

Aplicação de *Machine Learning* em Geotecnia: Previsão de Comportamento de Areia com Ensaios de Cisalhamento Direto

Gleyce de Souza Baptista, M.Eng

Eng^a Geotécnica, PUC-Rio, Rio de Janeiro, Brasil, gleycesouza@aluno.puc-rio.br

Marina Bellaver Corte, D.Eng

Professora Assistente, PUC-Rio, Rio de Janeiro, Brasil, marina@puc-rio.br

RESUMO: Este artigo explora a aplicação de técnicas de aprendizado de máquina — especificamente, *Random Forest* (RF), *Support Vector Regression* (SVR) e *Feedforward Neural Networks* (FNN) — para estimar curvas de tensão-deslocamento em ensaios de cisalhamento direto, um aspecto crítico em análises e projetos geotécnicos. A investigação centra-se na capacidade desses modelos de capturar a complexidade do comportamento do solo, como ferramenta alternativa complementar às análises geotécnicas, se opondo às limitações de métodos tradicionais baseados em correlações empíricas e teóricas. A pesquisa iniciou-se com uma seleção de 7 trabalhos acadêmicos que forneceram gráficos de tensão-deslocamento, do total de 205 ensaios, que foram digitalizados e formaram a base experimental que irá alimentar e treinar os algoritmos de Inteligência Artificial (IA) propostos. Os modelos foram avaliados utilizando métricas de desempenho como o *Mean Absolute Error* (MAE), a *Root Mean-Square Error* (RMSE) e o coeficiente de determinação (R^2), com o RF demonstrando superioridade em precisão e generalização. Além disso, a validação das previsões demonstrou potencial conservadorismo das estimativas, essencial para a segurança em projetos geotécnicos. Os resultados reafirmam o potencial das técnicas de aprendizado de máquina como ferramentas auxiliares valiosas para a engenharia geotécnica, mesmo com conjuntos de dados limitados. Este estudo sublinha a importância e a viabilidade de integrar abordagens computacionais avançadas na prática geotécnica.

PALAVRAS-CHAVE: Inteligência Artificial; Aprendizado de Máquina; Curva Tensão-Deslocamento; Ensaio de Cisalhamento Direto.

ABSTRACT: This article explores the application of machine learning techniques—specifically, Random Forest (RF), Support Vector Regression (SVR), and Feedforward Neural Networks (FNN)—to estimate stress-displacement curves in direct shear tests, a critical aspect in geotechnical analysis and projects. The investigation focuses on these models ability to capture the soil behavior's complexity, serving as a complementary alternative tool to geotechnical analyses, countering the limitations of traditional methods based on empirical and theoretical correlations. The research began with the selection of 7 academic papers that provided stress-displacement graphs, out of a total of 205 tests, which were digitized and formed the experimental base to feed and train the proposed Artificial Intelligence (AI) algorithms. The models were evaluated using performance metrics such as the Mean Absolute Error (MAE), the Root Mean-Square Error (RMSE), and the coefficient of determination (R^2), with RF demonstrating superiority in precision and generalization. Moreover, the validation of the predictions showed potential conservativeness of the estimates, essential for safety in geotechnical projects. The results reaffirm the potential of machine learning techniques as valuable auxiliary tools for geotechnical engineering, even with limited data sets. This study underscores the importance and feasibility of integrating advanced computational approaches into geotechnical practice.

KEYWORDS: Artificial intelligence; Machine Learning; Stress-Displacement Curve; Direct Shear Test.

1 INTRODUÇÃO

Entender a resistência ao cisalhamento do solo é essencial para a engenharia geotécnica, pois influencia diretamente a estabilidade e a segurança de estruturas civis e construções. Lambe e Whitman (1969) destacam que a resistência ao cisalhamento do solo granular seco é determinada por uma variedade de fatores, agrupados em dois conjuntos: um relacionado a características intrínsecas do solo, como índice de vazios, tensão

confinante e taxa de carregamento, e outro associado a fatores que diferenciam um solo de outro, tais como tamanho e forma das partículas e distribuição granulométrica.

Durante um cisalhamento drenado, identificam-se dois comportamentos volumétricos nas areias: contrátil, com redução de volume, e dilatante, com aumento de volume. Esses fenômenos dependem de uma série de fatores, incluindo o índice de vazios inicial (e_0) ou, mais especificamente, a compacidade relativa (CR), bem como a tensão normal (σ_v), no início do cisalhamento (Pinto, 2009).

A precisão na determinação dos parâmetros de resistência ao cisalhamento é crucial para análises e projetos geotécnicos. Tradicionalmente, esses parâmetros são estimados por meio de ensaios de campo e de laboratório, complementados por relações matemáticas empíricas e/ou teóricas. No entanto, essas abordagens podem não capturar toda a complexidade do comportamento do solo. Nesse contexto, a inteligência artificial (IA) emerge como uma ferramenta promissora. Recentemente, técnicas de computação, incluindo IA, têm sido usadas para resolver problemas variados em geociências e engenharia geotécnica, segundo Kanungo et al. (2014). Zhang et al. (2022) examinaram o avanço da utilização do aprendizado de máquina, do inglês *Machine Learning* (ML), na previsão de propriedades do solo, Figura 1, destacando sua crescente aplicação em engenharia e geologia desde 2018. O estudo mostra um aumento significativo no uso dessas técnicas, evidenciado por um gráfico com uma linha vermelha que representa o crescimento acumulado de publicações na área.

Nesse contexto, o presente trabalho tem como objetivo desenvolver e avaliar modelos preditivos baseados em técnicas de ML, como *Random Forest* (RF), *Support Vector Regression* (SVR) e *Feedforward Neural Networks* (FNN), para estimar as curvas de tensão-deslocamento a partir de ensaios de cisalhamento direto em areia.

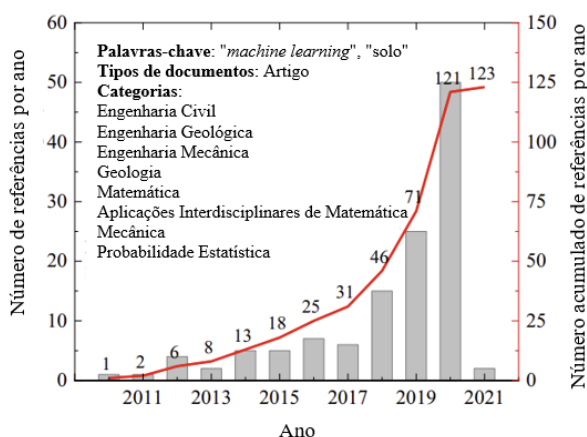


Figura 1 Número de publicações relacionadas ao desenvolvimento e aplicação de algoritmos de ML para previsão de propriedades do solo (Adaptado de Zhang et al., 2022)

1.1 Aprendizado de Máquina

De acordo com Goodfellow et al. (2016), a IA é um campo com muitas aplicações práticas e tópicos de pesquisa ativos. *Softwares* inteligentes são procurados para automatizar o trabalho de rotina, compreender a fala ou imagens, fazer diagnósticos em medicina e apoiar a investigação científica básica. Um modelo de ML pode ser definido, de acordo com Mohri et al. (2018), como métodos computacionais que utilizam a experiência (dados passados ao modelo) para melhorar o desempenho ou fazer previsões precisas.

O objetivo do modelo de ML é aprender uma função (h) que seja uma aproximação próxima da função desconhecida (f). O desafio é encontrar um equilíbrio entre a simplicidade do modelo e sua acurácia nas previsões, considerando as características dos dados e as necessidades específicas da aplicação (Baranauskas et al., 2000). Na Figura 2 (a), vemos um modelo que se ajusta demais aos dados de treinamento, incluindo possíveis ruídos, o que é um exemplo de *overfitting*. Já a Figura 2 (c) mostra *underfitting*, onde o classificador falha em se ajustar até mesmo aos dados de treinamento. A Figura 2 (b) ilustra um equilíbrio, com um preditor de complexidade média que classifica a maioria dos dados corretamente sem focar excessivamente em pontos isolados (Lorena e Carvalho, 2007).

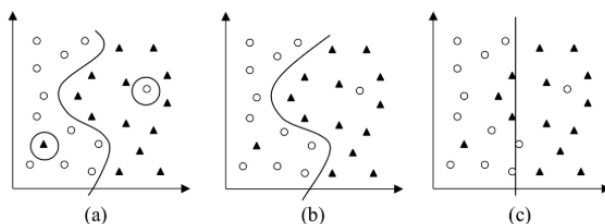


Figura 2 Conjunto de treinamento e três hipóteses diferentes (Lorena e Carvalho, 2007)

1.1.1 Decision Trees (DT) e Random Forest (RF)

Árvores de decisão (*Decision Trees* – DT) são técnicas de aprendizado de máquina que segmentam os dados de previsão em regiões simplificadas, utilizando a média ou a moda das respostas das observações de treinamento para fazer previsões (James et al., 2023).

Breiman (2001) descreve a *Random Forest* (RF) como um conjunto de DT que trabalham juntas como classificadores. Cada árvore usa vetores aleatórios para dar um voto à classe mais frequente para uma dada entrada em problemas de classificação. Zhang et al. (2022) ampliam essa definição, explicando que a floresta aleatória combina várias DT, cada uma atuando como um modelo preditivo independente, para formar um "comitê" que faz decisões coletivas. Em tarefas de regressão, o resultado é a média das previsões de todas as árvores. A criação de uma RF: inicia com a base de treino, cria-se subconjuntos para treinar árvores de decisão individuais e, finalmente, agrega-se os resultados dessas árvores para obter a predição final. Raschka e Mirjalili (2017) afirmam que, na última década, as RF se destacaram no campo do aprendizado de máquina por sua eficiência em classificação, escalabilidade e facilidade de uso.

Zhang et al. (2022) destacam que na configuração de RF, dois hiperparâmetros são essenciais: 'ntree', o número de DTs, e 'mtry', o número de atributos considerados em cada divisão de nó (eq. 1). Esses parâmetros determinam a estrutura e eficácia do modelo. A construção das DTs usa o método de *bagging* (*bootstrap aggregating*), gerando múltiplos conjuntos de treinamento através de amostragem com substituição. Geralmente, cada conjunto inclui dois terços dos dados originais, deixando um terço "out of the bag" para teste e avaliação de erro e desempenho da árvore.

1.1.2 Support Vector Regression (SVR)

As máquinas de vetores de suporte, *Support Vector Machine* (SVM), são embasadas pela teoria de aprendizado estatístico (TAE). Essa teoria estabelece uma série de princípios que devem ser seguidos na obtenção de classificadores com boa generalização, definida como a sua capacidade de prever corretamente a classe de novos dados do mesmo domínio em que o aprendizado ocorreu (Lorena e Carvalho, 2007). A Figura 3 apresenta uma representação do modelo e seus elementos. O SVR trabalha a partir da minimização do erro, maximizando a margem e considerando a flexibilidade através das variáveis de folga (ξ), que permitem erros dentro de um certo limiar (ϵ). Os vetores de suporte são os pontos de dados mais próximos do hiperplano e têm uma função importante na definição da margem, que é a distância entre o hiperplano e esses pontos mais próximos.

No treinamento do SVR, o objetivo é descobrir um hiperplano que não apenas separe todos os conjuntos de dados, mas que também mantenha a maior "folga" ou margem possível entre as diferentes categorias, matematicamente ele pode ser expresso pela eq. 2 (Zhang et al., 2022). Onde m é um total de conjuntos de dados; ξ é o parâmetro de folga, ou margem, indicando o erro permitido; C é a regularização do parâmetro para impor uma penalidade em caso de erro. A função *kernel* comumente utilizada na base pode ser obtida a partir da eq. 3, onde γ_{kernel} é o coeficiente *kernel*.

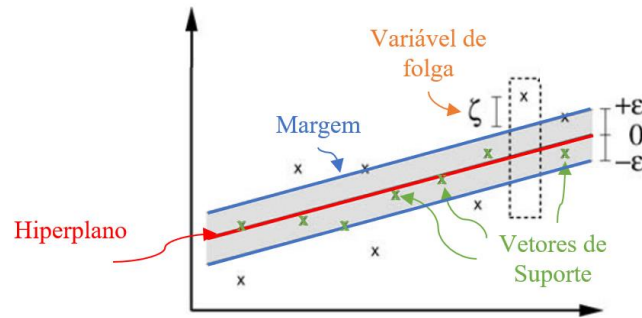


Figura 3 Representação do SVR (Adaptado de Scholkopf e Smola, 2002 apud Smola e Scholkopf, 2004)

$$\min_{\xi, w, b} \frac{1}{2} \|W\|^2 + C \sum_{i=1}^m \xi_i, \text{ s. a. } y_i (W^T \phi(X) + b) \geq 1 - \xi_i, \quad i = 1, 2, \dots, m, \quad \xi_i \geq 0 \quad (1)$$

$$\phi(x_i, x_j) = \exp\left(-\gamma_{kernel} \|x_i - x_j\|^2\right) \quad (2)$$

1.1.3 Feedforward Neural Network (FNN)

Conforme Fielding (1999) citou em seu trabalho, um modelo simplificado de neurônio em uma Rede Neural Artificial (RNA) replica a maneira como os neurônios biológicos recebem estímulos, processam sinais através de pesos sinápticos e disparam impulsos elétricos dependendo do nível de ativação. Russell e Norvig (2020) destacam a interação entre camadas das RNAs como um diferencial importante desses modelos em comparação a métodos lineares ou de decisão. Existem vários tipos de RNAs, incluindo as Redes Neurais *Feedforward*, que processam informações em uma única direção sem laços, as Convolucionais, especializadas em classificação de imagens, e as Recorrentes, adequadas para séries temporais e sequências (Russell e Norvig, 2020; James et al., 2023). *Feedforward Neural Network* (FNN) é o algoritmo de ML mais comumente usado, segundo Zhang et al. (2022), além de ser um dos modelos mais simples consistindo em camadas de entrada, ocultas e de saída e este será o modelo utilizado na presente pesquisa.

Uma rede neural processam vetores de entrada X_1, X_2, \dots, X_p usando uma função $f(X)$ para prever a resposta Y , operando numa estrutura de rede *feedforward* com p preditores. As ativações na camada oculta, A_k , são calculadas a partir das entradas através de funções de ativação $g(z)$ não lineares, como sigmoid ou ReLU, gerando $h_k(X)$. Estas são essenciais para a produção da resposta na camada de saída. James et al. (2023) argumentam a favor da ReLU em redes modernas devido à sua eficiência. O modelo realiza regressão linear sobre as ativações N , ajustando parâmetros, incluindo vieses (β_0, \dots, β_k) e pesos (W_{10}, \dots, W_{kp}), durante o treinamento. Aplicando funções de ativação $g(\cdot)$ a X , o modelo introduz uma não linearidade chave para capturar complexidades nos dados, permitindo a representação de relações complexas (James et al., 2023).

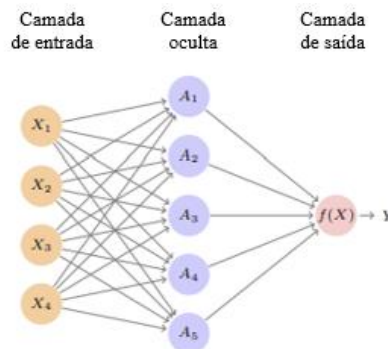


Figura 4 Exemplo de rede FNN com uma camada oculta (Adaptado de James et al., 2023)

$$f(X) = \beta_0 + \sum_{k=1}^K \beta_k A_k, \quad (3)$$

$$A_k = h_k(X) = g\left(W_{k0} + \sum_{j=1}^p W_{kj} X_j\right) \quad (4)$$

1.2 Medidas de Desempenho

Duas métricas comuns em problemas de regressão são o Erro Médio Absoluto (*Mean Absolute Error* - MAE) e a Raiz do Erro Quadrático Médio (*Root Mean Square Error* - RMSE). O MAE, eq. 6, calcula a média das diferenças absolutas entre previsões e valores reais, enquanto o RMSE, eq. 7, calcula a raiz quadrada da média dos erros ao quadrado, dando mais peso a erros maiores. O RMSE é sensível a *outliers*, diferentemente do MAE, que se assemelha à distância percorrida em quarteirões ortogonais em uma cidade, conhecida como norma Manhattan. Ambas as métricas fornecem uma medida da distância entre o vetor de previsões e o de valores reais (Baranauskas e Monard, 2000; Géron, 2019).

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - h(x_i)| \quad (5)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - h(x_i))^2} \quad (6)$$

2 MATERIAIS E MÉTODOS

2.1 Materiais

Este estudo se iniciou com uma pesquisa na literatura geotécnica em bases de dados, concentrando-se em ensaios de cisalhamento direto em areia com uso de palavras-chave direcionadas. Publicações, artigos, conferências e teses com dados experimentais em gráficos tensão-deslocamento foram selecionados. Excluindo materiais distintos de areia, selecionou-se 7 estudos com um total de 205 ensaios e 7.185 pares de dados de tensão-deslocamento. Dos trabalhos, 3 ensaios completos foram destinados para validação. Após interpolação e aplicação de um limite de deslocamento de 5 mm, resultaram 3.434 pares de dados. Estes foram divididos em conjuntos de treinamento (80%) e teste (20%). Os estudos selecionados são apresentados na Tabela 1, com especificação de autores, ano, título do trabalho, localização e quantidade de ensaios. Os parâmetros e as estatísticas descritivas dos dados usados na modelagem são apresentados na Tabela 2, que resume os valores mínimos e máximos para granulometria específica (G_s), porosidade inicial (e_0), taxa de consolidação (CR %), tensão vertical (σ_v) em kPa, e deslocamento horizontal (δ_h) em mm.

Tabela 1 Trabalhos utilizados para formação do banco de dados

Autor, Ano (ID)	Título do trabalho	Localização da areia	Quantidade de ensaios
Coutinho, 2021 (C21)	Ensaio de Cisalhamento Direto na Areia da Praia de Ipanema	Ipanema	71
Marques, 2009 (M09)	Ensaio de resistência ao cisalhamento com areia de Hokksund para projeto de revitalização da câmara de calibração	Hokksund	20
Monteiro et al., 2023 (M23)	Caracterização Geotécnica da Areia do Porto do Açú	Porto do Açú	16
Nunes, 2014 (N14)	Ensaio de caracterização geotécnica da areia da praia de Itaipuaçu	Itaipuaçu	29
Pinheiro, 2018 (P18)	Caracterização geotécnica em laboratório da areia da praia dos Cavaleiros – Macaé/RJ	Cavaleiros	24
Simões, 2015 (S15)	Caracterização geotécnica da areia da praia de Ipanema	Ipanema	30
Teles, 2013 (T13)	Estudo sobre os parâmetros de resistência e deformabilidade da areia de Hokksund	Hokksund	15

Tabela 2 Dados estatísticos dos *inputs* utilizados

-	G_s	e_0	CR (%)	σ_v (kPa)	δ_h (mm)
Mín	2.643	0.428	8	12.50	0
Máx	2.763	0.726	99	1600.00	5

2.2 Métodos

A metodologia adotada neste estudo consistiu em quatro etapas fundamentais: coleta de dados, pré-processamento, definição e treinamento das arquiteturas, e finalmente, a análise e validação.

Durante a coleta de dados, empregou-se uma abordagem sistemática para obter gráficos de tensão-deslocamento dos estudos selecionados. Os dados foram digitalizados utilizando o *software PlotDigitizer* (2024), facilitando sua conversão para formatos numéricos aptos para análise estatística. Além disso, informações complementares relevantes aos materiais e métodos de ensaio foram detalhadamente analisadas. Este processo resultou na criação de um banco de dados unificado.

O pré-processamento dos dados envolveu a limpeza, organização e transformação dos dados para garantir sua adequação para análises subsequentes, refletindo as condições experimentais.

A seleção da arquitetura dos modelos foi realizada através da técnica de busca em grade (*Grid Search*), explorando sistematicamente várias combinações de parâmetros para identificar a configuração mais eficiente. A implementação dessa técnica utilizou a ferramenta *GridSearchCV* da biblioteca *sklearn*, com validação cruzada para uma avaliação consistente dos modelos. A seleção dos melhores parâmetros baseou-se em métricas de desempenho como MAE, RMSE e o coeficiente de determinação (R^2), orientando o treinamento dos modelos.

Por fim, a análise de desempenho e validação dos modelos foi realizada utilizando MAE, RMSE e R^2 , para avaliar a acurácia e a capacidade de generalização dos modelos com dados de teste. Avaliações adicionais em um conjunto de validação buscaram verificar a aplicabilidade prática dos modelos, com uma análise comparativa entre eles visando identificar o desempenho superior com base nas métricas selecionadas.

3 RESULTADOS

Os resultados da avaliação de desempenho para as fases de treino e teste são fornecidos na Tabela 3. Na análise de desempenho, o modelo RF se destacou, apresentando as melhores métricas: MAE e RMSE menores (0.01 e 0.02, respectivamente) e um R^2 de 0.99 durante o treino, indicando alta precisão. No teste, o aumento do MAE e RMSE foi modesto (0.03 e 0.05, respectivamente), com um R^2 mantendo-se alto em 0.96, sugerindo boa generalização. Os modelos de SVR e FNN não alcançaram desempenho comparável. O SVR teve MAE e RMSE maiores em treino e teste em comparação com o RF, enquanto o FNN mostrou precisão reduzida no treino (MAE de 0.09 e RMSE de 0.13) e um R^2 mais baixo (0.74), performance que se manteve estável no teste, refletindo uma capacidade de generalização inferior.

Tabela 3 Avaliação de desempenho dos modelos

Modelo	Treino			Teste		
	MAE	RMSE	R^2	MAE	RMSE	R^2
RF	0.01	0.02	0.99	0.03	0.05	0.96
SVR	0.05	0.10	0.86	0.06	0.10	0.85
FNN	0.09	0.13	0.75	0.09	0.13	0.74

Na sequência do estudo, foram analisadas as curvas de resistência geradas pelos modelos para os conjuntos de validação, complementadas por uma avaliação do conservadorismo dessas previsões. Tal avaliação, conforme eq. 8, se baseia na comparação dos valores máximos previstos nos eixos y das curvas dos modelos com os observados em ensaios reais de validação. Uma diferença positiva indica uma previsão conservadora, potencialmente aumentando a segurança ao superestimar a resistência.

A Figura 5, demonstra os resultados dos modelos na previsão a relação tensão-deslocamento nos ensaios de validação, com previsões em vermelho e dados reais em azul. É possível perceber que o modelo de RF se alinhou bem com os dados, indicando uma modelagem precisa para diferentes condições do material, embora com ligeira subestimação da tensão máxima, variando de 0.01 a -0.03. O modelo SVR mostrou alta precisão em casos específicos, como no ID M23_27_300, mas apresentou divergências em outros, como no ID S15_88_150, refletindo variações na capacidade de captura de picos de tensão. As diferenças máximas entre as previsões e os dados reais foram moderadas, de -0.01 a 0.03. Por outro lado, o modelo de FNN teve menor aderência aos dados de validação, com diferenças variando de -0.03 a 0.11. Em dois dos três casos, o FNN tendeu a superestimar a tensão máxima, indicando uma propensão a favorecer a segurança. Conclui-se, a partir

dos resultados de validação, que os modelos RF e SVR oferecem uma combinação equilibrada de precisão e conservadorismo, adequados para prever curvas de tensão-deslocamento em cisalhamento direto. O FNN, embora apresente utilidade, requer uma avaliação cuidadosa.

$$dif = máx(y_{previsão}) - máx(y_{real}) \quad (7)$$

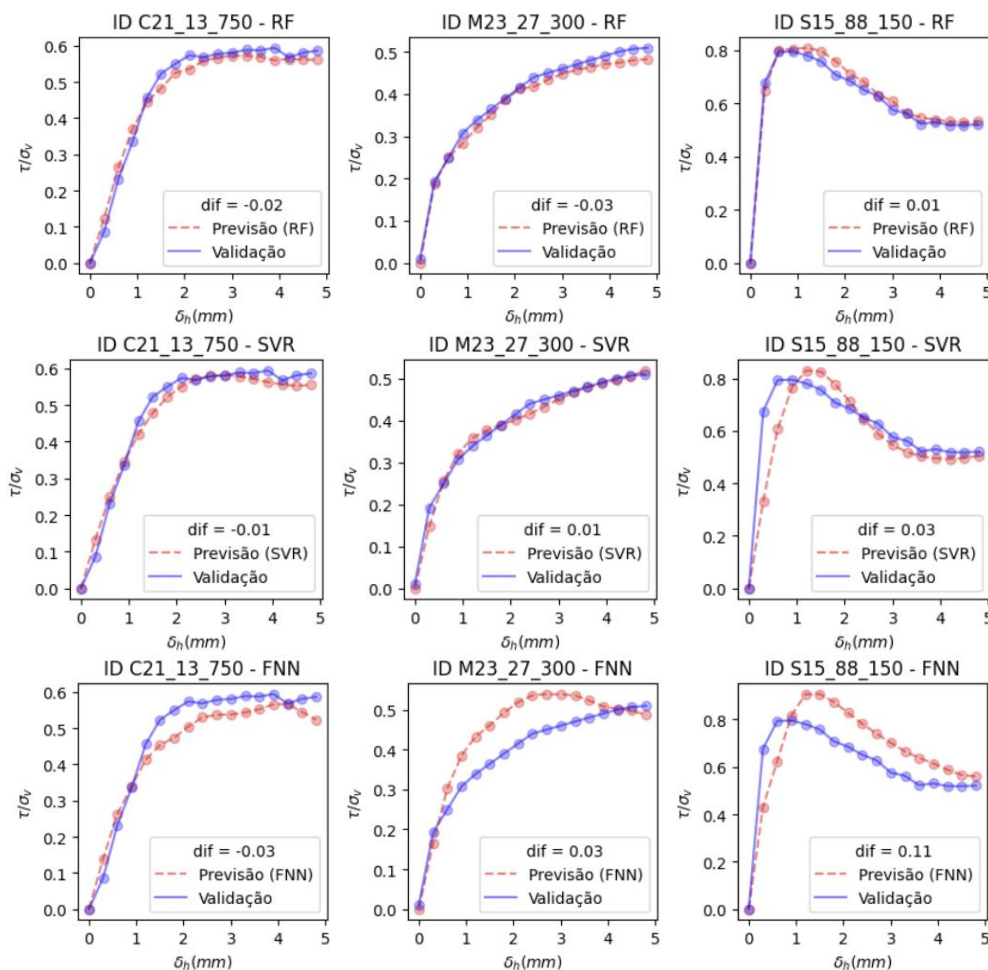


Figura 5 Curvas de previsão nos ensaios de validação

4 CONCLUSÕES

A avaliação dos modelos indicou que o modelo de RF apresentou desempenho superior aos demais, evidenciando elevada precisão e capacidade de generalização notável. Durante a fase de validação das previsões, o conservadorismo das estimativas foi examinado mediante a comparação entre os valores máximos previstos e os valores máximos observados, processo no qual o RF demonstrou consistência com as métricas de avaliação, refletindo uma boa modelagem. Em contrapartida, os modelos de SVR e FNN exibiram variações em sua habilidade de capturar e generalizar os dados. Mesmo diante de um conjunto de dados de dimensões limitadas, os resultados alcançados são promissores, sublinhando a viabilidade da aplicação dessas técnicas avançadas como ferramentas complementares na engenharia geotécnica. Este estudo reforça, portanto, o potencial significativo do emprego de métodos de aprendizado de máquina em melhorar a precisão e a confiabilidade das análises em projetos geotécnicos.

AGRADECIMENTOS

As autoras agradecem à CAPES (Coordenação de Aperfeiçoamento de Pessoal do Nível Superior) e à FAPERJ (Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro) pelo apoio financeiro.

REFERÊNCIAS BIBLIOGRÁFICAS

- BARANAUSKAS, J. A.; MONARD, M. C. *Reviewing Some Machine Learning Concepts and Methods*. São Carlos. 2000.
- BREIMAN, L. *Random Forests*. Kluwer Academic Publishers, v. 45, p. 5–32, 2001.
- COUTINHO, J. V. M. *Ensaio de Cisalhamento Direto na Areia da Praia de Ipanema*. 2021. Dissertação de Mestrado—Rio de Janeiro: Pontifícia Universidade Católica do Rio de Janeiro, 2021.
- FIELDING, A. H. *Machine Learning Methods for Ecological Applications*. Nova York: Springer US, 1999.
- GÉRON, A. *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. Second Edition ed. [s.l.] O'Reilly Media, Inc., 2019.
- GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep Learning*. [s.l.: s.n.].
- JAMES, G.; WITTEN, D.; HASTIE, T.; TIBSHIRANI, R.; TAYLOR, J. *An Introduction of Statistical Learning with Applications in Python*. [s.l.: s.n.].
- KANUNGO, D. P.; SHARMA, S.; PAIN, A. *Artificial Neural Network (ANN) and Regression Tree (CART) applications for the indirect estimation of unsaturated soil shear strength parameters*. *Frontiers of Earth Science*, v. 8, n. 3, p. 439–456, 20 set. 2014.
- LAMBE, T. WILLIAM.; WHITMAN, R. V. *Soil Mechanics*. Massachusetts: Massachusetts Institute of Technology, 1969.
- LORENA, A. C.; DE CARVALHO, A. C. P. L. F. *Uma Introdução às Support Vector Machines*. *Revista de Informática Teórica e Aplicada*, v. 14, n. 2, p. 43–67, 20 dez. 2007.
- MARQUES, F. DE L. *Ensaio de resistência ao cisalhamento com areia de Hokksund para projeto de revitalização da câmara de calibração*. 2009. Iniciação Científica—Rio de Janeiro: Universidade Federal do Rio de Janeiro (UFRJ) - Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do RJ (FAPERJ), 2009.
- MOHRI, M.; ROSTAMIZADEH, A.; TALWALKAR, A. *Foundations of Machine Learning second edition*. Second Edition ed. Cambridge: Massachusetts Institute of Technology, 2018. ISBN(9780262039406).
- MONTEIRO, D. P.; DANZIGER, B. R.; LIMA, B. T. *Caracterização Geotécnica da Areia do Porto do Açú*. 2023, Anais... 10º Seminário de Engenharia de Fundações Especiais e Geotecnia, 2023
- NUNES, V. P. *Ensaio de Caracterização Geotécnica da Areia da Praia de Itaipuaçu*. 2014. Projeto de Graduação—Rio de Janeiro: Universidade Federal do Rio de Janeiro, 2014.
- PINHEIRO, G. P. *Caracterização Geotécnica em Laboratório da Areia da Praia dos Cavaleiros-Macaé/RJ*. 2018. Macaé: Universidade Federal do Rio de Janeiro, 2018.
- PINTO, C. DE S. *Curso Básico de Mecânica dos Solos*. 3ª Edição ed. São Paulo: Oficina de Textos, 2009.
- PlotDigitizer*. Disponível em: <<https://plotdigitizer.com>>. Acesso em: 25 out. 2023.
- RASCHKA, SEBASTIAN.; MIRJALILI, VAHID. *Python machine learning: machine learning and deep learning with Python, scikit-learn, and TensorFlow*. [s.l.] Packt Publishing, 2017. 622 p. ISBN(9781787125933).
- RUSSELL, S.; NORVIG, P. *Artificial Intelligence: A Modern Approach*. Fourth Edition ed. [s.l.] Pearson, 2020. ISBN(9780134610993).
- SIMÕES, F. B. *Caracterização Geotécnica da Areia da Praia de Ipanema*. 2015. Projeto de Graduação—Rio de Janeiro: Universidade Federal do Rio de Janeiro, 2015.
- SMOLA, A. J.; SCHOLKOPF, B. *A tutorial on support vector regression **. *Statistics and Computing*, v. 14, p. 199–222, 2004.
- TELES, G. L. V. *Estudo Sobre os Parâmetros de Resistência e Deformabilidade da Areia de Hokksund*. 2013. Projeto de Graduação—Rio de Janeiro: Universidade Federal do Rio de Janeiro, 2013.
- ZHANG, P.; YIN, Z. Y.; JIN, Y. F. *Machine Learning-Based Modelling of Soil Properties for Geotechnical Design: Review, Tool Development and Comparison*. *Archives of Computational Methods in Engineering*, v. 29, n. 2, p. 1229–1245, 1 mar. 2022.