



Comparación de los algoritmos random forest y gradient boosting para una estimación global del índice de compresión

Comparison of random forest and gradient boosting algorithms for global compression ratio estimation

Comparação de algoritmos de floresta aleatória e de aumento de gradiente para estimativa da taxa de compressão global

ARTÍCULO ORIGINAL

Jaime Yelsin Rosales Malpartida

jrosalesm@uni.pe

<https://orcid.org/0000-0003-4574-5172>

Facultad de Ingeniería Civil

Universidad Nacional de Ingeniería, Lima, Perú

César Loo Gil

cesarlo@biofab.com.pe

<https://orcid.org/0000-0001-8396-5972>

Científico Investigador de Biofab Inc. y del Centro de

Investigación y Producción Científica IDEOs, Lima - Perú

Recibido 02 de Mayo 2025 | Arbitrado y aceptado 02 de Junio 2025 | Publicado el 16 de Julio 2025

RESUMEN

El asentamiento de las estructuras está determinado por la rigidez del suelo donde se construyen. El índice de compresión (Cc) cuantifica la compresibilidad del suelo y es un parámetro clave en el diseño de estructuras geotécnicas, por lo que desarrollar un modelo de Machine Learning para estimar el índice de compresión Cc podría ser una solución muy valiosa en el campo de la geotecnia que permitiría a los ingenieros obtener estimaciones rápidas y fiables sin la necesidad de esperar los 15 días que generalmente tarda un ensayo de consolidación. Por lo tanto, la presente investigación tiene como objetivo estimar el índice de compresión, por lo que se recopiló y analizó una base de datos de 230 puntos de datos obtenidas del Laboratorio N°2 de Mecánica de Suelos de la Facultad de Ingeniería Civil de la UNI. Se comparó dos modelos de Machine Learning basado en Random Forest (RF) y Gradient Boosting Machine (GBM) para estimar el Cc a partir de las variables de entradas las cuales son límite líquido, el índice de plasticidad, relación de vacíos inicial y el contenido de agua natural. Los resultados indicaron que el modelo más adecuado para estimar el Cc fue el random forest que tuvo menores errores en la fase de entrenamiento y prueba con respecto al GBM; para optimización de hiperparámetros se utilizó el método de grid search con validación cruzada.

Palabras claves: Machine Learning, Índice de compresión, Límite líquido, Índice de plasticidad, Contenido de agua natural y relación de vacíos inicial.

ABSTRACT

The settlement of structures is determined by the stiffness of the soil on which they are built. The compression index (Cc) quantifies soil compressibility and is a key parameter in the design of geotechnical structures. Therefore, developing a Machine Learning model to estimate the compression index Cc could be a highly valuable solution in the field of geotechnics, allowing engineers to obtain quick and reliable estimates without having to wait the typical 15 days required for a consolidation test. Accordingly, the present study aims to estimate the compression index by compiling and analyzing a dataset of 230 data points obtained from Soil Mechanics Laboratory No. 2 of the Faculty of Civil Engineering at UNI (FIC-UNI). Two Machine Learning models—Random Forest (RF) and Gradient Boosting Machine (GBM)—were compared to estimate Cc using input variables including liquid limit, plasticity index, initial void ratio, and natural water content. The results indicated that the Random Forest model was more suitable for estimating Cc, achieving lower error values in both training and testing phases compared to GBM. Hyperparameter optimization was performed using the grid search method with cross-validation.

Keywords: Artificial intelligence, XGBoost, Soil mechanical properties, Roads

RESUMO

O recalque das estruturas é determinado pela rigidez do solo onde são construídas. A taxa de compressão (Cc) quantifica a compressibilidade do solo e é um parâmetro-chave no projeto de estruturas geotécnicas, portanto, desenvolver um modelo de aprendizado de máquina para estimar a taxa de compressão Cc pode ser uma solução muito valiosa no campo da geotecnia, que permitiria aos engenheiros obter estimativas rápidas e confiáveis sem a necessidade de esperar os 15 dias que um teste de consolidação geralmente leva. Portanto, a presente pesquisa tem como objetivo estimar o índice de compressão, para o qual foi coletado e analisado um banco de dados de 230 pontos de dados obtidos no Laboratório de Mecânica dos Solos nº 2 da Faculdade de Engenharia Civil da UNI. Dois modelos de aprendizado de máquina baseados em Random Forest (RF) e Gradient Boosting Machine (GBM) foram comparados para estimar o Cc a partir das variáveis de entrada que são limite de liquidez, índice de plasticidade, índice de vazios inicial e teor de água natural. Os resultados indicaram que o modelo mais adequado para estimar o Cc foi o random forest, que apresentou menores erros na fase de treinamento e teste quando comparado ao GBM; Para otimização de hiperparâmetros, foi utilizado o método de busca em grade com validação cruzada.

Palavras-chave: Aprendizado de máquina, índice de compressão, limite de liquidez, índice de plasticidade, teor de água natural e índice de vazios inicial.

1. INTRODUCCIÓN

El índice de compresibilidad (C_c) que es obtenida del gráfico e vs $\text{Log } \sigma'$, y el coeficiente de consolidación (cv) que es obtenida del gráfico δ vs $\sqrt{t}$, definen primordialmente las propiedades de compresibilidad de los suelos de grano fino. (Craig 2004). El comportamiento de la compresibilidad y deformación bajo carga de un suelo es muy usado en la ingeniería geotécnica para estimar el asentamiento y la deformación de estructuras del suelo, como cimientos y terraplenes (Das, 2021). Los suelos con un cc más alto tiende a ser más compresibles y deformables bajo carga, mientras que los suelos con un cc más bajo son menos compresibles y tienen una mayor capacidad para soportar la deformación cuando están cargados. Ambos valores cc y cv son determinados a partir de ensayos edométricos en la primera representa la pendiente de la parte líneas de la curva de compresión en un gráfico antes mencionado y en la segunda la tasa de cambio del asentamiento con respecto al tiempo. Estos parámetros de estado se han utilizado para establecer correlaciones con diferentes propiedades de la ingeniería en los suelos. Además, la plasticidad está influenciada por el comportamiento electroquímico de minerales arcillosos (Carter y Bentley, 1991).

2. ANTECEDENTES

Hay diversos autores que proporcionaron ecuaciones lineales que relacionan C_c con el límite líquido (LL) de suelos (como, Azzouz et al., Bowles, 1979; Park y Lee, 2011; Sridharan y Nagaraj, 2000; Terzaghi et al., 1967; Tsuchida, 1991). Por otro lado, el índice de plasticidad (PI), también se relaciona con cc (por ejemplo, Sridharan y Nagaraj, 2000; Wroth and Wood, 1978). Además, muchas correlaciones basado en una relación lineal con el contenido de agua natural (w) fueron propuestos (por ejemplo, Azzouz et al., 1976; Koppula, 1981; Rendon-Herrero, 1980).

Se analizó la relación con la relación de vacíos inicial e_0 por Nishida (1956), Hough (1957) y Bowles (1979). Otros estudios incluyeron más de una propiedad índice en la estimación de cc , como w y LL (Koppula, 1981) o e_0 y LL (Al-Khafaji y Andersland, 1992). Sin embargo, cuando estas correlaciones se probaron con nuevos datos, mostraron una dispersión significativa, con desviaciones que alcanzaron el 30% (Spagnoli y Shimobe, 2020), lo que nos hace pensar en una falta de universalidad y validez aplicable. Sin embargo, son aplicables dentro de límites específicos y deben

estar restringidos al tipo de suelo o ubicación donde fueron validados (Verbrugge y Schroeder, 2018). Utilizando estas correlaciones en diferentes condiciones pueden conducir a resultados insatisfactorios (Onyejekwe et al., 2016).

Para abordar los límites de los enfoques de regresión clásicos en ingeniería geotécnica, la aplicación de algoritmos de aprendizaje automático (ML) se han desarrollado ampliamente demostrando un rendimiento mejorado para predecir varias propiedades de ingeniería del suelo en comparación a los métodos estadísticos tradicionales (por ejemplo, Dam Nguyen et al; Bardhan et al., 2023; 2022 Díaz y Tomás, 2021; Singh et al, 2023). Sin embargo, es de suma importancia ser consciente de las limitaciones e incertidumbres asociadas con enfoques de ML antes de aplicarlos a la geotecnia de proyectos de ingeniería en el mundo real. Numerosos estudios (Baghbani et al., 2022; Zhang et al., 2023) han expuesto estas limitaciones, que son principalmente: a) la escasez de datos de alta calidad, b) la dificultad para interpretación de los modelos, y c) la falta de generalización. Muy aparte también para la disponibilidad de datos, estos pueden llegar a ser muy costosos y a menudo incompletos o inciertos.

Esto puede llevar a modelos de ML que no sean tan precisos o confiables como se desea. Otra limitación del ML es la interpretabilidad de los modelos. Los modelos de ML suelen ser complejos y no lineales, lo que dificulta comprender la relación entre los datos de entrada y predicciones de salida. Finalmente, debido a la heterogeneidad inherente y variabilidad espacial de los depósitos del suelo, es difícil para los modelos empíricos capacitado en conjuntos de datos limitados para extrapolar de manera confiable más allá del ámbito geográfico representado por esos datos. Por lo tanto, mientras ML muestra gran potencial para complementar los enfoques tradicionales en ingeniería geotécnica, es crucial abordar estas limitaciones antes de la implementación de cualquier modelo de ML.

Recientemente, numerosos estudios han empleado algoritmos de ML para predecir cc de algunos parámetros relacionados con esta propiedad que se muestran prometedores resultados (por ejemplo, Desai et al., 2009; Kalantary y Kordnaei, 2012; Kurmar y Rani, 2011; Nesamatha y Arumairaj, 2015).

Por otro lado, Alam et al. (2014) construyó una base de datos de 125 muestras de arcilla, donde incluyó w , LL , e_0 y PI como variables de entrada y creó un modelo de red neuronal artificial (RNA) para predecir C_c . Kumar y Rani (2011) también utilizó una RNA para predecir cc considerando 41 muestras con las siguientes variables de entrada: contenido fino, LL , PI , densidad seca máxima y óptimo contenido de humedad. Por otra parte, Park y Lee (2011) desarrollaron una RNA utilizando 947 pruebas de consolidación realizadas en muestras de suelo recolectadas de 67 zonas de construcción en la República de Corea, y consideraron variables de entrada como w , LL , PI , e_0 , gravedad específica de partículas del suelo (G_s), y porcentaje en peso de arena, limo y arcilla. Benbouras et al. (2019) desarrollaron un modelo RNA con 373 edómetros, muestras de prueba para correlacionar cc con densidad húmeda, w , e_0 , contenido fino, LL , PI , y tipo de suelo.

Zhang et al. (2021) utilizaron un algoritmo de random forest utilizando una base de datos de 311 muestras con tres variables de entrada (LL , PI , e_0).

Cabe mencionar que la relación de vacíos en LL es un parámetro que no suele estar disponible en las fases de diseño de proyectos geotécnicos, mientras que e_0 se encuentra con mayor frecuencia. Todos los estudios presentados confían en un conjunto de datos limitado o las muestras recolectadas tienen una naturaleza local, es decir que provengan de la misma zona o país. Por lo tanto, el objetivo principal de este trabajo es encontrar el mejor algoritmo entre Gradient Boosting machine y Random Forest basado en un conjunto de datos con 230 muestras. El presente artículo aprovechó los estudios de conjunto, para desarrollar y validar un modelo de predicción para estimar el índice de compresión de suelos para 230 datos recopilados por muestras recogidos de diversos proyectos ejecutados en Perú en el Laboratorio de Mecánica de Suelos de la Facultad de Ingeniería Civil de la Universidad Nacional de Ingeniería.

3. Metodología

a. Base de datos experimental

La recopilación del conjunto de datos es el primer paso en la construcción de un modelo de Machine Learning (ML). En el presente trabajo, la base de datos experimental de cc (230 muestras) son datos propios del autor correspondientes a las

muestras generadas en el Laboratorio de Ingeniería Geotécnica de la Facultad de Ingeniería Civil en la Universidad Nacional de Ingeniería. Se pudo apreciar que la base de datos tiene suelos desde baja plasticidad hasta alta plasticidad. De hecho, LL oscila entre 20% y 88% e IP entre 3% y 54%. Por otro lado, la base de datos incluía suelos con alta compresibilidad y suelos con baja compresibilidad según valores de e_0 (que van desde 0.295 a 1.425). Este hecho también se corrobora al examinar los valores de Cc variando estos entre 0.009 y 0.438. Finalmente, los valores de w, también varían ampliamente, desde el 1.18% hasta 50.7%. Con base en todo lo anterior, la base de datos incluye una cantidad considerable de datos, haciéndolo adecuado para un estudio provisional. El conjunto de datos final se puede encontrar en tabla s1 (accesible en línea), que incluye detalles como referencias, tipos de suelo, mineralogía y origen cuando estén disponibles.

La figura 1 muestra diagramas de caja para LL, IP, e_0 , w y cc de la base de datos. Los gráficos se utilizan para visualizar la dispersión de los datos, que se dividió en cuartiles; esto se utiliza para detectar valores atípicos, simetría de datos, dispersión y asimetría (Reagan y Kiemele, 2008). El cuadro en un diagrama de caja muestra el rango intercuartil (IQR), donde la parte inferior y superior del cuadro representan los percentiles 25 y 75, respectivamente. La altura de la caja representa el rango intercuartil. Los valores atípicos se definen como puntos de datos que se extienden hasta $3 \times \text{IQR}$ (Reagan y Kiemele, 2008). Algunos puntos en la Fig. 1 se identifican como valores atípicos, que corresponden principalmente a valores elevados de las variables; este análisis revela el rango de variación y la elevada dispersión de las variables estudiadas, en clara relación con el carácter mundial de la base de datos compilada. Para una mejor interpretación de la Fig. 1, se pueden apreciar los datos incluidos en la Tabla I.

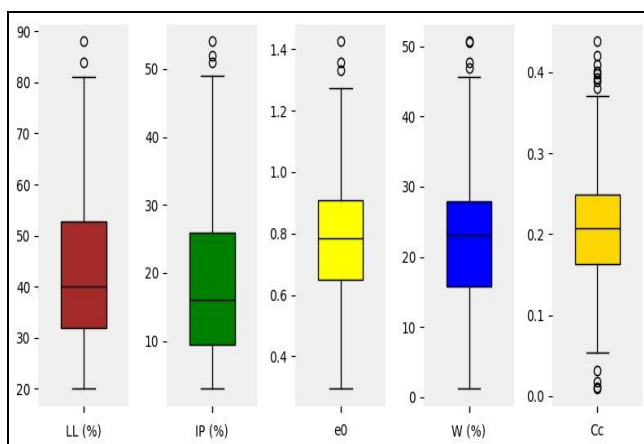


Fig. 1. Diagramas de caja de las variables consideradas.

Tabla I

Estadística descriptiva de los datos analizados

	LL (%)	IP (%)	W (%)	e0	Cc
Muestra	230	230	230	230	230
Media	43.86	18.46	22.63	0.79	0.21
Desviación estándar	14.81	11.49	9.69	0.20	0.08
Mínimo	20	3	1.18	0.295	0.009
Percentil 25	32	9.53	15.8	0.65	0.163
Mediana	40	16	23.15	0.7855	0.2075
Percentil 75	52.75	26	27.9	0.91	0.249
Máximo	88	54	50.7	1.425	0.438

b. Proceso de selección de modelo

Según el análisis de datos realizado en el anterior apartado, todos tienen la misma cantidad de datos (230). En este caso no se usó la técnica llamada imputación de datos más que para el análisis de datos atípicos, la cual es una técnica ampliamente utilizada en algoritmos de ML para tratar con valores faltantes y ha sido utilizado en temas geotécnicos satisfactoriamente, por ejemplo Aydin et al., 2023; Díaz et al., 2023.

Luego se realizó un análisis de detección de valores atípicos, utilizando el algoritmo SVM de una clase desarrollado por Scholkopf et al. (1999) y empleado con éxito en trabajos similares (por ejemplo Díaz et al., 2023). Este algoritmo es un método

diseñado para identificar valores atípicos y anomalías dentro de un conjunto de datos, utilizando los principios de las máquinas de vectores de soporte tradicionales (SVM).

Los valores atípicos son valores que se destacan significativamente del resto de los datos de un conjunto de datos. Pueden indicar una variabilidad extrema, errores de medición o, en algunos casos, fenómenos únicos que merecen especial atención. Identificar estos valores atípicos es importante porque pueden distorsionar los resultados estadísticos, como medias y varianzas, y afectar la validez de las conclusiones. Estos valores se identificaron con el diagrama de caja, que visualiza la distribución de un conjunto de datos. Los valores atípicos generalmente se representan como puntos aislados fuera de los “bigotes” del diagrama de caja. Los bigotes suelen extenderse hasta 1.5 veces el rango intercuartil (IQR) de los cuartiles (Q1 y Q3).

El proceso de selección de algoritmo de Machine Learning se llevó a cabo utilizando la técnica de validación cruzada de k veces (con k=5). Los algoritmos fueron realizados con Random Forest Regressor (Ho, 1995) y Gradient Boosting Regressor (Friedman, 2001).

Este proceso se realizó con las variables sin normalización. Pero cabe señalar que el mismo proceso de selección de algoritmos también se realizó la normalización de las variables mediante el método min-máx método, que escala cada variable individualmente entre cero y uno. Los resultados de este proceso de normalización de las variables, fueron exactamente lo mismo en ambos modelos con mejor desempeño.

c. Desarrollo de modelos

Para garantizar una adecuada generalización de los algoritmos, es buena práctica evaluar su desempeño con datos desconocidos. Para este propósito, el conjunto de datos se dividió en dos grupos (entrenamiento y prueba) con una partición 80% y 20% respectivamente. luego ambos algoritmos con mejor desempeño fueron sometidos a un proceso de ajuste de sus hiperparámetros para maximizar su desempeño. Para ello se utilizó la optimización de Grid search con validación cruzada (Yang L. y Shami A, 2020), esta técnica evalúa todas las combinaciones posibles de hiperparámetros en una cuadrícula predefinida y realizar múltiples rondas de validación para cada combinación. Los resultados de esta optimización están en la tabla V, en términos RMSE, MAE,

MAPE y R^2 , con el resultado tanto en el conjunto de entrenamiento como en el de la prueba.

Tabla II.

Matriz de correlación de las variables consideradas.

	LL (%)	IP (%)	e0	W (%)	Cc
LL (%)	1.00				
IP (%)	0.93	1.00			
e0	0.41	0.29	1.00		
W (%)	0.46	0.38	0.75	1.00	
Cc	0.33	0.29	0.65	0.49	1.00

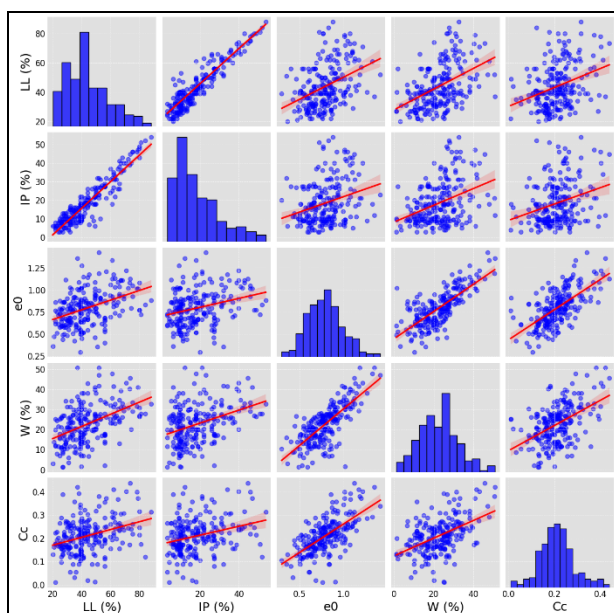


Fig. 2. Gráficos de dispersión e histogramas de distribución de las variables.

4. ANÁLISIS DE RESULTADOS

a. Estimación del índice de compresión (Cc).

En las tablas III y IV se muestran los hiperparámetros de los algoritmos de Gradient Boosting Machine y Random Forest respectivamente optimizados con el método Grid Search con validación cruzada considerando las variables de límite líquido, el índice de plasticidad, relación de vacíos inicial y el contenido de agua natural para estimar el Cc.

TABLA III

Hiperparámetros ajustados u optimizados de Gradient Boosting Machine para estimar el índice de compresibilidad.

Parámetros	Descripción	Valor	Rango
leargning_rate	Para reducir el paso de gradiente	0.01	0.01-0.1
max_depth	Profundidad del árbol	3	1 - 5
n_estimators	Construir el máximo número de árboles posible.	200	100-300
min_samples_split	número mínimo de muestras que se requieren para dividir un nodo en un árbol de decisión.	2	2-10
min_samples_leaf	Número mínimo de muestras que un nodo hoja debe tener después de que se realice una división.	1	1-4
subsample	Proporcion de muestras de entrenamiento que se utilizan para entrenar cada árbol.	0.8	0.8-1

TABLA IV

Hiperparámetros ajustados u optimizados de Random forest para estimar el índice de compresibilidad.

Parámetros	Descripción	Valor	Rango
max_depth	Profundidad del árbol	3	1 - 30
n_estimators	Construir el máximo número de árboles posible.	200	100 - 300
min_samples_split	número mínimo de muestras que se requieren para dividir un nodo en un árbol de decisión.	5	2-10
min_samples_leaf	Número mínimo de muestras que un nodo hoja debe tener después de que se realice una división.	2	1-4
Bootstrap	Proporcion de muestras de entrenamiento que se utilizan para entrenar cada árbol.	1	1 - 0

De tabla V se determinan los valores de las métricas de desempeño (RMSE, MAE, MAPE y R^2) para el Cc, mediante el algoritmo Gradient Boosting Machine y Random Forest para el escenario de entrenamiento (training) y escenario de prueba (testing). Se pudo apreciar que en el escenario de entrenamiento y prueba el modelo

Random Forest tuvo un $R^2 = 0.74$ y $R^2 = 0.38$ respectivamente, para el Gradient Boosting Machine se pudo apreciar que en el escenario de entrenamiento y prueba se tuvo un $R^2 = 0.67$ y $R^2 = 0.33$ respectivamente; y dado que los valores de RMSE, MAE y MAPE para el escenario de entrenamiento y prueba son menores ligeramente para RF con respecto al GBM, por lo tanto, se concluye que es el mejor modelo para estimar el índice de compresión Cc para el presente estudio.

Tabla V.

Resumen de las métricas de rendimiento.

Modelo	Conjunto	RMSE	MAE	MAPE	R^2
Random Forest	Training	0.039	0.030	31.811	0.74
	Test	0.060	0.042	38.056	0.38
Gradient Boosting Machine	Training	0.046	0.036	35.725	0.67
	Test	0.062	0.043	38.585	0.33

CONCLUSIONES

- Para la estimación del índice de compresión (Cc), el modelo Random Forest obtuvo su configuración óptima con una profundidad máxima de 3, 200 árboles, un mínimo de 5 muestras para dividir un nodo, 2 muestras por nodo hoja, y utilizando la técnica de bootstrap para seleccionar subconjuntos de entrenamiento en cada árbol. Esta configuración ofreció un balance adecuado entre complejidad y generalización del modelo.
- Al comparar ambos algoritmos, Random Forest demostró ser el más adecuado para la estimación del índice de compresión Cc, ya que presentó menores valores de error en las fases de entrenamiento y prueba respecto al modelo Gradient Boosting. Esto sugiere una mayor estabilidad y robustez en su capacidad predictiva bajo las condiciones de la base de datos analizada.
- Aunque los coeficientes de determinación R^2 obtenidos en la fase de prueba fueron inferiores a 0.5 para ambos modelos, lo que sugiere una capacidad predictiva aún limitada, este resultado no invalida el aporte del presente trabajo. Más bien, subraya la complejidad inherente al fenómeno estudiado y la

necesidad de ampliar la base de datos, así como de incorporar nuevas variables que capturen mejor el comportamiento del índice de compresión. En este marco, el estudio representa un avance significativo en la incorporación de técnicas de machine learning en la ingeniería geotécnica, al proponer un enfoque alternativo para la estimación del C_c que podría, en el futuro, reducir la dependencia de los ensayos de consolidación convencionales y optimizar los tiempos de análisis en proyectos de ingeniería.

REFERENCIAS

- [1] Craig, R.F., 2004. Craig's Soil Mechanics. CRC press.
- [2] Das, B.M., 2021. Principles of Geotechnical Engineering. Cengage learning.
- [3] Carter, M., Bentley, S.P., 1991. Correlations of Soil Properties. Pentech press publishers.
- [4] Azzouz, A.S., Krizek, R.J., Corotis, R.B., 1976. Regression analysis of soil compressibility. Soils Found. 16, 19–29.
https://doi.org/10.3208/sandf1972.16.2_19.
- [5] Bowles, J.E., 1979. Physical and Geotechnical Properties of Soils.
- [6] Park, H.I., Lee, S.R., 2011. Evaluation of the compression index of soils using an artificial neural network. Comput. Geotech. 38, 472–481.
- [7] Sridharan, A., Nagaraj, H., 2000. Compressibility behaviour of remoulded, fine-grained soils and correlation with index properties. Can. Geotech. J. 37, 712–722.
- [8] Terzaghi, K., Peck, R.B., Mesri, G., 1967. Soil Mechanics in Engineering Practice, 2nd edition. John Wiley and Sons, New York.
- [9] Tsuchida, T., 1991. A new concept of e-logp relationship for clays. In: Proce, 9th Asian Regional Conference on Soil Mechanics and Foundation Engineering, pp. 87–90.

- [10] Sridharan, A., Nagaraj, H., 2000. Compressibility behaviour of remoulded, fine-grained soils and correlation with index properties. *Can. Geotech. J.* 37, 712–722.
- [11] Wroth, C., Wood, D., 1978. The correlation of index properties with some basic engineering properties of soils. *Can. Geotech. J.* 15, 137–145.
- [12] A.S., Krizek, R.J., Corotis, R.B., 1976. Regression analysis of soil compressibility. *Soils Found.* 16, 19–29. https://doi.org/10.3208/sandf1972.16.2_19.
- [13] Koppula, S., 1981. Statistical estimation of compression index. *ASTM. Geotech. Test. J.* 4.
- [14] Rendon-Herrero, O., 1980. Universal compression index equation. *J. Geotech. Eng. Div.* 106, 1179–1200.
- [15] Nishida, Y., 1956. A brief note on compression index of soil. *J. Soil Mech. Found. Div.* 82, 1027-1021-1027-1014. Onyejekwe, S., Kang, X., Ge,
- [16] Hough, B.K., 1957. *Basic Soils Engineering*.
- [17] Al-Khafaji, A., Andersland, O., 1992. Equations for compression index approximation. *J. Geotech. Eng.* 118, 148–153. [https://doi.org/10.1061/\(ASCE\)0733-9410\(1992\)118:1\(148\)](https://doi.org/10.1061/(ASCE)0733-9410(1992)118:1(148)). Apostolopoulou,
- [18] Spagnoli, G., Shimobe, S., 2020. Statistical analysis of some correlations between
- [19] Verbrugge, J.-C., Schroeder, C., 2018. *Geotechnical Correlations for Soils and Rocks*. John Wiley & Sons.
- [20] Onyejekwe, S., Kang, X., Ge, L., 2016. Evaluation of the scale of fluctuation of geotechnical parameters by autocorrelation function and semivariogram function. *Eng. Geol.* 214, 43–49
- [21] Bardhan, A., Alzo'ubi, A.K., Palanivelu, S., Hamidian, P., GuhaRay, A., Kumar, G., Tsoukalas, M.Z., Asteris, P.G., 2023. A hybrid approach of ANN and improved PSO for estimating soaked CBR of subgrade soils of heavy-haul

railway corridor. Int. J. Pavem. Eng. 24, 2176494.
<https://doi.org/10.1080/10298436.2023.2176494>.

- [22] Dam Nguyen, D., Roussis, P.C., Thai Pham, B., Ferentinou, M., Mamou, A., Quang, Vu, Thi Bui, Q.-A., Kien Trong, D., Asteris, P.G., 2022. Bagging and Multilayer Perceptron Hybrid Intelligence Models predicting the Swelling potential of Soil. Transp. Geotech. 36, 100797
<https://doi.org/10.1016/j.trgeo.2022.100797>.
- [23] Díaz, E., Tomás, R., 2021. Upgrading the prediction of jet grouting column diameter using deep learning with an emphasis on high energies. Acta Geotech. 16, 1627–1633. <https://doi.org/10.1007/s11440-020-01091-8>.
- [24] Singh, M.J., Kaushik, A., Patnaik, G., Xu, D.-S., Feng, W.-Q., Rajput, A., Prakash, G., Borana, L., 2023. Machine learning-based approach for predicting the consolidation characteristics of soft soil. Mar. Georesour. Geotechnol. 1-15
<https://doi.org/10.1080/1064119X.2023.2193174>.
- [25] Baghbani, A., Choudhury, T., Costa, S., Reiner, J., 2022. Application of artificial intelligence in geotechnical engineering: A state-of-the-art review. Earth Sci. Rev. 228, 103991. Zhang, W., Gu, X., Hong, L., Han, L., Wang, L., 2023. Comprehensive review of machine learning in geotechnical reliability analysis: Algorithms, applications and further challenges. Appl. Soft Comput. 136, 110066 <https://doi.org/10.1016/j.asoc.2023.110066>.
- [26] Desai, V.M., Desai, V., Rao, D., 2009. Prediction of compression index using artificial neural networks. In: Indian Geotechnical Conference (IGC-2009), Guntur, India, pp. 614–617.
- [27] Kalantary, F., Kordnaeij, A., 2012. Prediction of compression index using artificial neural network. Sci. Res. Essays 7, 2835–2848.
- [28] Kumar, V.P., Rani, C.S., 2011. Prediction of compression index of soils using artificial neural networks (ANNs). Int. J. Eng. Res. Appl. 1, 1554–1558.

- [29] Nesamatha, R., Arumairaj, P., 2015. Numerical modeling for prediction of compression index from soil index properties. *Electron. J. Geotech. Eng.* 20, 4369–4378.
- [30] Alam, S., Khuntia, S., Patra, C., 2014. Prediction of compression index of clay using artificial neural network. In: *International Conference on Industrial Engineering Science and Applications-NIT, Durgapur*.
- [31] Park, H.I., Lee, S.R., 2011. Evaluation of the compression index of soils using an artificial neural network. *Comput. Geotech.* 38, 472–481.
- [32] Benbouras, M.A., Kettab Mitiche, R., Zedira, H., Petrisor, A.-I., Mezouar, N., Debiche, F., 2019. A new approach to predict the compression index using artificial intelligence methods. *Mar. Georesour. Geotechnol.* 37, 704–720.
- [33] Zhang, P., Yin, Z.-Y., Jin, Y.-F., Chan, T.H.T., Gao, F.-P., 2021. Intelligent modelling of clay compressibility using hybrid meta-heuristic and machine learning algorithms. *Geosci. Front.* 12, 441–452. <https://doi.org/10.1016/j.gsf.2020.02.014>.
- [34] Reagan, L.A., Kiemele, M.J., 2008. *Design for Six Sigma: The Tool Guide for Practitioners*.
- [35] Aydın, Y., Isıkdağ, Ü., Bekdas, G., Nigdeli, S.M., Geem, Z.W., 2023. Use of Machine Learning Techniques in Soil Classification. *Sustainability* 15, 2374.
- [36] Díaz, E., Salamanca-Medina, E.L., Tomás, R., 2023. Assessment of compressive strength of jet grouting by machine learning. *J. Rock Mech. Geotech. Eng.* <https://doi.org/10.1016/j.jrmge.2023.03.008>.
- [37] Schölkopf, B., Williamson, R.C., Smola, A., Shawe-Taylor, J., Platt, J., 1999. Support vector method for novelty detection. *Adv. Neural Inf. Proces. Syst.* 12.
- [38] Ho, T.K., 1995. Random decision forests. In: *Proceedings of 3rd International Conference on Document Analysis and Recognition*. IEEE, pp. 278–282.
- [39] Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine, pp. 1189–1232.

- [40] Yang, L., & Shami, A. (2020). On hyperparameter optimization of machine learning algorithms: Theory and practice. *Neurocomputing*, 415, 295-316.

Financiamiento de la investigación

Con recursos propios.

Declaración de intereses

Declaro no tener ningún conflicto de intereses, que puedan haber influido en los resultados obtenidos o las interpretaciones propuestas.

Declaración de consentimiento informado

El estudio se realizó respetando el Código de ética y buenas prácticas editoriales de publicación.

Derechos de uso

Copyright© 2025 por **Jaime Yelsin Rosales Malpartida, César Loo Gil**



[Este texto está protegido por la Licencia Creative Commons Atribución 4.0 Internacional.](https://creativecommons.org/licenses/by/4.0/)

Usted es libre para compartir, copiar y redistribuir el material en cualquier medio o formato y adaptar el documento, remezclar, transformar y crear a partir del material para cualquier propósito, incluso comercialmente, siempre que cumpla la condición de atribución: usted debe reconocer el crédito de una obra de manera adecuada, proporcionar un enlace a la licencia, e indicar si se han realizado cambios. Puede hacerlo en cualquier forma razonable, pero no de forma tal que sugiera que tiene el apoyo del licenciante o lo recibe por el uso que hace.