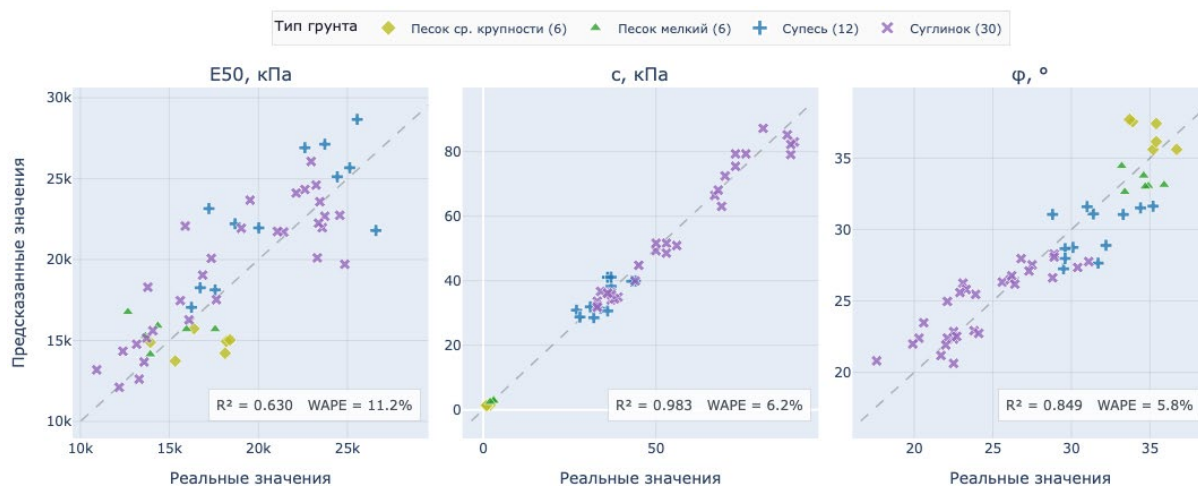


Рисунок 3 – Распределение предсказанных и реальных значений значений в валидационном объекте

Figure 3 – Distribution of predicted and observed values in the validation site

На объекте получены усредненные метрики $\overline{R^2} = 0,821$, $\overline{WAPE} = 7,75\%$, $\overline{Acc_{\pm 20\%}} = 92,59\%$. Детализация по параметрам приведена в Таблице 4.

Таблица 4 – Метрики валидационного объекта

Table 4 – Performance metrics for the validation site

Таргет	R^2	$RMSE$	$WAPE, \%$	$Acc_{\pm 20\%}, \%$
E_{50}	0,6297	2600,11	11,24	85,2
c	0,9833	3,4	6,21	92,6
ϕ	0,8493	2,01	5,81	100,0

На внешнем объекте наиболее надежно восстанавливается сцепление c (высокий R^2 при малой относительной ошибке), что указывает на хорошо прогнозируемый моделью режим связных грунтов и достаточность индексов состояния/плотности/классификационных признаков для реконструкции c в масштабе инженерных допусков. Для угла внутреннего трения ϕ достигнута строгая допусковая согласованность ($Acc_{\pm 20\%} = 100\%$) при высоком R^2 , то есть модель стабильно воспроизводит как вариативность, так и абсолютный уровень ϕ на объекте. Наиболее сложным параметром остается E_{50} : хотя доля попаданий в инженерный допуск на данном объекте высокая (85,2%), объясненная дисперсия заметно ниже ($R^2 \sim 0,63$), что типично для деформационных модулей из-за их повышенной чувствительности к структурным факторам и нюансам лабораторного протокола (а также к неизбежным объектным отличиям среды измерения по сравнению с обучающим массивом). Практически это означает, что ϕ и c на данном внешнем объекте допустимо использовать как восстанавливаемые параметры в допусковой постановке, тогда как E_{50} требует более осторожного применения: контроля применимости по типам/состояниям грунта и, при необходимости, локальной калибровки по небольшому числу опорных испытаний на конкретной площадке (особенно при переходе между песчаными и связными грунтами).

Для обеспечения проверяемости модели выполнена интерпретация предсказаний на валидационном объекте: для CatBoost рассчитаны SHAP-значения, для FT-Transformer получена важность признаков методом перестановок, после чего сформирована интерпретация ансамбля по линейному правилу $SHAP_{ens} = 0,9 \times$

$\times SHAP_{cat} + 0,1 \cdot SHAP_{ft}$. Дополнительно проверена аддитивность объяснений (сумма вкладов признаков + базовое значение воспроизводит предсказание), что является обязательным условием корректного использования SHAP в регрессионной постановке для инженерных выводов.

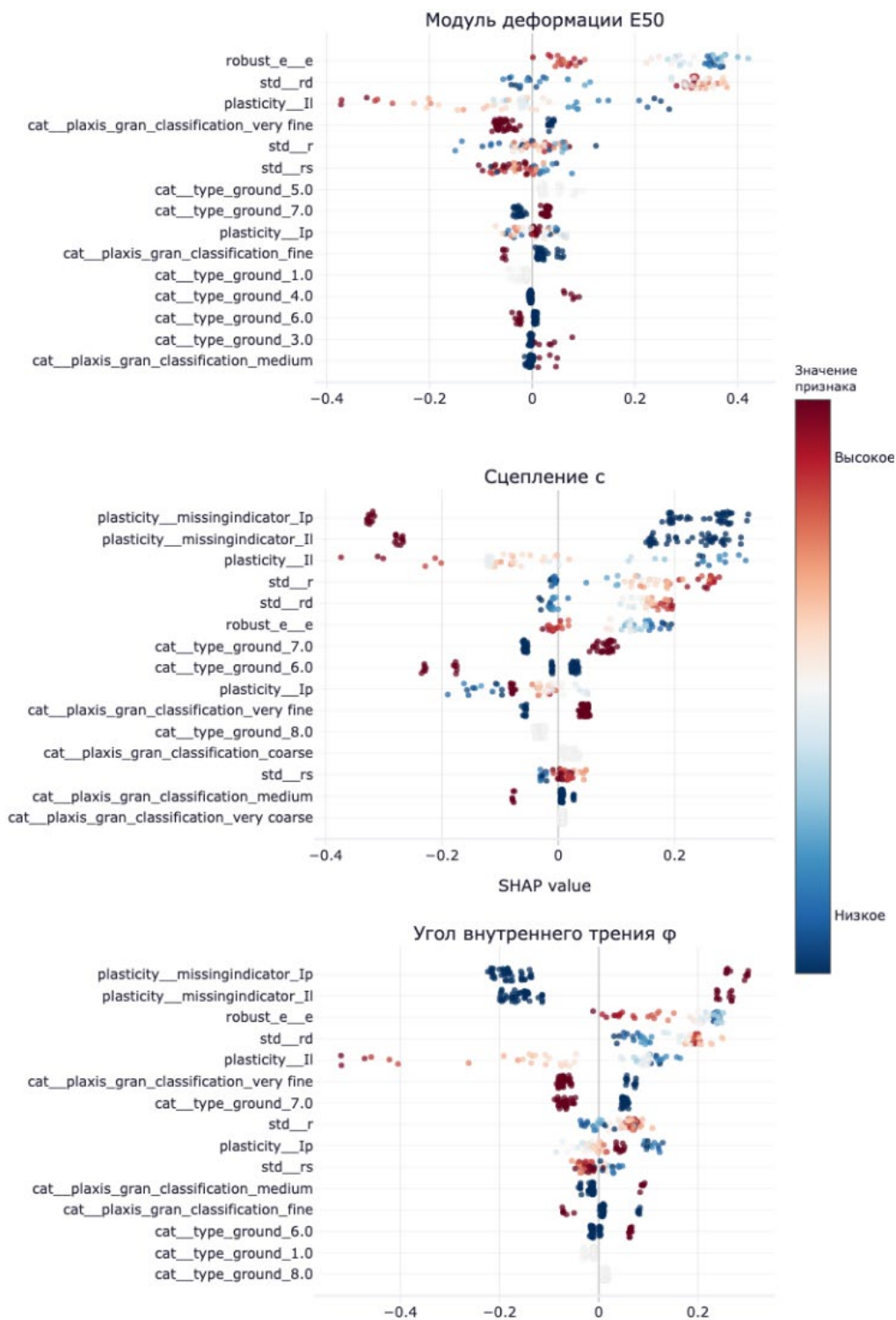


Рисунок 4 – SHAP summary plot ансамбля на валидационном объекте: глобальная важность признаков для таргетов E_{50} , c и φ

Figure 4 – SHAP summary plot for the ensemble on the validation site: global feature importance across the targets E_{50} , c and φ

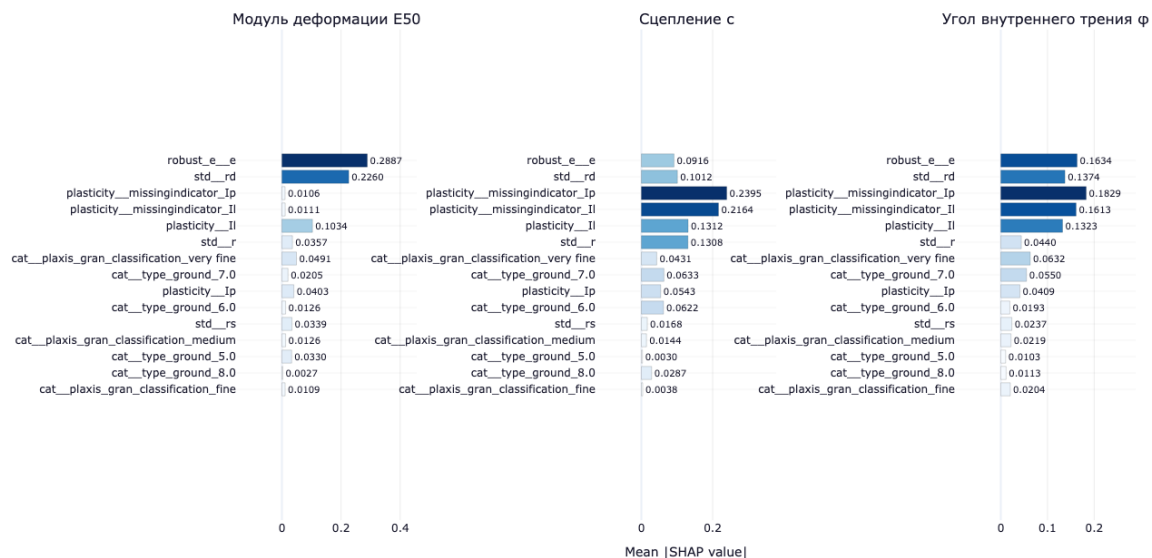


Рисунок 5 – SHAP bar plot ансамбля на валидационном объекте: средняя абсолютная важность признаков (mean | SHAP |)

Figure 5 – SHAP bar plot for the ensemble on the validation site: mean absolute feature importance values (mean | SHAP |)

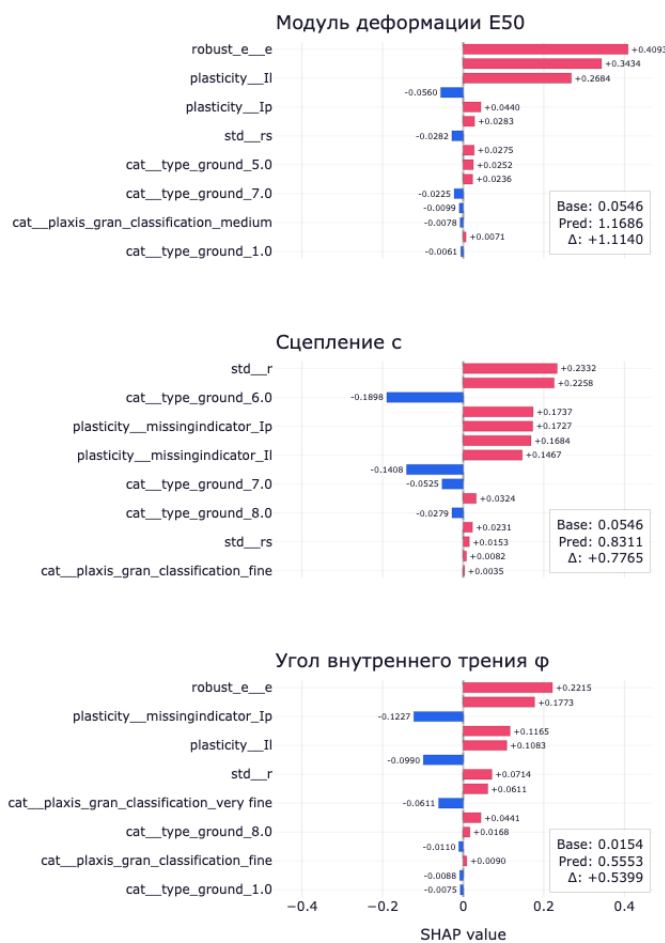


Рисунок 6 – Локальная интерпретация ансамбля для одной пробы (waterfall-объяснение): вклады признаков в предсказания E_{50} , c и ϕ ; base value и Δ приведены в легенде

Figure 6 – Local interpretation of the ensemble for a single sample (waterfall plot): feature contributions to the predictions of E_{50} , c and ϕ , with the base value and Δ reported in the legend

В сводной важности признаков на ансамбле доминируют факторы с прямым физическим смыслом: коэффициент пористости e , сухая плотность r_d , индексы пластичности/текучести (I_l , I_p) и индикаторы их пропусков, а также категориальные маркеры класса гранулометрии PLAXIS (особенно very fine) и типа грунта. Такой профиль важности согласуется с геотехнической логикой: именно состояние (пористость/плотность/пластичность) и типизация материала определяют прочность и деформируемость. Глобальные SHAP-графики и локальные waterfall-объяснения (Рисунки 4–6) показывают, что модель не является «черным ящиком» в инженерном смысле: вклад ключевых признаков трассируется, а выводы допускают экспертную проверку перед использованием в расчетных постановках.

Обсуждение и заключение

В статье решена прикладная задача инженерной геологии: восстановление ключевых механических параметров грунтов E_{50} , c и φ по набору физических и классификационных признаков в условиях типичной неполноты лабораторных протоколов. На массиве 16660 записей сформирован воспроизводимый пайплайн контроля качества и предобработки: после фильтрации получено 12596 наблюдений, а для моделирования с полными таргетами использовано подмножество $N = 9029$ с фиксированным разбиением Train/Validation/Test = 70/15/15. Дополнительно обеспечена сопоставимость деформационного таргета за счёт формирования E_{50} , пересчитанного на опорное давление (пересчёт выполнен для 9029 записей), что снижает методическую неоднородность исходных данных и задает корректную постановку задачи.

Проведено сравнение трех семейств моделей для табличных данных CatBoost, FT-Transformer и многозадачной нейросети (MTL) и показано, что на тестовой выборке лучшая одиночная модель CatBoost достигает $\overline{R^2} = 0,878$ и $\overline{Acc_{\pm 20\%}} = 76,3\%$ по агрегированным метрикам, превосходя MTL и FT-Transformer. При этом выявлена неодинаковая предсказуемость таргетов: φ восстанавливается наиболее устойчиво (для CatBoost $R^2 = 0,957$, $Acc_{\pm 20\%} = 98,9\%$), c – на высоком уровне ($R^2 = 0,936$, $Acc_{\pm 20\%} = 69,3\%$), тогда как E_{50} остается наиболее сложным параметром ($R^2 = 0,738$, $Acc_{\pm 20\%} = 60,7\%$), что согласуется с физической чувствительностью модулей деформации к протоколу испытаний и напряженному состоянию.

Для повышения устойчивости выполнено ансамблирование и дискретный подбор весов на симплексе (шаг 0,05; 231 комбинация). Оптимальным по $\overline{WAP\overline{E}}$ оказался двухкомпонентный ансамбль FT-Transformer + CatBoost с весами 0,10/0,90, обеспечивший на тестовой выборке $\overline{WAP\overline{E}} = 13,16\%$, $\overline{R^2} = 0,877$, $\overline{Acc_{\pm 20\%}} = 76,36\%$. Небольшой прирост относительно лучшей одиночной модели интерпретируется как ожидаемый результат при доминировании сильного базового алгоритма и служит аргументом в пользу использования ансамбля как более устойчивого решения.

Критически важным результатом является внешняя проверка на независимом объекте ($N = 54$), не участвовавшем в обучении и подборе весов: достигнуты $\overline{R^2} = 0,821$, $\overline{WAP\overline{E}} = 7,75\%$, $\overline{Acc_{\pm 20\%}} = 92,59\%$, причем φ демонстрирует строгую допусковую согласованность ($Acc_{\pm 20\%} = 100\%$), а E_{50} сохраняет инженерную применимость в допусковой постановке при ожидаемо меньшей объясненной дисперсии (порядка $R^2 = 0,63$). Это позволяет обоснованно рассматривать метод как инструмент «доставки» параметров при дефиците испытаний, с более осторожным применением для E_{50} (контроль области применимости и при необходимости локальная калибровка).

Проверяемость прогнозов обеспечена интерпретацией (SHAP для CatBoost, permutation importance для FT-Transformer и взвешенная агрегация для ансамбля): ведущие факторы (ухая плотность r_d , коэффициент пористости e , индекс пластичности I_p и индекс текучести I_t , а также типизация/классификация грунта и индикаторы пропусков) физически согласуются с механизмами прочности и деформируемости. В совокупности результаты подтверждают, что сочетание большого репрезентативного датасета, строгого протокола оценки, современных моделей для tabular-данных, ансамблирования и внешней валидации формирует методически устойчивую основу для практического внедрения в контур инженерно-геологического моделирования и параметризации расчетных моделей.

СПИСОК ИСТОЧНИКОВ / REFERENCES

1. Schanz T., Vermeer P.A., Bonnier P.G. The Hardening Soil Model: Formulation and Verification. In: *Beyond 2000 in Computational Geotechnics: 10 years of PLAXIS International: Proceedings of the International Symposium beyond 2000 in Computational Geotechnics, 18–20 March 1999, Amsterdam, The Netherlands*. Rotterdam: A. A. Balkema; 1999. P. 281–296.
2. Zhao T., Shen F., Xu L. Review and comparison of machine learning methods in developing optimal models for predicting geotechnical properties with consideration of feature selection. *Soils and Foundations*. 2024;64(6). <https://doi.org/10.1016/j.sandf.2024.101523>
3. Yuan B., Choo Ch.S., Yeo L.Y., et al. Physics-informed machine learning in geotechnical engineering: a direction paper. *Geomechanics and Geoengineering*. 2025;20(5):1128–1159. <https://doi.org/10.1080/17486025.2025.2502029>
4. Alzighaibi W.A., Daghistani F. Machine Learning in Geotechnical Engineering: A State-of-the-Art Review of Research Progress and Barriers to Real-World Implementation. In: *2025 IEEE International Conference on Emerging Trends in Engineering and Computing (ETECOM), 29–30 October 2025, Riffa, Bahrain*. IEEE; 2025. <https://doi.org/10.1109/ETECOM66111.2025.11319124>
5. Bozorgzadeh N., Feng Y. Evaluation structures for machine learning models in geotechnical engineering. *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*. 2024;18(1):52–59. <https://doi.org/10.1080/17499518.2024.2313485>
6. Lei D., Zhang Y., Lu Zh., et al. A machine learning framework for predicting shear strength properties of rock materials. *Scientific Reports*. 2025;15(1). <https://doi.org/10.1038/s41598-025-91436-8>
7. Ahmad M., Zubi M.A., Almujiabah H., et al. Improved prediction of soil shear strength using machine learning algorithms: interpretability analysis using SHapley Additive exPlanations. *Frontiers in Earth Science*. 2025;13. <https://doi.org/10.3389/feart.2025.1542291>
8. Yuke W., Shuang F., Yanhui Zh., Bei Zh., et al. A data-driven model for predicting shear strength indexes of normally consolidated soils. *Chinese Journal of Geotechnical Engineering*. 2023;45(S2):183–188. <https://doi.org/10.11779/CJGE2023S20025>
9. Lundberg S., Lee S.-I. *A Unified Approach to Interpreting Model Predictions*. arXiv. URL: <https://doi.org/10.48550/arXiv.1705.07874> [Accessed 10th March 2026].
10. Jas K., Dodagoudar G.R. Explainable machine learning model for liquefaction potential assessment of soils using XGBoost-SHAP. *Soil Dynamics and Earthquake Engineering*. 2023;165. <https://doi.org/10.1016/j.soildyn.2022.107662>

11. Prokhorenkova L., Gusev G., Vorobev A., et al. *CatBoost: unbiased boosting with categorical features*. arXiv. URL: <https://doi.org/10.48550/arXiv.1706.09516> [Accessed 10th March 2026].
12. Gorishniy Y., Rubachev I., Khrulkov V., Babenko A. *Revisiting Deep Learning Models for Tabular Data*. arXiv. URL: <https://doi.org/10.48550/arXiv.2106.11959> [Accessed 10th March 2026].
13. Somepalli G., Goldblum M., Schwarzschild A., Bruss C.B., Goldstein T. *SAINTE: Improved Neural Networks for Tabular Data via Row Attention and Contrastive Pre-Training*. arXiv. URL: <https://doi.org/10.48550/arXiv.2106.01342> [Accessed 10th March 2026].
14. Hollmann N., Müller S., Eggensperger K., Hutter F. *TabPFN: A Transformer That Solves Small Tabular Classification Problems in a Second*. arXiv. URL: <https://doi.org/10.48550/arXiv.2207.01848> [Accessed 11th March 2026].
15. Chen T., Guestrin C. XGBoost: A Scalable Tree Boosting System. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 13–17 August 2016, San Francisco, CA, USA*. New York: ACM; 2016. P. 785–794. <https://doi.org/10.1145/2939672.2939785>
16. Ke G., Meng Q., Finley Th., et al. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In: *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, 04–09 December 2017, Long Beach, CA, USA*. 2017. P. 3146–3154.
17. Friedman J.H. Greedy Function Approximation: A Gradient Boosting Machine. *Annals of Statistics*. 2001;29(5):1189–1232. <https://doi.org/10.1214/aos/1013203451>
18. Breiman L. Stacked Regressions. *Machine Learning*. 1996;24:49–64. <https://doi.org/10.1007/BF00117832>
19. Caruana R. Multitask Learning. *Machine Learning*. 1997;28:41–75. <https://doi.org/10.1023/A:1007379606734>
20. Salih A.M., Raisi-Estabragh Z., Galazzo I.B., et al. A Perspective on Explainable Artificial Intelligence Methods: SHAP and LIME. *Advanced Intelligent Systems*. 2024. <https://doi.org/10.1002/aisy.202400304>
21. Ankah M.L.Y., Adjei-Yeboah Sh., Ziggah Y.Y., Asare E.N. Advanced hybrid machine learning models with explainable AI for predicting residual friction angle in clay soils. *Scientific Reports*. 2025;15. <https://doi.org/10.1038/s41598-025-05962-6>
22. Wu P., Chen J., Huang J., et al. Interpretable machine learning approach for predicting lunar soil shear strength parameters based on data imputation techniques. *Advances in Space Research*. 2025;76(2):1091–1115. <https://doi.org/10.1016/j.asr.2025.04.071>
23. Molnar Ch. *Interpretable Machine Learning*. Lulu.com; 2020. 320 p.
24. Arik S.Ö., Pfister T. TabNet: Attentive Interpretable Tabular Learning. *Proceedings of the AAAI Conference on Artificial Intelligence*. 2021;35(8):6679–6687. <https://doi.org/10.1609/aaai.v35i8.16826>

ИНФОРМАЦИЯ ОБ АВТОРЕ / INFORMATION ABOUT THE AUTHOR

Тишин Никита Романович, аспирант, **Nikita R. Tishin**, Postgraduate, Bauman Московский государственный технический университет им. Н.Э. Баумана, Москва, the Russian Federation.
Московская Федерация.
e-mail: tnick1502@mail.ru
ORCID: [0009-0008-2484-591X](https://orcid.org/0009-0008-2484-591X)

*Статья поступила в редакцию 25.02.2026; одобрена после рецензирования 07.04.2026;
принята к публикации 30.04.2026.*

*The article was submitted 25.02.2026; approved after reviewing 07.04.2026;
accepted for publication 30.04.2026.*