

A large open multi-source soil compaction corpus and pretrained tabular estimators from index properties

Sompote Youwai

sompote.you@kmutt.ac.th

King Mongkut's University of Technology Thonburi

Chana Phutthananon

King Mongkut's University of Technology Thonburi

Warat Kongkitkul

King Mongkut's University of Technology Thonburi

Research Article

Keywords: maximum dry density, optimum moisture content, degree of saturation at optimum, compaction energy level, tabular foundation model, in-context learning

Posted Date: October 15th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11115086/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

A large open multi-source soil compaction corpus and pretrained tabular estimators from index properties

Sompote Youwai^{1*}, Chana Phutthananon¹ and Warat Kongkitkul¹

^{1*}AI Research Group, Department of Civil Engineering, King
Mongkut's University of Technology Thonburi, 126 Pracha Uthit Road,
Bangkok, 10140, Thailand.

*Corresponding author(s). E-mail(s): sompote.you@kmutt.ac.th;

Abstract

Maximum dry density and optimum moisture content require a full Proctor test, whereas published correlations generally rely on small, single-laboratory datasets. We assemble an open corpus of 4,878 laboratory compaction tests from six deposits and 48 studies, with per-record source provenance. The corpus spans four Proctor energy levels and fines contents from 0.1 to 100 %. Records were audited against their stated methods without inferring unreported energy levels. Zero-air-voids screening removed 11.8 % of harmonised records; the rejection rate varied from 19.8 % to 8.2 % across assumed specific gravities. Frozen tabular foundation models predicted both parameters from consistency limits, gradation, compactive energy, and specific gravity without corpus-specific hyperparameter tuning. Because every source occurs in both training and held-out folds, random five-fold scores measure within-corpus interpolation; generalisation to unseen sources remains untested. TabPFN 3.5 achieved density R^2 of 0.817 and mean absolute error of 0.067 Mg/m³, with corresponding water-content values of 0.801 and 1.67 percentage points. Ablation showed that removing gradation reduced density R^2 by 0.229, while removing consistency limits reduced it by 0.080. For water content, the corresponding losses were 0.112 and 0.109. Symbolic regression produced a dimensionless density relation with R^2 of 0.733, while saturation at optimum could not be predicted better than a constant. The corpus provides a benchmark dataset for future geotechnical research to develop, calibrate, and evaluate new machine-learning models for soil compaction.

Keywords: maximum dry density, optimum moisture content, degree of saturation at optimum, compaction energy level, tabular foundation model, in-context learning

1 Introduction

1.1 Compaction parameters and the cost of determining them

Compaction control governs the strength, stiffness, compressibility and permeability of every engineered fill, and two parameters specify it. The maximum dry density (MDD) is the highest dry density reached under a stated compactive effort, and the optimum moisture content (OMC) is the water content at which that maximum occurs. Both come from the test of Proctor (1933), standardised as ASTM D698 and D1557 and as AASHTO T99 and T180. It compacts four to six specimens at different water contents, each oven-dried for 24 h, so one determination takes about a working day. Specifications require one for every change of material, so the number of tests follows the variability of the deposits rather than the volume of fill placed.

The optimum matters in its own right because it governs soil fabric. Lambe (1958) showed that a clay compacted dry of optimum develops a flocculated particle arrangement, and the same soil compacted wet of optimum a dispersed one. Seed and Chan (1959) showed that the resulting differences in strength, volume change and permeability persist at equal dry density. The compaction water content also sets the soil-water potential of the fill, and with it the susceptibility to swelling or collapse on wetting (Mitchell and Soga 2005). Daniel and Benson (1990) accordingly proposed that acceptance be defined as a zone in the dry density–water content plane rather than as independent bounds, which makes the position of the optimum a design quantity. Continuous- and intelligent-compaction systems now report a mechanistic response across an entire lift (Baek et al. 2025), and the laboratory optimum remains the reference against which those measurements are interpreted.

The optimum is bounded above by the zero-air-voids (ZAV) line, along which the remaining voids are entirely water-filled, and no compactive effort can produce a state above it. That condition is a criterion of physical admissibility, used below both as the screen applied to every harmonised record (Section 3.4) and as the consistency check applied to every predicted pair (Section 6.1). Raising the compactive effort raises MDD and reduces OMC. The peaks obtained at different energy levels trace a locus approximately parallel to the ZAV line, because for a given soil the degree of saturation at the optimum is nearly invariant (Li and Sego 2000; Horpibulsuk et al. 2008, 2009).

1.2 Previous work and the gaps it leaves

That cost has long motivated estimating the two parameters from index properties already measured on the same sample. The correlations proposed for the purpose divide into two groups. The first regresses on a single consistency limit, following Sridharan and Nagaraj (2005), who argued that the plastic limit governs the position of the peak (Pandian et al. 1997; Gurtug and Sridharan 2004; Nagaraj et al. 2015; Sharma and Sridharan 2016; Vinod and Sridharan 2016), with Blotz et al. (1998) the standard reference for estimating both parameters across compactive efforts from the liquid limit alone. The second applies regression or machine learning to a wider set of

index properties, from the artificial neural networks of Günaydın (2009) to recent tree-ensemble, deep-learning and hybrid formulations (Khatti and Grover 2023; Rehman et al. 2025), reviewed by Verma and Sivapullaiah (2019). Reported coefficients of determination are frequently above 0.90.

The datasets behind those figures are of two kinds, and neither is a database in the sense used here. The first is the set a study assembles for its own paper, typically one to four hundred specimens, rarely deposited, and confined to plastic material wherever the correlation is keyed on a consistency limit. The largest is that of Rehman et al. (2025), who model 1,001 observations across compactive energy levels with gradient-boosted trees; it is the closest antecedent of this study and the point from which it departs, but its scope is cohesive soils. The second kind is the deposited dataset, which Section 3 enumerates in full. The retained deposits hold between 7 and 431 records, each from one laboratory at a single compactive energy. Beside them sits the Long-Term Pavement Performance release, large but a single programme following one protocol at standard effort. No public source holds the individual points of a compaction curve, and none holds a field density.

Six gaps follow from that literature, and they are the gaps this paper addresses.

1. *Scale and scope.* The correlations rest on one to four hundred specimens and the largest compilation is restricted to cohesive soils, so no correlation has been tested on the granular fills a project also meets.
2. *A single compactive effort.* Almost every dataset fixes the effort, so the reported correlations hold only at the standard the study used.
3. *Provenance discarded.* Datasets are pooled without retaining the origin of each record, so no study can score a model on a laboratory it has not seen, which is the question an engineer meeting a new borrow material asks.
4. *Accuracy reported as a single confounded figure.* Weights are fitted on the study’s own records, hyperparameters are searched on those same records, and missing entries are filled first. A reported figure is therefore the joint product of the architecture, the search effort and the imputation rule. Two studies reporting 0.92 and 0.88 may differ in the model, in how long the search ran, or in nothing but how a blank was filled.
5. *Absent measurements treated as missing data.* A non-plastic soil has no liquid or plastic limit, yet standard practice imputes one, supplying the model with a fabricated measurement.
6. *Physical admissibility unchecked.* MDD and OMC are predicted independently, so a predicted pair may imply a saturation above unity, and no study reports how often that occurs.

A class of model that addresses the fourth and fifth gaps has become available since most of that literature was written. A tabular foundation model is pre-trained once on synthetic data and then frozen. It predicts by a forward pass in which the training records are supplied as context, so nothing is fitted on the data the model is judged on (Müller et al. 2022; Hollmann et al. 2025; Grinsztajn et al. 2025; Qu et al. 2025). Section 2 develops the mechanism.

In-context learning has reached geotechnical engineering, but on other problems. Saito et al. (2026a, 2026b) applied TabPFN to site characterization at an offshore airport, and Schmidinger et al. (2026) combined it with kriging-derived features for soil mapping. Each works inside a single site or field, where position carries the information and the nearest measurement is the strongest predictor available. A compaction corpus has no such structure: its records come from different laboratories, decades and continents, so the mapping must run from index properties alone and hold across sources. To the authors’ knowledge in-context learning has not been applied to compaction parameters, nor to a geotechnical corpus assembled from independent sources on this scale.

1.3 Objectives and scope

The estimation problem is stated in full in Section 5.1. One routine classification suite on a disturbed sample gives the plastic limit, the plasticity index, the fines content, the sand content and the specific gravity, and the compaction standard is also known. The task is to estimate MDD and OMC from those inputs, and no quantity derived from the compaction curve itself is used as a predictor. This study has three objectives, which address the six gaps of Section 1.2 in turn.

1. To establish a corpus of laboratory compaction tests, to verify it, and to release it for open access, which addresses gaps 1 to 3. The corpus holds 4,878 tests from six deposits and 48 open-access studies, spans four Proctor energy levels, crosses the coarse–fine boundary, and is an order of magnitude larger than the datasets on which the published correlations rest. Every record carries the compactive energy its source states and the identity of that source. Verification is explicit rather than assumed. The harmonisation conventions are audited and their consequences measured, and every record is screened against the zero-air-voids condition with the failures reported rather than removed silently. The release carries a checksum manifest, a data dictionary and a validation script that asserts every invariant claimed for the data.
2. To estimate the two parameters by in-context learning and by symbolic regression, which addresses gaps 4 and 5. The original comparison applies two pretrained tabular foundation models with no fitting on these records, differing in architecture and in how each treats an absent cell, and compares them against a gradient-boosted ensemble and a perceptron on one fold assignment. A dated extension evaluates two later foundation-model releases on the identical partition. Symbolic regression is then run on dimensionless variables, so that the resulting expression is unit-free and invertible through the phase relation. Every predicted pair is checked against the zero-air-voids bound, which addresses gap 6, and the symbolic route is the only one that can be made to satisfy that bound by construction.
3. To quantify the contribution of each input, so that a testing programme can be told which measurements it must include. Two analyses are reported: an ablation that withdraws each input group in turn and measures the resulting loss in accuracy, and a Shapley decomposition that attributes an individual prediction to the inputs which produced it.

The paper follows that order. Section 2 sets out what a pretrained tabular model is and why this corpus calls for one. Section 3 gives the search, the retention rules, the harmonisation conventions and the admissibility screen. Section 4 describes the corpus that survives them. Section 5 states the estimation problem, the learners compared and the scoring design. Section 6 reports the accuracy, the input ablation, the Shapley attributions and the closed form. Section 7 sets those results against the published correlations and states what the corpus cannot support. Section 8 concludes.

2 Tabular foundation models and in-context learning

The estimators used here are pretrained, and that word carries a specific meaning which the geotechnical literature has not yet had cause to define. Figure 1 sets it out against the conventional alternative. A conventional model is fitted on the data it is judged on. The gaps in the inputs are filled first, because a network or a regression needs a number in every cell. Weights are then estimated by gradient descent on the corpus, and hyperparameters are searched by scoring candidate settings on the same records. Two choices are therefore made on the data that also measures the result: how the gaps were filled, and how long the search ran.

A tabular foundation model reverses the order. The network is trained once, offline, on synthetic datasets drawn from a prior over data-generating processes, and is then frozen (Müller et al. 2022; Hollmann et al. 2025; Qu et al. 2025). It never sees this corpus during training. At prediction time the training fold and the record to be predicted are passed through the frozen network together, the fold acting as context for the query. Nothing is fitted, and the operation is a forward pass rather than an optimisation. The procedure is in-context learning: the training fold plays the part the worked examples play in a language-model prompt, and inference involves no gradient and no weight change. The analogy covers that mechanism and nothing else. These networks are trained on synthetic tables rather than on text, and neither the scale nor the generality of a language model carries over. No hyperparameter is therefore searched on this corpus, so where a pretrained model and a refitted ensemble are compared below on identical folds, the comparison is of two mappings and not of two search budgets. That is the answer to gap 4 of Section 1.2, and Section 7.1 tests it.

Two properties of compaction data, rather than any general claim about architectures, make a pretrained model the appropriate tool here.

The corpus is small. 4,878 records with six inputs is an order of magnitude above the datasets behind the published correlations and still small in machine-learning terms. A network trained from scratch on a table of that size has more freedom than the data can constrain. A pretrained model moves the burden: the prior is learned from millions of synthetic tables, and these 4,878 records are used only as context.

Nearly half the records lack an input, and the missing ones are missing by construction. The mapping has five property columns: the plastic limit, the plasticity index, the fines content, the sand content and the specific gravity. The liquid limit is not among them, being withheld under Section 5.1, and the specific gravity is never empty, a substituted value and a flag being entered where a source reports none. 2,295 of the 24,390 cells are therefore empty, 9.4 %, and they are 2,029 absent

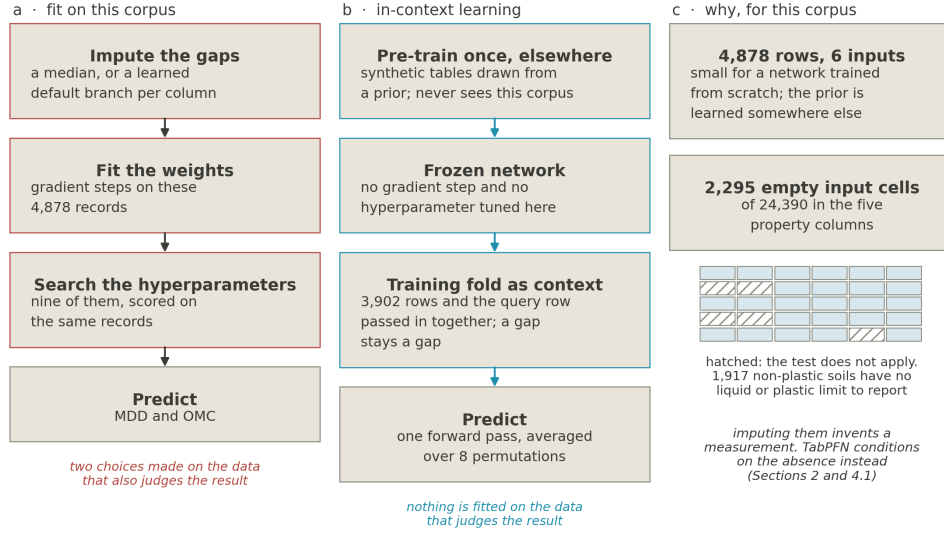


Fig. 1 In-context learning against the conventional route: (a) a model fitted on the corpus it is judged on; (b) a pretrained network, frozen, with the training fold supplied as context; (c) the two properties of this corpus that make the second route the appropriate one.

plastic limits and 266 absent sand contents (Table 4). 2,279 records are short of at least one. Those cells are not incomplete measurements. The liquid and plastic limits of a clean sand do not exist, and a plasticity index of zero already records the fact. Neither that absence nor a missing sand fraction is a number waiting to be recovered.

The two model families part company here. A regression, a support vector machine or a neural network needs a number in every cell, so the gap must be filled before the model sees it, and whatever is put there becomes a measurement. A column median substituted for the plastic limit of a sand is a plastic limit of about 20 % on a soil that has none, indistinguishable from the 20 % of a real silt. The remedy in the literature is a was-missing indicator beside each imputed column, which asks the model to learn what the corpus already records.

TabPFN instead takes the absence itself as the input, having been pre-trained on synthetic tables that contain gaps, so an absent cell is a state it has already learned to condition on. TabICL substitutes the column mean, so the two in-context models are not a single treatment of the gap and Section 6.1 reports them separately. Section 4.1 gives the encoding used for every fitted learner and what it costs, which under the random design is 0.0002 in R^2 for density and 0.0006 for water content.

The cost is computational rather than statistical: the whole training fold is re-supplied for every prediction, so inference takes minutes where the ensemble takes seconds.

3 Assembling and screening the corpus

The corpus was built in four steps. Two channels were searched and their candidates screened on retention conditions (Section 3.1). The surviving sources were audited one by one (Section 3.2). Three columns that do not mean the same thing in every source were harmonised (Section 3.3), and every remaining record was tested for physical admissibility (Section 3.4). Each step is reported with the number of records it removed.

3.1 Search and retention

Two channels were searched in August 2026, following PRISMA 2020 (Page et al. 2021) and PRISMA-S (Rethlefsen et al. 2021), with the result of each in Table 1. Deposited datasets were sought through the DataCite REST interface (api.datacite.org), which indexes Zenodo, figshare, Dryad, Mendeley Data and the Open Science Framework among others (Brase 2009). DataCite registers most research datasets, so that search approximates a census rather than a sample. Open-access articles were located through the OpenAlex interface (api.openalex.org), an open index of scholarly works (Priem et al. 2022). Both channels reach only open-access material and query in English, so the corpus samples the openly available literature rather than compaction testing at large.

A deposited dataset was retained only if it met five conditions. It needed an open licence permitting redistribution, both MDD and OMC on every record, at least one index property on every record, provenance identifiable at record level, and targets that are the peak of a laboratory compaction test. The last sets the scope of everything that follows, since it puts field densities, nuclear-gauge readings and optima back-calculated from field data outside the corpus.

Table 1 reports both channels. Every cell carries its own unit, because the two streams are counted in different things. The two indented DataCite rows are exclusions within the 618 candidates and retain nothing. The OpenAlex stream retains nothing as published, and the row beneath it is that same stream after the source PDFs were re-read, not a further set of studies. The LTTP programme is the sixth deposit. The retained records sum to the 4,878 of the corpus. The article stream was screened on completeness, not on quality.

Table 1 Search and retention. Both channels were run in August 2026.

Channel	Screened	Retained	Reason for exclusion
DataCite deposits	618 candidates	5 deposits, 961 records	licence, missing target, or no record-level provenance
— of which, integrity	2 deposits, 502 records	0 deposits, 0 records	near-duplicate records perturbed in the fourth to sixth decimal
— of which, curve points	all candidates	0 deposits, 0 records	no repository holds individual compaction-curve points
OpenAlex articles and theses, as published	11,119 documents, 4,370 parsed, 2,746 records	0 studies, 0 records	finer on 3.7 %, sand on 4.8 %, energy on none
— the same stream, after re-reading the PDFs	453 source PDFs, 2,301 cells filled	48 studies, 100 records	the remainder still lack a gate field
LTPP Standard Data Release 39	1 programme	1 programme, 3,817 records	—
Total retained		6 deposits + 48 studies, 4,878 records	

Two exclusions were decisions rather than filters. Two figshare deposits of 502 records met all five conditions but were excluded on integrity, each following every apparently genuine specimen with about nine near-copies perturbed in the fourth to sixth decimal place. The article and thesis stream was screened rather than admitted whole. Extraction recovers both targets on all 2,026 records but fines on 3.7 %, sand on 4.8 % and compactive energy on none, and admitting it in that state costs 0.283 in R^2 for density.

The stream was completed rather than dropped. Re-reading the source PDF behind each of its 453 references filled 2,301 cells. The compaction standard matters most, being a sentence in a methods section rather than a table: 256 papers name exactly one method, and a paper naming two is left alone, so no energy is inferred. The stream contributes 100 records over 48 references, 4.9 % of what it holds.

3.2 The retained sources

Table 2 lists the sources. Every retained record is the peak of a laboratory compaction curve, and no source reports a field density, so every statement in this paper about MDD, OMC and the saturation at optimum concerns laboratory compaction. The origin of each record was audited separately for each source. For the deposits that audit is the deposit’s own documentation, which states the method and the units. Two sources needed a decision beyond that, and both are set out below.

Table 1 The retained sources. n is records retained after screening, the fines and OMC columns are medians over those records, and G_s meas. counts records reporting a measured specific gravity rather than the substituted median.

Source	n	Groups	Laboratory method	Fines (%)	OMC (%)	G_s meas.
LTPP SDR 39 (FHWA 2025)	3,817	62	AASHTO T99, standard	20.7	10.0	1,647
Erden (2025), literature	431	5	standard Proctor	34.8	13.1	0
Kokcam et al. (2026), Türkiye	309	1	standard Proctor	61.0	18.1	0
Yogeshraj Urs (2026), multi-energy	175	83	ASTM D698 and D1557, four levels	60.0	15.1	175
open-access articles (Table A1)	100	48	per article, from its methods section	63.5	18.0	57
Soranzo (2024), Austria	39	12	Proctor, standard not named	85.2	18.7	39
Baghbani et al. (2026), parent soils	7	7	standard Proctor	95.0	22.0	6
Total	4,878	218		25.7	11.0	1,924

The first is LTPP, which supplies 78 % of the corpus. Its documentation describes the source table as holding standard Proctor results (Federal Highway Administration 2003), so every LTPP record is AASHTO T99 at 592.5 kJ/m³. Its density field carries no unit and spans 75 to 154, which admits two readings. Read as lb/ft³ it gives a median degree of saturation at optimum of 0.792, against 0.843 for the 1,061 records whose sources state their units. Read as kg/m³ the same numbers give 0.013, a saturation no compacted soil can reach. The lb/ft³ reading is therefore taken. Magnitude alone excludes Mg/m³ and kN/m³, whose plausible ranges are 1.2 to 2.5 and 12 to 25, and every density in the corpus is stored in Mg/m³. The second is the article row, which is not a source in the sense the others are but a pool of 48 independent studies linked only by the search that found them. Table A1 names them all and reports the long tail: the largest supplies 11 records and 27 of the 48 supply one.

Each source is named by its depositor and listed in full in the reference list, and three carry a qualification that matters later. Erden (2025) is a compilation from the published literature rather than one laboratory’s own testing, so its 431 records inherit the conventions of the papers behind them. Yogeshraj Urs (2026) is the only deposit carrying more than one compactive energy, so it supplies almost all of the multi-energy

evidence. And Baghbani et al. (2026) is a biopolymer-stabilisation study, from which only the untreated parent soils are taken, since a treated mix is not a natural soil.

The Groups column counts a finer provenance identifier: a state for LTPP, the originating reference for the two compilations, a site for the Austrian deposit and a soil for the smallest. It is released for reuse, and used once below for the cluster-robust interval of Section 6.1.

Two properties of Table 2 govern how the results should be read. LTPP dominates. Its 3,817 records are coarser and drier than every other source, at a median fines content of 20.7 % against 34.8 to 95.0 %. A pooled accuracy figure is therefore weighted towards LTPP soils, and the corpus median of 25.7 % is an LTPP figure more than a corpus one. And the non-standard efforts are concentrated, since 105 of the 124 come from Yogeshraj Urs (2026) and 19 from the article stream, so the effect of the effort cannot be fully separated from the deposit that applied it (Section 7).

3.3 Harmonisation

Three columns need not mean the same thing in every source. Each was audited record by record, and Table 3 states the convention adopted and its measured consequence. **Table 1** The three harmonisation audits. The effect column reports a measurement, not an assumption.

Column	The problem	Convention adopted	Measured effect
Particle size	ASTM D2487 divides gravel from sand at 4.75 mm; AASHTO and European systems at 2.00 mm, and most sources do not state theirs	LTPP fractions recomputed at 4.75 mm; other records carry their source’s boundary	Restating LTPP at 2.00 mm moves MDD R^2 from 0.789 to 0.794, inside the 0.005 fold-to-fold dispersion
Compactive energy	Sources name a standard, not a value	No value read from a source; energy follows from the blow-count specification of the named standard	Four levels: 355, 593, 1,347 and 2,693 kJ/m ³
Specific gravity	59.8 % of retained records report none	Reported value used where given; corpus median of measured values, 2.68, substituted and flagged otherwise	Nearly inert as an input; decisive in the screening rule (Section 3.4)

The energy convention makes the corpus multi-energy in a defensible sense. No compactive energy is inferred from the data. A record whose source does not identify its Proctor method is excluded rather than assigned a value. The four levels are reduced standard, standard, reduced modified and modified Proctor.

3.4 Physical admissibility

Admissibility was checked on every harmonised record, using quantities the record already contains. With ρ_w taken as 1 Mg/m³,

$$e = \frac{G_s}{\text{MDD}} - 1 \quad (1)$$

$$S_{opt} = \frac{\text{OMC} \cdot G_s}{e} \quad (2)$$

S_{opt} is the degree of saturation at the compaction peak. A value above unity is impossible as published; a very low value points to a quantity scaled wrongly. Records were kept where S_{opt} lies between 0.20 and 1.0, MDD between 0.8 and 2.6 Mg/m³ and OMC between 1 and 70 %. The window on S_{opt} is wide because the quantity is derived and inherits any error in an assumed specific gravity.

Screening removed 969 of the 8,242 harmonised records, or 11.8 %, and 94 % of those removals were above the zero-air-voids line. Taken at face value, about one published compaction record in nine describes a soil that cannot exist. The rate depends on how specific gravity was obtained, and that dependence sets the confidence the figure deserves. Where a measured value is reported the rate is 5.7 %, on 2,737 records, and where the median was substituted it is 14.8 %, on 5,505, but that comparison confounds two things because the two subsets are different records. Re-running the screen on all 8,242 with only the substituted value changed isolates the effect. The rejection rate falls from 19.8 % at $G_s = 2.60$ to 14.1 % at 2.65, 11.8 % at the adopted 2.68, 10.5 % at 2.70 and 8.2 % at 2.75. The finding that a material share of published compaction data is physically impossible is therefore robust across any plausible substitution. The figure of one in nine is not.

4 The corpus

4.1 Composition and the two deliberate gaps

The records surviving the screening of Section 3 form the corpus modelled here. It holds 4,878 laboratory compaction tests. A record is included when it carries a fines content, a declared compactive energy and a plasticity index. The mapping cannot do without those three fields and cannot impute them. Every other column may be absent, and Table 4 gives the composition by compactive effort and the state of each.

The inclusion rule needs stating plainly, because it decides the size of the corpus. A non-plastic soil has no liquid limit and no plastic limit. The tests do not apply to it, so the correct entry is an absent value rather than a number, and the plasticity index of zero already records the fact. Requiring the two limits instead would delete every non-plastic soil in the corpus: 1,917 records, almost all of them LTPP granular subgrade and base layers. They are kept here. That decision is worth 1,917 records, and it is concentrated at the coarse end where the corpus was previously thinnest (Section 3.4).

Table 4 gives the composition. Its lower block reports what a source failed to report, which is not the same as what the mapping receives as an empty cell. The mapping has five property columns, the plastic limit, the plasticity index, the fines content, the sand content and the specific gravity, so 24,390 cells in all. The liquid limit is listed

because the released corpus carries it, but it is withheld from the mapping and is not one of those columns. The specific gravity is never empty, a substituted value and a flag being entered instead. The empty cells are therefore the plastic limits and the sand contents alone.

Table 1 Composition of the modelling set, $n = 4,878$, by compactive effort and by incomplete column.

Compactive effort	Standard	Energy (kJ/m ³)	Records	Share (%)
Reduced standard	ASTM D698 A, 15 blows	355	29	0.6
Standard	ASTM D698 A / AASHTO T99	593	4,754	97.5
Reduced modified	ASTM D1557 A, half effort	1,347	8	0.2
Modified	ASTM D1557 A / AASHTO T180	2,693	87	1.8

Column	Not reported	Share on (%)	What absence means	How it reaches the model
Liquid limit	1,922	39.4	not applicable: the soil is non-plastic	it does not: withheld under Section 5.1, Step 4, and not a column of the mapping
Plastic limit	2,029	41.6	not applicable, on 1,917 of them	left empty; the fitted learners receive the training-fold median plus an indicator
Sand content	266	5.5	not reported: the source publishes fines but no grading curve	left empty; the same encoding
Specific gravity	2,954	60.6	not reported: corpus median 2.68 substituted and flagged	filled, so no cell is left empty: the substituted value plus a flag
Empty cells in the mapping	2,295	9.4	2,029 plastic limits and 266 sand contents	of the 24,390 cells in the five property columns

The gaps are kept rather than filled, and the reason differs by column. Requiring sand would delete every record from Yogeshraj Urs (2026) and so fix energy at a single level, which removes the variable of principal interest. The consistency limits are absent where the tests do not apply. 1,918 records report neither limit. 1,917 of those are non-plastic soils carrying a plasticity index of zero, and the remaining one is an article that published a plasticity index without either limit. A further 115 records report one limit and not the other, 111 a plastic limit without a liquid limit and 4 the reverse, and those are treated as two independent gaps rather than reconstructed from the plasticity index. The counts in Table 4 follow: $1,918 + 4 = 1,922$ liquid limits absent, and $1,918 + 111 = 2,029$ plastic limits absent.

Every learner here except TabPFN receives the training-fold median together with a was-missing indicator, and the indicator rather than the imputed number carries the information. A gradient-boosted tree can instead send a blank down a learned default branch, and under the random design the two encodings are equivalent on

both targets: they differ by 0.0002 in R^2 for density and 0.0006 for water content. The median-plus-indicator encoding is used for the gradient-boosted model throughout, because it is the encoding the perceptron and the refitted architectures also require, so one preprocessing serves every fitted learner. TabPFN takes the gaps natively.

4.2 Provenance and repeated records

Every record carries the identifier of the source it came from, and two properties of the corpus qualify what a random partition of it can show. The first is that records sharing a source share much more than a label. They share a sampling campaign, a laboratory protocol and a reporting convention, and they are often repeat tests on one pavement layer. Under a random partition every held-out record has its own source in the training fold, by construction. The second is that the corpus contains exact repetitions. 2,813 of the 4,878 records carry an input vector that occurs at least twice, leaving 3,421 distinct vectors, and under a random five-fold partition 46.7 % have their exact vector in the training fold. A non-plastic soil has two of its six inputs absent, so two soils of similar gradation at the same declared energy can present the same vector. Only two vectors span more than one source, so repetition is almost entirely within a source.

Both properties place the corpus in the class of datasets for which a random partition is known to return an optimistic score. Kaufman et al. (2012) set out the general form, in which information about the held-out records reaches the model through the way the data were assembled rather than through the inputs, and Roberts et al. (2017) showed that folds blocked on a temporal, spatial or hierarchical structure are required for a figure describing prediction rather than interpolation. Ploton et al. (2020) measured the effect on pooled ecological data, and Kapoor and Narayanan (2023) identified leakage of this kind across 294 papers as a principal cause of irreproducible results. The compaction literature has not examined the question because its datasets are usually single-source. It arises here, and Section 6.1 labels the figures accordingly.

4.3 The corpus in the compaction plane

Figure 2 shows the corpus. Panel (a) places every record on the MDD–OMC plane, marked by the compactive effort its source states and drawn with the zero-air-voids line and two constant-saturation contours at $G_s = 2.68$. Panel (b) sets the sources side by side on five quantities.

The records in panel (a) occupy a band beneath the zero-air-voids line rather than a cloud. The band is the near-invariance of the saturation at the peak seen directly: a soil reaches its maximum at an approximately fixed degree of saturation, so MDD and OMC are two coordinates of one point rather than two independent quantities. The band is closed above by the screening rule of Section 3.4 and thins below $S_r = 0.6$, and its width is what a model of these two targets has to work with.

The four effort classes overlap rather than separate. Modified-effort records do not sit above the standard ones; between 5 and 10 % water content their median density

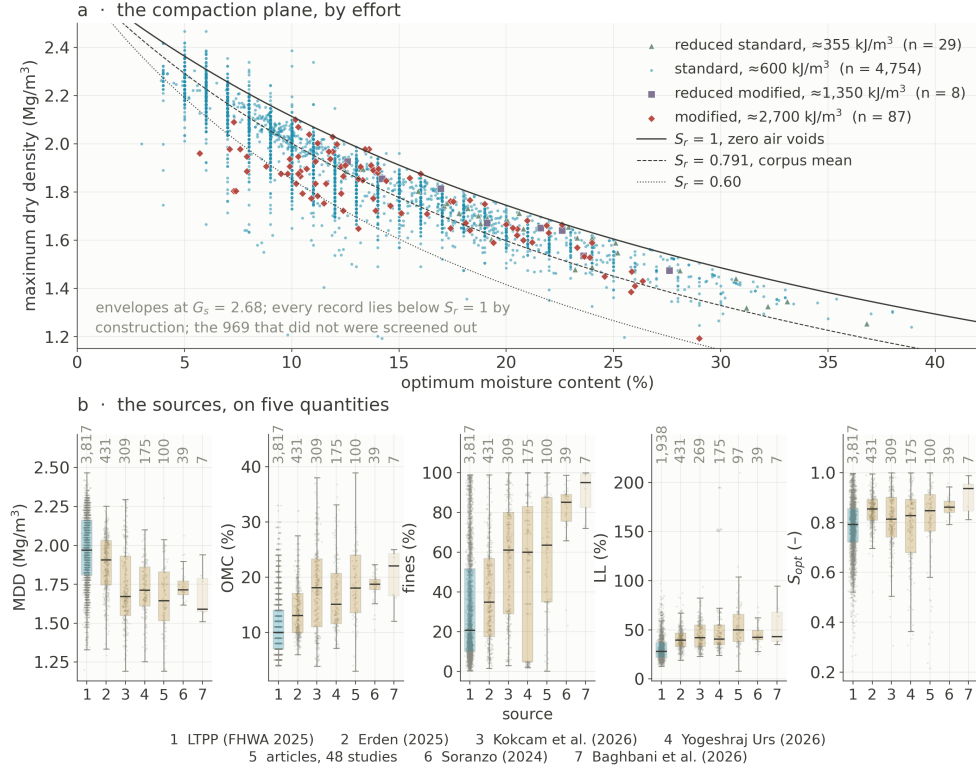


Fig. 2 The corpus: (a) the compaction plane by effort, with envelopes at $G_s = 2.68$; (b) the sources of Table 2, in that order and numbered as in the key, on five quantities. The count above each box is the records carrying that quantity, so it falls below the source total where a quantity is absent.

is 0.192 Mg/m^3 below the standard records, because the soils compacted harder are also the more plastic. Effort cannot be read off the plane.

The vertical striping is a reporting convention and not a property of the soils. LTPP records the optimum to the nearest whole per cent on all 3,817 of its records and reports density in pounds per cubic foot, so its records fall on a lattice. The density grid is the finer of the two, at 80 levels against 29, which is why only the water-content spacing is visible. The other six sources report both quantities to one or two decimals and form the off-lattice scatter. Rounding the optimum to a whole per cent destroys 0.21 % of the variance the optimum carries across the corpus, so the convention is conspicuous in the figure and negligible in every quantity computed from it.

Panel (b) shows that the sources sample different material rather than the same material twice. Across them in the order of Table 2 the median fines content rises from 20.7 % to 95.0 % and the median density falls from 1.970 to 1.590 Mg/m^3 . The corpus is therefore a large body of coarse pavement-subgrade material with a finer tail from six other places. The distributions overlap without any source being disjoint: LTPP's interquartile range for density is 1.81 to 2.16 Mg/m^3 against 1.52 to 1.83 for

the article stream. Pooling is therefore not a concatenation of unrelated populations, but neither is any one source a substitute for the rest.

The saturation at the peak is the exception that holds the corpus together. Its median moves only from 0.792 to 0.936 across the seven sources, a range of 0.14 against the 0.38 Mg/m³ spanned by the density medians. Seven independent sources, spanning clean sand to soils of 95 % fines under at least four laboratory protocols, agree on the state a soil reaches at its optimum far more closely than on any other quantity in the figure. That agreement is what permits a single zero-air-voids screen (Section 3.4) across such heterogeneous provenance.

The liquid limit panel carries 2,956 records of 4,878, because 1,917 are non-plastic soils for which the limit does not exist.

Figure 3 sets the same eight quantities against one another: each variable’s distribution on the diagonal, the pairwise scatter below it and the Spearman rank correlation above. A rank statistic is used because several of these quantities are bounded and skewed, and it does not assume the relation is linear.

The strongest relation in the grid is between the two targets, at $\rho = -0.94$. The correlation is the band of Figure 2(a) expressed as a number. It has a consequence for the estimation problem: ranked, MDD and OMC are very nearly the same quantity, so predicting them independently is close to predicting one of them twice. Section 6.3 uses that, recovering the optimum from the estimated density through the phase relation rather than fitting it separately.

Among the inputs, the fines content is the strongest single predictor of both targets, at -0.61 with density and $+0.71$ with the optimum, and the two consistency limits follow: w_P at -0.57 and $+0.56$, I_P at -0.56 and $+0.65$. The sand fraction is weaker at $+0.25$ and -0.35 , and its panel against fines shows why the two cannot be treated as independent evidence: a hard diagonal boundary, because the two fractions cannot sum above 100 %. That closure runs through the whole input set. Fines correlates with I_P at $+0.68$ and with sand at -0.61 , and I_P with sand at -0.54 , so the inputs overlap substantially and no one of them carries information the others lack. Section 6.2 returns to this.

Two inputs are almost orthogonal to everything else. No correlation involving G_s exceeds 0.11 and none involving $\ln \bar{E}$ exceeds 0.12. The specific gravity reaches only $+0.08$ with density, although density is $G_s \rho_w / (1 + e)$ and depends on it directly, because the void ratio varies far more across this corpus and dominates the product. The compactive energy correlates with nothing because it takes four discrete values and 97.5 % of records lie on one of them. Both results anticipate the ablation of Section 6.2, which prices these two inputs at 0.010 and 0.010 in R^2 against 0.229 for gradation.

The diagonal shows that the plasticity index is not a continuous variable over this corpus but a point mass at zero, the 1,917 non-plastic records, with a tail.

Spearman rank correlation (upper), pairwise scatter (lower), distribution (diagonal)

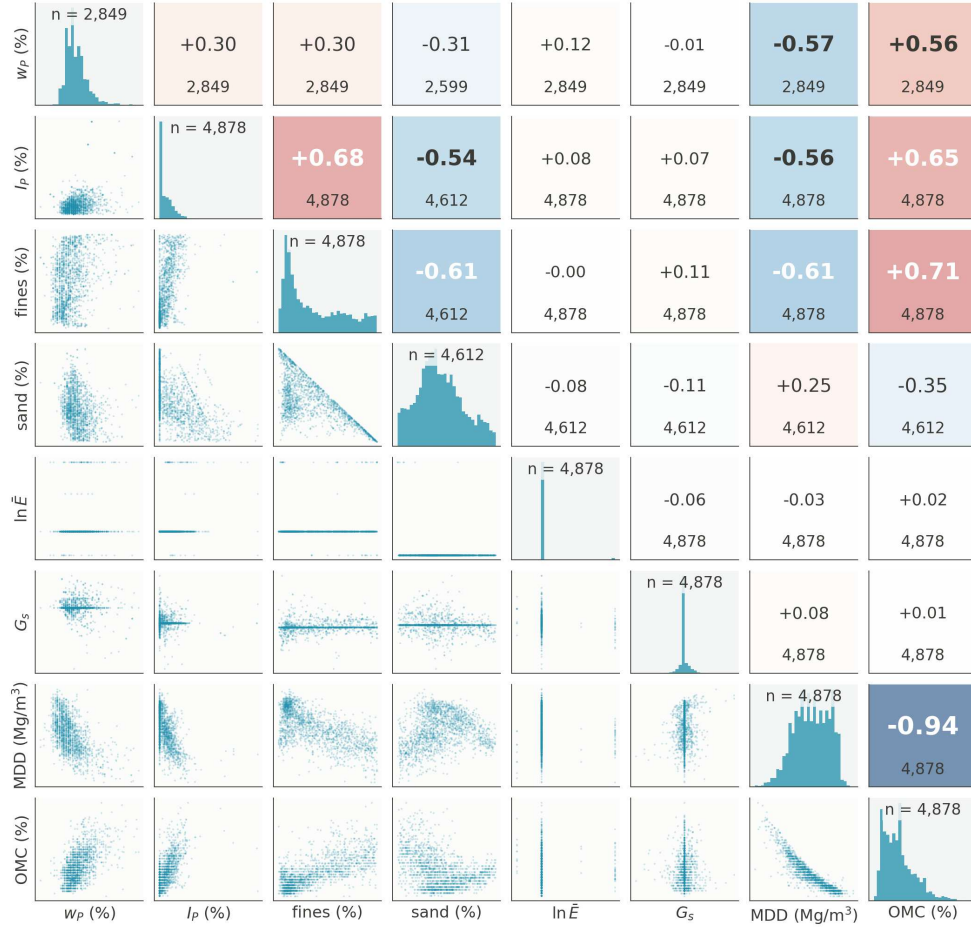


Fig. 3 The six inputs and the two targets against one another: distribution (diagonal), pairwise scatter (lower), Spearman rank correlation shaded by magnitude (upper). The small number in each upper cell is the records carrying both variables, which falls to 2,599 where two gapped inputs meet.

5 Methodology

5.1 The estimation problem

The problem is stated here in five steps: what a laboratory measures, what follows from those measurements by arithmetic, what the mapping receives and returns, what is deliberately withheld from it, and what the estimate is for.

Step 1, what the laboratory measures. One routine classification suite on a disturbed sample, and nothing more: the Atterberg limits to ASTM D4318, a sieve and hydrometer analysis, and a specific-gravity determination to ASTM D854. The

compaction standard is also known, being specified rather than measured. No quantity taken from the compaction curve itself enters the predictors.

Step 2, what follows by arithmetic. Three inputs need no further test. The plasticity index is the liquid limit less the plastic limit, and is zero where the limits cannot be run. The fines content is the silt plus the clay. The compactive energy is the logarithm of the value the named Proctor standard specifies, so it is declared rather than measured, which is what lets the corpus carry four levels without inferring any of them.

Step 3, the mapping. The six inputs are the plastic limit w_P and plasticity index I_P in per cent, the fines content F and sand content s in per cent of the total mass, the logarithm of the compactive energy $\ln \bar{E}$, and the specific gravity G_s . The targets are the maximum dry density MDD in Mg/m^3 and the optimum moisture content OMC, modelled as a decimal fraction and reported as a percentage. Writing $\mathbf{x} = (w_P, I_P, F, s, \ln \bar{E}, G_s) \in \mathbb{R}^6$, two regressors are fitted independently, $f_1(\mathbf{x}) \rightarrow \text{MDD}$ and $f_2(\mathbf{x}) \rightarrow \text{OMC}$, on identical inputs and identical partitions. The two targets are physically coupled, being the coordinates of a single curve peak and correlated at $\rho = -0.94$ in Figure 3, but no coupling is imposed on the estimators. Whether they nevertheless respect it is measured in Section 6.1, and Section 6.3 gives an alternative in which the coupling is imposed by construction.

Step 4, what is withheld. Three quantities are kept out of the mapping. The saturation at the peak is derived from both targets, so supplying it would leak them. The liquid limit is the third of three consistency limits, of which only two are independent. The source identifier assigns the folds, and a new material has no place in the corpus, so a model using it could not be applied.

Step 5, what the estimate is for. Three uses are intended: ranking borrow sources before any is tested, deciding which material warrants a Proctor determination first, and setting the water content of its first specimen. They are not a basis for accepting placed fill, for a reason of arithmetic rather than caution. The lowest mean absolute error obtained here, 0.067 Mg/m^3 , is about 3.5 % of the median maximum dry density. A field specification requiring 95 % of MDD allows a shortfall of 5 %, so the mean error alone is about three quarters of that tolerance. An acceptance decision would spend the margin on the prediction and leave nothing for the variability of the fill.

5.2 How the models are scored

A model must be scored on records it has not seen, and the way those records are chosen decides what the score means. Figure 4(a) shows the scheme. The 4,878 records are divided at random into five equal parts; each part is held out in turn while the other four are given to the model, and the held-out part is predicted. After five rounds every record has exactly one prediction, made by a model that did not read it. The partition is drawn once by scikit-learn's `KFold(5, shuffle=True, random_state=0)` and reused for every learner, so no difference between learners comes from the split. That seed governs every fold assignment in the paper, and every model is seeded at 0 as well.

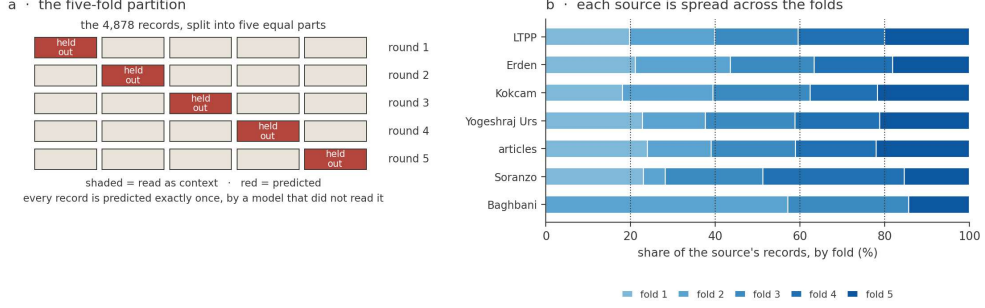


Fig. 4 Scoring: (a) the five-fold partition, in which each part is held out in turn and predicted by a model given the other four; (b) the share of each source's records falling in each fold. The smallest deposit has seven records and so cannot reach all five folds.

The four learners of Table 5 are specified here. The two in-context models are used as released, with nothing set beyond the device and the seed. Each prediction is then the mean of eight forward passes, each over a different permutation of the input columns and a different normalisation. A transformer reads its context as a sequence, so its output depends on the order the columns and rows arrive in, and averaging over permutations removes that dependence. The gradient-boosted model is `XGBRegressor` with 600 trees, learning rate 0.04, maximum depth 5, subsample and column subsample 0.85, minimum child weight 5 and L2 penalty 1.0, all fixed before any result was seen. The perceptron has three hidden layers of 256, 128 and 64 units with ReLU activation. It is trained by Adam at an initial learning rate of 10^{-3} , batch size 128 and L2 penalty 10^{-3} , for at most 600 epochs, stopping early when a tenth of the training fold held out for validation fails to improve for 25 epochs. Its targets are standardised on the training fold and its predictions rescaled. Both fitted learners receive the nine-column median-plus-indicator encoding of Section 4.1.

Two foundation models released after the original four-model analysis are evaluated as a dated extension on the identical folds, predictors and targets. `TabPFN 3.5` uses `tabpfn 9.0.0` with `ModelVersion.V3_5`, eight estimators, seed 0, fitted preprocessors and every eligible training-fold row as context (Prior Labs 2026). Google `TabFM` uses its v1.0.0 PyTorch regression checkpoint through `tabfm 1.0.1`, with eight estimators, seed 0, automatic mixed precision, batch size one and at most 512 context rows per estimator (Google Research 2026). Their weights remain frozen: `fit` constructs preprocessing and the in-context prediction state, and does not update the pretrained network. Neither model receives response correction or the physical decoder of Section 6.3. These configurations were fixed before their predictions were inspected, but the decision to extend the original comparison was made after Table 5 was known; the two later releases are therefore identified as a dated extension in Table 5 rather than presented as part of the original prospective comparison.

6 Results

Two things govern how every figure in this section should be read. Figure 4(b) shows the first. Each source's records are scattered across the folds, so holding out a fold

does not hold out a source, and 46.7 % of records share their exact input vector with a record in the training part. The second is that no untouched test set exists, since every figure below is cross-validated within the same 4,878 records and the input set and the ablations were chosen after inspecting them. The results therefore describe interpolation within this corpus. They are what the comparative literature reports and may be set beside published values, but no claim is made about a soil from an unseen source, and Section 7 states what that costs.

Three metrics are reported: the coefficient of determination, the mean absolute error and the root-mean-square error, each computed once over the pooled predictions rather than averaged across folds, because the admissibility check needs one prediction per record. The floor in every comparison is a mean predictor computed on the four parts and scored on the held-out one. Since the two targets are predicted separately a predicted pair may be physically impossible, so the proportion above the zero-air-voids line is reported beside the accuracy. Bold marks the best value in the comparison each table makes. The gradient-boosted model is run with hyperparameters fixed in advance rather than searched, and that costs it nothing. A nested random search over nine hyperparameters, with 60 candidates scored on a three-fold inner cross-validation inside each training part, returned R^2 0.764 for density and 0.755 for water content. The fixed values return 0.786 and 0.771, so the comparison below is not one the ensemble loses for want of tuning.

6.1 Predictive accuracy

Table 5 reports all six learners on the six inputs under the exact random five-fold design, together with the training-fold mean predictor as a floor. TabPFN 3.5 gives the best accuracy on both targets: R^2 0.817 for density at MAE 0.067 Mg/m³, and R^2 0.801 for water content at MAE 1.67 percentage points. These errors are reductions of 65 % and 62 % against the mean predictor. The density error is 3.5 % of the median density, and the water-content error is comparable with between-laboratory reproducibility of the Proctor test.

The predictions are not independent across records, so uncertainty is estimated by resampling the 218 provenance groups of Table 2 with replacement 5,000 times. TabPFN 3.5 has a 95 % interval of 0.789–0.837 for density R^2 and 0.766–0.823 for water-content R^2 . Relative to TabICL, its R^2 gains are 0.0154 for density and 0.0148 for water content, with paired 95 % intervals of 0.0086–0.0234 and 0.0087–0.0203. Every resample favours TabPFN 3.5. Its MAE reductions are 0.0039 Mg/m³ (95 % interval 0.0028–0.0050) and 0.081 percentage points (0.052–0.112). The gain is therefore resolved under this sampling model, although its magnitude is small.

Table 0 Out-of-fold performance, six inputs, exact random five-fold design, 4,878 records. MDD errors are Mg/m^3 ; OMC errors are percentage points. TabPFN 3.5 and Google TabFM are the dated extension described in Section 5.2. ZAV is the number of predicted pairs implying a degree of saturation above unity.

Model	MDD R^2	MDD MAE	MDD RMSE	OMC R^2	OMC MAE	OMC RMSE	ZAV, n
TabPFN 3.5	0.817	0.067	0.097	0.801	1.67	2.50	9
TabICL v2	0.802	0.071	0.101	0.786	1.75	2.59	4
TabPFN 2.5	0.789	0.074	0.104	0.773	1.82	2.67	1
XGBoost	0.786	0.076	0.105	0.771	1.84	2.68	10
Google TabFM 1.0.0	0.768	0.080	0.109	0.753	1.92	2.79	5
Perceptron	0.744	0.084	0.115	0.735	2.00	2.89	73
Mean predictor	-0.001	0.190	0.227	-0.001	4.42	5.61	39

Bold marks the best learner value in each column. The mean predictor is a floor and is excluded from the model ranking. XGBoost and the perceptron use the median-plus-indicator encoding of Section 4.1.

The random partition contains an exact training-fold match for the six-feature vector of 2,276 records (46.7 %). On the 2,602 records without such a match, TabPFN 3.5 retains a smaller advantage over TabICL: density R^2 0.800 against 0.795 and MAE 0.0749 against 0.0768 Mg/m^3 ; water-content R^2 0.798 against 0.785 and MAE 1.87 against 1.95 percentage points. The gain is therefore not confined to literal repetition of a predictor vector. It remains an in-corpus result, because the training part still contains other records from every represented source. TabFM finishes below TabICL in this configuration; because it samples 512 context rows per estimator, the result does not characterize every larger context or ensemble configuration.

Figure 5 draws the TabPFN 3.5 predictions of Table 5 against the measurements that produced them, with the effort classes marked as in Figure 2, so that the pooled figures can be read against the strata they average over.

The scatter is symmetric about the 1:1 line and shows no trend, so the predictions are not biased high or low over the range, and what lies outside the band of one mean absolute error lies outside in both directions. The error within an effort class must, however, be read as a mean absolute error and not as an R^2 , because each class has its own spread and the four are not on one scale. The difference is not academic. On density the modified-Proctor records score R^2 0.699 against 0.814 for standard Proctor, a gap that might be read as failure at the higher effort. Their mean absolute errors are 0.0649 and 0.0677 Mg/m^3 , respectively, so the modified-effort error is slightly lower. The standard deviation of density is 0.174 Mg/m^3 on the modified records against 0.226 on the standard ones, and that ratio explains why similar absolute errors produce different R^2 values. The model is not materially worse there; it is scored against a narrower spread. The vertical banding of the measured axis in panel (b) is likewise a reporting convention rather than a property of the soils. LTTP records the optimum to the nearest whole per cent and supplies 78 % of the corpus, and the predictions are continuous, so the columns appear on one axis only.

Two readings follow from Table 5. First, model choice changes the random-fold score, but by less than the input ablations of Section 6.2. The six learners span 0.073

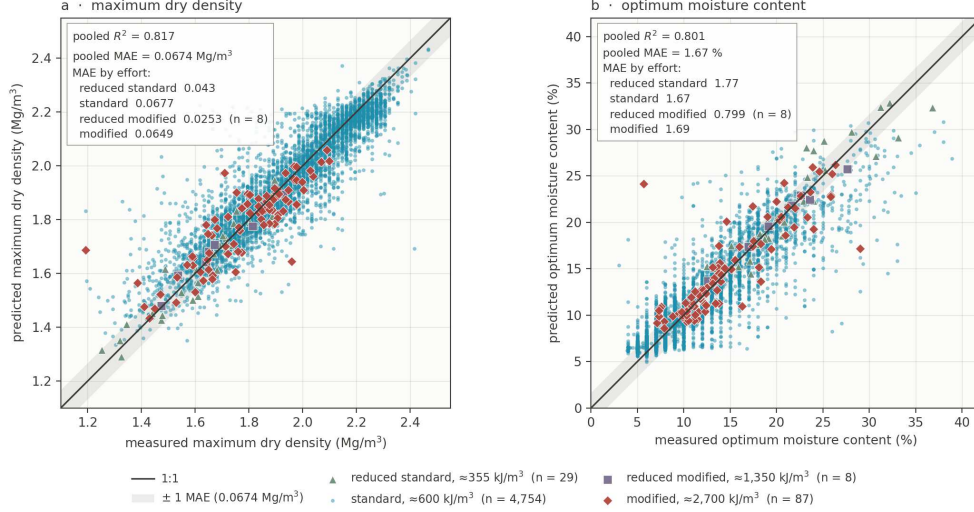


Fig. 5 Out-of-fold predictions of TabPFN 3.5 against measurement, by compactive effort: (a) maximum dry density; (b) optimum moisture content. The band is one mean absolute error either side of the 1:1 line. The annotation gives the pooled R^2 once, then the mean absolute error within each effort class.

in R^2 for density and 0.066 for water content. TabPFN 3.5 improves on the original best model by 0.015 on both targets, whereas removing gradation later costs 0.229 and 0.112. The frozen foundation models take no gradient step on these records, yet the best exceeds the ensemble refitted on every training fold. The learned prior matters, although the available soil description remains the larger limitation.

The second reading is that how a blank is handled makes little difference under this design. TabPFN reads an absent cell as absent and TabICL substitutes the column mean, yet they finish 0.013 apart in R^2 for density, inside the 0.005 the ensemble varies across folds. The result is not evidence that the encoding is unimportant in general. Under a random partition a model can learn where the gaps fall in this corpus and be scored on records whose gaps fall in the same places. Section 7 records the unseen-source case as untested.

Accuracy is not the only criterion. Because the two targets are predicted independently, a predicted pair may imply a degree of saturation above unity, and the learners separate far more sharply on that criterion than on accuracy. TabPFN 2.5 places 1 pair of 4,878 above the line, TabICL 4, TabFM 5 and TabPFN 3.5 9, against 10 for the ensemble and 73 for the perceptron, which is worse than the 39 a mean predictor places there. TabPFN 3.5 is first on accuracy but fourth on admissibility, so the two rankings differ. The rate cannot be driven to zero by improved fitting, because nothing in the formulation forbids a violation; Section 6.3 gives a predictor that does, at a cost Section 7.4 sets out.

Figure 5 reports TabPFN 3.5, the most accurate random-fold predictor. Table 6 therefore repeats the input ablation with TabPFN 3.5. Figure 6 and Section 6.3 retain TabICL, the model for which the exact Shapley and symbolic analyses were specified.

The two model-specific input analyses are identified separately below. TabICL costs 236 s per fold assignment against 2 s for the ensemble.

6.2 What the estimate rests on

Two questions are asked of the input set: which input groups govern the prediction, and what each input contributes to an individual prediction. The first is answered by ablation, in which Table 6 withdraws each input group in turn and rescores TabPFN 3.5 under the random design. The second is answered by the previously specified exact Shapley analysis of TabICL in Figure 6. These analyses address the same input question for two frozen in-context estimators; their absolute values should therefore be compared only within each model.

Table 0 Input ablation, TabPFN 3.5, scored out of fold on the exact random five-fold partition of Section 5.2. Δ is the loss in R^2 against the six-input reference in the first row. Gradation is fines and sand together.

Inputs	MDD R^2 (Δ)	OMC R^2 (Δ)
All six (reference)	0.817	0.801
Without the consistency limits	0.737 (0.080)	0.692 (0.109)
Without gradation entirely	0.588 (0.229)	0.689 (0.112)
Without sand only	0.719 (0.098)	0.774 (0.027)
Without compactive energy	0.807 (0.010)	0.789 (0.012)
Without specific gravity	0.807 (0.010)	0.795 (0.006)

The reference row of Table 6 is the TabPFN 3.5 entry in Table 5, and each Δ is computed from the unrounded pooled R^2 values. Every reduced-input condition uses the same package version, checkpoint, eight-estimator configuration and 4,878 out-of-fold assignments as that reference.

Table 6 separates the inputs into two tiers. Gradation is the load-bearing group for density: withholding both grading terms costs 0.229 in R^2 and withholding sand alone costs 0.098. For water content, gradation and the consistency limits are nearly level, at 0.112 and 0.109, so neither group alone governs that target. Specific gravity and compactive energy cost at most 0.012, an order of magnitude less than the grouped soil-description variables. Compactive energy is nearly constant over the corpus, at 97.5 % standard Proctor, so its pooled figure says little about what the variable is worth where the effort does change.

The reason gradation leads for density is the composition of the corpus. 1,917 of the 4,878 records are non-plastic soils with no consistency limits at all, so for two records in five gradation is the only description of the soil the mapping receives. On the plastic records alone the limits lead instead. Read together: the limits carry the fine-grained end, gradation carries the coarse-grained end, and a corpus that spans the boundary needs both.

Attribution. Table 6 reports what TabPFN 3.5 cannot recover once a group is removed. Removal is not the same question as use, since a group can appear cheap because another input substitutes for it. Figure 6 asks the second question directly for TabICL. It gives Shapley attributions for that model, computed on records held

out of the fold that supplied the context, so the attributions describe a prediction the model had not seen. The Shapley value is the unique division of a prediction among its inputs that is efficient, symmetric and additive, properties established for cooperative games by Shapley (1953) and brought to model explanation by Lundberg and Lee (2017). A Shapley decomposition suits collinear inputs, which Figure 3 shows these to be: it shares credit among correlated inputs by construction rather than assigning it to whichever the model happened to split on.

Six inputs admit the exact calculation: all $2^6 = 64$ coalitions are enumerated, so these are Shapley values and not the sampled approximation a wider table would force, and the efficiency property closes to machine precision. Where a coalition withholds an input the value is drawn from a background sample of the training fold, and a gap is carried through as a gap.

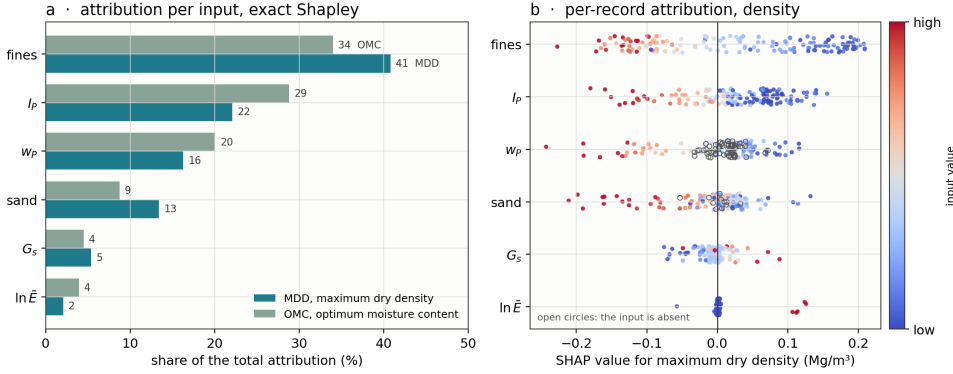


Fig. 6 Exact Shapley attributions (Shapley 1953; Lundberg and Lee 2017) for TabICL, 150 held-out records, 64 coalitions, 8 background rows: (a) each input’s share of the total attribution, for both targets; (b) the attribution to every explained record for density, coloured by the value of the input, with an absent input drawn as an open circle.

The two models agree on the broad density ranking. In the TabICL attribution, gradation leads at 54 % of the total, 41 % from the fines content and 13 % from the sand, against 38 % for the two limits together and 2 % and 5 % for the compactive energy and specific gravity. In Mg/m^3 the mean absolute attributions are 0.112, 0.061, 0.045, 0.037, 0.015 and 0.006 in the order plotted. The TabPFN 3.5 ablation puts the corresponding grouped losses at 0.229 and 0.080, and the energy and specific-gravity losses at 0.010 each. Agreement across two estimators supports the qualitative ranking, but it does not establish that either set of values is model independent.

Panel (b) adds what a mean cannot show, and two features of it are physical rather than statistical.

Every attribution has the sign soil mechanics requires. The rank correlation between an input and its own attribution runs from -0.96 for the plastic limit to -0.60 for sand, so a finer or more plastic soil is pushed towards lower density. It is $+0.66$ for the specific gravity and $+0.34$ for the compactive energy. None of this was imposed: the model was given no phase relation and no sign constraint.

Each direction has a mechanism, and the largest attribution has the best established one. A compacted soil is a packing whose dry density is set by how completely the solid phase fills the total volume. A small fraction of fines densifies a coarse skeleton, because the fine particles occupy voids between the coarse grains without displacing them. Beyond a threshold the coarse grains lose contact and float in a fine matrix, and density falls because the looser fine fraction now governs (Chang and Deng 2020; El-Husseiny 2020). The corpus reproduces that curve without being asked to: median density rises from 1.94 Mg/m³ below 5 % fines to 2.16 Mg/m³ at 5 to 10 %, then falls monotonically to 1.64 Mg/m³ above 90 %. Its median fines content is 25.7 %, so most records lie on the descending limb and the attribution to fines is negative on balance.

The consistency limits act through a different mechanism in the same direction. A plastic soil holds water in the diffuse double layer around its particles, and that adsorbed water cannot lubricate rearrangement under the rammer. A higher plastic limit or plasticity index therefore means more water is needed to reach any given state of packing (Mitchell and Soga 2005). The optimum moves wetter and the peak density falls, which is the pairing seen in Figure 3 and the reason the older single-limit correlations key on a consistency limit at all (Sridharan and Nagaraj 2005). Specific gravity enters through the phase relation rather than packing, since $MDD = G_s \rho_w / (1 + e)$, and its attribution is small only because G_s varies little while the void ratio varies greatly.

The absence of an input is used, and used correctly. On the 63 explained records with no plastic limit the mean attribution is +0.012 Mg/m³, against −0.003 where the limit exists. The model reads a missing plastic limit as evidence of a non-plastic soil and raises the density accordingly. The corpus supports that reading: records without a plastic limit average 2.034 Mg/m³ against 1.855 for those with one, their median fines content being 11 % against 50 %. The attribution makes the mechanism of Section 2 visible. The absence is not a hole to be filled but a state the model conditions on, and the attribution to it is smaller than to a measured value, at 0.020 against 0.063 Mg/m³.

6.3 A dimensionless closed form

The estimators of Table 5 are frozen networks, and a network cannot be written down. Objective 2 therefore also asks for a closed form, and the search is run on dimensionless variables. The choice is not presentational. Every closed form in this literature carries its units inside its coefficients, having been fitted on a density in Mg/m³ and an energy in kJ/m³. It therefore cannot be checked against a phase relation, or applied to data recorded in pcF, without silently changing meaning.

The inputs are $w_P/100$, $I_P/100$, the fines and sand fractions, $\hat{E} = \bar{E}/592.5$ and G_s . The targets are the solid volume fraction at the peak and the degree of saturation at the peak,

$$\eta = \frac{MDD}{G_s \rho_w} = \frac{1}{1 + e} \quad (3)$$

$$S_{opt} = \frac{w_{opt}G_s}{e} \quad (4)$$

both dimensionless and both exactly invertible, so the physical parameters return by identity rather than by a second fit.

The search is carried out with PySR (Cranmer 2023), which evolves a population of candidate expressions represented as trees. Each is scored on the mean squared error it leaves and on its complexity, taken as the number of operators, variables and constants. Candidates are recombined and mutated over successive generations and the constants of each survivor are optimised against the data. The procedure returns a Pareto front rather than a single expression, giving the most accurate expression at every complexity, so the accuracy bought by each additional term can be read off. The operators offered were addition, subtraction, multiplication, division, square, square root, logarithm and exponential, over 300 iterations and 24 populations with the complexity ceiling at 40. The search runs on the 2,599 records complete in all six inputs, since symbolic regression cannot take a gap, and accuracy is out of fold on the folds of Section 5.2. Every figure in this section is therefore computed on that subset, and the comparison values for the fitted models are recomputed on the same records.

The search finds a usable law for the solid volume fraction. Its Pareto front flattens at about twelve terms, and the expression there is

$$\eta = \exp\left[-0.126\left(\sqrt{F} + 1.369 I_P + 1.114 w_P G_s\right)\right]^2 \quad (5)$$

$$\text{MDD} = G_s \eta \rho_w \quad (6)$$

$$w_{opt} = S_{opt} \left(\frac{1}{\eta} - 1\right) / G_s, \quad S_{opt} = 0.822 \quad (7)$$

The three equations hold over the range of the records they were fitted on, and nothing here supports extrapolation beyond it. On those 2,599 records w_P runs from 1 to 49 %, I_P from 0 to 71 %, F from 0.3 to 100 % and G_s from 2.34 to 2.94. The targets span MDD 1.25 to 2.42 Mg/m³ and OMC 4.0 to 38.0 %, and the saturation that Equation (7) replaces by a constant spans 0.42 to 1.00.

Equation (5) is the only fitted relation; Equations (6) and (7) are the identities of Equations (3) and (4) rearranged, so the two physical parameters follow from η without a second regression. The density reaches R² 0.733 at a mean absolute error of 0.085 Mg/m³. Refitting its three constants inside each fold changes that by 0.0006, so the form is not carrying the records it was scored on. Letting the search run to its own preferred size returns R² 0.751 at a complexity of 37, in an expression with five levels of nested exponentials and square roots. The extra 0.018 does not justify an expression that cannot be written down, and Equation (5) is reported instead. Its structure is interpretable: the density falls as the soil becomes finer, more plastic or denser in its solid phase, and the three terms enter as a sum inside a single exponential decay.

The optimum is not fitted at all. The search found no expression for the saturation at the peak that improves on a constant, so S_{opt} enters Equation (7) as its mean over

these records. That returns R^2 0.697 at 2.23 % water content. The constant 0.822 is the mean over this subset and not over the corpus, whose mean is the 0.791 of Figure 2. The subset is a different mix of soils: requiring a plastic limit keeps only 23 of the 2,004 non-plastic records, and its median fines content is 50 % against 25.7 % for the corpus. Per-source medians span 0.792 to 0.936 (Section 4.3), so both values sit inside the spread between laboratories.

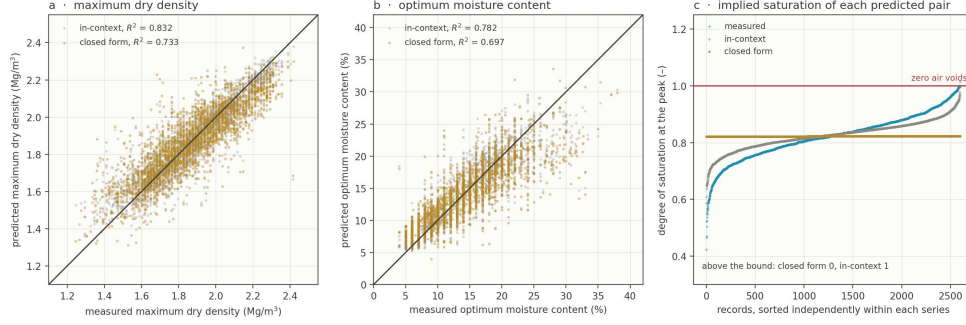


Fig. 7 The closed form against the in-context model on the 2,599 records complete in the six dimensionless inputs: (a) maximum dry density; (b) optimum moisture content, recovered through the phase relation; (c) the degree of saturation each predicted pair implies, sorted within each series, against the measured saturation and the zero-air-voids bound.

Figure 7 draws both routes against measurement on the same records, and panels (a) and (b) show the closed form to be the less accurate. On these 2,599 records TabICL reaches 0.832 for density and 0.782 for water content, against 0.733 and 0.697 for Equation (5) and the constant, so the closed form gives up about 0.10 in R^2 on each target. It gains admissibility by construction. S_{opt} is the degree of saturation and the zero-air-voids line is $S_{opt} = 1$, so a predicted saturation at or below unity cannot produce a pair above the line. The route returns no impossible pairs at all, against one for TabICL and one for TabPFN on the same records.

Panel (c) shows what that guarantee is and is not. The measured saturation runs from 0.42 to the bound and the in-context model reproduces most of that spread, crossing once, whereas the closed form is a horizontal line because it predicts the same saturation for every soil. It is admissible in the weak sense that a constant below unity cannot exceed unity, not in the strong sense of tracking the saturation a soil reaches, and the guarantee is bought by declining to model the quantity at all. Section 8 records the trade this represents: for screening a borrow source the fitted model is the better estimator, and for a calculation that must not return an impossible state the closed form is.

7 Discussion

7.1 Model differences are smaller than information losses

The six learners were evaluated on identical random folds, so their scores can be compared directly within this corpus. TabPFN 3.5 improves on TabICL by 0.0154 in density R^2 and 0.0148 in water-content R^2 ; the paired provenance-group bootstrap intervals exclude zero (Table 5). The separation is therefore resolved under the stated resampling scheme, although it is small in absolute terms. The comparison identifies the performance of the released systems as configured here. It does not isolate a causal effect of architecture or pretraining, because the foundation models also differ in their learned priors, context construction and treatment of absent values.

The fitted baselines do not lose through an obviously inadequate search budget. The gradient-boosted ensemble uses values fixed before inspection of the results, and a nested search over 60 alternatives selected a slightly poorer configuration. Conversely, the foundation models take no gradient step on this corpus. Their advantage therefore concerns the complete pretrained inference systems rather than additional optimisation on the evaluation records.

Input information has the larger association with performance. For density, removing gradation from TabPFN 3.5 reduces R^2 by 0.229 and removing the consistency limits reduces it by 0.080, compared with the 0.015 advantage over TabICL. For water content the corresponding ablation losses are 0.112 and 0.109 (Tables 5 and 6). These are predictive ablations rather than causal effects: correlated inputs can substitute for one another, and the losses are specific to this model and corpus. The independent TabICL Shapley analysis nevertheless gives the same broad density ranking. Agreement across models and methods supports the qualitative conclusion that gradation and consistency limits carry more information than the remaining measured inputs.

7.2 Interpreting the benchmark against published results

The compaction literature frequently reports R^2 values above 0.90, whereas the best values here are 0.817 and 0.801. These numbers cannot be ranked without their data and evaluation protocols. R^2 depends on the variance of the target in the evaluated sample, and most published correlations are fitted and scored on one to four hundred specimens from a single laboratory. The present corpus pools six deposits and 48 studies across the coarse-fine boundary. Its absolute errors of 0.067 Mg/m³ and 1.67 percentage points provide a second scale, but they too remain conditional on the materials represented.

The random folds do not measure transfer to a new source. Every held-out record has its source represented in the training part, and 46.7 % have an identical six-input vector there. Removing records with an exact input match reduces but does not remove the dependence, because records from the same programme retain shared laboratory and reporting conventions. The reported scores should therefore be read as a reproducible measure of interpolation within this corpus, not as an estimate of performance for an unseen borrow source, laboratory or field-compacted lift. This

distinction also prevents a direct claim that 0.82 here is better or worse than 0.90 under a different experimental design.

As a benchmark, the corpus supplies a common table, fixed folds, source labels, record-level predictions and physical checks with which new machine-learning methods can be developed, calibrated and compared. Repeated optimisation against the same folds can itself overfit a benchmark, however. A future source added after model development should therefore be retained as an external test set, with source-level separation preserved.

7.3 Implications for a testing programme

The ablation and attribution analyses identify a testing priority within this corpus. A preliminary predictive workflow should retain both gradation and the consistency limits: removing either group produces a substantially larger loss than removing compactive energy or specific gravity. The two groups are complementary. Gradation describes the 1,917 non-plastic records for which consistency limits do not exist, while the limits are informative for the fine-grained, plastic records.

The small pooled ablation loss for compactive energy is not evidence that energy is physically unimportant. Standard effort accounts for 97.5 % of the corpus, and 105 of the 124 non-standard records come from one deposit. Energy variation is therefore too sparse and too confounded with source to estimate a general energy response. Likewise, the small predictive loss after removing specific gravity does not make G_s dispensable for quality control. A median substitution may be adequate for preliminary prediction within these folds, but G_s is the denominator of the phase relation and materially changes the zero-air-voids screen. It should be measured when physical admissibility is to be assessed.

These results also delimit correlations based on a single consistency limit. Such a relation is undefined for non-plastic material unless a value is fabricated. Retaining an explicit absence state permits the model to represent those soils and avoids extending a fine-soil correlation to granular records outside its domain.

7.4 Accuracy and physical admissibility answer different questions

The highest-accuracy learner is not the most physically admissible. TabPFN 3.5 places 9 of 4,878 predicted pairs above the zero-air-voids line, whereas TabPFN 2.5 places one above it (Table 5). These counts are small but non-zero, and the independent-target formulation provides no guarantee that tuning or a larger training set will eliminate them. Accuracy and admissibility must therefore be reported separately.

The dimensionless closed form enforces the bound through the phase relation and returns no impossible pairs on its complete-case subset, but gives up about 0.10 in R^2 on each target. Its saturation component is a constant, so the guarantee is obtained without representing the observed variation in saturation at optimum. The trade-off is therefore explicit: the fitted model is more accurate for within-corpus ranking, while the closed form supplies a weak admissibility guarantee. A joint predictor that

incorporates the phase relation could enforce the bound while retaining flexible input dependence, but its accuracy and transfer performance require separate evaluation.

7.5 What the corpus cannot support

Table 0 What the corpus cannot support.

Limitation	Evidence	What it restricts
Laboratory-only	no source reports a field density	every accuracy and every admissibility figure is a laboratory statement
Peak only	no public source retains curve points	the curve away from the optimum is not modelled
Thin energy coverage	124 non-standard records, 105 from one deposit	no comparison across compactive efforts is drawn; energy is carried as an input only
Coarse source composition	six deposits, one holding 78 %, three below 40 records	a source-held-out design would confound transfer with training-set size, so none is reported
No laboratory identifier	no source publishes the laboratory of each test	no laboratory-held-out evaluation is possible
Limit set fixed in advance	w_P with I_P chosen before the results were seen, with no comparison of alternatives	claims that this pair is optimal among possible consistency-limit representations
Sand boundary not uniform	3,817 records at 4.75 mm, 1,061 as published	the sand column mixes two definitions
G_s substituted on 60.6 %	screen rate moves 19.8 % to 8.2 % over 2.60–2.75	the 11.8 % rejection figure, not the finding behind it
No admissibility constraint	every learner puts some pairs above the bound (Table 5)	the predicted pair, which is not guaranteed to be physical
Gap encoding untested across sources	every held-out record’s source is in the training part, so the pattern of absence is never unfamiliar	the finding that the encoding of a blank barely matters (Section 6.1)
No independent test set	all models cross-validated within 4,878 records	every predictive performance estimate in this paper
Symbolic search not exhaustive	six inputs, complexity ceiling of 40, 300 iterations, and the reported structure chosen on all records before its constants were refitted per fold	the closed form of Section 6.3, which is a lower bound on what a symbolic form can reach

The corpus is laboratory-only and contains peak values only. Every accuracy in Section 6 describes the MDD and OMC obtained from a laboratory compaction curve. No source reports a field density or retains the individual points of that curve.

The benchmark therefore supports prediction of laboratory peak parameters, not field density or the moisture–density relation away from its maximum.

The source composition is as coarse as the sources allow. A single programme contributes 78 % of the records and three of the six deposits hold fewer than 40. Holding a source out would therefore change the training-set size as much as the material, so no source-held-out figure is reported. No source publishes the laboratory that performed each test, so a laboratory-held-out evaluation cannot be run on any public corpus known to the authors.

No independent test set exists, and every fold is random. All models are cross-validated within the same 4,878 records. The input representation, analysis choices and reported comparisons have therefore all been examined on this corpus, even where constants are refitted inside each fold. The figures of Section 6 measure interpolation within the benchmark and should not be used as an unbiased estimate for an unseen source. The released folds make comparisons reproducible; credible claims of generalisation require a later source or laboratory that played no role in constructing the corpus or selecting the analysis.

7.6 A starting point for an open geotechnical benchmark

The corpus contributes beyond the two prediction scores reported here. ImageNet and COCO advanced computer-vision research by pairing open data with common tasks and evaluation conventions (Deng et al. 2009; Lin et al. 2014). The analogy is organisational rather than numerical: 4,878 compaction tests do not match the scale or task breadth of those image collections. They do provide soil-compaction research with the same basic infrastructure—a public, versioned dataset, explicit inputs and targets, fixed reference folds, common metrics and an auditable baseline.

The release contains more than MDD and OMC values. Each record retains its source and provenance group, Proctor method and energy, index properties, missing-value state, specific-gravity substitution flag, harmonisation decisions and physical-screening outcome. These fields support source-aware evaluation, alternative preprocessing, independent checking of the inclusion rules and traceable future additions.

The corpus can serve as an initial calibration and evaluation resource for tabular foundation, deep-learning and ensemble models, physics-constrained joint predictors, and empirical or symbolic equations. Its size does not justify training an unconstrained deep network from scratch, and the random folds do not establish external generalisation. Its longer-term value depends on community extension with balanced compactive-energy data, complete moisture–density curves, field measurements and laboratory identifiers, while new sources are reserved for external testing. The corpus is therefore a starting point for cumulative open geotechnical data research, not a definitive dataset for soil compaction.

8 Conclusions

This study establishes an open, multi-source corpus for laboratory soil compaction and demonstrates how it can support reproducible model comparison. Its principal

contribution is the combination of measurements with source provenance, declared test methods, harmonisation decisions and physical-quality checks. The screening results show that assembling published values is not sufficient by itself: phase consistency and the status of substituted properties must remain part of the data record. In this corpus, the zero-air-voids screen removed 11.8 % of the harmonised records.

Among the evaluated methods, TabPFN 3.5 gave the strongest random-fold predictions, reaching R^2 values of 0.817 for maximum dry density and 0.801 for optimum moisture content, with mean absolute errors of 0.067 Mg/m³ and 1.67 percentage points, respectively. The comparison also showed that the information supplied to the model matters more than small differences among high-performing learners. Gradation and consistency limits were the principal input groups, and the absence of a consistency limit should be retained as evidence of non-plastic material rather than replaced by a fabricated measurement.

Predictive accuracy and physical admissibility remain separate requirements. Models that estimate density and water content independently can cross the zero-air-voids bound even when their average errors are small. The dimensionless closed form illustrates the alternative: physical consistency can be enforced through the phase relation, but at a loss of accuracy and with a simplified representation of saturation. Future models should treat the two targets jointly if both flexibility and guaranteed admissibility are required.

The reported scores quantify interpolation within the present corpus; they do not establish transfer to an unseen source, laboratory or field-compacted material. The release should therefore be viewed as a starting benchmark for developing, calibrating and evaluating machine-learning models and empirical equations. Its value will increase as future contributions add balanced energy levels, complete compaction curves, field measurements and laboratory identifiers while preserving new sources for external testing.

References

- AASHTO T 180. *Standard Method of Test for Moisture–Density Relations of Soils Using a 4.54-kg (10-lb) Rammer and a 457-mm (18-in.) Drop*. American Association of State Highway and Transportation Officials, Washington, DC.
- AASHTO T 99. *Standard Method of Test for Moisture–Density Relations of Soils Using a 2.5-kg (5.5-lb) Rammer and a 305-mm (12-in.) Drop*. American Association of State Highway and Transportation Officials, Washington, DC.
- Abdul Hudoos A, Premalatha K (2020) Stabilization of CL and CI Soil Using Sugar-cane Bagasse Ash and Quarry Dust. *International Journal of Innovative Science and Research Technology* 5(7):1360-1364. <https://doi.org/10.38124/ijisrt20jul851>.
- Adedokun SI, Ariye YK, Olagunju BY, Apata AC, Ganiyu AA (2025) Geotechnical Behaviour of Steel Slag–Lateritic Soil Reinforced with Shredded Plastic Waste for Road Foundations. *FUOYE Journal of Engineering and Technology* 10(4):697-705. <https://doi.org/10.4314/fuoyejet.v10i4.20>.

- Adunoye GO, Osineye TO, Ezech HC (2022) Investigation of properties of soils stabilised with groundnut waste. *International Journal of Physical Research* 10(1):33-38. <https://doi.org/10.14419/ijpr.v10i1.31881>.
- Al-Kifae A, Hamad MA (2017) Inspecting the effects of organic content on compaction and consolidation characteristics of organic soil models. *American Scientific Research Journal for Engineering, Technology, and Sciences*, 31(1), 109-123.
- Ale TO (2022) Effect of drying temperature on the engineering properties of stabilised and natural soils for road construction. *Engineering Heritage Journal* 6(1):01-05. <https://doi.org/10.26480/gwk.01.2022.01.05>.
- Amu O, Adetayo O, Faluyi F, Akinyele E (2021) Experimental Study of Improving the Properties of Lime-Stabilized Structural Lateritic Soil for Highway Structural Works using Groundnut Shell Ash. *Walailak Journal of Science and Technology (WJST)* 18(9). <https://doi.org/10.48048/wjst.2021.9475>.
- Arega A (2020) The effect of random inclusion of teff straw on shear strength characteristics of Addis Ababa expansive soil. MSc thesis, Addis Ababa University.
- ASTM D1557-12 (2021) *Standard Test Methods for Laboratory Compaction Characteristics of Soil Using Modified Effort (56,000 ft-lbf/ft³ (2,700 kN·m/m³))*. ASTM International, West Conshohocken, PA. <https://doi.org/10.1520/D1557-12R21>
This is the edition the corpus maps to. The compactive effort is fixed by the title and is unchanged across editions, so a source citing an earlier one carries the same 2,693 kJ/m³.
- ASTM D2487-17 (2025) *Standard Practice for Classification of Soils for Engineering Purposes (Unified Soil Classification System)*. ASTM International, West Conshohocken, PA.
- ASTM D4318-17e1. *Standard Test Methods for Liquid Limit, Plastic Limit, and Plasticity Index of Soils*. ASTM International, West Conshohocken, PA.
- ASTM D698-12 (2021) *Standard Test Methods for Laboratory Compaction Characteristics of Soil Using Standard Effort (12,400 ft-lbf/ft³ (600 kN·m/m³))*. ASTM International, West Conshohocken, PA. <https://doi.org/10.1520/D0698-12R21>
This is the edition the corpus maps to. The compactive effort is fixed by the title and is unchanged across editions, so a source citing an earlier one carries the same 592.5 kJ/m³.
- ASTM D854-14. *Standard Test Methods for Specific Gravity of Soil Solids by Water Pycnometer*. ASTM International, West Conshohocken, PA. Withdrawn in 2023 and replaced by ASTM D854-23, *Standard Test Methods for Specific Gravity of Soil Solids by the Water Displacement Method*.
- Ayodele FO, Popoola OO (2019) Potential of snail shell and palm kernel shell powders in improving engineering properties of clay. *Journal of Applied Sciences and Environmental Management* 23(8):1437. <https://doi.org/10.4314/jasem.v23i8.5>.

- Baek S-H, Kim J, Park K, Choi H (2025) Field study on intelligent compaction for compaction quality control of subgrade bases. *Canadian Geotechnical Journal* 62:1-14. <https://doi.org/10.1139/cgj-2024-0300>.
- Baghbani A, Bahrampour D, Moballegh A, Daghistani F (2026) *Dataset of UCS and CBR results for biopolymer-stabilised soils from multi-additive experimental studies* [data set]. Zenodo. <https://doi.org/10.5281/zenodo.19242689>
- Begashaw S, Sorsa A, Alemu C (2025) Experimental study on the suitability of high-land bamboo sawdust in improving expansive soil. *Sustainable Structures* 5(3). <https://doi.org/10.54113/j.sust.2025.000083>.
- Bhatt A, Suman S (2022) Evaluation of the Effect of RAP and Lime on Geotechnical Properties of Clayey Soil. *International Journal for Research in Applied Science and Engineering Technology* 10(4):2531-2541. <https://doi.org/10.22214/ijraset.2022.41839>.
- Birhan A, Alelgn D, Teshome W, Dessie A (2026) Sustainable Stabilization of Expansive Soil for Subgrade Applications Using Waste Glass Powder and Steel Slag. <https://doi.org/10.21203/rs.3.rs-8918321/v1>
- Blotz LR, Benson CH, Boutwell GP (1998) Estimating optimum water content and maximum dry unit weight for compacted clays. *Journal of Geotechnical and Geoenvironmental Engineering* 124(9):907-912. [https://doi.org/10.1061/\(ASCE\)1090-0241\(1998\)124:9\(907\)](https://doi.org/10.1061/(ASCE)1090-0241(1998)124:9(907)).
- Brase J (2009) DataCite: a global registration agency for research data. *Fourth International Conference on Cooperation and Promotion of Information Resources in Science and Technology* 257–261. <https://doi.org/10.1109/COINFO.2009.66>.
- Chang CS, Deng Y (2020) Packing potential index for binary mixtures of granular soil. *Powder Technology* 372:148-160. <https://doi.org/10.1016/j.powtec.2020.06.005>.
- Chilukwa NN (2013) Vibratory hammer compaction of granular materials. MSc thesis, University of Stellenbosch.
- Cranmer M (2023) *Interpretable machine learning for science with PySR and SymbolicRegression.jl*. arXiv preprint. <https://doi.org/10.48550/arXiv.2305.01582>
- Daniel DE, Benson CH (1990) Water content–density criteria for compacted soil liners. *Journal of Geotechnical Engineering* 116(12):1811-1830. [https://doi.org/10.1061/\(ASCE\)0733-9410\(1990\)116:12\(1811\)](https://doi.org/10.1061/(ASCE)0733-9410(1990)116:12(1811)).
- Deng J, Dong W, Socher R, Li L-J, Li K, Li F-F (2009) ImageNet: A large-scale hierarchical image database. *2009 IEEE Conference on Computer Vision and Pattern Recognition* 248–255. <https://doi.org/10.1109/CVPR.2009.5206848>.
- El-Husseiny A (2020) Improved packing model for functionally graded sand-fines mixtures: incorporation of fines cohesive packing behavior. *Applied Sciences* 10(2):562. <https://doi.org/10.3390/app10020562>.

- Emmanuel UO, Ogbonnaya I, Uche UB (2021) An investigation into the cause of road failure along Sagamu-Papalanto highway southwestern Nigeria. *Geoenvironmental Disasters* 8(1):3. <https://doi.org/10.1186/s40677-020-00174-8>.
- Erden C (2025) *Geotechnical properties of soils: a dataset from literature* [data set]. figshare. <https://doi.org/10.6084/m9.figshare.28681187>
- Evriana T, Satrya TR, Maulana MA, Winurseto WS (2024) Correlation of Initial Soil Density and Maximum Soil Density Under Drying-Wetting Cycles and Their Soil Erodibility. *Journal of Infrastructure & Facility Asset Management* 6(0). <https://doi.org/10.12962/jifam.v6i0.19993>.
- Farooq K, Mujtaba H, Ashiq SZ, Shah MM, Mohamed A (2024) Amelioration of fat clay treated with cohesive and cohesion less soils. <https://doi.org/10.21203/rs.3.rs-4115772/v1>
- Fattah MY, Al-Lami AH, Ahmed MD (2015) Effect of Initial Water Content on the Properties of Compacted Expansive Unsaturated Soil. *Journal of Engineering* 21(3):93-108. <https://doi.org/10.31026/j.eng.2015.03.06>.
- Federal Highway Administration (2003) *Introduction to the LTPP Information Management System (IMS)*. Report FHWA-RD-03-088. Federal Highway Administration, McLean, VA. <https://www.fhwa.dot.gov/publications/research/infrastructure/pavements/ltp/reports/03088/>
- Federal Highway Administration (2025) *Long-Term Pavement Performance Standard Data Release 39*, Material Test volume. InfoPave. <https://infopave.fhwa.dot.gov/Data/StandardDataRelease>
- Fentaw M, Alemayehu E, Geremew A (2021) Experimental study of stabilization of expansive soil using the mixture of marble dust, rice husk ash and cement for sub-grade road construction: a case study of woldia town. *Journal of Civil Engineering, Science and Technology* 12(2):141-159. <https://doi.org/10.33736/jcest.3977.2021>.
- Google Research (2026) *TabFM: tabular foundation models* [software repository]. <https://github.com/google-research/tabfm>
- Grinsztajn L, Flöge K, Key O, Birkel F, Jund P, Roof B, Jäger B, Safaric D, Alessi S, Hayler A, Manium M, Yu R, Jablonski F, Hoo SB, Garg A, Robertson J, Bühler M, Moroshan V, Purucker L, Cornu C, Wehrhahn LC, Bonetto A, Schölkopf B, Gambhir S, Hollmann N, Hutter F (2025) TabPFN-2.5: advancing the state of the art in tabular foundation models. arXiv preprint. <https://doi.org/10.48550/arXiv.2511.08667>
- Günaydin O (2009) Estimation of soil compaction parameters by using statistical analyses and artificial neural networks. *Environmental Geology* 57(1):203-215. <https://doi.org/10.1007/s00254-008-1300-6>.
- Gurtug Y, Sridharan A (2004) Compaction behaviour and prediction of its characteristics of fine grained soils with particular reference to compaction energy. *Soils and Foundations* 44(5):27-36. https://doi.org/10.3208/sandf.44.5_27.

- Audu HAP, Okovido JO (2019) Experimental Study on Effect of Lime and Cement Stabilizers on Geotechnical Properties of Lateritic Soil. *International Journal of Engineering Research and Advanced Technology* 5(4):09-15. <https://doi.org/10.31695/ijerat.2019.3409>.
- Hasan R (2021) Evaluating the impact of light crude oil contaminations on geotechnical behaviors of the soil. *Anbar journal of agricultural sciences* 19(2):143-157. <https://doi.org/10.32649/ajas.2021.175869>.
- Hollmann N, Müller S, Purucker L, Krishnakumar A, Körfer M, Hoo SB, Schirrmeister RT, Hutter F (2025) Accurate predictions on small data with a tabular foundation model. *Nature* 637(8045):319-326. <https://doi.org/10.1038/s41586-024-08328-6>.
- Horpibulsuk S, Katkan W, Apichatvullop A (2008) An approach for assessment of compaction curves of fine grained soils at various energies using a one point test. *Soils and Foundations* 48(1):115-125. <https://doi.org/10.3208/sandf.48.115>.
- Horpibulsuk S, Katkan W, Naramitkornburee A (2009) Modified Ohio's curves: a rapid estimation of compaction curves for coarse- and fine-grained soils. *Geotechnical Testing Journal* 32(1):64-75. <https://doi.org/10.1520/GTJ101659>.
- Hossain MB, Roknuzzaman M, Biswas MA, Islam M (2021) Evaluation of engineering properties of thermal power plant waste for subgrade treatment. *Journal of Civil Engineering, Science and Technology* 12(2):112-123. <https://doi.org/10.33736/jcest.3975.2021>.
- Issac S (2018) Stabilization of Expansive Kuttanad clay using Lime treated Rice straw fibres. *International Journal for Research in Applied Science and Engineering Technology* 6(2):580-586. <https://doi.org/10.22214/ijraset.2018.2082>.
- Jafer HM, Atherton W, Al-Dulaimi A (2017) The characterisation of the strength development of a cement-stabilised soft soil treated with two different types of fly ashes. Proceedings of the LJMU 16th Annual International Conference on Asphalt, Pavement Engineering and Infrastructure, Liverpool, UK.
- Jumaa RM, Saeed KA, Al-Jaberi LA (2026) Effect of Fly Ash and Nano-Silica Fume on Soft Clay: Atterberg Limits, MDD, and OMC. *Civil Engineering Journal* 12(2):627-637. <https://doi.org/10.28991/cej-2026-012-02-012>.
- Kapoor S, Narayanan A (2023) Leakage and the reproducibility crisis in machine-learning-based science. *Patterns* 4(9):100804. <https://doi.org/10.1016/j.patter.2023.100804>.
- Kaufman S, Rosset S, Perlich C, Stitelman O (2012) Leakage in data mining: formulation, detection, and avoidance. *ACM Transactions on Knowledge Discovery from Data* 6(4):1-21. <https://doi.org/10.1145/2382577.2382579>.
- Kaushik D, Pandey S, Sharma G, Kushwah VK (2026) Experimental Investigation of Lime and Cement Stabilization for Sustainable Pavement Subgrade Applications. *International Journal of Aquatic Research and Environmental Studies*. <https://doi.org/10.70102/ijares/v6s4/6-s4-1055>

- Khatti J, Grover KS (2023) Prediction of compaction parameters for fine-grained soil: critical comparison of the deep learning and standalone models. *Journal of Rock Mechanics and Geotechnical Engineering* 15(11):3010-3038. <https://doi.org/10.1016/j.jrmge.2022.12.034>.
- Khokar IA, Al-Saidi KS, Ali SA, Hayder G (2025) Utilizing Recycled Vegetable Plastic Bags as an Innovative and Sustainable Material for Soil Reinforcement Applications. *International Journal of Environmental Impacts* 8(5):1093-1100. <https://doi.org/10.56578/ije080520>.
- Kokcam AH, Dagdeviren U, Kurnaz TF, Demir AS, Erden C (2026) *Dataset: geotechnical properties and California bearing ratio (CBR) values of soil samples from Türkiye* [data set]. Zenodo. <https://doi.org/10.5281/zenodo.20737270>
- Krishna DKVM (2012) Strength and F O S Performance of Black Cotton Soil Treated with Calcium Chloride. *IOSR Journal of Mechanical and Civil Engineering* 2(6):21-25. <https://doi.org/10.9790/1684-0262125>.
- Kumar R (2020) Compaction Characteristics of BC Soil Reinforced with Untreated and Treated Coir Fibers. *International Journal of Engineering Research and Technology* 9(6). <https://doi.org/10.17577/ijertv9is060678>.
- Kushwah SS, Kaur S (2026) Geotechnical Characterization of Black Cotton Soil Stabilized with Incinerated Municipal Solid Waste Bottom Ash. *Dandaao Xuebao/Journal of Ballistics* 38(2):1140-1149. <https://doi.org/10.52783/dxjb.v38.462>.
- Lambe TW (1958) The structure of compacted clay. *Journal of the Soil Mechanics and Foundations Division* 84(SM2). <https://doi.org/10.1061/JSFEAQ.0000114>.
- Leon LP (2024) Effects of Quarry and Recycled Concrete Waste as Rapid Use Stabilizers for Clay Soils. *Proceedings of the 22nd LACCEI International Multi-Conference for Engineering, Education and Technology (LACCEI 2024): "Sustainable Engineering for a Diverse, Equitable, and Inclusive Future at the Service of Education, Research, and Industry for a Society 5.0."* <https://doi.org/10.18687/laccei2024.1.1.1618>
- Li H, Sego DC (2000) Equation for complete compaction curve of fine-grained soils and its applications. In *Constructing and Controlling Compaction of Earth Fills*, ASTM STP 1384, 113–125. <https://doi.org/10.1520/STP15277S>
- Lin T-Y, Maire M, Belongie S, Hays J, Perona P, Ramanan D, Dollár P, Zitnick CL (2014) Microsoft COCO: Common objects in context. In *Computer Vision – ECCV 2014*, 740–755. https://doi.org/10.1007/978-3-319-10602-1_48
- Lundberg SM, Lee S-I (2017) A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems* 30:4765-4774. <https://doi.org/10.48550/arXiv.1705.07874>.
- Majeed DZH, Aubais EKJ, Taha DMR (2020) Using Nanomaterials in Stabilization of Soil for Oil Infrastructures. *Journal of Petroleum Research and Studies* 10(3):36-53. <https://doi.org/10.52716/jprs.v10i3.329>.

- Mallikarjuna V, Mani B (2016) Soil stabilization using plastic waste. *International Journal of Research in Engineering and Technology* 05(05):391-394. <https://doi.org/10.15623/ijret.2016.0505074>.
- Mitchell JK, Soga K (2005) *Fundamentals of Soil Behavior* (3rd ed.). John Wiley & Sons.
- Müller S, Hollmann N, Arango SP, Grabocka J, Hutter F (2022) Transformers can do Bayesian inference. *International Conference on Learning Representations (ICLR)*. <https://doi.org/10.48550/arXiv.2112.10510>
- Nagaraj HB, Reesha B, Sravan MV, Suresh MR (2015) Correlation of compaction characteristics of natural soils with modified plastic limit. *Transportation Geotechnics* 2:65-77. <https://doi.org/10.1016/j.trgeo.2014.09.002>.
- Nik Daud NN, Muhammed AS, Md. Yusoff Z (2018) Geotechnical assessment of palm oil fuel ash (pofa) mixed with granite residual soil for hydraulic barrier purposes. *Malaysian Journal of Civil Engineering* 28. <https://doi.org/10.11113/mjce.v28.15990>.
- Oke JA, Obaji NO, Osinubi KJ (2020) Oil palm empty fruit bunch ash stabilized laterite as a fill material for low-volume pavement. *Nigerian Journal of Technology* 39(3):721-731. <https://doi.org/10.4314/njt.v39i3.10>.
- Okunade EA (2010) Geotechnical Properties of Some Coal Fly Ash Stabilized South-western Nigeria Lateritic Soils. *Modern Applied Science* 4(12). <https://doi.org/10.5539/mas.v4n12p66>.
- Ola SA, Usifoh OA, Nnochiri ES (2023) The effects of potassium permanganate on the geotechnical properties of soils. *Acta Polytechnica* 63(4). <https://doi.org/10.14311/ap.2023.63.0260>.
- Olofinyo OO, Franklin OO, Oluwatoyin FI, Odebode OS (2024) Geotechnical investigation for subsoil hazard during earth fill dam construction. *Malaysian Journal of Geosciences* 8(1):17-25. <https://doi.org/10.26480/mjg.01.2024.17.25>.
- Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD (2021) The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, 372, n71. <https://doi.org/10.1136/bmj.n71>
- Pal S, Adhikari K, Mukherjee S, Ghosh S (2015) Potential of Silty Clay Soil as an Attenuation Material for Containment of Phenolic Wastewater Outfall Site. *International Journal of Environmental Science and Development* 6(12):895-900. <https://doi.org/10.7763/ijesd.2015.v6.718>.
- Pandian NS, Nagaraj TS, Manoj M (1997) Re-examination of compaction characteristics of fine-grained soils. *Géotechnique* 47(2):363-366. <https://doi.org/10.1680/geot.1997.47.2.363>.
- Ploton P, Mortier F, Réjou-Méchain M, Barbier N, Picard N, Rossi V, Dormann C, Cornu G, Viennois G, Bayol N, Lyapustin A, Gourlet-Fleury S, Péliissier R

- (2020) Spatial validation reveals poor predictive performance of large-scale ecological mapping models. *Nature Communications* 11(1):4540. <https://doi.org/10.1038/s41467-020-18321-y>.
- Pratap A, Kumar P, Kumar R (2026) Synergistic Geotechnical Enhancement of Expansive Regur Soil through Sustainable Composite Stabilization Using Fly Ash, Kota Stone Powder, and Polyolefin Fibre. *International Journal of Aquatic Research and Environmental Studies*. <https://doi.org/10.70102/ijares/v6s1/6-s1-1722>
- Priem J, Piwowar H, Orr R (2022) *OpenAlex: a fully-open index of scholarly works, authors, venues, institutions, and concepts*. arXiv preprint. <https://doi.org/10.48550/arXiv.2205.01833>
- Prior Labs (2026) *TabPFN-3.5 model card*. Hugging Face. https://huggingface.co/Prior-Labs/tabpfn_3_5
- Proctor RR (1933) Fundamental principles of soil compaction. *Engineering News-Record* 111(9):245-248.
- Qu J, Holzmüller D, Varoquaux G, Le Morvan M (2025) TabICL: a tabular foundation model for in-context learning on large data. *Proceedings of the 42nd International Conference on Machine Learning (ICML)* PMLR 267. <https://doi.org/10.48550/arXiv.2502.05564>.
- Rajakumar C, Rama Rao PK (2019) Compaction Characteristics of Swelling Soil Stabilized with Bottom Ash and Geogrid. *International Journal of ChemTech Research* 12(6):157-162. <https://doi.org/10.20902/ijctr.2019.120620>.
- Rajwade P (2019) Study of Compaction Characteristics of Black Cotton Soil Reinforced with Natural Coir Fibre. *International Journal for Research in Applied Science and Engineering Technology* 7(10):369-375. <https://doi.org/10.22214/ijraset.2019.10056>.
- Rakaraddi PG (2015) Establishing relationship between cbr with different soil properties. *International Journal of Research in Engineering and Technology* 04(02):182-188. <https://doi.org/10.15623/ijret.2015.0402023>.
- Rao DDK (2025) Stabilization of Expansive Soil Using Bagasse Ash and Geopolymers to Suit as Subgrade in Flexible Pavement. *International Journal for Research in Applied Science and Engineering Technology* 13(11):2018-2025. <https://doi.org/10.22214/ijraset.2025.75528>.
- Rashid Z, Mahmood C (2026) Characterization and Volumetric Stability of High-Plasticity Clay Stabilized by Fly Ash and Metakaolin Based Geopolymers. *Sulaimani Journal for Engineering Sciences* 12(2):312-328. <https://doi.org/10.17656/sjes.10213>.
- Rehman ZU, Khalid U, Ijaz N, Ijaz Z (2025) Big data-driven global modeling of cohesive soil compaction across conceptual and arbitrary energies through machine learning. *Transportation Geotechnics* 50:101470. <https://doi.org/10.1016/j.trgeo.2024.101470>.

- Rethlefsen ML, Kirtley S, Waffenschmidt S, Ayala AP, Moher D, Page MJ, Koffel JB (2021) PRISMA-S: an extension to the PRISMA statement for reporting literature searches in systematic reviews. *Systematic Reviews* 10:39. <https://doi.org/10.1186/s13643-020-01542-z>.
- Roberts DR, Bahn V, Ciuti S, Boyce MS, Elith J, Guillerá-Arroita G, Hauenstein S, Lahoz-Monfort JJ, Schröder B, Thuiller W, Warton DI, Wintle BA, Hartig F, Dormann CF (2017) Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography* 40(8):913-929. <https://doi.org/10.1111/ecog.02881>.
- Saito T, Otake Y, Wu S (2026a) Applying a tabular foundation model to geotechnical site characterization. *Geodata and AI* 6:100040. <https://doi.org/10.1016/j.geoai.2025.100040>.
- Saito T, Otake Y, Mizutani D, Wu S (2026b) TabPFN extensions for interpretable geotechnical modelling. *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards* 1–20. <https://doi.org/10.1080/17499518.2026.2707500>.
- Sapta W, Harianto Y, Gofar N (2023) CBR Correlation with Index and Compaction Properties of Soil. *Indonesian Geotechnical Journal* 2(3):179-188. <https://doi.org/10.56144/igj.v2i3.72>.
- Schmidinger J, Barkov V, Vogel S, Atzmueller M, Heuvelink GBM (2026) Kriging prior regression: a case for kriging-based spatial features with TabPFN in soil mapping. *Computers and Electronics in Agriculture* 243:111352. <https://doi.org/10.1016/j.compag.2025.111352>.
- Seed HB, Chan CK (1959) Structure and strength characteristics of compacted clays. *Journal of the Soil Mechanics and Foundations Division* 85(SM5):87-128. <https://doi.org/10.1061/JSFEAQ.0000229>.
- Shapley LS (1953) A value for n-person games. In H. W. Kuhn & A. W. Tucker (Eds.), *Contributions to the Theory of Games, Volume II* (pp. 307–317). Princeton University Press. <https://doi.org/10.1515/9781400881970-018>
- Sharma B, Sridharan A (2016) Static method to determine compaction characteristics of fine-grained soils. *Geotechnical Testing Journal* 39(6):1048-1055. <https://doi.org/10.1520/GTJ20150221>.
- Shukla V (2019) Improvement of Pavement Soil Subgrade by using Cement Kiln Dust. *International Journal for Research in Applied Science and Engineering Technology* 7(4):2341-2345. <https://doi.org/10.22214/ijraset.2019.4422>.
- Singh A (2023) Shredded Waste Tyre General Approach as a Construction Material. *International Journal For Multidisciplinary Research* 5(5):6515. <https://doi.org/10.36948/ijfmr.2023.v05i05.6515>.

- Soranzo E (2024) *Dataset for “Machine learning predictions on an extensive geotechnical dataset of laboratory tests in Austria”* [data set]. Zenodo. <https://doi.org/10.5281/zenodo.14251190>
- Sridharan A, Nagaraj HB (2005) Plastic limit and compaction characteristics of fine-grained soils. *Proceedings of the Institution of Civil Engineers – Ground Improvement* 9(1):17-22. <https://doi.org/10.1680/grim.2005.9.1.17>.
- Thakur Y (2018) Geotechnical Characteristic of Compacted Sand Bentonite Mixture. *International Journal for Research in Applied Science and Engineering Technology* 6(3):2102-2108. <https://doi.org/10.22214/ijraset.2018.3497>.
- Trinh MT (2006) *Shear strength and volume change relationship for an unsaturated soil*. PhD thesis, Nanyang Technological University. <https://doi.org/10.32657/10356/12084>
- Verma G, Sivapullaiah PV (2019) Prediction of compaction parameters for fine-grained and coarse-grained soils: a review. *International Journal of Geotechnical Engineering* 14(8):970-977. <https://doi.org/10.1080/19386362.2019.1595301>.
- Vinod P, Sridharan A (2016) Toughness limit: a useful index property for prediction of compaction parameters of fine-grained soils at any rational compactive effort. *Indian Geotechnical Journal* 47(1):107-114. <https://doi.org/10.1007/s40098-016-0194-6>.
- Yogeshraj Urs C (2026) *Compaction characteristics data* [data set]. figshare. <https://doi.org/10.6084/m9.figshare.32955851>

Statements and Declarations

Funding

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Competing interests

The authors have no relevant financial or non-financial interests to disclose.

Author contributions

Sompote Youwai contributed to conceptualization, methodology, software, formal analysis, data curation, visualization, and writing the original draft. Chana Phut-thananon contributed to validation, investigation, and review and editing. Warat Kongkitkul contributed to supervision, resources, and review and editing. All authors reviewed and approved the manuscript.

Data availability

The open corpus, provenance information, and validation material are available at https://github.com/Sompote/open_compaction under the CC BY 4.0 licence.

Code availability

The analysis code required to reproduce the reported screening, model evaluation, and statistical analyses is available in the same public repository. The foundation-model comparison used TabPFN 2.5 with `tabpfn` 6.4.1, TabICL v2, TabPFN 3.5 with `tabpfn` 9.0.0, and Google TabFM 1.0.0 with `tabfm` 1.0.1.

Ethics approval, consent to participate, and consent for publication

Not applicable.

Appendix A Studies behind the article records

Every article record carries the study it came from, and each is included in the reference list. Table A1 lists them with the records each supplied, so a prediction can be traced to the laboratory that made the measurement. The list is long-tailed, the largest study supplying 11 records and 27 of the 48 supplying one, and it holds 19 of the 124 non-standard-effort records. Two corrections were needed: four studies reached the corpus with no DOI and were identified from the source document, and five more carry Crossref metadata naming an affiliation, an initial or a punctuation mark where the author should be. Both are recorded with the release.

One venue is flagged rather than corrected. Kushwah and Kaur (2026) appears in *Dandao Xuebao / Journal of Ballistics* under ISSN 1004-499X, which belongs to an established Chinese ordnance journal, but the DOI is registered by a different publisher and the subject is soil stabilisation. That combination is the signature of a hijacked title. The record is retained because the corpus is frozen and every figure in this paper is computed on it, and because the entry supplies one record of 4,878. A reader excluding it should expect no change beyond the fourth decimal. The screening of Section 3 tests completeness and physical admissibility, not the standing of the venue, and this case marks the limit of that screen.

Table A1 The 48 open-access studies behind the article row of Table 2. Study is surname and year, and effort is the Proctor method the study names.

Study	DOI	<i>n</i>	Compactive effort
Ale (2022)	10.26480/gwk.01.2022.01.05	11	standard
Ola et al. (2023)	10.14311/ap.2023.63.0260	7	standard
Shukla (2019)	10.22214/ijraset.2019.4422	7	standard
Rakaraddi and Gomarsi (2015)	10.15623/ijret.2015.0402023	5	standard
Abdul Hudoos and Premalatha (2020)	10.38124/ijisrt20jul851	4	standard
Nik Daud et al. (2018)	10.11113/mjce.v28.15990	4	standard
Kumar (2020)	10.17577/ijertv9is060678	3	standard
Majeed et al. (2020)	10.52716/jprs.v10i3.329	3	standard
Mallikarjuna and Mani (2016)	10.15623/ijret.2016.0505074	3	modified
Olofinyo et al. (2024)	10.26480/mjg.01.2024.17.25	3	standard
Rao (2025)	10.22214/ijraset.2025.75528	3	modified
Amu et al. (2021)	10.48048/wjst.2021.9475	2	standard
Begashaw et al. (2025)	10.54113/j.sust.2025.000083	2	modified
Chilukwa (2013)	—	2	standard
Emmanuel et al. (2021)	10.1186/s40677-020-00174-8	2	standard
Farooq et al. (2024)	10.21203/rs.3.rs-4115772/v1	2	modified
Fentaw et al. (2021)	10.33736/jcest.3977.2021	2	standard
Hossain et al. (2021)	10.33736/jcest.3975.2021	2	modified
Adunoye et al. (2022)	10.14419/ijpr.v10i1.31881	2	standard
Okunade (2010)	10.5539/mas.v4n12p66	2	standard
Rajwade (2019)	10.22214/ijraset.2019.10056	2	standard
Adedokun et al. (2025)	10.4314/fuoyejet.v10i4.20	1	modified
Pratap et al. (2026)	10.70102/ijares/v6s1/6-s1-1722	1	standard
Al-Kifae (2017)	—	1	standard
Arega (2020)	—	1	modified
Ayodele and Popoola (2019)	10.4314/jasem.v23i8.5	1	standard
Bhatt and Suman (2022)	10.22214/ijraset.2022.41839	1	standard
Birhan et al. (2026)	10.21203/rs.3.rs-8918321/v1	1	modified
Kaushik et al. (2026)	10.70102/ijares/v6s4/6-s4-1055	1	standard
Evriana et al. (2024)	10.12962/jifam.v6i0.19993	1	standard
Fattah et al. (2015)	10.31026/j.eng.2015.03.06	1	standard
Audu and Okovido (2019)	10.31695/ijerat.2019.3409	1	standard
Hasan (2021)	10.32649/ajas.2021.175869	1	standard
Issac (2018)	10.22214/ijraset.2018.2082	1	standard
Jafer (2017)	—	1	standard
Jumaa et al. (2026)	10.28991/cej-2026-012-02-012	1	modified
Khokar et al. (2025)	10.56578/ijei080520	1	standard
Krishna (2012)	10.9790/1684-0262125	1	standard
Kushwah and Kaur (2026)	10.52783/dxjb.v38.462	1	modified
Leon (2024)	10.18687/laccei2024.1.1.1618	1	standard
Oke et al. (2020)	10.4314/njt.v39i3.10	1	standard

Study	DOI	<i>n</i>	Compactive effort
Pal et al. (2015)	10.7763/ijesd.2015.v6.718	1	standard
Rajakumar and Rama Rao (2019)	10.20902/ijctr.2019.120620	1	standard
Rashid and Mahmood (2026)	10.17656/sjes.10213	1	standard
Sapta et al. (2023)	10.56144/igj.v2i3.72	1	standard
Singh (2023)	10.36948/ijfmr.2023.v05i05.6515	1	modified
Thakur (2018)	10.22214/ijraset.2018.3497	1	modified
Trinh (2006)	10.32657/10356/12084	1	standard
Total		100	