Proceedings of the 21st ICSMGE, Vienna, Austria, 14 – 19 June 2026. Pistrol, Adam & Schweiger (eds.)
Published by: ÖGG, Austrian Society for Geomechanics, Salzburg, Austria, ISBN 978-3-9503898-4-5
https://doi.org/10.53243/ICSMGE2026-2121

Machine learning based verification of dry deep mixing

Hilde Aas Nøst, Stefan Ritter, Priscilla Paniagua, Marit Skaug Løyland
Norwegian Geotechnical Institute, NGI, Norway, hilde.nost@ngi.no

ABSTRACT: Dry Deep Mixing (DDM) is a widely used technique to improve the properties of soft ground to execute underground construction projects including deep excavations. The practical application of DDM typically relies on resource-intensive laboratory and full-scale field trials to estimate the future performance of the improved ground. Quality control in the field has been restricted to destructive techniques such as, for example, core sampling, pre-installed reverse sounding tests, push-in resistance tests and cone penetration tests. All these field verification methods have several shortcomings including varying core sampling quality and uncertainty when estimating engineering properties from recorded probing resistance. In addition, only a small percentage of the executed stabilized soil columns are tested in the field. Hence, the properties of the untested soil columns remain unknown. During DDM column installation, substantial amounts of data are generated as part of the execution process, which are typically used for qualitative documentation and quality control. However, these datasets are often stored in file formats that limit their usability for advanced analysis and practical verification due to the interdependent nature of various parameters. This study investigates the potential of machine learning (ML) based methods to leverage DDM datasets, offering a novel approach to map spatial patterns in DDM applications. This contribution describes the initial application, using datasets from one DDM field tests at the Norwegian GeoTest Sites (NGTS) Tiller-Flotten, and performance evaluation of different ML techniques to detect weakness zones in terms of strength. The aim of the work is to establish a methodology for further evaluation of the spatial variability in DDM execution and its effects on the performance of the ground improved soil.

KEYWORDS: Dry deep mixing, machine learning, regression.

1 INTRODUCTION

Dry Deep Mixing (DDM) is a widely used ground improvement technique in the Nordic region to improve the properties of the challenging soils, such as soft marine clay, sensitive quick clay and peat. The DDM method mechanically mixes dry cementitious binders, typically cement and lime, with the natural soil using specifically designed DDM rigs. More recent applications adopt new binders consisting of industrial-by-products to reduce the environmental impact of the binder production. Common applications of DDM are embankment and foundation support, slope stabilization and deep excavations.

A main challenge of the DDM is the inherent spatial variability of stabilized soil columns (Al-Naqshabandy et al., 2013; Savila et al., 2025; Wong et al., 2024). In addition, methods to quantify the in-situ properties of DDM improved soils have several uncertainties. For example, widely applied methods, such as column penetration tests, rely on a constant bearing capacity factor to estimate the strength of a stabilized soil column. Recent research, however, showed that the bearing capacity factor varies with the degree of confinement (Hov et al., 2025a) and the undrained strength of the stabilized soil (Timoney & McCabe, 2017). Furthermore, only a small percentage of the executed stabilized soil columns can be tested in the field. The properties of the untested soil columns remain unknown which results in uncertainty about the performance of the improved ground.

DDM rigs record a wide range of execution parameters including revolutions per minute and binder content, among many others, during the drilling and mixing phases. These parameters are usually plotted and shown in file formats like PDFs to verify the field execution. For large projects, extensive datasets are produced but these data have so far not been sufficiently analyzed to find patterns between the engineering properties of the stabilized soil and the machine data. In particular, the application of Machine Learning (ML) algorithms including linear regression, lasso, ridge regression and other regression models, which have recently been adopted to predict the performance of DDM improved soil, has, to the author's knowledge, not yet been carried out.

This paper explores the application of the above-mentioned models to detect correlations between DDM rig data and the strength of stabilized soil columns with the main aim of detecting weakness zones that may cause performance issues. First, this paper introduces a recently conducted test field of DDM in which both execution data and strength properties of stabilized soil columns were acquired. It will then move on to describe the adopted ML techniques, followed by a detailed presentation of the results obtained. Finally, the paper discussed the practical implications of the investigation carried out.

2 DRY DEEP MIXING TEST FIELD

The DDM work was carried out at the NGTS site Tiller-Flotten close to Trondheim, Norway. The site is mainly comprised of marine sensitive clay deposits of more than 50 m depth. The upper 7-8 m is mainly medium sensitive clay followed by a quick clay with sensitivity reaching up to 200 (L'Heureux et al. 2019).

The DDM included the installation of 91 columns with 800 mm diameter and 15 m depth in a pattern as shown in Figure 1. 28 columns were installed in an overlapping pattern to create an improved soil block of 10 m x 3 m size. The rest of the columns were installed as single columns in a quadratic pattern with a center-to-center distance of about 1.39 m.

Two types of binders were applied: a paper sludge ash (PSA) from paper production and recycling, and a lime-cement (LC) mixture of 35% quick lime and 65% standard Portland cement CEM II. The binders were applied in quantities of 90 kg/m^3 and 50 kg/m^3, respectively, aiming to reach 300 kPa at 28 days of curing. Conventional cone penetration tests (CPTs) were conducted in 9 cured columns after 7–15 days to assess the strength development. In addition, the site was heavily instrumented with fibre optics, geophones, pore pressure sensors and inclinometers to assess other parameters which are explained elsewhere (Hov et al. 2025b).

During DDM, several parameters are recorded during drilling and mixing phases. Some of these parameters are predefined by the machine operator, some are measured and some are calculated. There is no standard on which parameters should be recorded, and this varies from company to company. In general, the most common parameters include coordinates, tilt, length of the column, drilling duration, mixing duration, total binder content, revolutions per minute for the rotating mixing tool, and withdrawal rate, among others. Table 1 summarizes some of these parameters. The blade rotation number (BRN), which represents the total rotations of the mixing blades during the installation of one-meter stabilized soil column, is a widely used execution parameter.

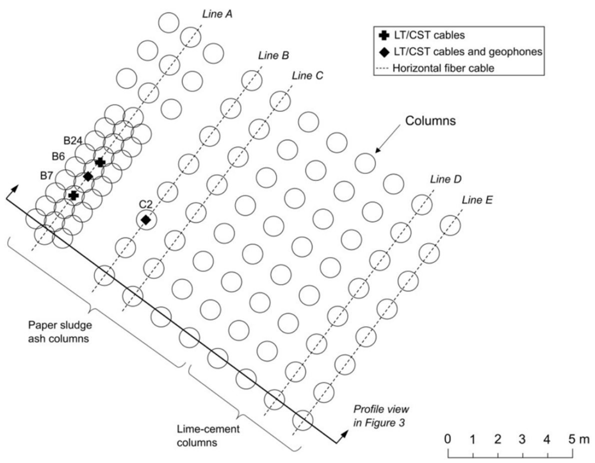

Figure 1. Plan view of the deep mixed columns at Tiller-Flotten site. Taken from Hov et al. (2025b).

3 MACHINE LEARNING TECHNIQUES

3.1 *Dataset description*

The dataset consists of nine CPTs with their corresponding DDM rig data. Of these, five used LC and the remaining four used PSA. The CPT parameters that are used in this study are both derived from raw data. The curing time is the time in days between the DDM installation and the CPT. The undrained shear strength c_u is computed from

$$c_u = \frac{q_c}{N_c} \tag{1}$$

where q_c is the cone tip penetration resistance and N_c is the constant bearing capacity factor. A constant $N_c = 10$ was used throughout this study.

0 shows the mean and standard deviation of 4 selected features: revolutions per minute (rpm), binder content per cubic meter (kg/m^3), undrained shear strength (kPa) and curing time (days). The actual installed binder content per cubic meter does align well with the pre-set target values (see Section 2).

Table 1. Dataset summary statistics separated by binder type (LC or PSA) reported with mean ± standard deviation.

Parameter	Origin	LC	PSA
Revolutions per minute (rpm)	DDM rig	158.5 ± 18.3	174.1 ± 13.8
Binder content (kg/m^3)	DDM rig	50.8 ± 6.4	91.0 ± 5.2
Undrained shear strength (kPa)	CPT (derived)	218.7 ± 89.3	326.8 ± 150.9
Curing time (days)	Time between CPT and binder installation	10.7 ± 3.1	11.2 ± 3.3

3.2 *Data Preprocessing*

The data from the DDM installation process and the post-stabilization CPT have records per depth, but at varying depth intervals. To perform the types of machine learning regression methods selected here, on the combination of these two datasets, these records must be aligned. Based on experience, the DDM rig data should be grouped on a 1-2 m scale due to large volatility and uncertainties during installation. Consequently, both the CPT data and the DDM rig data were grouped and averaged over 1 m depth intervals. The maximum depth of a column is floored to the nearest depth interval boundary to allow for full intervals only. The records at greater depth are discarded from the study.

As presented both in Section 2 and 0, the two binder type groups do differ significantly in the parameters binder content and undrained shear strength. Ideally, the two groups should be trained and tested within each group with the assumption that binder type affects the undrained shear strength. There is ongoing research to assess the effect of binder type on shear strength, but that is beyond the scope of this paper. The two parameters could be separately normalized per binder group to get the same mean across the binder type groups, but with the original variance kept. However, since this is a preliminary study, the parameters are not scaled. Selected features are rpm, pull rate, lift coefficient, binder content per cubic meter, curing time, and depth. The target parameter is the undrained shear strength.

3.3 *Linear regression models*

With the assumption that the machine data as described in Section 2 can be linked to the strength of the stabilized soil, several linear regression machine learning models have been applied to the dataset. With such a small data set, emphasis was placed on the simpler regression models to prevent overfitting at this stage of analysis.

The following machine learning models were explored: linear regression, lasso (L1 as regularizer), ridge regression (linear least squares with L2 regularizer) and regression models such as Decision Tree Regressor and Random Forest Regressor. Lasso and ridge regression differs by their choice of regularizer, that is, their approach to prevent overfitting.

3.4 *Scoring methods*

For assessing the performance of the regression models, the models will be scored using the coefficient of determination, R^2, in addition to the median absolute error. The coefficient of determination indicates the goodness of fit of the model, where the best possible score is 1 and a score of 0 means that the model is unable to describe any of the variations observed in the target variable. The median absolute error metric give a result in the unit of interest which aids in interpreting them, in addition to being robust against outliers.

In addition to the metrics mentioned above, a visual inspection of the residual plots will be performed. A good model or regression analysis will have a randomly scattered residual plot without any visible trend, indicating that all relations have been accounted for in the model.

4 RESULTS

The dataset was analyzed for each stabilized soil column. Furthermore, to prevent selecting particularly favorable or unfavorable stabilized soil columns for the testing data, the so-called leave-p-out cross-validation was performed (Celisse & Robin 2008). The chosen hold out size was p = 3, or a third of the boreholes, and the scores aggregated over all 84 results.

There is no extra test set, as hyperparameter tuning was not carried out in this study.

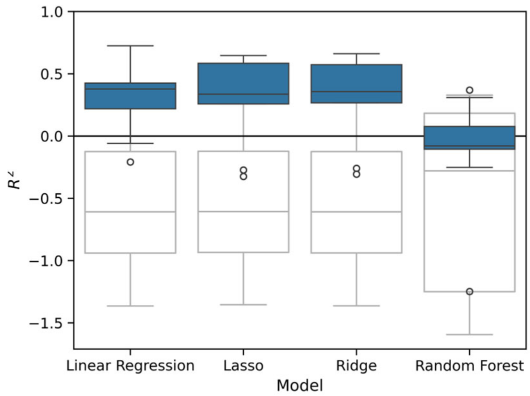

Figure 2. The coefficient of determination scores for all included regression models. The filled boxes use all defined features as described in Section 3.1, while the empty boxes only use curing time and depth.

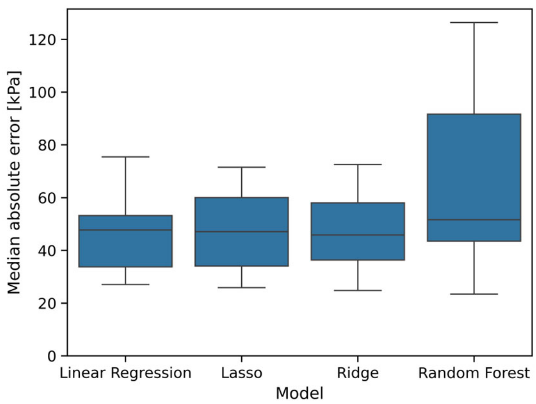

Figure 3. Median absolute error for all included regression models.

Dummy models were trained to check the baseline performance, using only the depth and the curing time as predicting features. The results of this trial run are shown in Figure 2 as the outlined boxes. The coefficient of determination lies around 0, meaning that those features alone do not explain the shear strength.

In general, the coefficient of determination lies slightly under 0.5 for the best models, which are the linear regression models. This indicates that the models capture some of the data variability, but not all of it. With the inherent noisiness and uncertainties related to CPTs and DDM, this is an expected value. Random Forest does not perform well at all, considering that the score signifies that it performs worse than a constant model, that is, a model without any possibility to explain the resulting variance of the undrained shear strength.

The median absolute error is shown in Figure 3. Random forest has a larger variance of the distribution of the error, but all models have a mean median absolute error around 50 kPa. Compared to the scale of the CPT derived undrained shear strength which is around 200 kPa and 330 kPa respectively for the LC and PSA columns, this error is acceptable.

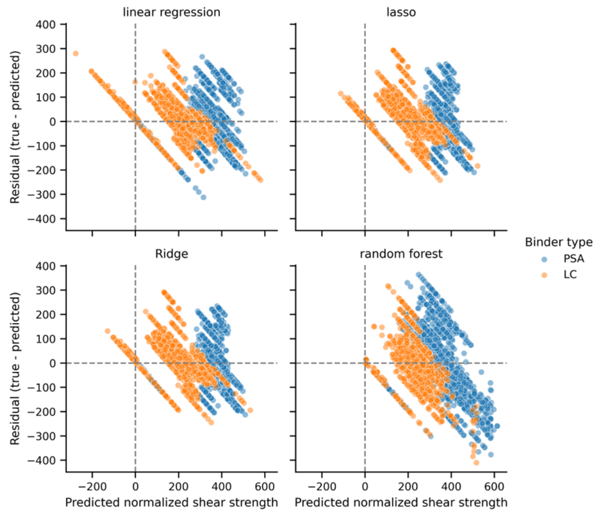

Figure 4. Residual plots of the resulting predictions of undrained shear strength of the selected regression models.

Figure 4 shows the residual plots of the predictions for all tested models. In all models there is a clear outlier with a negative linear trend. While most models do show an evenly distributed scatter of the residuals, there is a slanted negative linear trend. This might be due to treating each record as independent, when they are not. That is, the shear strength at depth 2 m depends on the shear strength at depth 2 m.

Figures 6-8 show strength profiles with depth for selected columns, of which 2 had LC as binder and 1 had PSA as binder. The figures showcase how the models struggle with capturing the profiles with larger variance. However, the resulting strength profiles from the models are all in the vicinity of the actual strength profile. Figure 7 shows that the trained models might skip possible weaker zones by overestimating the undrained shear strength.

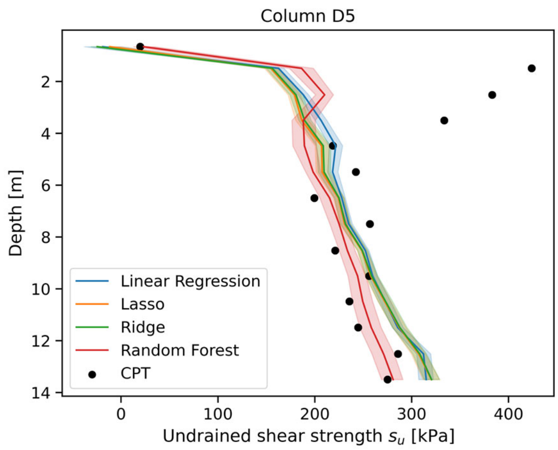

Figure 5. Undrained shear strength with depth for column D5. Black circles indicate undrained shear strength derived from CPT results. This column used LC as binder.

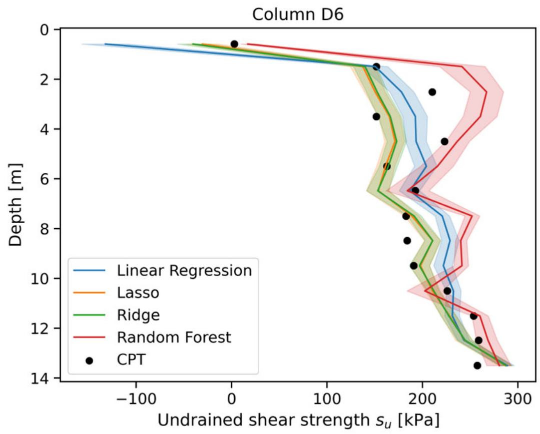

Figure 6. Undrained shear strength with depth for column D6. Black circles indicate undrained shear strength derived from CPT results. This column used LC as binder.

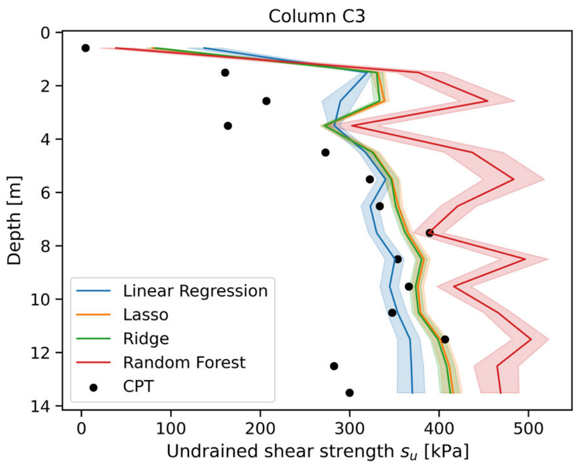

Figure 7. Undrained shear strength with depth for column C3. Black circles indicate undrained shear strength derived from CPT results. This column used PSA as binder.

5 CONCLUSIONS

This study explored the potential of machine learning techniques to utilize DDM datasets for mapping spatial patterns in DDM applications. It presents an initial implementation using data from a field test conducted at the Norwegian GeoTest Sites (NGTS) Tiller-Flotten. The study also evaluates the performance of various ML methods in identifying zones of weakness based on strength characteristics.

The main drawback of this analysis is the limited dataset with very few datapoints and groups. This limits the ability to draw any concrete conclusions at this stage, but preliminary results are promising, and when coupled with other methods, which is part of ongoing work, the regression models could be a useful tool for a quick assessment of the predicted shear strength after having performed the DDM stabilization.

There are negative linear trends in the residuals. This might indicate that the individual records per depth are dependent on each other. Another set of models might be more suitable for this kind of problem.

The coefficient of determination and median absolute error exhibit moderate performance, neither good nor bad, and together with the predicted strength profiles per column, show that the models perform well enough to be of practical use. However, further work is required before applying the discussed models in practice.

Further work would include applying a trained regression model on the stabilized soil columns that do not have CPT data to derive potential strength profiles and obtain zones with potential weakness zones.

6 ACKNOWLEDGEMENTS

The authors are grateful to Keller Geoteknikk AS for providing and explaining the drill rig data. Funding was provided by the basic research funding to NGI from The Research Council of Norway and the GOAL project (RCN 328767).

7 REFERENCES

Al-Naqshabandy, M.S., Bergman, N. and Larsson, S., 2012. Strength variability in lime-cement columns based on cone penetration test data. *Proceedings of the Institution of Civil Engineers-Ground Improvement*, *165*(1), pp.15-30.

Celisse, A., Robin, S., 2008. Nonparametric density estimation by exact leave-p-out cross-validation, *Computational Statistics & Data Analysis*, 52(5), pp. 2350-2368.

Hov, S., Edvin Moe, E., Haraldsen, G.H., Paniagua, P. and Larsson, S. 2025a. Laboratory scale column penetration tests for deep mixing purposes. *Canadian Geotech Journal* 62, 1-14.

Hov, S., Trafford, A., Rørstadbotnen, R.A., Dong, H., Long, M., Landrø, M. and Paniagua, P. 2025. Distributed acoustic sensing in ground improvement by dry deep mixing. Proceedings of the Institution of Civil Engineers - Ground Improvement https://doi.org/10.1680/jgrim.25.00033

L'Heureux, J.-S., Lindgård, A., & Emdal, A. (2019). The Tiller-Flotten research site: Geotechnical characterization of a very sensitive clay deposit. AIMS Geosciences, 5(4), 831–867.

Savila, I., Korkiala-Tanttu, L., Forsman, J. and Löfman, J., *Mechanical properties of stabilized soil: study on recovered field samples from deep stabilization sites. Transportation Geotechnics 51 (2025) 101540* [online]

Timoney M.J., McCabe B.A. 2017. Strength verification of stabilized soil–cement columns: a laboratory investigation of the push-in resistance test (PIRT). *Canadian Geotechnical Journal* 54 (6), 789–805.

Wong, D.Y.C., Sadasivan, V., Isaksson, J., Karlsson, A. and Dijkstra, J., 2024. Trans-scale spatial variability of lime-cement mixed columns. *Construction and Building Materials*, *417*, p.135394.