



15º SIMPÓSIO NACIONAL DE
GEOMORFOLOGIA

MODELAGEM ESPACIAL DA DISTRIBUIÇÃO DO pH DO SOLO BASEADA EM APRENDIZADO DE MÁQUINA E VARIÁVEIS GEOMORFOMÉTRICAS DERIVADAS DE LiDAR

Rone José dos Santos Costa ¹

Fabiano Souza de Macedo ²

Renato Batista Cavalcante ³

Ana Livia Araújo de Azevedo ⁴

Francisco das Chagas Araújo de Paiva ⁵

João Santiago Reis ⁶

RESUMO

A relação entre solo e relevo exerce influência direta na distribuição espacial dos solos e em suas propriedades, como o pH. Com os avanços tecnológicos, dados de alta resolução vem sendo utilizados em estudos ambientais, destacando-se o *Light Detection and Ranging* (LiDAR) como ferramenta eficaz para análise detalhada da paisagem. Associado a isso, o uso de algoritmos de aprendizado de máquina tem se mostrado promissor na predição de atributos do solo, oferecendo maior precisão e escalabilidade. Neste estudo, foi realizada a modelagem espacial dos valores de pH em horizontes superficiais (0–20 cm) em um assentamento rural no semiárido do Rio Grande do Norte, utilizando os algoritmos Random Forest (RF), Cubist, Support Vector Machine (SVM) e Gradient Boosting Machines (GBM). A performance dos modelos foi avaliada por meio do coeficiente de determinação (R^2), erro médio absoluto (MAE) e raiz do erro quadrático médio (RMSE). A amostragem do solo baseou-se no método do Hipercubo Latino Condicionado e o pH foi quantificado em laboratório. Foram utilizadas variáveis preditoras geomorfométricas derivadas de dados LiDAR, cuja importância relativa também foi avaliada. O modelo Cubist apresentou o melhor desempenho ($R^2 = 0,34$; MAE = 0,46; RMSE = 0,57). As cinco variáveis mais relevantes foram: Índice Multi-resolução de Planicidade de Topos de Vale, Modelo Digital de Terreno, Altura Padronizada, Índice de Umidade Topográfica e Índice de Rugosidade do Terreno. Os mapas gerados indicaram que os maiores valores de pH ocorrem em áreas de acumulação, geralmente mais baixas, enquanto os menores valores concentram-se em encostas íngremes. A maior incerteza foi observada em regiões com pH elevado. Apesar do R^2 relativamente baixo, os padrões espaciais foram

Graduando de Licenciatura em Geografia da Universidade Federal do Rio Grande do Norte - UFRN, rone.jose.costa.018@ufrn.edu.br;

² Graduando de Licenciatura em Geografia da Universidade Federal do Rio Grande do Norte - UFRN, fabiano.souza.103@ufrn.edu.br;

³ Graduando de Bacharelado em Geografia da Universidade Federal do Rio Grande do Norte - UFRN, renatocavalcantetreze@hotmail.com;

⁴ Mestranda do Programa de Pós-Graduação em Geografia (GEOCERES) da Universidade Federal do Rio Grande do Norte - UFRN, analivia.ob@hotmail.com;

⁵ Mestrando do Programa de Pós-Graduação em Manejo do Solo e Água da Universidade Federal Rural do Semiárido - UFERSA, franciscochagas621@gmail.com;

⁶ Programa de Pós-Graduação em Geografia (CERES) da Universidade Federal do Rio Grande do Norte - UFRN, joao.reis@ufrn.com.



compatíveis com os dados de campo, evidenciando a utilidade da geomorfometria e do aprendizado de máquina na modelagem espacial de atributos do solo.

INTRODUÇÃO

O solo é um recurso natural indispensável, mesmo indiretamente, no ciclo de vida dos seres vivos, cumprindo papel decisivo na ciclagem de nutrientes, no armazenamento de carbono, e na nutrição humana através dos cultivos e filtragem da água (Lu et al., 2023). Dessa forma, monitorar as propriedades físicas e químicas é fundamental para questões regionais e locais, como degradação da terra, sequestro de carbono e aplicação de técnicas de manejo (Liu, et al., 2022). Diante desse cenário, o estudo e entendimento detalhado da variabilidade espacial do pH do solo através de um mapeamento preditivo eficaz é importante para o manejo, uso da terra e classificação do solo (Öztürk et al., 2024).

O relevo, juntamente com os demais fatores de formação (clima, material de origem, organismos e tempo) interagem entre si e resultam nos processos pedogenéticos (Jenny, 1994; Lehmann; Rillig., 2015), Knott et al. (1980) afirmam que há coincidência, até certo ponto, nas limitações envolvendo o mapeamento geomorfológico e o limite do material do solo, sendo possível estabelecer correlação entre a geomorfologia e o solo.

O grande aprimoramento tecnológico, nas últimas décadas, permitiu um significativo avanço nas diferentes áreas da ciência, no ramo do mapeamento a laser – *Light Detection and Ranging* (LiDAR) – permitiu a modernização na captura de dados da superfície terrestre. As informações coletadas pelo sistema LiDAR possibilitam a automação do mapeamento, garantido pela elevada precisão e velocidade de execução, proximidade entre o valor obtido na experiência e o valor verdadeiro, voo abaixo das nuvens, densidade de pontos tridimensionais, seleção dos arquivos por planos de informação – layers (vegetação, solo, edificações) entre outros (Carneiro, 2018).

Neste contexto, os levantamentos de solos e de seus atributos se apresentam como ferramentas eficazes no planejamento de uso da terra de novas áreas a serem utilizadas por populações atingidas por grandes obras hídricas, desde que as populações atingidas tenham acesso a este tipo de conhecimento. Técnicas de machine learning têm sido amplamente utilizadas para observação, interpretação e entendimento da distribuição espacial de variáveis ambientais nos últimos anos (McCabe et al., 2017; Mostafa et al., 2023).



15º SIMPÓSIO NACIONAL DE GEOMORFOLOGIA

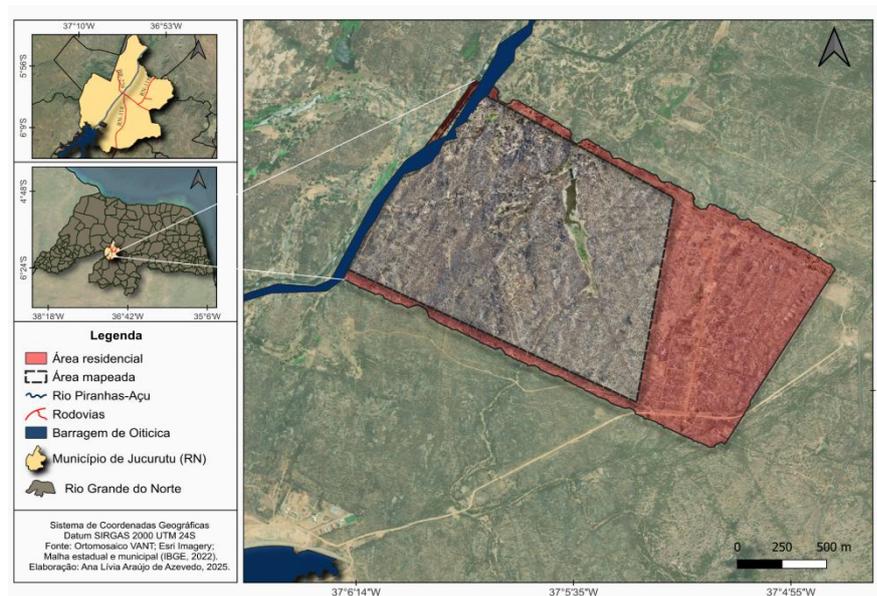
Considerando a ausência de informações de distribuição espacial de atributos dos solos em escalas detalhadas no estado do Rio Grande do Norte, este trabalho tem como objetivo realizar a predição espacial de valores de pH em um assentamento rural no semiárido potiguar, combinando produtos derivados de levantamento de LiDAR com técnicas de aprendizado de máquina.

MATERIAL E MÉTODOS

Área de estudo

A Agrovila Raimundo Nonato, situada no município de Jucurutu (RN), foi criada para suprir as demandas dos agricultores que estavam nas regiões que seriam alagadas pela Barragem de Oiticica. A área total da Agrovila abrange 275 hectares, entretanto, a extensão que será mapeada corresponde à área utilizada pelos assentados para atividades agropecuárias, totalizando cerca de 189 hectares (Figura 1).

Figura 1: Localização da área de estudo.



Fonte: Elaborado pelos autores, 2025.

O local está localizado em uma região semiárida, com um clima classificado como Semiárido quente (BS_h) segundo a classificação de Köppen. Neste cenário, a temperatura média registrada é de 28,1°C, e a precipitação anual gira em torno de 861,0 mm. A vegetação predominante é do tipo Caatinga hiperxerófila, com características arbóreo-arbustivas, destacando-se principalmente as espécies de jurema-preta (*Mimosa*



tenuiflora), mufumbo (*Combretum leprosum* Mart), faveleiro (*Cnidioscolus quercifolius*), marmeleiro (*Croton sonderianus*) e xique-xique (*Pilosocereus gounellei*) (CPRM, 2005).

Amostragem e variáveis preditoras

Para a amostragem e as variáveis preditoras, foi aplicado o método do Hipercubo Latino Condicionado – HLC (Carvalho Filho et al., 2014), em que utilizamos dados das camadas de variáveis de relevo para selecionar pontos amostrais mais representativos da área de estudo. Um total de 100 pontos amostrais foi estabelecido, sendo utilizados para a coleta de amostras para análises laboratoriais, cujos resultados foram usados em modelos preditivos sobre a distribuição espacial do pH do solo.

Para a geração das variáveis de relevo, foi utilizado um Modelo Digital de Terreno (MDT) gerado com a técnica LiDAR, capturado por um sensor DJI Zenmuse L1 acoplado a um veículo aéreo não tripulado (VANT). Utilizando o MDT, foi realizada a extração da geomorfometria em ambiente SAGA (Conrad et al., 2015).

As variáveis geomorfométricas utilizadas foram: declividade (DECL), orientação das vertentes (OV), índice de posição topográfica (IPT), perfil de curvatura (PC) e classificação do relevo baseada no IPT (CR) (Gallant & Wilson, 2000); modelo digital de terreno (MDT), altura da declividade (ADECL), posição de meia encosta (PME), índice de balanço de massa (BM), altura padronizada (AP), elevação normalizada (EN), e índice de convexidade (CONV) (Conrad et al., 2015); profundidade do vale (PF) (Rodriguez, F. et al., 2002), índice de rugosidade do terreno (IRT) (Riley; De Gloria; Elliot, 1999), índice de umidade topográfica (IUT) (Boehner, 2002); índice multirresolução de planicidade do topo do vale (IMPTV) e índice multirresolução de planicidade do fundo do vale (IMPFV) (Gallant e Dowling, 2003).

As análises laboratoriais do pH foram realizadas seguindo o método de Teixeira et al. (2017), utilizado para mensurar a atividade de íons H^+ na solução do solo. O procedimento tem como princípio a mensuração do pH por meio de eletrodo combinado imerso em suspensão solo:água na proporção 1:2,5 (Teixeira et al., 2017).

Modelagem espacial e predição dos valores de pH do solo

A predição da distribuição espacial dos valores de pH foi conduzida por meio de uma sequência estruturada de procedimentos. Inicialmente, as variáveis



15º SIMPÓSIO NACIONAL DE GEOMORFOLOGIA

geomorfométricas foram organizadas em camadas raster e seus valores extraídos para os pontos amostrais de solo. Em seguida, quatro algoritmos de aprendizado de máquina foram empregados para o treinamento dos modelos preditivos. Cada modelo foi submetido a 100 execuções independentes. Posteriormente, procedeu-se à validação e análise comparativa do desempenho preditivo dos algoritmos. Para tal, foram observados os valores do coeficiente de determinação (R^2), erro médio absoluto (MAE) e raiz do erro quadrático médio (RMSE). Por fim, foram gerados mapas representando a média das predições e a incerteza associada a cada local estimado. Os algoritmos aplicados foram Support Vector Machine (SVM), Random Forest (RF), Gradient Boosting Machines (GBM) e Cubist, sendo toda a rotina implementada no ambiente R com o suporte do pacote *caret* (Kuhn, 2008).

RESULTADOS E DISCUSSÃO

A Tabela 1 apresenta a estatística descritiva dos valores de pH encontrados nas 100 amostras utilizadas para treinamento dos algoritmos. Observa-se que apesar da amplitude ser relativamente alta para valores de pH (Mín = 5.03; Máx. = 8.45), os dados não apresentam significativa variabilidade, conforme indicado pelos baixos valores de desvio padrão (DP) e coeficiente de variação (CV).

Tabela 1: Estatística descritiva valores de pH nas amostras utilizadas na modelagem.

| Variável | Parâmetros | | | | | Quartis | | |
|----------|------------|---------|------|--------|------|---------|------|------|
| | Média | Mediana | DP | CV (%) | Min. | Máx. | Q1 | Q3 |
| pH | 6.12 | 6.03 | 0.68 | 11.11 | 5.03 | 8.45 | 5.65 | 6.50 |

DP - Desvio-padrão; CV - Coeficiente de variação; Min - Mínimo; Máx - Máximo; Q1 - primeiro quartil; Q3 - terceiro quartil.

As métricas utilizadas para avaliar o desempenho dos modelos foram o Coeficiente de Determinação (R^2), Erro Médio Absoluto (MAE) e a Raiz do Erro Quadrático Médio (RMSE), apresentadas na Tabela 2 e nas Figuras 2 e 3.

De maneira geral, os valores do coeficiente de determinação (R^2) foram baixos para a maioria dos modelos avaliados, no entanto, é possível perceber similaridades significativas entre as métricas utilizadas na predição. Considerando que o modelo com melhor desempenho é aquele em que o R^2 apresenta maior valor e menores valores de

MAE e RMSE, o Cubist destacou-se com a melhor performance, apresentando um valor médio de R^2 de 0.342, e mediana similar.

A distribuição dos valores de R^2 , considerando as 100 execuções, é apresentada em forma de *boxplots* na Figura 2. É possível observar que o Cubist, apesar de ter o R^2 , apresenta uma maior dispersão nos dados. Em contrapartida, o modelo SVM encontra-se com valores significativamente reduzidos (0,051), mas com uma menor dispersão dos dados entre os modelos, o que sugere uma maior precisão.

Tabela 2: Estatística descritiva das métricas de performance dos modelos de predição dos teores de pH.

| Parâmetros | | Modelos | | | |
|------------|---------|---------|-------|-------|--------|
| | | SVM | RF | GBM | Cubist |
| R^2 | Média | 0,051 | 0,202 | 0,192 | 0,342 |
| | Mediana | 0,051 | 0,202 | 0,192 | 0,342 |
| | Mín | 0,037 | 0,167 | 0,154 | 0,237 |
| | Máx | 0,065 | 0,237 | 0,23 | 0,447 |
| | Dp | 0,02 | 0,05 | 0,054 | 0,149 |
| MAE | Média | 0,53 | 0,518 | 0,516 | 0,469 |
| | Mediana | 0,53 | 0,518 | 0,516 | 0,469 |
| | Mín | 0,526 | 0,515 | 0,486 | 0,444 |
| | Máx | 0,535 | 0,521 | 0,546 | 0,495 |
| | Dp | 0,006 | 0,004 | 0,042 | 0,037 |
| RMSE | Média | 0,693 | 0,638 | 0,637 | 0,579 |
| | Mediana | 0,693 | 0,638 | 0,637 | 0,579 |
| | Mín | 0,675 | 0,636 | 0,611 | 0,526 |
| | Máx | 0,71 | 0,64 | 0,662 | 0,632 |
| | Dp | 0,024 | 0,003 | 0,036 | 0,075 |

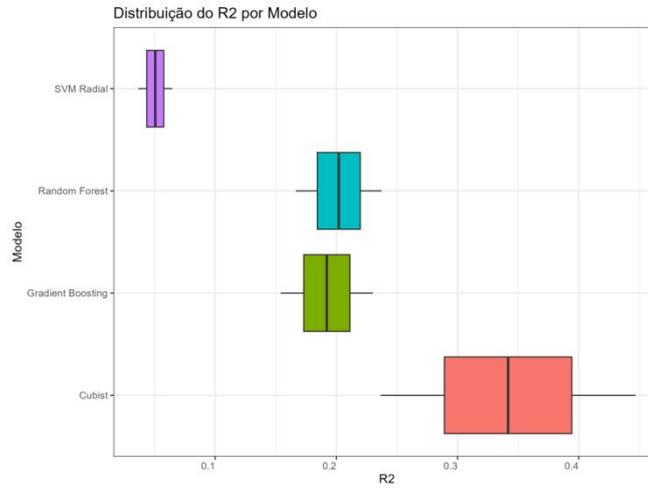
DP - Desvio-padrão; Min - Mínimo; Máx – Máximo.

Fonte: Elaborado pelos autores, 2025.

A partir da Figura 3, pode ser observado que o menor MAE foi do modelo Cubist, com um valor de 0.469 (Tabela 2). Os modelos SVM e RF têm o pior desempenho, com MAE de 0.53 e 0.518, respectivamente. No entanto, esses modelos apresentam a menor variabilidade nos dados, sendo, portanto, os mais consistentes.

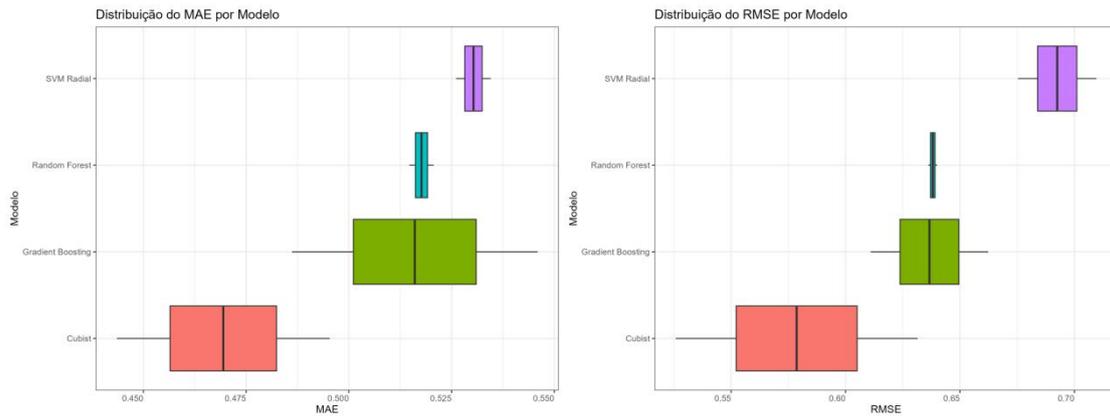
Os valores de RMSE apresentam uma maior concentração nos valores mais elevados (Figura 3), variando entre 0.526 e 0.71. O modelo RF se destaca pela baixa dispersão de dados, sugerindo uma maior precisão. O Cubist se destacou como o modelo mais eficiente para prever o pH, apresentando um RMSE de 0.579, no entanto, é o modelo de maior dispersão.

Figura 2: Coeficiente de Determinação (R^2) de cada modelo para a predição dos teores de pH.



Fonte: Elaborado pelos autores, 2025.

Figura 3: Erro Médio Absoluto (MAE) e Raiz do Erro Quadrático Médio (RMSE), de cada modelo para a predição dos teores de pH.



Fonte: Elaborado pelos autores, 2025.

Na Figura 4 são apresentados os gráficos de importância das variáveis geomorfométricas em cada modelo para a modelagem do pH. Nos modelos Cubist, RF e SVM, a variável Índice de Multi-Resolução de Planicidade dos Topos de Vale (IMPTV) apresentou-se como a principal variável preditora na modelagem deste atributo. No GBM, a variável com maior importância preditiva foi o Modelo Digital de Terreno (MDT). Nesse sentido, observa-se que o GBM apresenta uma distribuição mais uniforme da importância das variáveis, sem conferir destaque a uma única variável.

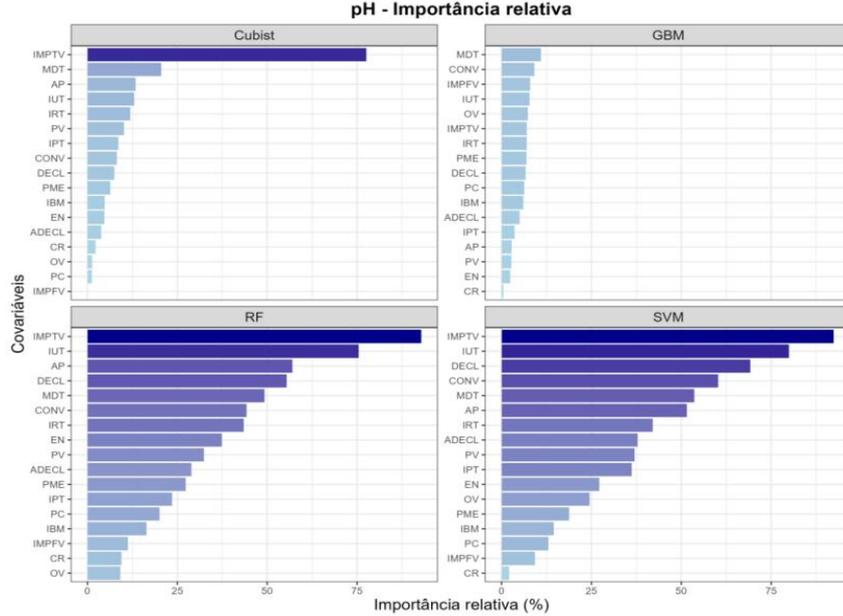
Quando observada a distribuição espacial dos valores de pH, gerada pela predição em todos os modelos, identifica-se uma tendência de valores mais altos em



15º SIMPÓSIO NACIONAL DE
GEOMORFOLOGIA

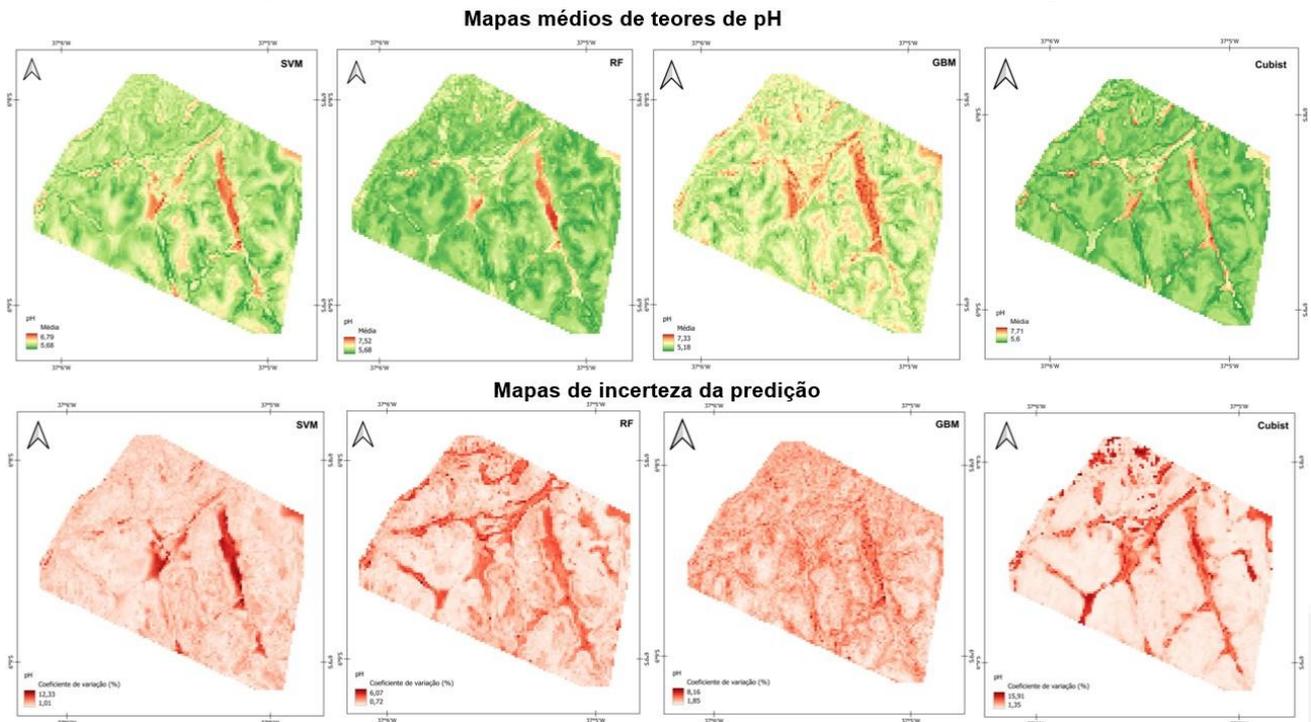
superfícies aplainadas em fundo de vale, assim como em topos de morro, e valores mais baixos em áreas de maior declividade nos terços médio e inferior de encosta (Figura 5). Os maiores valores de pH são observados nos fundos de vale, corroborando comportamento esperado para a dinâmica de acumulação de nutrientes/sais em solos de ambientes semiáridos e estudos similares de predição de pH (Suleymanov et al., 2023).

Figura 4: Importância relativa das variáveis geomorfométricas para a predição de pH.



Fonte: Elaborado pelos autores, 2025.

Figura 5: Distribuição espacial da predição dos valores de pH para cada algoritmo.





CONSIDERAÇÕES FINAIS

Os resultados obtidos demonstram o potencial da combinação entre variáveis geomorfométricas derivadas de LiDAR e algoritmos de aprendizado de máquina na predição espacial do pH do solo em ambientes semiáridos. Apesar dos baixos valores de R^2 observados, os padrões espaciais mostraram-se coerentes com a distribuição topográfica e com o comportamento esperado em áreas de acúmulo, como fundos de vale.

O modelo Cubist apresentou desempenho superior em relação aos demais algoritmos testados. As variáveis mais influentes na modelagem foram associadas à planicidade (IMPTV), altitude relativa (AP) e umidade potencial (IUT).

Este estudo reforça a viabilidade do mapeamento digital de atributos do solo, utilizando abordagens baseadas em aprendizado de máquina. Recomenda-se, em estudos futuros, o incremento de variáveis preditoras associadas aos demais fatores de formação de solos, a fim de aumentar a acurácia e estabilidade dos modelos.

Palavras-chave: Mapeamento de solos; Aprendizado de máquina; Acidez do solo; Solos do semiárido; Pedologia; Pedometria.

REFERÊNCIAS BIBLIOGRÁFICAS

BOEHNER, J., et al. Soil Regionalisation by Means of Terrain Analysis and Process Parameterisation. In: Micheli, E., Nachtergaele, F., Montanarella, L. [Ed.]: Soil Classification 2001. **European Soil Bureau**, Research Report, n. 7, EUR 20398 EN, Luxembourg, p.213-222, 2002.

CARNEIRO, M. C. D. M. et al. Light Detection and Ranging in the Geomorphological Characterization of the Coast Dunes Field of Rio Do Fogo, Rio Grande do Norte, Brazil. **Revista Geográfica de América Central**, v. 2, n. 61, p. 315-348, 2018.

CARVALHO FILHO, W. et al. Método do Hipercubo Latino Condicionado para a amostragem de solos na presença de covariáveis ambientais visando o mapeamento digital de solos. **Revista Brasileiro de Ciência do Solo**, v. 38, p. 386-396, 2014.

COMPANHIA DE PESQUISA DE RECURSOS MINERAIS – CPRM. Serviço Geológico do Brasil **Projeto cadastro de fontes de abastecimento por água subterrânea**. Diagnóstico do município de Jucurutu, estado do Rio Grande do Norte. Org. MASCARENHAS, J. C. et al. Recife: CPRM/PRODEEM, 2005.



15º SIMPÓSIO NACIONAL DE
GEOMORFOLOGIA

CONRAD, O., et al. System for Automated Geoscientific Analyses (SAGA) v. 2.1.4, **Geosci. Model Dev.**, v. 8, n. 1991–2007, 2015.

GALLANT, J.C.; DOWLING, T.I. A multiresolution index of valley bottom flatness for mapping depositional areas. **Water Resources Research**, p. 1347-1359, 2003.

GALLANT, J. C.; WILSON, J. P. Primary topographic attributes. In: **Terrain Analysis: Principles and applications**. John Wiley & Sons, New York, 2000.

JENNY, Hans. **Factors of soil formation: a system of quantitative pedology**. Courier Corporation, 1994.

KNOTT, P.; DOORNKAMP, J.; JONES, R. The relationship between soils and geomorphological mapping units-a case study from northamptonshire. **Bulletin of Engineering Geology & the Environment**, v. 21, n. 1, 1980.

LEHMANN, Anika; RILLIG, Matthias C. Understanding mechanisms of soil biota involvement in soil aggregation: A way forward with saprobic fungi?. **Soil Biology and Biochemistry**, v. 88, p. 298-302, 2015.

LIU, Feng et al. Mapping high resolution national soil information grids of China. **Science Bulletin**, v. 67, n. 3, p. 328-340, 2022.

LU, Qikai; TIAN, Shuang; WEI, Lifei. Digital mapping of soil pH and carbonates at the European scale using environmental variables and machine learning. **Science of the Total Environment**, v. 856, p. 159171, 2023.

MCCABE, M. F. et al. The future of Earth observation in hydrology. *Hydrol. Earth Syst. Sci.*, v.21, p.3879-3914, 2017.

MOSTAFA, R. R. et al. Modeling Potential Evapotranspiration by Improved Machine Learning Methods Using Limited Climatic Data. *Water*, v.15, n.3, 2023.

ÖZTÜRK, Mustafa; KILIÇ, Miraç; GÜNAL, Hikmet. Digital Mapping of Soil pH and Electrical Conductivity: A Comparative Analysis of Kriging and Machine Learning Approaches. **MAS Journal of Applied Sciences**, v. 9, n. 4, p. 1168–1185, 2024.

RODRIGUEZ, F., MAIRE, É., COURJAULT-RADÉ, P., DARROZES, J. The Black Top Hat function applied to a DEM: A tool to estimate recent incision in a mountainous watershed (Estibère Watershed, Central Pyrenees). **Geophysical Research Letters**, v. 29, n. 6, 2002.

RILEY, S.J., DE GLORIA, S.D., ELLIOT, R. A Terrain Ruggedness that Quantifies Topographic Heterogeneity. **Intermountain Journal of Science**, v.5, n.1-4, p.23-27, 1999.

SULEYMANOV et al. Random Forest Modeling of Soil Properties in Saline Semi-Arid Areas. **Agronomy**, v. 13, n. 5, 2023.

TEIXEIRA, W. G. **Manual de métodos de análise de solo**. 3ª ed. Brasília, DF: EMBRAPA, 2017. 573p.