

UNIVERSIDADE FEDERAL DE JUIZ DE FORA
INSTITUTO DE CIÊNCIAS EXATAS - FACULDADE DE ENGENHARIA
PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM
COMPUTACIONAL

Ruan Rampinelli Ribeiro

Otimização de Redes Neurais Artificiais para a Predição Aprimorada de
Funções de Pedotransferência em Modelagem Computacional do Solo

Juiz de Fora

2025

Ruan Rampinelli Ribeiro

**Otimização de Redes Neurais Artificiais para a Predição Aprimorada de
Funções de Pedotransferência em Modelagem Computacional do Solo**

Dissertação apresentada ao Programa de Pós-Graduação em Modelagem Computacional da Universidade Federal de Juiz de Fora como requisito parcial à obtenção do título de Mestre em Modelagem Computacional. Área de concentração: Modelagem Computacional

Orientador: Prof. Dr. Leonardo Goliatt da Fonseca

Coorientadora: Prof. Dr. Priscila Vanessa Zabala Capriles Goliatt

Juiz de Fora

2025

Ficha catalográfica elaborada através do Modelo Latex do CDC da UFJF
com os dados fornecidos pelo(a) autor(a)

Rampinelli Ribeiro, Ruan.

Otimização de Redes Neurais Artificiais para a Predição Aprimorada
de Funções de Pedotransferência em Modelagem Computacional do Solo /
Ruan Rampinelli Ribeiro. – 2025.

93 f. : il.

Orientador: Leonardo Goliatt da Fonseca

Coorientadora: Priscila Vanessa Zabala Capriles Goliatt

Dissertação (Mestrado) – Universidade Federal de Juiz de Fora, Instituto
de Ciências Exatas - Faculdade de Engenharia. Programa de Pós-Graduação
em Modelagem Computacional, 2025.

1. Funções de Pedotransferência. 2. Redes Neurais Artificiais. 3.
Propriedades Hidráulicas do Solo. I. da Fonseca, Leonardo Goliatt, orient.
II. Goliatt, Priscila Vanessa Zabala Capriles, coorient. III. Título.

Ruan Rampinelli Ribeiro

Otimização de Redes Neurais Artificiais para a Predição Aprimorada de Funções de Pedotransferência em Modelagem Computacional do Solo

Dissertação apresentada ao Programa de Pós-Graduação em Modelagem Computacional da Universidade Federal de Juiz de Fora como requisito parcial à obtenção do título de Mestre em Modelagem Computacional. Área de concentração: Modelagem Computacional.

Aprovada em 06 de agosto de 2025.

BANCA EXAMINADORA

Prof. Dr. Leonardo Goliatt da Fonseca - Orientador

Universidade Federal de Juiz de Fora

Prof.^a Dr.^a Priscila Vanessa Zabala Capriles Goliatt

Universidade Federal de Juiz de Fora

Prof. Dr. Celso Bandeira de Melo Ribeiro

Universidade Federal de Juiz de Fora

Prof.^a Dr.^a Camila Martins Saporetti

Universidade do Estado do Rio de Janeiro

Juiz de Fora, 11/08/2025.



Documento assinado eletronicamente por **Leonardo Goliatt da Fonseca, Professor(a)**, em 14/08/2025, às 14:46, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Camila Martins Saporetti, Usuário Externo**, em 14/08/2025, às 18:06, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Priscila Vanessa Zabala Capriles Goliatt, Professor(a)**, em 15/08/2025, às 09:48, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Celso Bandeira de Melo Ribeiro, Professor(a)**, em 15/08/2025, às 16:08, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no Portal do SEI-Ufjf (www2.ufjf.br/SEI) através do ícone Conferência de Documentos, informando o código verificador **2545329** e o código CRC **DADAF7DA**.

Dedico este trabalho à minha família, pelo apoio incondicional, e a todos que acreditaram no meu potencial durante esta jornada acadêmica.

AGRADECIMENTOS

Agradeço primeiramente a Deus, pela força e perseverança durante esta caminhada. Aos meus orientadores, Prof. Dr. Leonardo Goliatt da Fonseca e Prof.^a Dra. Priscila Vanessa Zabala Capriles Goliatt, pela orientação técnica, paciência e confiança.

Aos colegas do Programa de Pós-Graduação em Modelagem Computacional, pelas discussões e aprendizados compartilhados. À Universidade Federal de Juiz de Fora, pelo suporte institucional e apoio a pesquisa científica no Brasil.

À minha família, pelo incentivo constante, carinho e compreensão diante das ausências necessárias para a realização deste trabalho.

Agradeço ainda ao suporte financeiro proporcionado pela bolsa de estudos ofertada pela Universidade Federal de Juiz de Fora (UFJF), cujo investimento em pesquisa e educação foi essencial para a realização deste trabalho, permitindo dedicação integral ao desenvolvimento acadêmico e científico aqui apresentado.

RESUMO

As funções de pedotransferência (PTFs) desempenham papel essencial na estimativa de propriedades hidráulicas do solo, como a curva de retenção de água, a partir de atributos físicos de obtenção mais simples e menos custosa. A caracterização hidráulica do solo é um desafio recorrente, especialmente em regiões tropicais, devido à heterogeneidade espacial e à dificuldade de medições diretas em laboratório ou campo. Nesse contexto, as redes neurais artificiais (RNAs) surgem como alternativa promissora, capazes de capturar relações não lineares complexas e superar limitações de métodos estatísticos tradicionais. Este trabalho teve como objetivo otimizar arquiteturas de RNAs para a predição de PTFs em três bases de dados com diferentes condições edafoclimáticas: B1 (Brasil – Nordeste), B2 (Brasil – Acre) e D1 (Dinamarca). Foram aplicadas técnicas de ajuste de hiperparâmetros e otimização bayesiana, considerando diferentes funções de ativação, número de camadas, neurônios e taxas de regularização. Os resultados demonstraram que o modelo final atingiu valores de desempenho expressivos, com coeficiente de determinação médio superior a 0,90, erro quadrático médio (RMSE) em torno de 0,03 e erro absoluto médio (MAE) próximo de 0,02, indicando elevada acurácia na predição da umidade volumétrica do solo. Conclui-se que as RNAs apresentam elevado potencial na modelagem de PTFs, constituindo alternativa robusta para aplicações em hidrologia, agricultura e manejo sustentável do solo, embora estudos futuros devam considerar maior diversidade de dados e comparações com outros algoritmos de aprendizado de máquina.

Palavras-chave: Modelagem Computacional. Funções de Pedotransferência. Redes Neurais Artificiais. Propriedades Hidráulicas do Solo. Aprendizado de Máquina.

ABSTRACT

Pedotransfer functions (PTFs) play a key role in estimating soil hydraulic properties, such as the soil water retention curve, from simpler and less expensive physical attributes. Hydraulic characterization of soils is a recurring challenge, especially in tropical regions, due to spatial heterogeneity and the difficulty of direct measurements in laboratory or field conditions. In this context, artificial neural networks (ANNs) emerge as a promising alternative, capable of capturing complex nonlinear relationships and overcoming the limitations of traditional statistical methods. This study aimed to optimize ANN architectures for the prediction of PTFs using three distinct datasets under different edaphoclimatic conditions: B1 (Brazil – Northeast), B2 (Brazil – Acre), and D1 (Denmark). Hyperparameter tuning and Bayesian optimization techniques were applied, considering different activation functions, number of layers, neurons, and regularization rates. The results demonstrated that the final model achieved expressive performance values, with an average coefficient of determination higher than 0.90, root mean square error (RMSE) around 0.03, and mean absolute error (MAE) close to 0.02, indicating high accuracy in predicting soil volumetric water content. It can be concluded that ANNs show strong potential in PTF modeling, constituting a robust alternative for applications in hydrology, agriculture, and sustainable soil management, although future studies should consider a greater diversity of data and comparisons with other machine learning algorithms.

Keywords: Computational Modeling. Pedotransfer Functions. Artificial Neural Networks. Soil Hydraulic Properties. Machine Learning.

LISTA DE ILUSTRAÇÕES

Gráfico de evolução da produção de artigos sobre "machine learning + pedotransfer function"(2014-2023)	18
Um diagrama ternário do triângulo textural do solo, traduzido para o português, que exibe as diferentes classificações de textura do solo segundo o Departamento de Agricultura dos Estados Unidos (USDA).	26
Gráfico típico da curva de retenção de água no solo para um solo arenoso usando uma equação proposta por van Genuchten, onde ψ_a é o valor de entrada de ar, θ_r e θ_s são os conteúdos volumétricos de água residual e saturada.	28
PTFs relacionam propriedades do solo simples de medir a parâmetros menos disponíveis de processos do sistema terrestre.	32
Conceito geral do desenvolvimento de PTFs, fundamentado em um banco de dados de calibração contendo tanto as propriedades básicas quanto as propriedades do solo a serem estimadas.	34
Esquema de operação da rede neural artificial (RNA) do tipo feed-forward generalizada (FFG).	42
Esquema autoral da rede neural utilizada neste trabalho	46
Esquema do processo de validação cruzada <i>k-fold</i>	50
Localização dos locais de amostragem para retenção de água (conjuntos de dados Brasil-NE e Dinamarca).	52
Localização geográfica da área de estudo no estado do Acre, Brasil. As fotografias ilustram os três tipos de uso do solo analisados: floresta nativa, pastagem e sistema integrado lavoura-pecuária.	54
Fluxograma que sintetiza o processo metodológico	66
Gráficos para resultado da base de dados B1 - M1 a M5	67
Gráficos para resultado da base de dados B1 - M6 a M12	68
Gráficos para resultado da base de dados B2 - M1 a M5	72
Gráficos para resultado da base de dados B2 - M6 a M12	73
Gráficos para resultado da base de dados D1 - M1 a M5	77
Gráficos para resultado da base de dados D1 - M6 a M12	78

LISTA DE TABELAS

Tabela 1	– Comparação entre tipos de modelos de funções de pedotransferência (PTFs), destacando base metodológica, vantagens, desvantagens e exemplos. . .	36
Tabela 2	– Conjunto de melhores hiperparâmetros testados no modelo para B1. . .	62
Tabela 3	– Conjunto de melhores hiperparâmetros testados no modelo para B2. . .	62
Tabela 4	– Conjunto de melhores hiperparâmetros testados no modelo para D1. . .	64
Tabela 5	– Resultados obtidos para B1	69
Tabela 6	– Dados de referência para B1	70
Tabela 7	– Variação de R^2 e RMSE para os melhores resultados apresentados de B1 em comparação com os dados de referência	70
Tabela 8	– Resultados obtidos para B2	74
Tabela 9	– Dados de referência para B2	75
Tabela 10	– Variação de R^2 e RMSE para os melhores resultados apresentados de B2 em comparação com os dados de referência	75
Tabela 11	– Resultados obtidos para D1	79
Tabela 12	– Dados de referência para D1	79
Tabela 13	– Variação de R^2 e RMSE para os melhores resultados apresentados de D1 em comparação com os dados de referência	80
Tabela 14	– Comparação dos resultados deste trabalho com estudos da literatura . .	84

LISTA DE ABREVIATURAS E SIGLAS

RNA	Rede Neural Artificial (<i>Artificial Neural Network</i>)
ANN	Rede Neural Artificial (<i>Artificial Neural Network</i>)
PTF	Função de Pedotransferência (<i>Pedotransfer Function</i>)
ML	Aprendizado de Máquina (<i>Machine Learning</i>)
RMSE	Erro Quadrático Médio (<i>Root Mean Square Error</i>)
MAE	Erro Absoluto Médio (<i>Mean Absolute Error</i>)
R^2	Coefficiente de Determinação
MLP	Perceptron Multicamadas (<i>Multi-Layer Perceptron</i>)
SVM	Máquina de Vetores de Suporte (<i>Support Vector Machine</i>)
USDA	Departamento de Agricultura dos Estados Unidos (<i>United States Department of Agriculture</i>)
SWRC	Curva de Retenção de Água no Solo (<i>Soil Water Retention Curve</i>)

LISTA DE SÍMBOLOS

θ	Conteúdo volumétrico de água no solo ($\text{cm}^3 \text{ cm}^{-3}$)
Ψ	Potencial mátrico (kPa)
n	Parâmetro de forma do modelo de Van Genuchten
α	Parâmetro de ajuste de sucção (cm^{-1})
m	Parâmetro relacionado à curva de retenção
W_s	Saturação do solo
N_s	Número de amostras de calibração
N_p	Número de parâmetros do modelo
t	Valor observado
t'	Valor previsto

SUMÁRIO

1	INTRODUÇÃO	13
1.1	CONSIDERAÇÕES INICIAIS	13
1.2	JUSTIFICATIVA	17
1.3	ELABORAÇÃO DOS OBJETIVOS	21
1.4	DEFINIÇÃO DA METODOLOGIA	22
1.5	ESTRUTURA DO TRABALHO	23
2	FUNDAMENTAÇÃO TEÓRICA	24
2.1	O SOLO E SUAS PROPRIEDADES	24
2.1.1	ESTRUTURA E COMPOSIÇÃO DO SOLO	24
2.1.2	RETENÇÃO DE ÁGUA E CONDUTIVIDADE HIDRÁULICA	27
2.1.3	IMPORTÂNCIA DAS PROPRIEDADES FÍSICAS DO SOLO NA MODELA- GEM	29
2.2	AS FUNÇÕES DE PEDOTRANSFERÊNCIA (PTF)	31
2.2.1	DEFINIÇÃO E APLICAÇÕES	31
2.2.2	TIPOS DE PTFs	34
2.2.3	DESAFIOS E LIMITAÇÕES	37
2.3	APRENDIZADO DE MÁQUINA E REDES NEURAI ARTIFICIAIS	39
2.3.1	FUNDAMENTOS DO APRENDIZADO DE MÁQUINA	39
2.3.2	ARQUITETURAS DE REDES NEURAI APLICADAS A PTFs	41
2.3.3	AJUSTE DE HIPERPARÂMETROS E OTIMIZAÇÃO	47
2.3.3.1	OTIMIZAÇÃO BAYESIANA	50
3	MATERIAIS E MÉTODOS	52
3.1	BASE DE DADOS	52
3.2	PRÉ-PROCESSAMENTO DOS DADOS	55
3.2.1	LIMPEZA DOS DADOS	55
3.2.2	ESTRUTURA DOS DADOS	55
3.2.3	TRANSFORMAÇÃO E NORMALIZAÇÃO DOS DADOS	56
3.2.4	DIVISÃO DOS DADOS	57
3.2.5	DEFINIÇÃO DAS VARIÁVEIS DE ENTRADA E SAÍDA	57
3.3	ARQUITETURA E PARAMETRIZAÇÃO DO MODELO DE REDE NEU- RAL	57
3.3.1	CONSTRUÇÃO E PARAMETRIZAÇÃO DA REDE NEURAL	58
3.3.2	OTIMIZAÇÃO DOS HIPERPARÂMETROS	60
3.3.2.1	BASE DE DADOS B1	61
3.3.2.2	BASE DE DADOS B2	62
3.3.2.3	BASE DE DADOS D1	63
3.4	CONSTRUÇÃO E EXECUÇÃO DO MODELO DE REDE NEURAL	64

3.5	FLUXO METODOLÓGICO	65
4	RESULTADOS E DISCUSSÕES	67
4.1	RESULTADOS	67
4.1.1	BASE DE DADOS B1	67
4.1.2	BASE DE DADOS B2	71
4.1.3	BASE DE DADOS D1	76
4.2	DISCUSSÕES	80
4.2.1	INTERPRETAÇÃO DOS RESULTADOS	81
4.2.2	COMPARAÇÃO COM OUTROS ESTUDOS	82
4.2.3	LIMITAÇÕES E MELHORIAS	84
4.2.4	PERSPECTIVAS FUTURAS	85
5	CONCLUSÃO	86
	REFERÊNCIAS	87

1 INTRODUÇÃO

1.1 CONSIDERAÇÕES INICIAIS

A condição de fertilidade do solo resulta de uma combinação de várias características físicas e químicas, as quais variam de acordo com as condições geográficas e climáticas. É uma exigência crescente para a humanidade preservar a composição biológica e química do solo, dada sua influência crucial na prática agrícola. Qualquer anomalia nas propriedades do solo não apenas impacta a agricultura, mas também afeta o ciclo da água na terra, podendo, em certa medida, contribuir para a ocorrência de desastres naturais. (ROSE et al., 2018)

Uma compreensão exata dos procedimentos e características do solo, juntamente com uma previsão precisa, é fundamental para analisar o sistema terrestre e avaliar os efeitos das mudanças climáticas e da utilização do solo. No entanto, alcançar isso demanda uma precisão na parametrização dos processos do solo, assim como a utilização de formas adequadas e confiáveis para representar a heterogeneidade espacial da superfície terrestre. Em muitas situações, a indisponibilidade de observações referentes às propriedades fundamentais do solo, estados e parâmetros que regulam os fluxos de água, energia e gás nos sistemas terrestres ocorre devido à impraticabilidade de realizar medições com cobertura espacial e temporal suficiente. Nas disciplinas das ciências do solo, hidrologia, superfície terrestre e clima, a precisão na obtenção de estimativas, como um exemplo, da umidade e temperatura do solo, torna-se crucial para reduzir a incerteza associada à previsão da respiração do solo e dos fluxos de calor e água. (LOOY et al., 2017)

As características hidráulicas do solo desempenham um papel crucial na simulação da disponibilidade e movimentação de água nos solos. Uma propriedade hidráulica significativa do solo é a sua capacidade de retenção de água, a qual impacta diretamente na produtividade e no manejo do solo. Ter conhecimento sobre a capacidade de retenção de água e compreender os efeitos do uso da terra nessa propriedade são aspectos essenciais para uma gestão eficaz do solo e da água, além de serem fundamentais para estimar o fornecimento de água para irrigação, sujeito a alterações no uso da terra. (D'EMILIO et al., 2018)

A obtenção de informações sobre propriedades hidráulicas do solo, que sejam facilmente acessíveis e representativas, frequentemente constitui um desafio significativo para compreender a dinâmica da água e dos solutos no solo não saturado. Isso também representa uma barreira à aplicação de modelos de simulação, visando prevenir e controlar a degradação do solo decorrente de práticas agrícolas intensivas. (D'EMILIO et al., 2018)

Diante desse desafio, as Funções de Pedotransferência (PTFs) surgem como uma alternativa valiosa para estimar as propriedades hidráulicas do solo a partir de dados mais facilmente disponíveis, como textura e densidade aparente. As PTFs tradicionais, embora

úteis, possuem limitações em capturar a complexidade e não-linearidade das relações entre as propriedades do solo. Nesse contexto, o uso de métodos de aprendizado de máquina, e em particular as Redes Neurais Artificiais (RNAs), tem se destacado como uma abordagem promissora para aprimorar as PTFs, oferecendo potencial para previsões mais precisas e robustas das propriedades hidráulicas do solo.

O interesse de pesquisa na dinâmica da umidade do solo está em constante crescimento, especialmente com a crescente ênfase na gestão ambiental. A automação do tratamento de informações do solo por meio de sistemas computacionais abre novas perspectivas na pesquisa agrícola. Progressos recentes na amostragem do solo, assim como na análise e interpretação das propriedades físicas do solo, proporcionam oportunidades inéditas para incorporar dados do solo em diversas aplicações, abrangendo áreas como agricultura, hidrologia, engenharia civil, geociências, engenharia de petróleo, ciências ambientais, entre outras. Isso tem gerado uma grande demanda por bancos de dados de solos. Embora as informações do solo sejam rotineiramente coletadas por meio de levantamentos em muitos países, é raro que esses dados incluam propriedades hidráulicas do solo, como a curva de retenção de água (WRC) e a condutividade hidráulica saturada (K_s). Portanto, há uma necessidade urgente de métodos econômicos e rápidos para estimar essas propriedades (SCHAAP et al., 2001; GIVI et al., 2004; ALVAREZ-ACOSTA et al., 2012 apud PATIL; SINGH, 2016). Além disso, em muitos países ao redor do mundo, a informação sobre o solo não está acessível em uma escala regional. Pesquisadores, planejadores e tomadores de decisão nesses países frequentemente lidam com a falta de dados adequados para conduzir seus trabalhos (PATIL et al., 2010; PATIL et al., 2011 apud PATIL; SINGH, 2016).

Por essas razões, têm sido sugeridos diversos métodos para a estimativa das propriedades hidráulicas do solo a partir de características facilmente mensuráveis, como textura e densidade aparente (BD), e/ou dados limitados coletados durante levantamentos de solo (PARSURAMAN et al., 2007; VEREECKEN et al., 2010 apud PATIL; SINGH, 2016). Uma expressão ou modelo criado para realizar a estimativa indireta das propriedades do solo é conhecida como função de pedotransferência (PTF) (BOUMA, 1989 apud PATIL; SINGH, 2016). Dada a morosidade associada à medição direta da condutividade hidráulica em campo e considerando a questão da variabilidade espacial em uma escala de campo, os modelos preditivos, incluindo aqueles que estimam as propriedades hidráulicas com base na textura do solo e dados correlatos, representam a única abordagem viável para caracterizar essas propriedades em extensas áreas de terreno. A medição direta, por sua vez, pode ser economicamente vantajosa somente para problemas específicos em locais específicos (WOSTEN et al., 2001; PATIL e CHATURVEDI, 2012 apud PATIL; SINGH, 2016).

A expressão "função de pedotransferência" foi descrita como uma maneira de descrever o processo de "traduzir os dados que temos naquilo que precisamos" (BOUMA,

1989 apud BOTULA; RANST; CORNELIS, 2014). As funções de pedotransferência são mecanismos preditivos que estabelecem relações entre dados de solo mais prontamente mensuráveis, como a textura do solo (teores de areia, silte e argila), densidade aparente (BD), teor de matéria orgânica (MO) ou carbono orgânico (CO), e/ou outros dados comumente coletados em levantamentos de solo, até parâmetros hidráulicos, como a curva de retenção de água no solo (SWRC, do inglês Soil Water Retention Curve) (BOUMA e VAN LANEN, 1987; BOUMA, 1989; VAN DEN BERG et al., 1997 apud BOTULA; RANST; CORNELIS, 2014).

A abordagem conhecida como função de pedotransferência (PTF), que utiliza funções preditivas e dados provenientes de levantamentos de solo, é amplamente reconhecida na comunidade científica e de engenharia do solo (DUAN et al., 2012, MINASNY e MCBRATNEY, 2000, YAO et al., 2015, CHAPUIS, 2012 apud ZHAO et al., 2016). A PTF representa uma ferramenta valiosa para avaliar a precisão e a confiabilidade necessárias em aplicações específicas (WÖSTEN et al., 2001 apud ZHAO et al., 2016). Em geral, as propriedades hidráulicas do solo são influenciadas pela combinação de efeitos dos materiais originais do solo, vegetação, topografia e clima (BRANTLEY, 2008 apud ZHAO et al., 2016).

Tradicionalmente, as funções de pedotransferência (PTFs) eram predominantemente construídas com base em equações de regressão, especialmente utilizando regressão linear (múltipla) (RAWLS et al., 1982 apud XU et al., 2017). Contudo, a regressão linear foi substituída por técnicas de regressão não linear, uma vez que esta última foi considerada mais precisa (Wösten et al., 2001 apud XU et al., 2017). Recentemente, métodos de aprendizado de máquina, como as Redes Neurais Artificiais (RNA), têm ganhado destaque no desenvolvimento de PTFs (TWARAKAVI et al., 2009 apud XU et al., 2017).

O emprego de técnicas de aprendizado de máquina (ML) em diversas áreas da ciência experimentou um crescimento rápido, especialmente ao longo da última década. Na pesquisa científica do solo, notadamente na pedometria, modelos estatísticos têm sido empregados para "aprender" ou compreender a distribuição espacial e temporal do solo com base nos dados disponíveis (MCBRATNEY et al., 2019 apud PADARIAN; MINASNY; MCBRATNEY, 2020). A ampla disponibilidade de dados do solo, que podem ser adquiridos de maneira eficiente tanto remotamente quanto em proximidade, aliada à acessibilidade de algoritmos de código aberto, tem conduzido a uma rápida adoção de técnicas de ML na análise de dados do solo. Diversas aplicações notáveis de ML na ciência do solo incluem a previsão de tipos e propriedades do solo por meio de mapeamento digital de solo (DSM) ou funções de pedotransferência, bem como a análise de dados espectrais infravermelhos para inferir propriedades do solo. A análise de aprendizado automático de dados do solo também é utilizada para deduzir informações sobre os controles que influenciam a distribuição do solo. (PADARIAN; MINASNY; MCBRATNEY, 2020)

Um exemplo é o modelo ROSETTA, que se baseia em Redes Neurais Artificiais (SCHAAP et al., 2001 apud XU et al., 2017). Por sua vez, já foi apresentado um método de programação genética para criar funções de pedotransferência (PTFs) com o objetivo de estimar a condutividade hidráulica saturada (K_s) a partir do teor de areia, silte e argila, além da densidade aparente (PARASURAMAN et al., 2007 apud XU et al., 2017). A vantagem dos modelos de RNA, em comparação com as PTFs convencionais, está na ausência da necessidade de conhecimento prévio de um conceito de modelo (SCHAAP et al., 1998 apud XU et al., 2017). Além disso, a estrutura de aprendizado dos modelos de RNA pode ser aplicada com êxito mesmo quando a relação entre variáveis dependentes e independentes é desconhecida (CHANG E ISLAM, 2000 apud XU et al., 2017). Adicionalmente, os modelos de RNA têm a capacidade de aproximar de maneira mais eficiente relações causais de variáveis não lineares em comparação com os modelos de regressão linear tradicionais. Já foi destacado que os modelos de RNA apresentam uma precisão de previsão superior quando comparados à análise de regressão múltipla (ERZIN et al., 2009 apud XU et al., 2017).

Recentemente, os algoritmos de regressão não paramétricos, que estão vinculados ao campo do aprendizado de máquina, surgiram como uma interface conveniente que combina a eficiência das técnicas estatísticas convencionais com a minúcia e complexidade das abordagens baseadas em princípios físicos. Esses algoritmos também oferecem a capacidade de processar de maneira eficiente um crescente volume de dados de observação da Terra (VERRELST et al., 2015; VERRELST et al., 2012 apud HOUBORG; MCCABE, 2018). Entre as abordagens comuns de aprendizado estão inclusas árvores de decisão, como as florestas aleatórias propostas por Breiman, redes neurais artificiais, conforme apresentadas por Haykin, kernels, incluindo as máquinas de vetores de suporte conforme proposto por Vapnik, e a regressão de processos gaussianos conforme descrita por Rasmussen e Williams (BREIMAN, 2001; HAYKIN, 1998; VAPNIK et al., 1996; RASMUSSEN E WILLIAMS, 2005 apud HOUBORG; MCCABE, 2018).

No contexto do aprendizado de máquina, esses algoritmos permitem desvendar associações complexas, como relacionamentos não lineares, entre uma propriedade alvo e um potencialmente ilimitado número de variáveis preditoras explicativas, tudo isso sem a necessidade de conhecimento explícito sobre os processos subjacentes. Isso se traduz em permitir que "os dados falem por si" (MCCABE et al., 2017 apud HOUBORG; MCCABE, 2018).

No entanto, dado que as aplicações de mapeamento baseadas em aprendizado de máquina geralmente dependem de dados coletados em campo para o treinamento de modelos (MUTANGA et al., 2012; VERRELST et al., 2015 apud HOUBORG; MCCABE, 2018), os desafios relacionados à transferibilidade do modelo provavelmente persistirão (VUOLO et al., 2013 apud HOUBORG; MCCABE, 2018). Além disso, as capacidades de generalização podem variar dependendo do algoritmo utilizado (VERRELST et al., 2013;

VERRELST et al., 2012 apud HOUBORG; MCCABE, 2018).

1.2 JUSTIFICATIVA

O presente trabalho de pesquisa se justifica pela crescente relevância do solo como elemento central na abordagem dos desafios ambientais globais contemporâneos.

O solo desempenha um papel fundamental na abordagem da maioria dos desafios ambientais globais atuais, incluindo questões relacionadas à segurança alimentar, disponibilidade de água, energia e diversidade biológica (BOUMA e MCBRATNEY, 2013 apud KESKIN; GRUNWALD; HARRIS, 2019).

Recentemente, as funções de transferência de pedo-transferência têm sido extensivamente utilizadas em modelos de processos do solo como uma forma de enfrentar desafios ambientais dentro do âmbito das ciências do sistema terrestre (VAN LOOY et al., 2017 apud SONG et al., 2020).

O conteúdo de água no solo (CAS) desempenha um papel crucial nos processos de troca entre o solo e o ar, influenciando diretamente os balanços de água e energia. É um conhecimento fundamental para a gestão de recursos hídricos em diversas áreas, incluindo a prevenção de desastres naturais como deslizamentos de terra e inundações, a previsão de fluxos por meio de modelos hidrológicos com diferentes objetivos, a estimativa da produção de sedimentos e a investigação do transporte de nutrientes e poluentes no solo. O CAS possui uma forte correlação com o fluxo dos rios e pode ser utilizado para sua estimativa (HUMPHREY et al., 2016 apud OLIVEIRA VANESSA SARI; PEDROLLO, 2017).

Compreender as características hidráulicas do solo é essencial para abordar uma variedade de desafios na gestão de solos e recursos hídricos em contextos agrícolas, ecológicos e ambientais. Essas características são cruciais para entender e antecipar a movimentação de água e substâncias dissolvidas, além de serem importantes para modelar os processos de transferência de calor e matéria próximos à superfície do solo (CORNELIS et al., 2001 apud MOHANTY; SINHA; PAINULI, 2015). As características hidráulicas do solo, como a umidade do solo no ponto de capacidade de campo (FC) e no ponto de murcha permanente (PWP), são dois elementos fundamentais necessários em diversos modelos biofísicos (RAB et al., 2011 apud MOHANTY; SINHA; PAINULI, 2015), além disso, são empregados para determinar a quantidade de água necessária para a irrigação (GIVI e PATEL, 2004 apud MOHANTY; SINHA; PAINULI, 2015) e também para avaliar a disponibilidade de água, um fator crucial na determinação da viabilidade de uma produção agrícola (SARMADIAN, 2010 apud MOHANTY; SINHA; PAINULI, 2015).

As características de fluxo de água no solo influenciam o movimento da água dentro do sistema solo-planta-atmosfera, regulando processos como a recarga de aquíferos e a transferência de nutrientes entre o solo e as plantas. Compreender a hidrodinâmica do solo

é crucial para simular fenômenos físicos relacionados ao teor de umidade do solo. Apesar dos avanços nos métodos de medição, determinar com precisão as propriedades hidráulicas do solo, especialmente em solos não perturbados e em condições secas, continua sendo um desafio. Além disso, a avaliação dessas propriedades é um processo lento, meticuloso e dispendioso (MERDUN, 2010 apud ESMAEELNEJAD et al., 2015).

Além disso, ao fazer uma pesquisa no Google Scholar procurando por "machine learning + pedotransfer function" que são temas relevantes neste trabalho, ao longo dos últimos dez anos, podemos observar um aumento significativo no número de artigos publicados com esses assuntos em cada ano, como mostra o gráfico abaixo da Figura 1:

Figura 1 – Gráfico de evolução da produção de artigos sobre "machine learning + pedotransfer function" (2014-2023)



Fonte: do autor

Pode-se observar que em 2014 houve 54 artigos publicados, já em 2023 houve 637, o que representa um aumento de aproximadamente onze vezes o valor anterior. Tal crescimento pode ser explicado pelo aumento da preocupação ambiental, pelos avanços em sistemas computacionais e na coleta de dados, e pela crescente demanda por soluções eficientes para a gestão do solo e da água.

O interesse pela dinâmica da umidade do solo está em constante crescimento, especialmente com o aumento da preocupação com a gestão ambiental. O desenvolvimento de sistemas computacionais para o processamento automatizado de informações do solo abriu novas perspectivas na pesquisa agrícola. Avanços recentes na coleta de amostras do solo e na análise das propriedades físicas do solo criaram oportunidades sem precedentes para integrar dados do solo em diversas aplicações, incluindo agricultura, hidrologia,

engenharia civil, geociências, engenharia de petróleo, ciências ambientais, entre outras, gerando uma grande demanda por bancos de dados de solos.

Embora a coleta de informações do solo seja rotineira em muitos países por meio de levantamentos de solo, esses dados geralmente não incluem propriedades hidráulicas fundamentais, como a curva de retenção de água (WRC) e a condutividade hidráulica saturada (K_s). Isso ressalta a necessidade de métodos rápidos e econômicos para estimar essas propriedades (Schaap et al., 2001; Givi et al., 2004; Alvarez-Acosta et al., 2012). Minasny e Hartemink (2011) observaram uma grande escassez de medições de propriedades do solo em muitos países tropicais. Além disso, a disponibilidade de informações sobre o solo em escala regional é limitada em muitos países ao redor do mundo. Isso frequentemente deixa pesquisadores, planejadores e tomadores de decisão nessas regiões com dados insuficientes para trabalhar (Patil et al., 2010, 2011).

O uso de técnicas de aprendizado de máquina (ML) em diversos campos científicos tem crescido significativamente, especialmente nos últimos 10 anos. Na investigação científica do solo, em particular na pedometria, modelos estatísticos têm sido empregados para "aprender" ou compreender a distribuição do solo no espaço e no tempo a partir dos dados disponíveis (MCBRATNEY et al., 2019 apud PADARIAN; MINASNY; MCBRATNEY, 2020). A disponibilidade crescente de dados do solo, obtidos de forma eficiente por meio de métodos remotos e próximos, juntamente com algoritmos de código aberto disponíveis gratuitamente, tem impulsionado a rápida adoção de técnicas de ML na análise de dados do solo. Diversas aplicações bem conhecidas de ML na ciência do solo incluem a previsão de tipos e propriedades do solo por meio de mapeamento digital do solo (DSM) ou funções de pedotransferência, assim como a análise de dados espectrais infravermelhos para inferir propriedades do solo. A análise automatizada de dados do solo por meio de aprendizado de máquina também é utilizada para deduzir os fatores que controlam a distribuição do solo (PADARIAN; MINASNY; MCBRATNEY, 2020).

Outro ponto a se destacar é o fato de recentemente no Prêmio Internacional de Arquitetura Paisagística Cornelia Hahn Oberlander 2023 foi premiado o arquiteto paisagista de Pequim, Kongjian Yu, pelo seu conceito de "cidades esponja".

A rápida urbanização tem levado à expansão de grandes áreas impermeáveis, aumentando significativamente o risco de inundações em diversas cidades ao redor do mundo (BOOTH e JACKSON, 1997 apud FU et al., 2020). Nas últimas décadas, tanto inundações urbanas quanto alagamentos tornaram-se eventos frequentes ao longo da costa sudeste da China, especialmente nas bacias hidrográficas, devido ao intenso volume de água das tempestades (LIU et al., 2020 apud FU et al., 2020). Dentro desse contexto, surgiu o conceito de Cidade Esponja (SPC), que visa restaurar os processos hidrológicos naturais, diminuindo o escoamento superficial por meio de medidas como infiltração, detenção e retenção (WANG et al., 2018 apud FU et al., 2020). Vários municípios têm

implementado tais projetos para aumentar a permeabilidade urbana, melhorar a infiltração do solo e reduzir o escoamento superficial (XIA et al., 2017 apud FU et al., 2020). Como resultado, cada vez mais projetos de infraestrutura estão incorporando a infiltração como uma alternativa aos sistemas de drenagem convencionais (FU et al., 2020).

As estratégias adotadas nas cidades esponja envolvem diversas ações, como o cultivo de pastagens, a instalação de valas de relva e a criação de espaços verdes submersos, além de outras medidas (SHAFIQUE e KIM, 2015 apud JIA et al., 2020). Ao implementar essas medidas, foi possível controlar as fontes de escoamento de inundações e modificar as características hidrológicas da superfície urbana, regulando a umidade do solo por meio de diferentes técnicas de esponja (WANG et al., 2009, 2017 apud JIA et al., 2020). Esse processo também permitiu a acumulação e redistribuição da água, contribuindo para a regulação dos processos hidrológicos urbanos, a redução das inundações e a melhoria da gestão da água na cidade. Portanto, é essencial investigar as alterações nas características hidrológicas do solo que ocorrem após a implementação de medidas de cidades esponja (JIA et al., 2020).

E através da utilização de um modelo acoplado (SWMM/HYDRUS-1D), FU et al., em seu artigo estuda os efeitos hidrológicos no escoamento de águas pluviais utilizando as funções de pedotransferência com previsão por redes neurais para calcular parâmetros como o teor de água residual, teor de água saturada, parâmetro de forma hidráulica, parâmetro hidráulico e condutividade hidráulica saturada de cada camada do solo (FU et al., 2020).

A implementação bem-sucedida do conceito de "Cidade Esponja" depende fundamentalmente de uma compreensão precisa das propriedades hidráulicas do solo urbano. Para planejar e otimizar medidas como bacias de retenção, telhados verdes e pavimentos permeáveis, é crucial estimar com acurácia a capacidade de infiltração, retenção de água e condutividade hidráulica dos solos urbanos. As Funções de Pedotransferência (PTFs), especialmente as aprimoradas por Redes Neurais Artificiais (RNAs), podem desempenhar um papel crucial nesse contexto, fornecendo estimativas mais precisas e confiáveis das propriedades hidráulicas do solo urbano, auxiliando no projeto de soluções de "Cidade Esponja" mais eficazes e sustentáveis. Assim, a busca por PTFs mais precisas, como as investigadas nesta dissertação, ganha relevância direta para o avanço de iniciativas de gestão urbana da água como as "Cidades Esponja".

Embora diversos métodos de aprendizado de máquina tenham sido aplicados na modelagem de funções de pedotransferência, como regressão múltipla, árvores de decisão e máquinas de vetor de suporte (SVM), esses algoritmos apresentam limitações na captura de relações altamente não lineares e em cenários de dados heterogêneos. As Redes Neurais Artificiais (RNAs), por sua vez, destacam-se pela flexibilidade e pela reconhecida capacidade de aproximação universal de funções (HORNIK; STINCHCOMBE; WHITE,

1989), possibilitando a representação de interações complexas entre variáveis físicas do solo e suas propriedades hidráulicas.

Estudos recentes confirmam o desempenho superior das RNAs em relação a métodos tradicionais e a outros algoritmos de aprendizado de máquina na predição de propriedades hidráulicas do solo, como observado em Zhang e Schaap (2019), Montzka et al. (2017), Kotlar et al. (2019). Nesse contexto, a escolha pelas RNAs neste trabalho justifica-se não apenas pela robustez metodológica, mas também pela oportunidade de explorar estratégias modernas de ajuste de hiperparâmetros, como a otimização bayesiana, pouco aplicadas em estudos de pedotransferência, o que caracteriza o diferencial desta pesquisa.

Diante da relevância e dos desafios apresentados, o escopo deste trabalho se concentrará no desenvolvimento e avaliação de um modelo de PTF baseado em Redes Neurais Artificiais para aprimorar a estimativa da retenção de água no solo, conforme detalhado no próximo capítulo.

1.3 ELABORAÇÃO DOS OBJETIVOS

O presente trabalho tem como objetivo geral avaliar um modelo de aplicação de uma função de pedotransferência através da utilização de redes neurais parametrizadas para estimar propriedades do solo por meio de três conjuntos de dados coletados na literatura.

Esse objetivo geral é atingível através do desenvolvimento de alguns pontos específicos, que são:

- Realizar uma revisão sistemática e crítica da literatura científica para identificar e analisar os estudos mais recentes sobre a modelagem de Funções de Pedotransferência (PTFs) utilizando Redes Neurais Artificiais (RNAs) para a estimativa da retenção de água no solo, com foco nas arquiteturas de RNA, técnicas de otimização de hiperparâmetros, métricas de avaliação, principais resultados e lacunas existentes.
- Selecionar, compilar e preparar três conjuntos de dados secundários de propriedades do solo, provenientes da literatura científica e representativos de diferentes regiões geográficas e tipos de solo, para serem utilizados no treinamento, validação e teste dos modelos de Redes Neurais Artificiais (RNAs).
- Implementar e parametrizar modelos de Redes Neurais Artificiais (RNAs) utilizando a biblioteca Keras em Python, explorando diferentes arquiteturas e hiperparâmetros para a predição da retenção de água no solo.
- Otimizar os hiperparâmetros dos modelos de RNA através de técnicas de busca Bayesiana e validação cruzada, visando maximizar o desempenho preditivo e a capacidade de generalização dos modelos.

- Avaliar o desempenho preditivo dos modelos de RNA utilizando métricas estatísticas e validação cruzada, para quantificar a precisão, robustez e capacidade de generalização dos modelos em diferentes bases de dados.
- Analisar e discutir os resultados obtidos, interpretando os gráficos de desempenho, métricas de avaliação e comparando o desempenho dos modelos de RNA com modelos de referência e resultados de estudos prévios presentes na literatura.

1.4 DEFINIÇÃO DA METODOLOGIA

A metodologia empregada neste trabalho se caracteriza por uma abordagem quantitativa e computacional, focada no desenvolvimento e avaliação de um modelo preditivo baseado em redes neurais artificiais (RNA). A pesquisa segue as seguintes etapas principais:

1. **Revisão Bibliográfica:** Realização de um levantamento abrangente da literatura científica pertinente à modelagem de funções de pedotransferência (PTFs), aprendizado de máquina aplicado a propriedades do solo, e técnicas de otimização de modelos computacionais. Esta etapa visa fundamentar teoricamente o trabalho e identificar as metodologias e abordagens mais relevantes na área.
2. **Seleção e Preparação dos Dados:** Utilização de bases de dados secundárias, provenientes de estudos científicos publicados e repositórios de dados abertos. As bases de dados serão selecionadas com base em sua relevância para o estudo de propriedades hidráulicas do solo e sua adequação para treinamento e validação de modelos de aprendizado de máquina. A preparação dos dados envolverá a organização, limpeza e formatação dos conjuntos de dados para serem utilizados nos modelos de RNA.
3. **Desenvolvimento e Parametrização do Modelo de RNA:** Construção e implementação de modelos de redes neurais artificiais utilizando a biblioteca Keras em ambiente Python. A arquitetura da rede, incluindo o número de camadas, neurônios por camada, funções de ativação e outros hiperparâmetros, será definida e otimizada através de técnicas de busca Bayesiana e validação cruzada, visando alcançar o melhor desempenho preditivo.
4. **Treinamento e Validação do Modelo:** Divisão dos dados em conjuntos de treinamento, validação e teste. O modelo de RNA será treinado utilizando o conjunto de treinamento, com ajuste dos pesos sinápticos através do algoritmo de otimização Adam. A validação será realizada durante o treinamento para monitorar o desempenho e evitar o sobreajuste. A avaliação final do modelo será feita utilizando o conjunto de teste, através de métricas de desempenho como o coeficiente de determinação (R^2), erro quadrático médio (RMSE) e correlação de Pearson (R).

5. **Análise e Discussão dos Resultados:** Interpretação dos resultados obtidos, com análise dos gráficos de desempenho, métricas de avaliação e comparação com dados de referência e resultados de outros estudos presentes na literatura. A discussão abordará as potencialidades e limitações da metodologia proposta, bem como as implicações dos resultados para a área de modelagem de propriedades do solo.
6. **Documentação e Reprodutibilidade:** Registro detalhado de todas as etapas metodológicas, incluindo a descrição dos dados, código de programação, parâmetros utilizados e resultados obtidos, visando garantir a transparência e a reprodutibilidade da pesquisa.

Esta metodologia busca seguir um fluxo lógico e sistemático, característico de pesquisas em modelagem computacional, com ênfase na utilização de técnicas de aprendizado de máquina para aprimorar a estimativa de propriedades do solo e contribuir para o avanço do conhecimento na área.

1.5 ESTRUTURA DO TRABALHO

Para melhor organização e clareza, o presente trabalho foi estruturado em capítulos, cada um com um objetivo específico, que em conjunto visam apresentar de forma completa e detalhada a pesquisa realizada. A estrutura da dissertação é descrita a seguir:

O Capítulo 1 apresenta as considerações iniciais, a justificativa, os objetivos e a estrutura do texto. O Capítulo 2 descreve a fundamentação teórica, abordando conceitos de funções de pedotransferência, características hidráulicas do solo e técnicas de aprendizado de máquina. O Capítulo 3 detalha os materiais e métodos empregados, incluindo as bases de dados, o processo de construção dos modelos e os critérios de avaliação utilizados. O Capítulo 4 apresenta e discute os resultados obtidos. Por fim, o Capítulo 5 reúne as conclusões e sugestões para trabalhos futuros.

Esta estrutura visa guiar através do desenvolvimento da pesquisa, desde a contextualização e fundamentação teórica, passando pela metodologia e apresentação dos resultados, até as conclusões e perspectivas futuras, proporcionando uma compreensão clara e abrangente do trabalho realizado.

2 FUNDAMENTAÇÃO TEÓRICA

A fundamentação teórica apresentada a seguir reúne os principais conceitos relacionados às funções de pedotransferência (PTFs), às propriedades hidráulicas do solo e às técnicas de aprendizado de máquina, com ênfase em redes neurais artificiais (RNAs) e métodos de otimização de hiperparâmetros. Esses elementos fornecem a base necessária para compreender as escolhas metodológicas e análises realizadas neste trabalho.

O estudo das propriedades do solo e sua influência na dinâmica da água é um tema essencial na engenharia e nas ciências ambientais. Modelos preditivos, como as Funções de Pedotransferência (PTFs), têm sido amplamente utilizados para estimar características hidrofísicas do solo a partir de informações básicas, reduzindo a necessidade de medições diretas, que podem ser caras e demoradas.

Com o avanço do Aprendizado de Máquina, métodos baseados em Redes Neurais Artificiais vêm sendo explorados como alternativas promissoras para a modelagem das PTFs, permitindo capturar relações complexas entre variáveis do solo e melhorando a precisão das previsões. Entretanto, a eficácia dessas abordagens depende de diversos fatores, como a qualidade dos dados utilizados, a arquitetura da rede e os critérios de validação dos modelos.

Neste capítulo, são abordados os conceitos fundamentais para o entendimento do presente estudo. Primeiramente, são discutidas as propriedades físicas e hidráulicas do solo, essenciais para compreender a importância das PTFs. Em seguida, são apresentados os princípios das Funções de Pedotransferência, sua evolução e limitações. Por fim, explora-se o potencial das técnicas de Aprendizado de Máquina, com ênfase no uso de Redes Neurais Artificiais para a predição de propriedades do solo.

2.1 O SOLO E SUAS PROPRIEDADES

2.1.1 ESTRUTURA E COMPOSIÇÃO DO SOLO

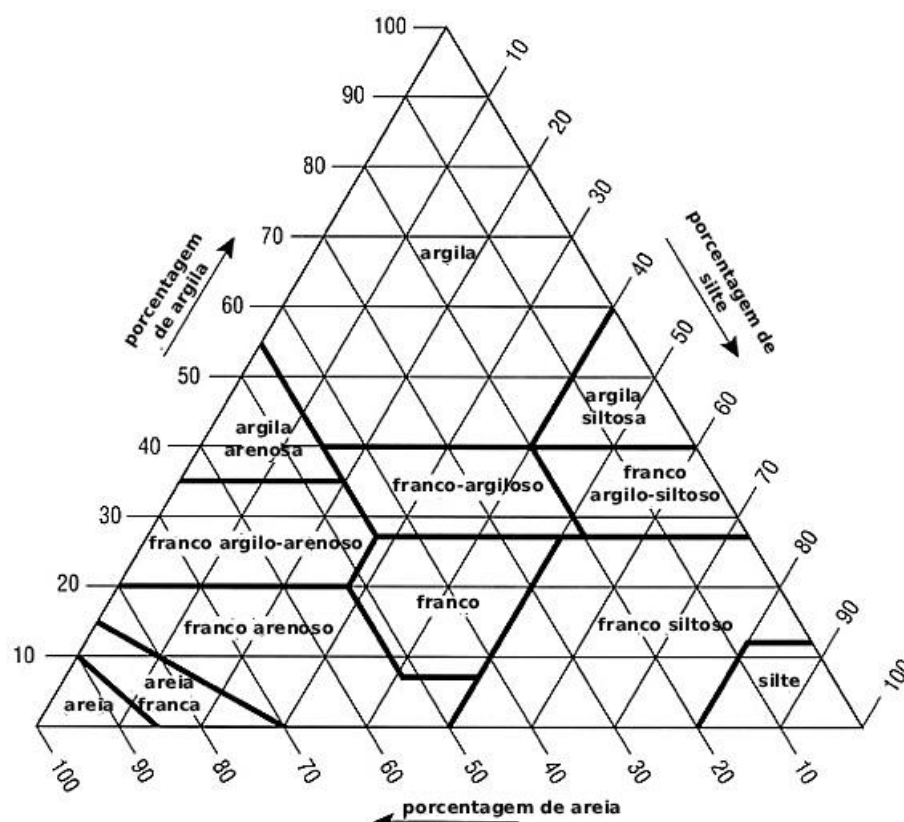
O solo desempenha um papel crítico nos ecossistemas da Terra, atuando como meio para o crescimento das plantas, regulador da dinâmica da água e habitat para diversas formas de vida. Sua definição evoluiu para abranger propriedades físicas, químicas e biológicas que determinam sua funcionalidade em ecossistemas naturais e manejados. A importância do solo é particularmente notável nas áreas de engenharia, agricultura e gestão ambiental. Pesquisadores reconhecem o solo como um recurso não renovável essencial para a manutenção da biodiversidade, a regulação da qualidade da água e o suporte aos sistemas de produção de alimentos (SCHOONOVER; CRIM, 2015; JIMÉNEZ; FLORES; DELGADO-INIESTA, 2023; KOCH et al., 2013). O conceito de "segurança do solo" destaca a necessidade de um manejo sustentável para preservar essas funções vitais (KOCH et al.,

2013; YAWSON et al., 2016).

A composição do solo resulta da interação complexa entre diferentes componentes, consistindo principalmente de partículas minerais, matéria orgânica, água e ar. As frações minerais incluem partículas de tamanhos variados — areia, silte e argila — cada uma contribuindo com propriedades físicas distintas. A areia proporciona excelente drenagem e aeração, o silte melhora a retenção de água e a capacidade de armazenamento de nutrientes, enquanto a argila influencia a plasticidade e a coesão do solo (ZHANG et al., 2022; CIORUTA; COMAN, 2022). A matéria orgânica, derivada da decomposição de plantas e animais, tem um papel central no enriquecimento do solo com nutrientes e na formação de agregados, essenciais para a estrutura do solo (WILPISZESKI et al., 2019). Além disso, o solo contém água e ar, ambos cruciais para a atividade biológica e o crescimento das plantas.

A Figura 2 apresenta o triângulo textural do solo utilizado pelo Departamento de Agricultura dos Estados Unidos (USDA). Esse diagrama permite classificar o solo em diferentes classes texturais a partir da proporção relativa de areia, silte e argila. Solos mais arenosos apresentam maior condutividade hidráulica e menor capacidade de retenção de água, enquanto os mais argilosos retêm mais água, mas com menor disponibilidade hídrica para as plantas. As classes intermediárias, como os francos, representam combinações que equilibram essas propriedades (TWARAKAVI; ŠIMŮNEK; SCHAAP, 2010).

Figura 2 – Um diagrama ternário do triângulo textural do solo, traduzido para o português, que exibe as diferentes classificações de textura do solo segundo o Departamento de Agricultura dos Estados Unidos (USDA).



Fonte: adaptado de (TWARAKAVI; ŠIMŮNEK; SCHAAP, 2010)

A estrutura do solo refere-se à disposição das partículas e à formação de agregados ou grumos, facilitando o movimento do ar e da água dentro do perfil do solo. Solos bem estruturados apresentam uma rede de espaços porosos que aumentam sua capacidade de retenção de umidade e sustentam o desenvolvimento das raízes (TUTOVA et al., 2022). A estrutura do solo é influenciada por fatores bióticos e abióticos, incluindo a atividade microbiana, as raízes das plantas e perturbações mecânicas. Por exemplo, minhocas e outros organismos do solo contribuem para a estabilização dos agregados por meio de suas atividades, promovendo a saúde e fertilidade do solo (TUTOVA et al., 2022). Processos químicos, como a ligação de partículas do solo pela matéria orgânica, e processos físicos, como a compactação e a carga exercida sobre o solo, também afetam significativamente a agregação e a estabilidade do solo (ZHANG et al., 2022; ZHOU et al., 2020).

A estrutura e a composição do solo influenciam profundamente a retenção de água e as propriedades mecânicas. Os agregados do solo afetam a infiltração de água e ajudam a reter a umidade, o que é essencial para a saúde das plantas. Estudos indicam que solos com boa estabilidade dos agregados tendem a resistir à erosão e a manter uma alta taxa de infiltração e retenção de água, sustentando a vegetação durante períodos de seca (ZHOU et al., 2020). A integridade mecânica do solo, incluindo sua resistência e capacidade de

suportar cargas, depende de sua organização estrutural. Arranjos desfavoráveis podem tornar o solo suscetível a deslizamentos ou colapsos sob carga (YIN et al., 2021; SUHAÏLA et al., 2021).

Em resumo, os conceitos de estrutura e composição do solo são fundamentais para compreender sua funcionalidade em diferentes contextos, desde áreas agrícolas até ambientes urbanos. A interação entre matéria orgânica, conteúdo mineral, atividade biológica e processos ambientais contribui para a natureza dinâmica dos solos, evidenciando a necessidade de práticas de manejo consciente para garantir sua sustentabilidade para as futuras gerações.

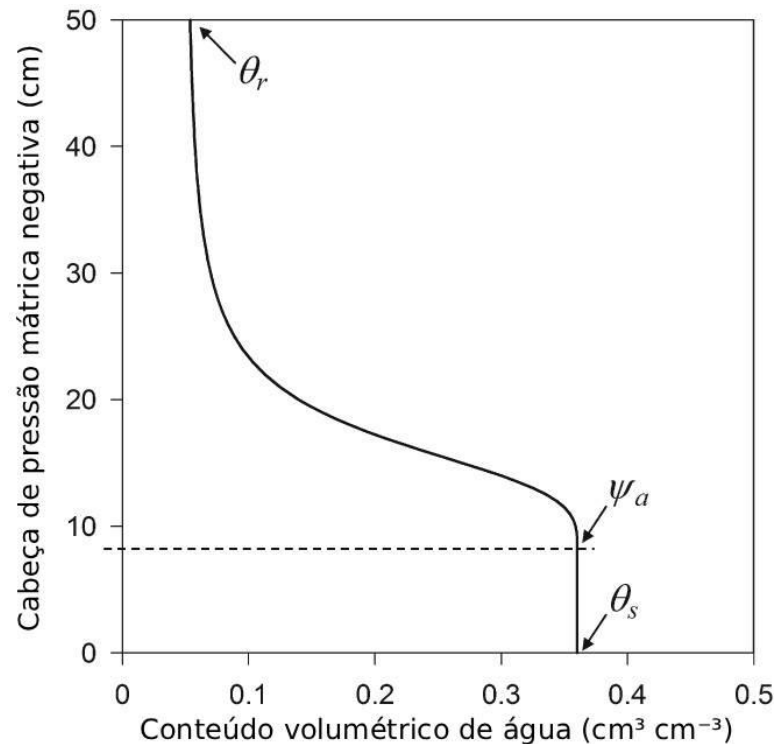
2.1.2 RETENÇÃO DE ÁGUA E CONDUTIVIDADE HIDRÁULICA

A retenção de água no solo é definida como a capacidade do solo de reter água contra as forças gravitacionais, uma propriedade crítica que afeta a produtividade agrícola, as dinâmicas ecológicas e as aplicações geotécnicas. Essa capacidade é fundamental para sustentar o crescimento das plantas e influencia processos como o manejo da irrigação e a recarga de águas subterrâneas (KUANG et al., 2020). A retenção adequada de água aumenta a disponibilidade hídrica para as plantas, ao mesmo tempo que reduz os riscos de seca e erosão, desempenhando assim um papel essencial na produtividade agrícola e na saúde do solo (GHANBARIAN et al., 2015).

A Curva de Retenção de Água no Solo (*Soil Water Retention Curve* — SWRC) é uma representação gráfica que descreve a relação entre o teor de umidade do solo e o potencial mátrico (o estado energético da água no solo) (KUANG; JIAO, 2014). Essa curva é crucial para estimar a retenção de água do solo em diferentes níveis de umidade. A forma e a posição da SWRC são influenciadas por diversos fatores, incluindo a textura e estrutura do solo, o teor de matéria orgânica, a compactação e a mineralogia (GHANBARIAN-ALAVIJEH; HUNT, 2012; WANG et al., 2015). Por exemplo, solos de textura mais grosseira (como os arenosos) apresentam drenagem rápida e menor capacidade de retenção de água em comparação com solos de textura mais fina (como os argilosos), que retêm água com maior eficiência devido à sua maior área superficial e ao menor tamanho de poros (WANG et al., 2016).

A curva característica de retenção de água no solo descreve a relação entre o conteúdo volumétrico de água e a pressão mátrica. Esse comportamento é representado pela Figura 3, em que os eixos foram traduzidos para o português: no eixo horizontal tem-se o *conteúdo volumétrico de água* ($\text{cm}^3 \text{cm}^{-3}$), enquanto o eixo vertical indica a *cabeça de pressão mátrica negativa* (cm). Além disso, são destacados parâmetros importantes: θ_r , corresponde à *umidade residual*, θ_s à *umidade de saturação* e Ψ_a ao ponto de *entrada de ar*, que indica a tensão na qual os primeiros macroporos começam a drenar.

Figura 3 – Gráfico típico da curva de retenção de água no solo para um solo arenoso usando uma equação proposta por van Genuchten, onde ψ_a é o valor de entrada de ar, θ_r e θ_s são os conteúdos volumétricos de água residual e saturada.



Fonte: adaptado de (KUANG; JIAO, 2014)

Observa-se que em valores próximos à saturação (θ_s), a maior parte da água encontra-se disponível nos macroporos do solo. Com o aumento da sucção, a água é gradualmente retida em microporos, reduzindo sua disponibilidade até atingir a umidade residual (θ_r), quando a água encontra-se fortemente adsorvida às partículas do solo e indisponível às plantas. O ponto de entrada de ar (Ψ_a) representa a tensão em que os primeiros macroporos começam a drenar (KUANG; JIAO, 2014).

A SWRC é, portanto, uma propriedade hidráulica fundamental e um alvo primário para a estimativa por meio de Funções de Pedotransferência (PTFs), especialmente aquelas aprimoradas por técnicas de aprendizado de máquina.

Os fatores que afetam a retenção de água incluem a textura e a estrutura do solo, sendo que texturas mais finas e uma estrutura bem desenvolvida normalmente resultam em maior retenção de água devido ao aumento do número de microporos (GHANBARIAN et al., 2015; RAFRAF et al., 2016). A matéria orgânica também desempenha um papel significativo na melhoria da retenção de água do solo, promovendo a agregação e modificando a estrutura dos poros, o que ajuda a reter a umidade (KUANG et al., 2020). Além disso, a compactação reduz o volume dos poros maiores, aumentando a predominância de microporos e impactando a acessibilidade da água para as plantas (KOZAITI; KOSTOPOULOU, 2019).

O conceito de condutividade hidráulica do solo é essencial para compreender o movimento da água dentro do perfil do solo. A condutividade hidráulica reflete a facilidade com que a água se desloca pelos poros do solo e é influenciada pela estrutura do solo e pelo grau de saturação (KATEB et al., 2019). A relação entre condutividade hidráulica e teor de água é descrita por modelos empíricos, sendo a equação de Van Genuchten um dos mais amplamente utilizados para descrever o comportamento hidráulico de solos não saturados (SAFFARI et al., 2019). Outra equação bem conhecida é a de Brooks-Corey, que caracteriza a curva de retenção de água e a condutividade hidráulica, enfatizando a natureza capilar do movimento da água no solo (WANG et al., 2015; KUANG; JIAO, 2014). Esses modelos fornecem informações essenciais sobre o comportamento do solo sob diferentes condições de umidade, facilitando a previsão da dinâmica do fluxo de água.

A modelagem matemática, especialmente por meio das equações de Van Genuchten e Brooks-Corey, permite uma compreensão quantitativa da retenção de água e da condutividade hidráulica em diferentes estados de umidade. O modelo de Van Genuchten caracteriza a curva ajustando parâmetros como α e n , permitindo calcular índices relacionados ao teor de água em diferentes potenciais mátricos e possibilitando a previsão das propriedades hidráulicas em diversos tipos de solo (KUANG et al., 2020; GHANBARIAN et al., 2015). Essas representações matemáticas também viabilizam o desenvolvimento de Funções de Pedotransferência (*Pedotransfer Functions* — PTFs), que servem como ferramentas práticas para estimar as propriedades hidráulicas do solo com base em atributos de fácil medição (GUZMÁN et al., 2019; TAO et al., 2019). As PTFs permitem uma modelagem eficiente em contextos agrícolas e ambientais, auxiliando na adoção de melhores práticas de manejo e conservação.

Em resumo, a compreensão da retenção de água e da condutividade hidráulica no solo é fundamental para o manejo eficiente da água em contextos agrícolas, ecológicos e geotécnicos. O uso de modelos como a SWRC e equações de condutividade hidráulica permite prever o comportamento do solo, possibilitando estratégias aprimoradas de irrigação, manejo do solo e restauração ambiental.

2.1.3 IMPORTÂNCIA DAS PROPRIEDADES FÍSICAS DO SOLO NA MODELAGEM

As propriedades físicas do solo englobam uma variedade de atributos, incluindo textura, estrutura, densidade e porosidade, que coletivamente influenciam os processos hidrológicos e são essenciais para o manejo eficiente da terra e da água. A textura do solo, definida pela proporção de areia, silte e argila, afeta significativamente a capacidade do solo de reter água e transmitir propriedades hidráulicas. A interação entre a textura e a estrutura do solo determina o tamanho dos poros e sua conectividade, influenciando o movimento da água e do ar ao longo do perfil do solo (TWARAKAVI; ŠIMŮNEK; SCHAAP, 2010; HARDIE et al., 2012). A densidade, especialmente a densidade aparente, é uma

medida da massa do solo por unidade de volume, enquanto a porosidade indica o volume de poros disponíveis para armazenamento de ar e água. Uma alta porosidade geralmente melhora a retenção de água, mas deve ser equilibrada com a densidade para garantir aeração e drenagem adequadas (PINHEIRO et al., 2021). Juntas, essas propriedades formam a base para a compreensão do comportamento do solo durante eventos de precipitação e irrigação.

A influência das propriedades físicas do solo na modelagem preditiva é substancial. A modelagem precisa da retenção de água, da condutividade hidráulica e da disponibilidade hídrica depende da representação exata dessas propriedades. As propriedades de retenção de água determinam quanta umidade pode ser armazenada no solo em diferentes níveis de tensão, enquanto a condutividade hidráulica descreve a facilidade com que a água pode fluir através da matriz do solo (LIAO et al., 2018; KANG et al., 2024). Há evidências substanciais de que variações na textura e na estrutura podem levar a discrepâncias significativas nas previsões, destacando a necessidade de uma caracterização detalhada dessas propriedades nos esforços de modelagem (BORMANN, 2010; DRAYE et al., 2010). O uso de Funções de Pedotransferência (*Pedotransfer Functions* — PTFs), que derivam parâmetros hidráulicos a partir de dados de solo de fácil medição, frequentemente depende do conhecimento preciso das propriedades físicas (BOHNE; RENGGER; WESSOLEK, 2019). Por exemplo, relações derivadas de dados de textura do solo podem fornecer estimativas para a condutividade hidráulica, ressaltando o papel crítico da textura nas previsões (DRAYE et al., 2010).

Abordagens de aprendizado de máquina surgiram como ferramentas valiosas para explorar as propriedades físicas do solo na melhoria da previsão de parâmetros hidráulicos. Esses modelos podem assimilar conjuntos de dados complexos para identificar padrões e aprimorar a precisão preditiva além dos métodos estatísticos tradicionais (LI; PAN; XU, 2013). A integração do aprendizado de máquina com PTFs permite atualizações dinâmicas das previsões à medida que novos dados se tornam disponíveis, promovendo um arcabouço robusto para a compreensão da dinâmica da água no solo em diferentes ambientes (HE et al., 2024). No entanto, apesar do seu potencial, desafios como a variabilidade espacial das propriedades físicas do solo complicam os esforços de modelagem. A heterogeneidade inerente do solo dificulta a obtenção de amostras representativas, podendo levar a imprecisões nas simulações (ZEINELDIN; BIRO; ALGHANNAM, 2021). Portanto, dados bem distribuídos e precisos são essenciais para superar essas limitações e garantir a robustez dos modelos preditivos.

Entre as técnicas de aprendizado de máquina, as Redes Neurais Artificiais (RNAs) têm se destacado por sua capacidade de modelar relações não lineares complexas e lidar com grandes volumes de dados, tornando-se particularmente promissoras para aprimorar as PTFs e a estimativa de propriedades hidráulicas do solo.

Técnicas avançadas, incluindo o uso de redes neurais e outras abordagens de aprendizado de máquina, oferecem caminhos promissores para aprimorar a modelagem das propriedades do solo. Esses métodos podem ser empregados para gerenciar melhor a incerteza nas previsões, uma vez que são capazes de capturar complexidades e variabilidades que métodos tradicionais podem não considerar (SPOKAS et al., 2016; PINNINGTON; QUAIFE; BLACK, 2018). Por exemplo, métodos estatísticos podem ser utilizados para quantificar a incerteza nas previsões de condutividade hidráulica, permitindo uma tomada de decisão mais informada sobre o manejo da terra e da água (GAMIE; SMEDT, 2017). Além disso, a integração de análises fractais para compreender as relações entre a estrutura do solo e suas propriedades hidráulicas pode fornecer insights mais profundos sobre a complexa interação entre solo e água (ARFEEN; KHAN, 2020). Ao abordar essas técnicas avançadas de modelagem e reconhecer os desafios associados, a literatura destaca a importância das propriedades físicas do solo na modelagem eficaz e na gestão dos recursos hídricos.

Em conclusão, a relevância das propriedades físicas do solo na modelagem não pode ser subestimada. Essas propriedades são componentes fundamentais para a compreensão e previsão do comportamento hidrológico. Sua integração em modelos aprimora a precisão das previsões e auxilia no desenvolvimento de estratégias sustentáveis para o manejo da terra. À medida que as técnicas de modelagem evoluem, especialmente com a incorporação do aprendizado de máquina e de abordagens estatísticas avançadas, uma compreensão mais detalhada das relações complexas entre as propriedades do solo e a dinâmica da água emergirá, permitindo uma gestão mais eficaz dos recursos naturais.

2.2 AS FUNÇÕES DE PEDOTRANSFERÊNCIA (PTF)

2.2.1 DEFINIÇÃO E APLICAÇÕES

As Funções de Pedotransferência (*Pedotransfer Functions* — PTFs) são modelos matemáticos utilizados para prever propriedades específicas do solo, particularmente propriedades hidráulicas, a partir de atributos do solo de fácil medição, como textura, densidade aparente e teor de carbono orgânico. Sua principal função é facilitar a estimativa de propriedades hidráulicas complexas, que muitas vezes são difíceis e custosas de medir diretamente em condições laboratoriais ou de campo. As PTFs atuam como uma ponte que converte dados simples em parâmetros hidrológicos essenciais para diversas aplicações na ciência do solo (OSTOVARI; ASGARI; CORNELIS, 2015; BOHNE; RENGGER; WES-SOLEK, 2019; QIAO et al., 2018). A Figura 4 apresenta uma visão esquemática dessa relação.

Figura 4 – PTFs relacionam propriedades do solo simples de medir a parâmetros menos disponíveis de processos do sistema terrestre.



Fonte: adaptado de (LOOY et al., 2017)

Observa-se que, a partir de informações relativamente simples, como textura e teor de matéria orgânica, é possível estimar parâmetros hidráulicos (ex.: retenção de água, condutividade), de transporte de solutos, térmicos e até biogeoquímicos. Essa capacidade torna as PTFs ferramentas fundamentais em hidrologia e agricultura, reduzindo a necessidade de ensaios experimentais demorados e custosos.

A importância das PTFs reside em sua capacidade de reduzir o tempo e os custos associados às técnicas de medição direta. Métodos tradicionais para determinar propriedades hidráulicas do solo podem ser intensivos em mão de obra e onerosos, especialmente em áreas espaciais extensas. Em contrapartida, as PTFs utilizam dados prontamente disponíveis, agilizando o processo para cientistas do solo e hidrólogos (CHIRICO; MEDINA; ROMANO, 2010; QIAO et al., 2018; TOBELLA et al., 2024). Estudos demonstraram que as PTFs podem estimar de forma eficaz parâmetros como curvas de retenção de água e condutividade hidráulica saturada, essenciais para a simulação do movimento da água e do transporte de solutos no solo (QIAO et al., 2018; OBIERO et al., 2013; GÜLSER; CANDEMIR, 2014). A precisão dessas funções pode aprimorar significativamente os modelos hidrológicos, melhorar as estratégias de manejo da paisagem e apoiar práticas agrícolas sustentáveis (GÜLSER; CANDEMIR, 2014; BENKE et al., 2020).

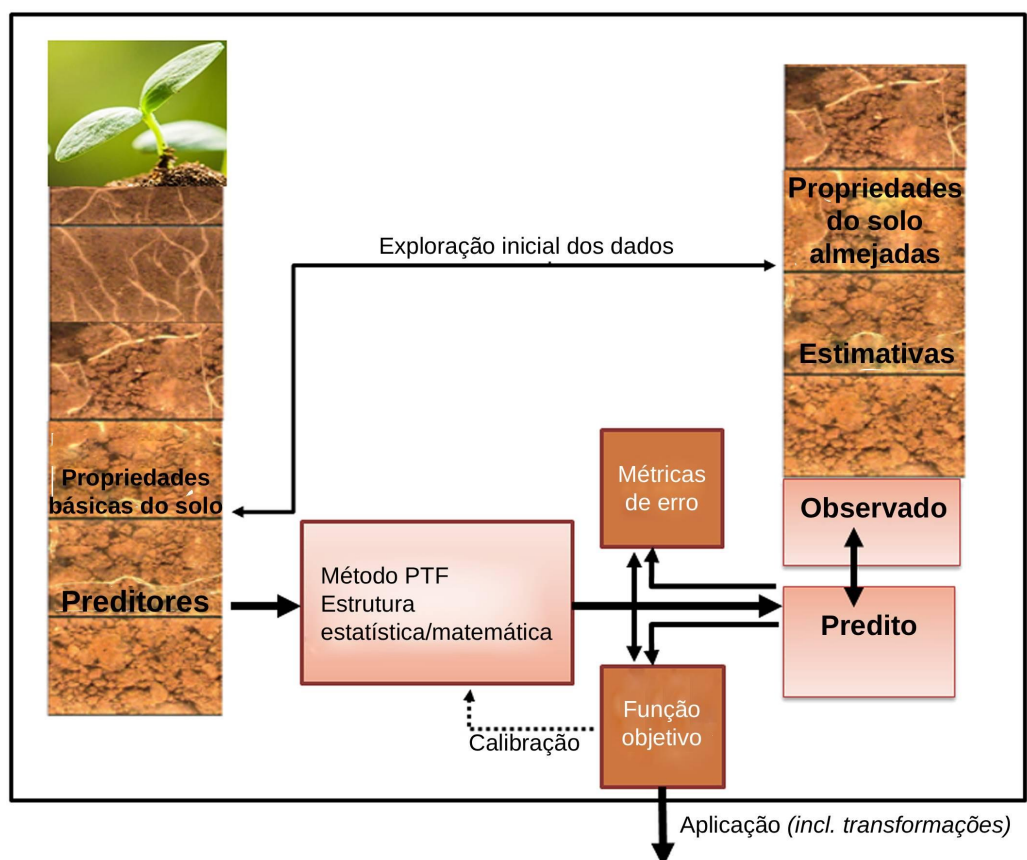
As PTFs têm ampla aplicação em diversos domínios, incluindo modelagem hidrológica, agricultura de precisão e estudos ambientais. Na hidrologia, elas são fundamentais para simular a dinâmica da água no solo e gerenciar recursos hídricos ao integrar proprie-

dades hidráulicas do solo em modelos hidrológicos (OSTOVARI; ASGARI; CORNELIS, 2015; CHIRICO; MEDINA; ROMANO, 2010; QIAO et al., 2018). Por exemplo, Ostovari et al. demonstraram a eficácia das PTFs na previsão da capacidade de campo e do ponto de murcha permanente usando diferentes conjuntos de dados, estabelecendo seu papel nas simulações da dinâmica da água (OSTOVARI; ASGARI; CORNELIS, 2015). Além disso, no contexto agrícola, as PTFs auxiliam no planejamento da irrigação e na modelagem do crescimento das culturas, fornecendo estimativas precisas das condições de umidade do solo (KALUMBA et al., 2020).

Ademais, as PTFs contribuem significativamente para práticas agrícolas sustentáveis e conservação do solo, ao esclarecer a relação entre a saúde do solo e suas propriedades hidrológicas. Elas permitem a otimização dos cronogramas de irrigação e a avaliação do risco de erosão do solo, prevendo como os solos se comportam sob diferentes condições de umidade (KALUMBA et al., 2020; ARSHAD et al., 2013). Essas informações são fundamentais para a tomada de decisões que aumentam a produtividade agrícola e, ao mesmo tempo, promovem a conservação ambiental (TOBELLA et al., 2024; BENKE et al., 2020). O desafio global de aumentar a segurança alimentar, ao mesmo tempo em que se conserva os recursos naturais, ressalta o papel crucial das PTFs no equilíbrio entre produtividade agrícola e sustentabilidade (BENKE et al., 2020; TÓTH et al., 2014).

Em suma, as Funções de Pedotransferência são ferramentas inestimáveis na ciência moderna do solo, permitindo a estimativa indireta de propriedades hidráulicas essenciais. Ao reduzir a necessidade de medições laboriosas, as PTFs aumentam a eficiência da modelagem hidrológica, da agricultura de precisão e de práticas ambientais sustentáveis. Seu desenvolvimento e aplicação continuarão a evoluir, incorporando técnicas avançadas de estatística e aprendizado de máquina para melhorar previsões e enfrentar desafios emergentes no manejo e conservação do solo. Portanto, o desenvolvimento de funções de pedotransferência (PTFs) pode ser compreendido como um processo que parte de variáveis básicas do solo, denominadas preditores, e busca estimar propriedades mais complexas, consideradas como variáveis-alvo. Esse fluxo é representado na Figura 5.

Figura 5 – Conceito geral do desenvolvimento de PTFs, fundamentado em um banco de dados de calibração contendo tanto as propriedades básicas quanto as propriedades do solo a serem estimadas.



Fonte: adaptado de (LOOY et al., 2017)

Observa-se que as PTFs utilizam atributos básicos como textura, densidade e matéria orgânica para estimar parâmetros hidráulicos ou de transporte. O processo envolve a definição de um modelo estatístico ou matemático, sua calibração e posterior validação por meio de métricas de erro e funções objetivo. A etapa final corresponde à aplicação do modelo, que pode incluir transformações adicionais, resultando em valores preditos que são comparados com os observados.

2.2.2 TIPOS DE PTFs

As Funções de Pedotransferência (PTFs) são modelos estatísticos utilizados para estimar propriedades hidráulicas do solo a partir de atributos do solo facilmente mensuráveis, melhorando significativamente o manejo do solo e a modelagem hidrológica. Os tipos de PTFs podem ser amplamente categorizados em PTFs empíricas, PTFs baseadas em aprendizado de máquina (ML), PTFs híbridas e considerações relacionadas aos desafios de cada método.

As PTFs empíricas são modelos estatísticos tradicionais baseados predominantemente em abordagens de regressão linear ou múltipla. Essas PTFs requerem calibração

específica para regiões e tipos de solo definidos a fim de garantir sua precisão e confiabilidade. Um exemplo proeminente de PTFs empíricas inclui os modelos desenvolvidos por Saxton e Rawls, que incorporam dados sobre textura e estrutura do solo para prever a retenção de água e a condutividade hidráulica (FIDALSKI; HELBEL, 2020; ZHANG; SCHAAP, 2017). O modelo Rosetta é amplamente reconhecido como uma PTF empírica projetada para estimar múltiplas propriedades hidráulicas do solo usando características básicas do solo; no entanto, frequentemente necessita de recalibração quando aplicado a novos conjuntos de dados ou tipos de solo (ZHANG; SCHAAP, 2017; BOHNE; RENGGER; WESSOLEK, 2019). As PTFs empíricas fornecem estimativas rápidas das propriedades hidráulicas, mas podem ter dificuldades para capturar relações não lineares complexas inerentes aos sistemas naturais do solo, levando a limitações em sua aplicabilidade generalizada em diversos tipos de solo e regiões (OBIERO et al., 2013; PACHEPSKY et al., 2014).

Ao transitar para as PTFs baseadas em aprendizado de máquina, esses modelos aproveitam algoritmos avançados para descobrir padrões e relações complexas nos dados do solo. Técnicas como Redes Neurais Artificiais (ANNs), Random Forest e Máquinas de Vetores de Suporte (SVMs) são comumente empregadas para produzir previsões mais robustas do que seus equivalentes empíricos. Por exemplo, Zhang e Schaap demonstraram que uma ANN poderia melhorar a estimativa de parâmetros hidráulicos ao acomodar interações não lineares dentro dos conjuntos de dados do solo (ZHANG; SCHAAP, 2017; JENA et al., 2021). O Random Forest tem demonstrado forte capacidade de modelar relações não lineares complexas em grandes conjuntos de dados, superando abordagens tradicionais de regressão linear (HENGL et al., 2017). Além disso, as aplicações de aprendizado de máquina têm demonstrado sucesso considerável na previsão da retenção de água e da condutividade hidráulica em diversos tipos de solo, incluindo estudos realizados nos solos tropicais do Sri Lanka (GUNARATHNA et al., 2019; JENA et al., 2021). Essas abordagens baseadas em ML exigem menos suposições sobre os dados e frequentemente produzem maior precisão nas previsões, tornando-se cada vez mais populares nas ciências do solo contemporâneas (JORDA et al., 2015).

O conceito emergente de PTFs híbridas busca aprimorar as vantagens das abordagens empíricas e de aprendizado de máquina, integrando modelos empíricos tradicionais com técnicas de aprendizado de máquina. Essa integração visa refinar previsões empíricas e corrigir vieses, ao mesmo tempo que melhora a generalização das PTFs em diversos contextos. Ao utilizar algoritmos de ML para ajustar ou reavaliar parâmetros empíricos, os modelos híbridos podem capitalizar o conhecimento existente e, simultaneamente, acomodar as complexidades de diferentes conjuntos de dados do solo (BAGNALL et al., 2022). O potencial das PTFs híbridas reside na sua maior precisão e adaptabilidade, o que pode melhorar significativamente as previsões das propriedades do solo em diferentes configurações ambientais.

Apesar dos avanços nas PTFs empíricas e baseadas em ML, ainda existem desafios em sua aplicação. As principais preocupações incluem a dependência da disponibilidade de conjuntos de dados representativos e de alta qualidade para os processos de treinamento e validação. Questões como a transferência surgem quando modelos calibrados para um tipo ou região de solo não apresentam bom desempenho em outros contextos, enfatizando a necessidade de abordagens robustas de validação cruzada na ciência do solo (PACHEPSKY et al., 2014; KOESTEL et al., 2018). Além disso, embora os modelos de ML geralmente apresentem bom desempenho, eles podem criar um efeito de "caixa-preta", tornando difícil interpretar as influências subjacentes de diferentes parâmetros do solo nas propriedades estimadas (MCBRATNEY; MINASNY; TRANTER, 2011; KOESTEL et al., 2018). Equilibrar a interpretabilidade das PTFs com seu poder preditivo é essencial para avançar suas aplicações na modelagem ambiental e no manejo do solo.

Diferentes tipos de modelos de funções de pedotransferência podem ser classificados em empíricos, baseados em aprendizado de máquina e híbridos. Cada abordagem apresenta vantagens e limitações específicas, conforme resumido na Tabela 1.

Tabela 1 – Comparação entre tipos de modelos de funções de pedotransferência (PTFs), destacando base metodológica, vantagens, desvantagens e exemplos.

Tipo de Modelo	Base Metodológica	Vantagens	Desvantagens	Exemplos
Empírico	Regressão Estatística	Simplicidade, facilidade de uso, baixo custo computacional	Requer calibração específica do local, aplicabilidade limitada a condições semelhantes às usadas para desenvolvimento	Saxton- Rawls, Cosby et al.
Aprendizado de Máquina (ML)	Algoritmos de Aprendizado de Máquina (ex.: redes neurais, árvores de decisão)	Alta precisão, capacidade de capturar relações não lineares complexas, generalização para uma ampla gama de condições	Natureza de “caixa preta” (baixa interpretabilidade), necessidade de grandes conjuntos de dados de treinamento, alto custo computacional potencial	Redes Neurais Artificiais (RNAs), Máquinas de Vetores de Suporte (SVMs), Random Forests
Híbrido	Combinação de abordagens empíricas e de ML	Pode oferecer o balanço de precisão e interpretabilidade, utiliza o conhecimento do domínio para melhorar o desempenho do modelo de ML	Complexidade no desenvolvimento e implementação, pode herdar desvantagens de ambos os métodos	Modelos que combinam regressão com pós-processamento de ML, uso de ML para selecionar variáveis em modelos empíricos

Observa-se que os modelos empíricos, embora simples e de baixo custo computacional, possuem aplicabilidade restrita a condições semelhantes às utilizadas no seu desenvolvimento. Já os modelos baseados em aprendizado de máquina apresentam maior precisão e capacidade de generalização, mas são frequentemente criticados pela baixa interpretabilidade e pela necessidade de grandes conjuntos de dados. Por sua vez, os modelos híbridos buscam conciliar as duas abordagens, equilibrando precisão e interpretabilidade,

ainda que com maior complexidade de implementação.

Em conclusão, a literatura apresenta uma ampla variedade de tipos de PTFs, desde modelos empíricos tradicionais até sofisticadas estruturas baseadas em aprendizado de máquina, cada uma com pontos fortes e limitações únicos. O desenvolvimento contínuo de PTFs híbridas oferece uma perspectiva promissora para aprimorar a precisão e a aplicabilidade das previsões das propriedades hidráulicas do solo, abordando alguns dos desafios enfrentados pelas metodologias de PTFs atuais.

2.2.3 DESAFIOS E LIMITAÇÕES

O desenvolvimento e a utilização de Funções de Pedotransferência (PTFs) proporcionaram um avanço significativo na estimativa das propriedades hidráulicas do solo a partir de características do solo de fácil mensuração. No entanto, apesar de sua utilidade, tanto as abordagens tradicionais quanto as modernas de modelagem enfrentam certas limitações e desafios que restringem sua ampla aplicabilidade e precisão. Esta revisão da literatura discute esses desafios principais, enfatizando as limitações das PTFs tradicionais, os desafios dos modelos baseados em aprendizado de máquina, questões de validação e aplicabilidade, bem como estratégias propostas para superar esses desafios.

As PTFs tradicionais dependem fortemente de modelos empíricos que exigem dados específicos de determinada região para calibração, criando dificuldades substanciais na extrapolação dos resultados para outras áreas ou tipos de solo. Essa dependência frequentemente resulta em baixa precisão preditiva em solos complexos e heterogêneos. Além disso, as PTFs empíricas são geralmente baseadas em pressupostos simplificados que podem não representar adequadamente as nuances das diversas condições do solo (LOOY et al., 2017). Por exemplo, solos com texturas extremas, como solos arenosos ou orgânicos, frequentemente apresentam baixa precisão quando avaliados por esses modelos, devido a comportamentos hidráulicos únicos que divergem das suposições preditivas típicas (BAGNALL et al., 2022; SCHARNAGL et al., 2011). Consequentemente, seu desempenho diminui significativamente em situações em que os solos não seguem os padrões estatísticos estabelecidos no conjunto de calibração.

Em contraste, abordagens baseadas em aprendizado de máquina (ML) surgiram como uma alternativa promissora para lidar com algumas das limitações das PTFs tradicionais. No entanto, elas também apresentam desafios. Um problema principal associado aos modelos de ML é a necessidade de grandes conjuntos de dados rotulados e de alta qualidade para um treinamento adequado. Essa exigência pode limitar sua aplicação em regiões ou contextos onde tais dados são escassos ou pouco confiáveis (MOEYS; BERGHEAUD; COQUET, 2011; ZHANG; SCHAAP, 2017). Além disso, muitos modelos avançados de ML, como redes neurais profundas, possuem altos custos computacionais, impondo limitações práticas quanto à acessibilidade e viabilidade operacional para usuários com recursos

limitados (STEENHUIS et al., 2013). Outro problema crítico é o risco de sobreajuste; modelos complexos de ML podem ter um desempenho excelente nos dados de treinamento, mas falhar na generalização para novos conjuntos de dados, comprometendo sua utilidade (MCBRATNEY; MINASNY; TRANTER, 2011). Ademais, enquanto as PTFs empíricas permitem uma interpretação relativamente direta, os modelos de ML geralmente apresentam menor interpretabilidade, tornando difícil compreender os mecanismos subjacentes às previsões das propriedades do solo (LOOY et al., 2017; JORDA et al., 2015).

Outro desafio relevante no campo diz respeito à validação e aplicabilidade das PTFs. Ainda há uma falta de metodologias padronizadas para avaliar o desempenho dos modelos, o que cria inconsistências na avaliação da eficácia das PTFs entre diferentes estudos (IRESON et al., 2022; JORDA et al., 2015). A incerteza associada às previsões também é uma questão pouco abordada, uma vez que análises de sensibilidade e métodos de propagação de incerteza são frequentemente subexplorados, gerando uma falta de confiança nas previsões (LOOY et al., 2017; PINNINGTON et al., 2021). Além disso, a dependência de variáveis de entrada específicas complica a aplicabilidade dos modelos, pois nem todas as variáveis estão amplamente disponíveis em bancos de dados públicos ou repositórios de dados de solos (SHUSTER et al., 2021). Essa limitação frequentemente restringe a adoção mais ampla das PTFs em diferentes contextos ambientais e geográficos.

Para lidar com esses desafios persistentes, diversas estratégias foram propostas. Uma abordagem promissora é o desenvolvimento de modelos híbridos que integram métodos empíricos e de ML, combinando as vantagens de ambos para aprimorar a precisão preditiva e a robustez dos modelos (IBRAHIM et al., 2014). Além disso, a aplicação de técnicas de aprendizado por transferência pode melhorar consideravelmente a aplicabilidade dos modelos em diferentes regiões, permitindo o uso de modelos pré-treinados em novos contextos sem necessidade de um retreinamento extenso (COOPER et al., 2020). Melhorar a interpretabilidade dos modelos por meio de métodos de explicabilidade em ML é outra via promissora, pois esses métodos podem ajudar a elucidar como características específicas do solo influenciam os resultados previstos, tornando os modelos mais acessíveis aos usuários (PINNINGTON et al., 2021; COOPER et al., 2021). Por fim, a construção de bancos de dados mais robustos e diversificados para o treinamento de PTFs é fundamental para garantir que os modelos sejam bem informados e capazes de generalizar para uma gama mais ampla de tipos e condições de solo (LOOY et al., 2017; TÓTH et al., 2014).

Embora as PTFs sejam uma ferramenta valiosa para estimar propriedades hidráulicas do solo, existem desafios e limitações significativas tanto nas abordagens tradicionais quanto nas modernas. Ao abordar questões relacionadas à necessidade de dados, interpretabilidade dos modelos, validação e generalização, os pesquisadores podem aprimorar significativamente a capacidade preditiva e a relevância das PTFs, contribuindo assim para a melhoria da gestão do solo e da modelagem hidrológica.

2.3 APRENDIZADO DE MÁQUINA E REDES NEURAIS ARTIFICIAIS

2.3.1 FUNDAMENTOS DO APRENDIZADO DE MÁQUINA

O Aprendizado de Máquina (Machine Learning – ML) é um ramo da Inteligência Artificial (IA) que se concentra no desenvolvimento e implementação de algoritmos que permitem que sistemas aprendam a partir de dados e façam previsões ou tomem decisões com base nesses dados. O objetivo central do ML é automatizar o processo de identificação de padrões sem a necessidade de programação explícita para realizar uma tarefa específica. Isso envolve a capacidade de melhorar o desempenho à medida que mais dados se tornam disponíveis, aprimorando assim o poder preditivo e as capacidades de tomada de decisão desses sistemas (OSTOVARI; ASGARI; CORNELIS, 2015; BOHNE; RENGGER; WESSOLEK, 2019).

Na ciência do solo, o ML tem se tornado cada vez mais relevante, pois oferece ferramentas poderosas para modelar as interações complexas e frequentemente não lineares entre diversas propriedades do solo. A aplicação do ML permite a avaliação eficiente de propriedades do solo que são difíceis de medir diretamente, como condutividade hidráulica, retenção de umidade e disponibilidade de nutrientes (QIAO et al., 2018). Por meio da análise de grandes conjuntos de dados, os métodos de ML podem identificar relações complexas entre atributos do solo, auxiliando em estratégias de manejo, previsões de produtividade agrícola e gestão dos recursos hídricos (CHIRICO; MEDINA; ROMANO, 2010).

Os algoritmos de ML podem ser classificados em três categorias principais. No aprendizado supervisionado, os modelos são treinados com conjuntos de dados rotulados, onde cada entrada tem uma saída correspondente. Essa abordagem permite que o algoritmo aprenda um mapeamento entre entradas e saídas, sendo amplamente utilizado em tarefas de regressão e classificação, como na estimação de propriedades hidráulicas do solo (TOBELLA et al., 2024; OBIERO et al., 2013). Já no aprendizado não supervisionado, os dados não possuem rótulos e o modelo busca identificar padrões ocultos ou agrupamentos. Técnicas como k-means e análise de componentes principais (PCA) têm sido aplicadas para classificar diferentes tipos de solo e entender sua variabilidade espacial (GÜLSER; CANDEMIR, 2014; BENKE et al., 2020). Além dessas abordagens, o aprendizado por reforço envolve um agente que interage com o ambiente, recebendo recompensas ou penalidades conforme suas ações. Embora menos comum na modelagem de propriedades do solo, seus princípios podem ser explorados para otimizar práticas agrícolas e de manejo sustentável (KALUMBA et al., 2020).

O desenvolvimento de modelos de ML exige uma estrutura metodológica bem definida. Uma prática fundamental é a divisão dos dados em conjuntos de treino, validação e teste. Essa separação permite ajustar o modelo de forma eficiente e avaliar sua capacidade

de generalização em dados não vistos, reduzindo o risco de viés e garantindo um desempenho robusto (ARSHAD et al., 2013). Além disso, a definição de funções de custo, como erro quadrático médio (MSE) e entropia cruzada, permite quantificar a discrepância entre previsões e valores reais. Técnicas de otimização, como gradiente descendente, são então utilizadas para minimizar esses erros e aprimorar o desempenho do modelo (TÓTH et al., 2014; FIDALSKI; HELBEL, 2020).

Ao longo do treinamento dos modelos, dois desafios comuns são o sobreajuste e o subajuste. O primeiro ocorre quando o modelo aprende padrões específicos do conjunto de treino, mas tem dificuldades em generalizar para novos dados. Já o subajuste acontece quando o modelo não consegue capturar as tendências subjacentes, resultando em previsões imprecisas. Para mitigar esses problemas, estratégias como validação cruzada, regularização e poda são frequentemente aplicadas (ZHANG; SCHAAP, 2017; BOHNE; RENGGER; WESSOLEK, 2019). Além disso, a avaliação do desempenho do modelo é realizada por meio de métricas como erro médio absoluto (MAE), raiz do erro quadrático médio (RMSE) e coeficiente de determinação (R^2), que fornecem indicadores sobre a precisão e a confiabilidade das previsões (OBIERO et al., 2013; PACHEPSKY et al., 2014).

O Aprendizado de Máquina tem demonstrado grande potencial para aprimorar as funções de pedotransferência (PTFs), utilizadas para prever propriedades do solo a partir de variáveis mais acessíveis. As PTFs tradicionais, geralmente baseadas em modelos estatísticos simples, podem apresentar limitações devido à sua natureza linear e necessidade de calibração regional. Em contrapartida, os modelos de ML são capazes de capturar relações não lineares e integrar múltiplas fontes de dados, resultando em previsões mais precisas e robustas (JENA et al., 2021).

Estudos recentes indicam que abordagens baseadas em aprendizado profundo, como redes neurais convolucionais (CNNs), têm apresentado desempenho superior na modelagem de curvas de retenção de água no solo. Essas técnicas permitem identificar padrões mais complexos nos dados, reduzindo erros e aumentando a confiabilidade das previsões (HENGL et al., 2017). Além disso, a combinação de ML com dados de sensoriamento remoto tem ampliado as possibilidades de predição, permitindo análises mais detalhadas da variabilidade espacial do solo e auxiliando na tomada de decisões agrícolas e ambientais (GUNARATHNA et al., 2019). Técnicas como Florestas Aleatórias (Random Forests) e Gradient Boosting também têm se destacado por sua capacidade de melhorar a precisão e a robustez dos modelos aplicados ao estudo do solo em diferentes condições ambientais (JORDA et al., 2015).

Em síntese, os fundamentos do aprendizado de máquina oferecem subsídios teóricos e práticos relevantes para suas aplicações na ciência do solo, especialmente no desenvolvimento de funções de pedotransferência. A compreensão dos distintos paradigmas de aprendizado de máquina, de seus conceitos fundamentais e das métricas utilizadas para

avaliação de desempenho, permite aos pesquisadores empregar essas técnicas de forma mais eficaz, contribuindo para o aumento da precisão e da confiabilidade na predição de propriedades do solo. A integração entre o aprendizado de máquina e as funções de pedotransferência representa um avanço significativo para a ciência do solo, promovendo decisões mais fundamentadas voltadas ao uso sustentável da terra e à gestão eficiente dos recursos naturais.

2.3.2 ARQUITETURAS DE REDES NEURAI APLICADAS A PTFs

As Redes Neurais Artificiais (RNAs) são modelos computacionais inspirados na estrutura e no funcionamento das redes neurais biológicas do cérebro. Compostas por grupos interconectados de unidades de processamento simples, ou neurônios, organizados em camadas, as RNAs são projetadas para reconhecer padrões complexos e relações nos dados. Elas podem aprender com exemplos por meio da experiência, ajustando os pesos das conexões com base nos erros das saídas, melhorando assim seu desempenho ao longo do tempo. Essa adaptabilidade as torna particularmente adequadas para aplicações que exigem modelagem complexa, como a estimativa de propriedades do solo por meio das Funções de Pedotransferência (PTFs) (OSTOVARI; ASGARI; CORNELIS, 2015; BOHNE; RENGGER; WESSOLEK, 2019).

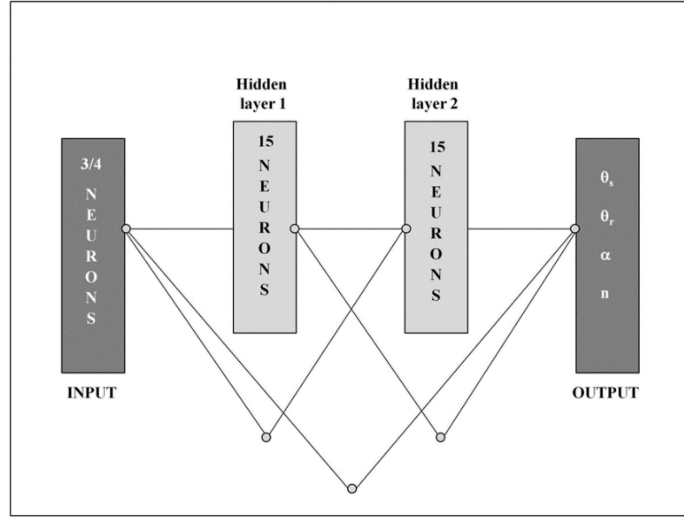
Dada a complexidade inerente dos sistemas do solo, que envolvem inúmeros parâmetros interativos, os métodos empíricos tradicionais usados para desenvolver PTFs muitas vezes se baseiam em relações lineares simplificadas, que podem não ser válidas para diferentes tipos de solo. Em contraste, as redes neurais oferecem um modelo não linear capaz de aprender a partir de grandes conjuntos de dados, capturando de maneira mais eficaz as relações entre diversos atributos do solo e suas propriedades hidráulicas associadas. Essa capacidade de modelar interações complexas de forma adaptativa permite que os pesquisadores melhorem a precisão preditiva e a robustez das PTFs em comparação com os métodos convencionais (QIAO et al., 2018; CHIRICO; MEDINA; ROMANO, 2010).

Pesquisas acadêmicas realizadas nos últimos anos têm explorado diferentes arquiteturas de redes neurais no desenvolvimento de funções de pedotransferência, evidenciando avanços e desafios. De acordo com estudos recentes como (LOOY et al., 2017; PADARIAN; MINASNY; MCBRATNEY, 2020; ZHANG; SCHAAP, 2019; KOTLAR; de Jong van Lier; de Souza Brito, 2020), as principais arquiteturas aplicadas em PTFs incluem:

1. Redes Neurais Feedforward (Perceptrons de Múltiplas Camadas - MLP): Os Perceptrons de Múltiplas Camadas (MLPs) são a estrutura de rede neural mais utilizada na modelagem de PTFs devido à sua flexibilidade e eficácia. Um MLP é composto por uma camada de entrada, uma ou mais camadas ocultas e uma camada de saída. Essa arquitetura permite simular relações complexas por meio das interconexões ponderadas entre essas camadas (TOBELLA et al., 2024).

Para ilustrar a lógica de funcionamento de uma RNA do tipo feed-forward, base da arquitetura MLP, apresenta-se a Figura 6.

Figura 6 – Esquema de operação da rede neural artificial (RNA) do tipo feed-forward generalizada (FFG).



Fonte: (D'EMILIO et al., 2018)

Esse esquema evidencia o princípio fundamental das RNAs aplicadas a PTFs: a capacidade de modelar relações não lineares entre atributos básicos do solo (entradas) e variáveis alvo, por meio de camadas ocultas ajustadas durante o treinamento.

Segundo (LOOY et al., 2017), o vetor de entrada de neurônios x_j ($j = 1 \dots J$) na rede é ponderado, somado e polarizado para produzir os neurônios ocultos y_k ($k = 1 \dots K$) :

$$y_k = \sum_{j=1}^J w_{jk}x_j + b_k, \quad (2.1)$$

onde J é o número de neurônios de entrada e k é o número de neurônios ocultos. Os neurônios ocultos consistem na entrada ponderada (w_{jk}) e um bias (b_k). Os neurônios ocultos y_k são então operados por uma função de ativação ou transferência f para produzir

$$r_k = f(y_k) \quad (2.2)$$

A saída dos neurônios ocultos é processada por um procedimento similar ao da equação 2.2, como segue:

$$v_l = \sum_{k=1}^K u_{kl}r_k + b_l, \quad (2.3)$$

e então são transformados por outra função de ativação F para produzir a saída z :

$$z_l = F(v_l) \quad (2.4)$$

Os pesos e biases são obtidos na RNA minimizando a seguinte função objetivo através de um procedimento iterativo,

$$O(w_{jk}, b_{jk}, u_{kl}, b_l) = \sum_{n=1}^{N_s} \sum_{m=1}^{N_p} [t_{n,m} - t'_{n,m}(w_{jk}, b_k, u_{kl}, b_l)]^2 \quad (2.5)$$

Onde N_s é o número de amostras de calibração, N_p o número de parâmetros, e t e t' as variáveis observadas e previstas. Esses elementos estão diretamente associados ao processo ilustrado na Figura 5, na qual se esquematiza a lógica de funcionamento das Funções de Pedotransferência. Observa-se que os preditores (atributos básicos do solo) são processados por um método estatístico/matemático, passando por etapas de calibração e avaliação por métricas de erro, até a obtenção das variáveis previstas, que são então comparadas com as variáveis observadas. Esse fluxo conceitual sintetiza a tradução de dados de fácil acesso em propriedades hidráulicas de maior complexidade, característica central das PTFs.

2. Redes Neurais Convolucionais (CNNs): As CNNs são redes neurais especializadas no processamento de dados organizados em grade, tornando-as particularmente úteis para tarefas que envolvem imagens ou dados espaciais. No contexto da ciência do solo, as CNNs podem ser aplicadas para analisar imagens do solo ou dados de sensoriamento remoto que fornecem informações relevantes sobre a textura e composição do solo (KALUMBA et al., 2020).
 - Aplicação no Mapeamento do Solo: Ao identificar características e relações espaciais, as CNNs podem melhorar o desempenho das PTFs na modelagem de propriedades do solo influenciadas pela variabilidade espacial. Por exemplo, a integração de dados de imagens do solo coletadas por drones ou sensoriamento remoto por satélite com CNNs pode levar a previsões espaciais mais precisas de propriedades como teor de carbono orgânico e salinidade, essenciais para a agricultura de precisão (ARSHAD et al., 2013).
3. Redes Neurais Recorrentes (RNNs) e Long Short-Term Memory (LSTM): As RNNs se destacam por sua capacidade de processar dados sequenciais, tornando-as adequadas para a análise de séries temporais ambientais relacionadas à dinâmica da umidade do solo e variações climáticas. As LSTMs, uma subclasse das RNNs, são projetadas para reter informações por longas sequências e são especialmente úteis em cenários onde existem dependências de longo prazo, como na previsão da umidade do solo ao longo do tempo (TÓTH et al., 2014; FIDALSKI; HELBEL, 2020).

- **Análise de Séries Temporais:** A capacidade das LSTMs de reter dados históricos permite melhorar previsões baseadas em condições meteorológicas, facilitando avaliações mais precisas sobre como essas mudanças afetam o desempenho e a saúde do solo ao longo do tempo. Ao incorporar dados temporais, as RNNs e LSTMs podem fornecer insights valiosos para a gestão eficaz da água e dos nutrientes no solo (ZHANG; SCHAAP, 2017; BOHNE; RENGGER; WESSOLEK, 2019).
4. **Redes Híbridas:** Arquiteturas híbridas que combinam múltiplos tipos de redes neurais podem gerar previsões mais precisas ao aproveitar os pontos fortes de cada modelo. Por exemplo, uma rede híbrida pode integrar CNNs para processar características espaciais enquanto utiliza RNNs para levar em conta elementos temporais, criando um modelo que explora de forma eficaz ambas as dimensões espacial e temporal dos dados do solo (OBIERO et al., 2013).
- **Aplicações Avançadas:** Essas abordagens híbridas são especialmente úteis no desenvolvimento de PTFs que exigem a compreensão das propriedades do solo ao longo do tempo e do espaço, proporcionando modelos aprimorados para a previsão da saúde do solo e da produtividade agrícola (PACHEPSKY et al., 2014).

Vários fatores técnicos críticos impactam a eficácia dos modelos de redes neurais aplicados a PTFs:

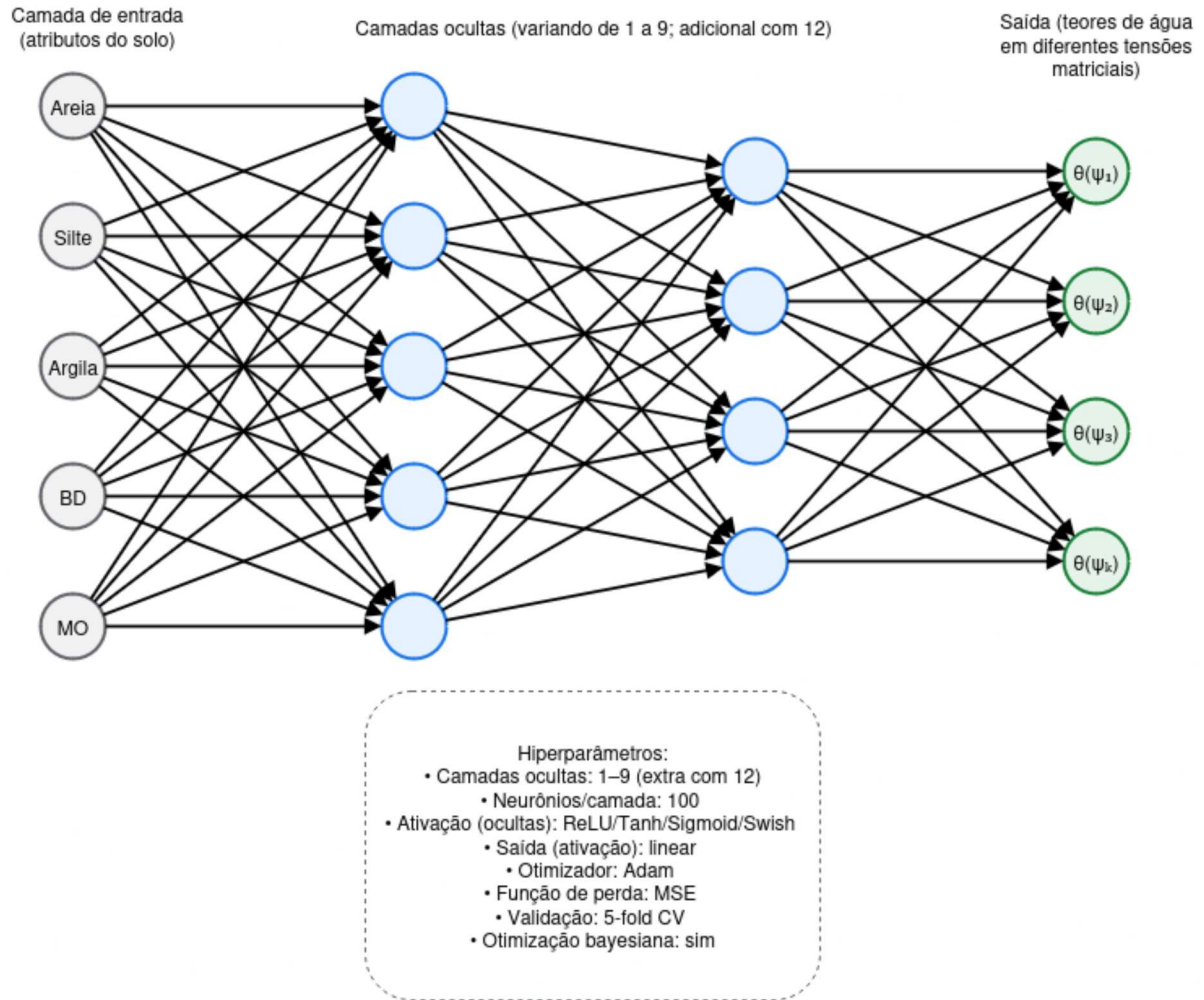
1. **Escolha das Funções de Ativação:** As funções de ativação são fundamentais para determinar a saída dos neurônios dentro de uma rede neural, influenciando significativamente o aprendizado e o desempenho do modelo. Algumas funções comuns incluem:
 - **ReLU:** Amplamente utilizada por sua simplicidade e eficácia, a ReLU reduz o problema do gradiente desaparecendo em redes profundas, permitindo uma convergência mais rápida durante o treinamento (JENA et al., 2021).
 - **Sigmoide:** Utilizada em problemas de classificação binária, gera saídas entre 0 e 1. No entanto, sua tendência à saturação pode desacelerar o treinamento, tornando-a menos favorável para camadas ocultas (HENGL et al., 2017).
 - **Tanh:** Produz saídas na faixa de -1 a 1 e frequentemente leva a uma convergência mais rápida do que a função sigmoide, sendo uma alternativa adequada em certos contextos (GUNARATHNA et al., 2019).
 - **Função de ativação linear:** Utilizada principalmente na camada de saída, a função linear pode produzir uma faixa contínua de resultados e é particularmente

útil para tarefas de regressão. No entanto, não é comumente empregada em camadas ocultas devido à sua ineficácia em introduzir não-linearidades (JENA et al., 2021).

- **Função de Ativação Swish:** Swish, definida como $(f(x) = x \cdot \text{sigmoid}(x))$, é uma função de ativação mais recente que, em alguns casos, demonstrou superar a ReLU em redes profundas. Sua natureza não monótona permite uma representação de aprendizado não linear, o que pode resultar em um desempenho superior em determinados cenários (HENGL et al., 2017).
2. **Impacto de Camadas Ocultas e Neurônios:** A complexidade da arquitetura, determinada pelo número de camadas ocultas e neurônios por camada, afeta diretamente a capacidade de uma rede de modelar padrões intrincados. Embora arquiteturas mais profundas forneçam flexibilidade aprimorada, elas também necessitam de ajuste cuidadoso para evitar overfitting. A escolha da profundidade da rede é frequentemente informada por testes empíricos, equilibrando o poder e a estabilidade do modelo (JORDA et al., 2015; BAGNALL et al., 2022).
 3. **Estratégias para Evitar o Sobreajuste:** O sobreajuste é um desafio crítico no treinamento de redes neurais complexas. Para melhorar a generalização do modelo, diversas técnicas são empregadas:
 - **Regularização L2:** Penaliza coeficientes elevados, desencorajando modelos excessivamente complexos (KOESTEL et al., 2018).
 - **Dropout:** Desativa aleatoriamente um subconjunto de neurônios durante o treinamento, forçando a rede a aprender representações redundantes e aumentando sua robustez (MCBRATNEY; MINASNY; TRANTER, 2011).
 - **Parada Antecipada:** Monitoramento do desempenho em um conjunto de validação para interromper o treinamento quando a performance começa a piorar (LOOY et al., 2017).

Enquanto a Figura 6 apresenta um esquema genérico, a Figura 7 sintetiza a arquitetura autoral proposta neste estudo, contemplando a estrutura de entrada, camadas ocultas e variáveis de saída correspondentes aos teores de água em diferentes tensões matriciais.

Figura 7 – Esquema autoral da rede neural utilizada neste trabalho



Fonte: do autor.

Observa-se que a rede recebe como entrada atributos físicos do solo (areia, silte, argila, densidade aparente e matéria orgânica), os processa em múltiplas camadas ocultas com funções de ativação não lineares e gera como saída um vetor de teores de água no solo em distintas condições de pressão matricial. O número exato de camadas e neurônios, bem como as funções de ativação utilizadas, foi definido a partir da otimização de hiperparâmetros via busca bayesiana, conforme detalhado no Capítulo 3. Ressalta-se, entretanto, que a Figura 7 apresenta apenas duas camadas ocultas de forma ilustrativa, visando facilitar a visualização. Na prática, foram avaliadas arquiteturas variando de 1 a 9 camadas ocultas, além de uma configuração adicional com 12 camadas, de modo a investigar a influência da profundidade da rede na capacidade de generalização do modelo.

Os avanços recentes na aplicação de redes neurais às Funções de Pedotransferência (PTFs) evidenciam tanto o progresso quanto os desafios em curso.

1. Treinamento Eficiente de Redes Profundas: O treinamento de redes neurais profundas exige recursos computacionais substanciais e algoritmos otimizados. Técnicas

como normalização em lote (batch normalization) e otimizadores avançados (como Adam) foram introduzidas para acelerar a convergência e estabilizar o treinamento, facilitando o desenvolvimento eficaz de arquiteturas profundas (LOOY et al., 2017; BAGNALL et al., 2022).

2. **Aprendizado por Transferência (Transfer Learning):** O aprendizado por transferência permite reutilizar modelos treinados em grandes conjuntos de dados para novas aplicações com bases de dados potencialmente menores. Essa abordagem tem se mostrado particularmente vantajosa na ciência do solo, onde conjuntos de dados abrangentes para propriedades específicas nem sempre estão disponíveis (SCHARNAGL et al., 2011; MOEYS; BERGHEAUD; COQUET, 2011)
3. **Inteligência Artificial Explicável (Explainable AI - XAI):** Com o aumento do uso de redes neurais para previsões complexas na ciência do solo, a necessidade de interpretabilidade torna-se crucial. As metodologias de XAI buscam esclarecer como as redes neurais chegam às suas previsões, aumentando a confiança e a compreensão em aplicações cujas decisões podem impactar o uso da terra e as práticas agrícolas. Técnicas como SHAP (SHapley Additive exPlanations) e LIME (Local Interpretable Model-agnostic Explanations) desempenham papéis fundamentais ao fornecer insights sobre as decisões do modelo (ZHANG; SCHAAP, 2017; STEENHUIS et al., 2013).
4. **Integração do Conhecimento Especializado:** A incorporação do conhecimento agrônomo e pedológico no design das redes neurais pode aprimorar seu poder preditivo. Essa abordagem permite um modelagem mais precisa das propriedades do solo, alinhando as arquiteturas de aprendizado de máquina com princípios científicos estabelecidos e observações da ciência do solo (MCBRATNEY; MINASNY; TRANTER, 2011).

As arquiteturas de redes neurais oferecem uma alternativa poderosa aos métodos tradicionais no desenvolvimento de Funções de Pedotransferência na ciência do solo. Sua capacidade de aprender relações complexas a partir de grandes conjuntos de dados, adaptar-se a sequências espaciais e temporais e empregar estruturas híbridas aprimora significativamente suas capacidades preditivas. Embora desafios como sobreajuste, interpretabilidade e custo computacional persistam, avanços recentes em otimização, aprendizado por transferência e inteligência artificial explicável estão facilitando aplicações mais eficientes das redes neurais na predição de propriedades do solo.

2.3.3 AJUSTE DE HIPERPARÂMETROS E OTIMIZAÇÃO

Os hiperparâmetros desempenham um papel fundamental no desenvolvimento e na otimização de modelos de aprendizado de máquina (ML), funcionando como configurações ajustáveis que ditam o processo de aprendizagem e a arquitetura do modelo antes da

introdução do conjunto de dados de treinamento. Diferentemente dos parâmetros, que são ajustados automaticamente durante o aprendizado a partir dos dados (como os pesos em uma regressão linear), os hiperparâmetros devem ser definidos antes da fase de treinamento e abrangem aspectos como taxas de aprendizado, tamanhos de lote e a escolha dos algoritmos. A importância da otimização de hiperparâmetros não pode ser subestimada, pois ela influencia significativamente o desempenho do modelo em dados não vistos, determinando, em última instância, a precisão preditiva e a capacidade de generalização do modelo (QIAN, 2024; GU et al., 2021).

Ao definir hiperparâmetros, é essencial reconhecer a distinção entre eles e os parâmetros dentro de frameworks de aprendizado supervisionado. Parâmetros, como aqueles obtidos durante o treinamento (por exemplo, os pesos em uma rede neural), são otimizados diretamente pelo algoritmo de aprendizado em correlação com os dados de entrada (QIAN, 2024). Por outro lado, os hiperparâmetros são valores estáticos que definem a estrutura e a estratégia de aprendizagem do algoritmo antes do início do treinamento (HOQUE; ALJAMAAN, 2021). Sua configuração adequada é crucial, pois determina como o algoritmo de treinamento será executado, impactando a velocidade de convergência e a eficácia geral no aprendizado da distribuição subjacente dos dados (MUGHNI; FAHRUDIN; KAMISUTARA, 2021).

A otimização de hiperparâmetros é um procedimento complexo, frequentemente abordado por meio de técnicas sistemáticas como *Grid Search* e *Random Search*. O *Grid Search* é um método clássico que busca exaustivamente dentro de um conjunto predefinido de hiperparâmetros, testando sistematicamente cada combinação para determinar o melhor desempenho (GHAWI; PFEFFER, 2019). Embora seja direto e informativo, seu caráter exaustivo pode tornar-se proibitivamente demorado, especialmente conforme o espaço de parâmetros cresce — o número de configurações avaliadas pode se tornar impraticavelmente alto. Em contraste, o *Random Search* tende a amostrar configurações de forma aleatória, permitindo explorar um conjunto mais diverso de combinações de hiperparâmetros. Pesquisas demonstram que o *Random Search* pode superar o *Grid Search* em termos de eficiência, pois não exige testar todas as combinações e pode obter resultados impressionantes avaliando um número menor de configurações (JIN, 2022).

Os *trade-offs* entre essas abordagens destacam considerações importantes de eficiência. O *Grid Search* garante que cada combinação especificada seja avaliada, mas sua natureza exaustiva pode levar a retornos decrescentes à medida que a dimensionalidade dos hiperparâmetros aumenta (LIN et al., 2021). O *Random Search*, embora potencialmente mais eficiente, apresenta variabilidade nos resultados, pois pode deixar de explorar regiões ótimas. Portanto, a escolha entre esses métodos deve considerar o equilíbrio entre os recursos computacionais disponíveis e o grau de completude desejado na busca (JIN, 2022).

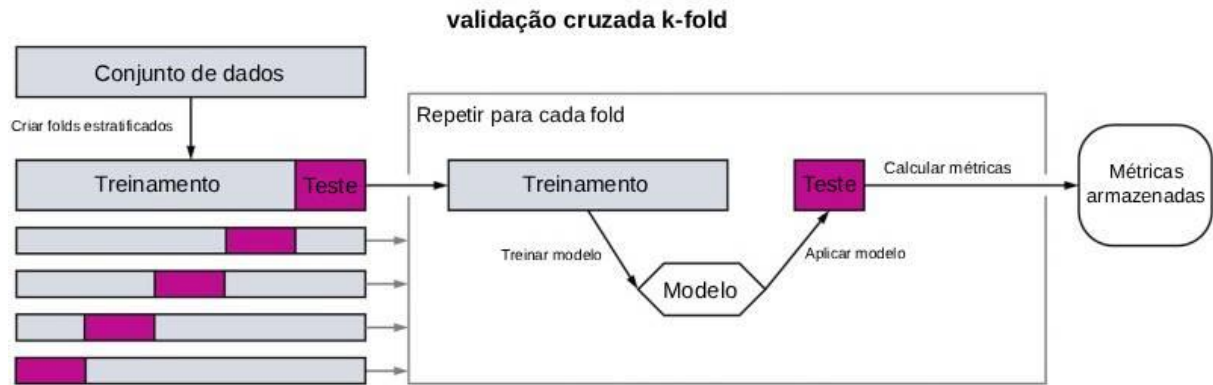
Para contextualizar o impacto do ajuste de hiperparâmetros em modelos específicos,

podemos observar algoritmos supervisionados como Máquinas de Vetores de Suporte (SVM) e XGBoost. Estudos demonstraram que a otimização de hiperparâmetros melhora significativamente o desempenho de modelos SVM. Por meio do ajuste cuidadoso de hiperparâmetros, como o custo e o tipo de *kernel*, pesquisadores observaram melhorias substanciais na acurácia da classificação, com modelos apresentando alta eficácia em diferentes conjuntos de dados (MUGHNI; FAHRUDIN; KAMISUTARA, 2021; QIAN, 2024). No caso do XGBoost, hiperparâmetros críticos como taxa de aprendizado e coeficientes de regularização influenciam diretamente suas capacidades preditivas; ajustá-los pode mitigar problemas como *overfitting*, levando a métricas de desempenho significativamente melhores, conforme demonstrado em estudos comparativos (HOQUE; ALJAMAAN, 2021).

Na interseção entre o ajuste de hiperparâmetros e a avaliação de modelos encontra-se a prática essencial da validação cruzada. Esse método é fundamental durante a otimização de hiperparâmetros, pois previne o *overfitting* ao garantir que as estimativas de desempenho sejam confiáveis e imparciais. A validação cruzada envolve o particionamento dos dados em subconjuntos, onde determinadas porções são usadas para treinamento e outras para validação, fornecendo assim um meio de avaliar o quão bem o modelo generaliza para novos dados (WAINER; CAWLEY, 2018).

Pesquisas indicam que a implementação de validação cruzada *k-fold* durante o ajuste de hiperparâmetros pode reduzir a probabilidade de *overfitting*, já que permite múltiplas avaliações do desempenho do modelo em diferentes divisões dos dados, estabelecendo um arcabouço robusto de validação (WAINER; CAWLEY, 2018). Nesse método, o conjunto de dados é dividido em k subconjuntos (folds), e a cada iteração um fold é usado para teste enquanto os demais são utilizados para treinamento. Após k iterações, cada subconjunto é empregado exatamente uma vez para validação, e a média dos erros fornece uma estimativa mais robusta da capacidade de generalização do modelo, como ilustrado na Figura 8.

Figura 8 – Esquema do processo de validação cruzada k -fold



Fonte: adaptado de (KUBBEN; DUMONTIER; DEKKER, 2019)

Esse procedimento garante que cada amostra seja utilizada tanto para treinamento quanto para validação, resultando em uma métrica de desempenho mais estável e representativa. No presente trabalho, adotou-se $k = 5$ folds, combinado com Otimização Bayesiana, contribuindo para uma avaliação mais robusta da capacidade de generalização das RNAs utilizadas.

Evidências empíricas também sustentam a eficácia dessas metodologias. Por exemplo, um estudo que investigou o papel do ajuste de hiperparâmetros na previsão de preços de ações com modelos de aprendizado de máquina demonstrou melhorias marcantes no desempenho com o ajuste ideal (HOQUE; ALJAMAAN, 2021). Em outro caso envolvendo detecção da Doença de Parkinson, configurações ideais de hiperparâmetros foram cruciais para melhorar o desempenho de modelos como Máquinas de Vetores de Suporte e Árvores de Decisão, evidenciando os benefícios tangíveis do ajuste cuidadoso de hiperparâmetros por meio de experimentações (ELGELDAWI et al., 2021).

2.3.3.1 OTIMIZAÇÃO BAYESIANA

Diante das limitações das técnicas tradicionais de otimização, a *Otimização Bayesiana* surge como um processo estatisticamente fundamentado que aprimora as metodologias de ajuste de hiperparâmetros. Esse método modela a função desconhecida que mapeia configurações de hiperparâmetros para o desempenho do modelo por meio de uma função substituta, tipicamente um processo gaussiano. Ao equilibrar exploração e exploração, a Otimização Bayesiana atualiza suas crenças sobre a função objetivo com base em avaliações anteriores, guiando buscas futuras em direção a áreas promissoras do espaço de hiperparâmetros (STUKE; RINKE; TODOROVÍĆ, 2020). Muitos estudos recentes ilustram a eficácia dessa estratégia, demonstrando maior eficiência na otimização em domínios desafiadores, como modelagem médica e design de algoritmos complexos (ELSHEWEY et al., 2023).

A Otimização Bayesiana (OB) tem como elemento central a função de aquisição,

responsável por indicar os próximos pontos do espaço de busca a serem avaliados. Entre as mais utilizadas destacam-se a *Expected Improvement* (EI), a *Probability of Improvement* (PI) e a *Upper Confidence Bound* (UCB). Essas funções equilibram a exploração de regiões pouco testadas com a exploração de regiões já conhecidas como promissoras, permitindo que a OB encontre combinações de hiperparâmetros próximas do ótimo global com um número reduzido de avaliações da função objetivo (SHAHRIARI et al., 2016).

No contexto desta pesquisa, a OB foi implementada por meio da biblioteca *Optuna*, que automatiza o processo de seleção de hiperparâmetros de redes neurais. Esse procedimento considerou aspectos como número de camadas ocultas, número de neurônios por camada, funções de ativação, taxa de aprendizado e parâmetros de regularização. A métrica de avaliação foi o erro quadrático médio (MSE), utilizando o otimizador Adam, em conjunto com validação cruzada *k-fold* para garantir robustez dos resultados. Dessa forma, a OB contribuiu para identificar arquiteturas mais eficientes, reduzindo o risco de sobreajuste e otimizando o desempenho preditivo das funções de pedotransferência.

Em síntese, a Otimização Bayesiana apresenta vantagens claras sobre métodos tradicionais: enquanto o *Grid Search* e o *Random Search* exploram o espaço de forma exaustiva ou aleatória, a OB faz uso de modelos probabilísticos que guiam a busca de forma inteligente e iterativa. Essa característica a torna particularmente adequada em cenários de alta dimensionalidade e custo computacional elevado, como no caso da calibração de redes neurais aplicadas à modelagem de propriedades hidráulicas do solo.

Em resumo, os hiperparâmetros são um aspecto crucial das metodologias de aprendizado de máquina, estabelecendo a estrutura fundamental que dita como os modelos aprendem com os dados. Sua otimização é essencial para o desempenho preditivo, estando diretamente correlacionada com o sucesso dos modelos em diversas aplicações. Embora métodos tradicionais como *Grid Search* e *Random Search* ofereçam mecanismos para esse ajuste, abordagens mais recentes como a Otimização Bayesiana utilizam estruturas estatísticas avançadas para aumentar a eficiência e a confiabilidade dos resultados. Estudos empíricos indicam que o ajuste cuidadoso de hiperparâmetros traz benefícios concretos de desempenho em diversos contextos de aprendizado de máquina, reforçando o papel indispensável dessas configurações no desenvolvimento de modelos preditivos robustos.

3 MATERIAIS E MÉTODOS

Este capítulo apresenta os materiais e métodos utilizados no desenvolvimento do presente estudo. Inicialmente, são descritas as bases de dados empregadas, seguidas pelos procedimentos de preparação dos dados. Na sequência, detalha-se a construção e parametrização dos modelos de redes neurais artificiais, incluindo a definição dos hiperparâmetros e as técnicas de otimização adotadas. Por fim, apresentam-se as métricas de avaliação e a estratégia de validação utilizada.

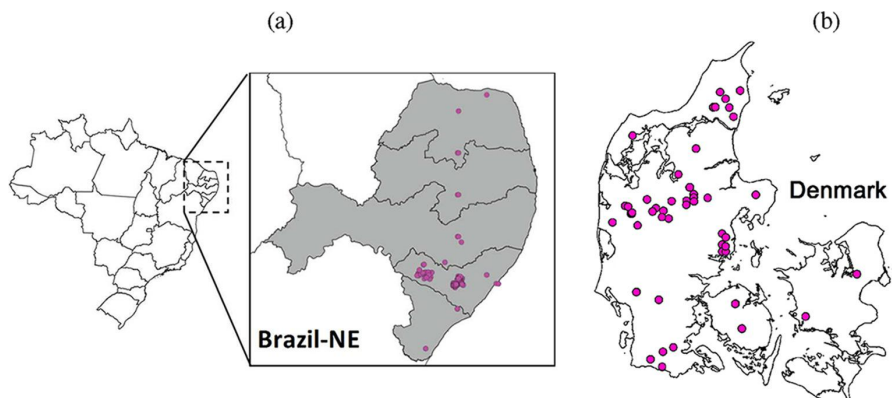
3.1 BASE DE DADOS

A escolha da base de dados utilizada nesta pesquisa foi um passo fundamental para garantir a qualidade e a confiabilidade dos resultados obtidos. Para isso, foram selecionadas três bases de dados distintas, denominadas B1: *Brasil-NE*, B2: *Brasil-AC* e D1: *Dinamarca*. Essas bases contêm informações essenciais para a aplicação e validação dos modelos desenvolvidos neste estudo.

As bases B1 e D1 foram extraídas do trabalho publicado por (KOTLAR et al., 2019). Os dados correspondentes a essas bases encontram-se disponíveis no repositório *Dryad*.

Os dados consistem em amostras coletadas na região Nordeste do Brasil (Brasil-NE, conforme mostrado na Figura 9a, os quais foram obtidos de (BARROS et al., 2013) e de um conjunto de dados da Embrapa, Empresa Brasileira de Pesquisa Agropecuária. O banco de dados consistiu em 838 amostras, nas quais os teores de água foram medidos em -0,6, -1, -3, -5, -10, -20 e -150 m de pressão de sucção (representados como $\theta_{0,6}$, θ_1 , θ_3 , θ_5 , θ_{10} , θ_{20} , θ_{150} , respectivamente), juntamente com informações sobre frações granulométricas (teores de areia, silte e argila), matéria orgânica (OM) e densidade do solo (BD).

Figura 9 – Localização dos locais de amostragem para retenção de água (conjuntos de dados Brasil-NE e Dinamarca).



Fonte: (KOTLAR et al., 2019)

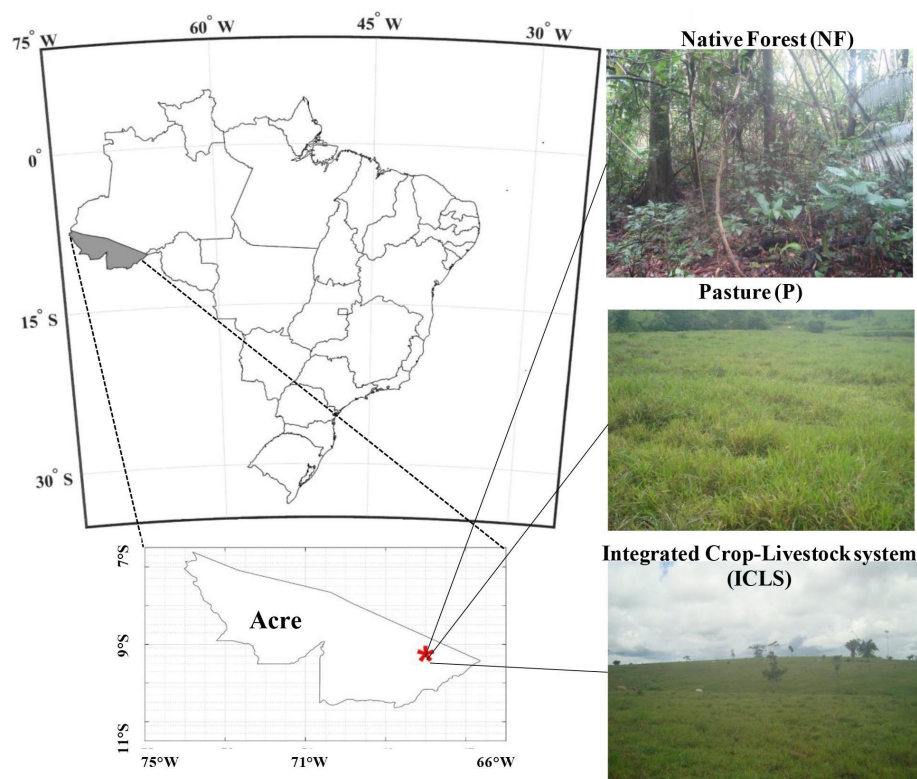
Ainda em (KOTLAR et al., 2019), os dados do solo de uma região temperada foram extraídos de um banco de dados dinamarquês que continha 186 amostras (conforme mostrado na Figura 9b, coletadas principalmente na península da Jutlândia, no oeste da Dinamarca, conforme relatado por (IVERSEN et al., 2011). Esses dados incluíam informações semelhantes, mas abrangiam o conteúdo de água em apenas quatro cabeças de pressão: -0,1, -1, -10 e -150 m (denotadas como $\theta_{0,1}$, θ_1 , θ_{10} e θ_{150} , respectivamente). Em ambas as bases de dados, as partículas foram classificadas de acordo com a Taxonomia de Solo do USDA como argila ($<0,002$ mm), silte (0,002–0,05 mm) e areia (0,05–2 mm). O conteúdo de água saturada (θ_s) foi considerado igual à porosidade total e foi calculado a partir da densidade do solo (BD) e da densidade das partículas (assumida como $2,65 \text{ g cm}^{-3}$).

Esses conjuntos de dados foram escolhidos por sua relevância na área de estudo, contendo medições detalhadas e de alta qualidade relacionadas às variáveis de interesse.

Por outro lado, a base B2 foi obtida a partir do estudo publicado em (KOTLAR; de Jong van Lier; de Souza Brito, 2020), estando disponível na plataforma *Mendeley Data*. Essa base foi selecionada por complementar as informações contidas nas bases B1 e D1, permitindo uma análise mais abrangente e generalizável dos modelos utilizados.

O conjunto de dados utilizados é composto de informações de solos siltosos do estado do Acre, Brasil. A pesquisa foi realizada em uma região situada a aproximadamente 50 km de Rio Branco, capital do Acre, Brasil (latitude $9^{\circ}56'40''\text{S}$, longitude $68^{\circ}14'10''\text{W}$, Fig. 10. O terreno apresenta altitude média de 194 metros, com um regime anual de chuvas variando entre 1877 e 1982 mm e temperatura média de 25°C . As amostras foram coletadas em três diferentes sistemas de uso da terra: Floresta nativa tropical (NF), Sistema integrado lavoura-pecuária (ICLS) e Pastagem convencional (P).

Figura 10 – Localização geográfica da área de estudo no estado do Acre, Brasil. As fotografias ilustram os três tipos de uso do solo analisados: floresta nativa, pastagem e sistema integrado lavoura-pecuária.



Fonte: (KOTLAR; de Jong van Lier; de Souza Brito, 2020)

A distância entre os diferentes sistemas de uso do solo (NF, ICLS e P) era de aproximadamente 500 metros. Em cada sistema, foram escavadas quatro trincheiras com 1,40 m de profundidade, separadas entre si por 40 metros. No total, foram coletadas 192 amostras indeformadas (anéis de 100 cm³) e um número equivalente de amostras deformadas (trado), abrangendo profundidades de 0–5, 5–10, 10–20, 20–30, 30–40, 40–60, 60–80 e 80–100 cm. As amostras foram analisadas quanto à distribuição do tamanho das partículas, resistência à penetração (RP), estabilidade dos agregados e características químicas como matéria orgânica (MO) e capacidade de troca de cátions (CTC).

A seleção dessas bases de dados foi realizada com o objetivo de garantir diversidade nos cenários analisados, abrangendo diferentes regiões e condições ambientais. Além disso, os trabalhos que deram origem a essas bases são amplamente reconhecidos na literatura, o que reforça a confiabilidade dos dados e a validade das análises realizadas. A integração dessas bases possibilita um estudo robusto e aprofundado, contribuindo para uma avaliação abrangente dos métodos aplicados.

A escolha pelas três bases (B1, B2 e D1) deve-se à intenção de contemplar cenários pedoclimáticos contrastantes e, ao mesmo tempo, metodologicamente relevantes. A base B1 representa solos tropicais do Nordeste brasileiro em condições semiáridas, frequentemente caracterizados por limitações hídricas e grande variabilidade textural. A base

B2, proveniente do estado do Acre, abrange solos tropicais úmidos da Amazônia, onde predominam condições de elevada precipitação e diferentes usos da terra, permitindo avaliar o desempenho dos modelos em contextos de manejo contrastantes. Já a base D1, originária da Dinamarca, foi selecionada por representar solos de clima temperado, amplamente utilizados em estudos internacionais, possibilitando a comparação dos resultados obtidos em condições tropicais e temperadas. Assim, a combinação dessas três bases permite avaliar a robustez e a capacidade de generalização dos modelos de pedotransferência em diferentes ambientes pedoclimáticos.

3.2 PRÉ-PROCESSAMENTO DOS DADOS

O pré-processamento dos dados constitui uma etapa fundamental para assegurar a qualidade, integridade e consistência dos modelos desenvolvidos. Esta etapa visa preparar os dados para treinamento e avaliação dos modelos de redes neurais, mitigando possíveis interferências causadas por dados inconsistentes ou ruídos.

3.2.1 LIMPEZA DOS DADOS

Inicialmente, realizou-se uma inspeção nos conjuntos de dados para identificação de registros incompletos. A estratégia adotada foi a remoção de linhas que apresentavam valores ausentes. Essa decisão se justifica pela baixa ocorrência de dados faltantes nos arquivos analisados e pela preferência por manter a integridade das medições experimentais, evitando imputações que poderiam introduzir vies.

3.2.2 ESTRUTURA DOS DADOS

Os conjuntos utilizados possuem as seguintes características:

- **Base B1 – Brasil-NE (838 amostras):** Inclui as variáveis físico-químicas: teor de areia (S), silte (Si), argila (Cl), matéria orgânica (OM) e densidade do solo (BD). As saídas correspondem ao teor de água em sete tensões: $\theta_{0,6}$, θ_1 , θ_3 , θ_5 , θ_{10} , θ_{20} e θ_{150} .
- **Base B2 – Brasil-AC (192 amostras):** Apresenta uma descrição mais detalhada dos atributos físicos do solo, incluindo argila (Clay), areia (Sand), densidade (BD), resistência à penetração (PR), diâmetro médio geométrico dos agregados (GMD), diâmetro médio ponderado (MWD), matéria orgânica (OM) e capacidade de troca de cátions (CEC). As saídas são os teores de água nas tensões: θ_0 , θ_{10} , θ_{60} , θ_{100} , θ_{330} , θ_{1000} , θ_{3000} , θ_{5000} e θ_{15000} .
- **Base D1 – Dinamarca (464 amostras):** Inclui matéria orgânica (OM), areia (S), silte (Si), argila (Cl) e densidade (BD). As variáveis de saída são os teores de água nas tensões: $\theta_{0,1}$, θ_1 , θ_{10} e θ_{150} .

3.2.3 TRANSFORMAÇÃO E NORMALIZAÇÃO DOS DADOS

Não foi aplicada nenhuma técnica explícita de normalização ou padronização dos dados. As variáveis foram utilizadas em suas escalas originais, o que significa que atributos como densidade do solo (valores entre aproximadamente 1.07 e 1.98 para a base B2, por exemplo) e frações texturais (valores entre 0 e 1) foram diretamente alimentados nos modelos.

Durante o desenvolvimento dos modelos de rede neural, foi avaliada a aplicação de técnicas de normalização, que realizam a padronização dos dados com base na média e desvio padrão. Esta prática é comum na literatura de ciência de dados e aprendizagem de máquina, especialmente em domínios onde os dados apresentam escalas muito heterogêneas.

Entretanto, testes preliminares realizados no presente trabalho indicaram que a aplicação da normalização não resultou em melhorias no desempenho dos modelos. Pelo contrário, observou-se uma leve degradação nas métricas de ajuste, como coeficiente de determinação (R^2) e erro quadrático médio (RMSE), quando comparados aos modelos construídos com dados em suas escalas originais.

Tal comportamento pode ser atribuído a três fatores principais:

1. As variáveis físicas do solo, como frações texturais (areia, silte, argila), densidade do solo (BD) e matéria orgânica (OM), já apresentam escalas numéricas relativamente alinhadas, não havendo discrepâncias significativas entre ordens de grandeza.
2. A utilização do otimizador **Adam** — que adapta dinamicamente a taxa de aprendizado com base nos momentos dos gradientes — reduz consideravelmente a sensibilidade do modelo a diferentes escalas nas variáveis.
3. As relações físico-mecânicas presentes nos atributos do solo são diretamente proporcionais às propriedades hidráulicas, e sua manipulação por normalização pode atenuar sinais importantes contidos nos dados.

Esta decisão metodológica está alinhada com práticas adotadas em trabalhos consolidados na literatura, como em (KOTLAR et al., 2019) e (WÖSTEN; PACHEPSKY; RAWLS, 2001), nos quais os modelos preditivos foram desenvolvidos preservando as escalas físicas originais dos dados, de modo a manter a coerência com os processos físicos que regem o comportamento do solo.

Diante disso, optou-se por conduzir o treinamento dos modelos sem aplicação de normalização explícita, utilizando os dados em suas escalas originais. Esta abordagem apresentou estabilidade, bom desempenho e foi considerada adequada para os objetivos do presente estudo.

3.2.4 DIVISÃO DOS DADOS

O particionamento dos dados seguiu duas abordagens complementares:

1. **Divisão Inicial – Treino/Teste:** Para cada base, 70% dos dados foram utilizados para o treinamento dos modelos, enquanto 30% foram destinados à avaliação final no conjunto de teste. Esta divisão foi implementada utilizando uma função de uma biblioteca, com controle de aleatoriedade por meio de sementes específicas para garantir reprodutibilidade.
2. **Validação Cruzada – K-Fold:** Durante o processo de otimização dos hiperparâmetros com a biblioteca `Optuna`, foi aplicada validação cruzada com cinco partições (`KFold` com $k = 5$), sem estratificação, devido à natureza do problema ser de regressão. Essa abordagem possibilitou uma avaliação robusta dos modelos, minimizando o risco de overfitting e proporcionando uma estimativa mais confiável de desempenho.

3.2.5 DEFINIÇÃO DAS VARIÁVEIS DE ENTRADA E SAÍDA

As variáveis de entrada e saída foram definidas conforme a disponibilidade de cada conjunto de dados:

- **Variáveis de entrada comuns:**
 - Textura do solo: areia (S ou Sand), silte (Si) e argila (Cl ou Clay).
 - Densidade do solo (BD).
 - Matéria orgânica (OM), quando presente.
- **Variáveis adicionais na base B2:**
 - Resistência à penetração (PR).
 - Estabilidade dos agregados: diâmetro médio geométrico (GMD) e diâmetro médio ponderado (MWD).
 - Capacidade de troca de cátions (CEC).
- **Variáveis de saída (targets):** Os teores de água no solo em diferentes tensões matriciais de pressão, que variam conforme a base, refletindo desde condições próximas à saturação até condições de capacidade de campo e ponto de murcha permanente.

3.3 ARQUITETURA E PARAMETRIZAÇÃO DO MODELO DE REDE NEURAL

A utilização de Redes Neurais Artificiais (RNA) para estimar funções de pedotransferência requer a definição de uma série de parâmetros que influenciam diretamente a

qualidade e a robustez das previsões do modelo. A escolha desses parâmetros impacta a capacidade da rede de capturar padrões complexos nos dados e de evitar problemas como sobreajuste ou subajuste.

A arquitetura da Rede Neural Artificial empregada neste estudo é ilustrada na Figura 7. O modelo utilizado foi uma rede do tipo perceptron de múltiplas camadas (MLP), composta por uma camada de entrada, múltiplas camadas ocultas e uma camada de saída. As variáveis de entrada corresponderam aos atributos físicos do solo (areia, silte, argila, densidade aparente e matéria orgânica), enquanto as variáveis de saída foram os teores de água no solo em diferentes tensões matriciais. O número exato de camadas ocultas e de neurônios por camada foi definido a partir do processo de otimização bayesiana, conforme descrito na Seção 2.2.3.

3.3.1 CONSTRUÇÃO E PARAMETRIZAÇÃO DA REDE NEURAL

O modelo de RNA foi construído utilizando a biblioteca Keras, uma ferramenta de alto nível em Python amplamente utilizada para desenvolvimento e experimentação com redes neurais. A arquitetura escolhida foi a de camadas densas totalmente conectadas (*Fully Connected Layers*), também conhecidas como *Multilayer Perceptrons* (MLPs). Esta arquitetura foi selecionada devido à sua versatilidade e capacidade comprovada em modelar relações complexas e não lineares, sendo particularmente eficaz em problemas de regressão e modelagem de dados tabulares, características inerentes ao problema de previsão de funções de pedotransferência (PTFs) para propriedades do solo.

- **Número de camadas ocultas:** Para explorar a influência da profundidade da rede na capacidade de aprendizado e generalização, testou-se um intervalo de camadas ocultas, variando de 1 a 9 camadas. Adicionalmente, incluiu-se uma configuração com 12 camadas para investigar o comportamento em arquiteturas ainda mais profundas. A variação no número de camadas permitiu avaliar o trade-off entre a capacidade de modelagem de relações complexas (associada a redes mais profundas) e o risco de overfitting (que pode ser exacerbado em redes mais complexas).
- **Número de neurônios por camada:** Dentro de cada camada oculta, o número de neurônios foi um parâmetro a ser ajustado. Foram testados dois valores principais: 50 e 100 neurônios por camada. A escolha entre um número menor ou maior de neurônios em cada camada também impacta a complexidade do modelo e sua capacidade de aprender padrões nos dados. Um número maior de neurônios aumenta a capacidade de representação da rede, mas também pode aumentar o risco de overfitting e o custo computacional do treinamento.
- **Função de ativação das camadas ocultas:** A função de ativação introduz não-linearidade na rede neural, permitindo que ela aprenda relações complexas. Foram

testadas cinco funções de ativação distintas para as camadas ocultas, buscando identificar a mais adequada para o problema em questão:

- **Linear** (representada pelo número 0): Utilizada como linha de base para comparação, embora não introduza não-linearidade nas camadas ocultas.
 - **ReLU** (*Rectified Linear Unit*) (representada pelo número 1): Uma função amplamente utilizada por sua simplicidade computacional e eficácia em evitar o problema do desaparecimento do gradiente em redes profundas.
 - **Tanh** (*Hyperbolic Tangent*) (representada pelo número 2): Outra função não linear que produz saídas entre -1 e 1, frequentemente associada a uma convergência mais rápida em alguns casos.
 - **Sigmoid** (representada pelo número 3): Tradicionalmente utilizada em redes neurais, produz saídas entre 0 e 1, mas pode sofrer com o problema da saturação e desaparecimento do gradiente em redes profundas.
 - **Swish** (representada pelo número 4): Uma função de ativação mais recente, que em alguns contextos demonstrou superar a ReLU, especialmente em redes mais profundas, devido à sua natureza não-monotônica.
- **Taxa de dropout:** Para mitigar o risco de overfitting, especialmente em redes mais profundas e complexas, foi implementada a técnica de dropout. A taxa de dropout define a probabilidade de um neurônio ser temporariamente desativado durante o treinamento. Foi testada uma variação da taxa de dropout de 0 a 0.5 para regularização. O dropout força a rede a aprender representações mais robustas e menos dependentes de neurônios individuais.
 - **Função de ativação da camada de saída:** foram testadas as mesmas funções da camada de entrada.
 - **Função de perda:** Para quantificar o erro entre as previsões do modelo e os valores reais durante o treinamento, foi utilizada a função de erro quadrático médio (MSE - Mean Squared Error). O MSE é uma métrica comum e adequada para problemas de regressão, penalizando grandes erros de previsão de forma quadrática.
 - **Otimizador:** O algoritmo de otimização escolhido para ajustar os pesos da rede durante o treinamento foi o Adam (*Adaptive Moment Estimation*). O Adam é um otimizador amplamente utilizado e eficiente, que combina as vantagens do *momentum* e do RMSprop, adaptando as taxas de aprendizado para cada parâmetro individualmente e demonstrando bom desempenho em problemas não estacionários, como é o caso do treinamento de redes neurais com dados complexos.
 - **Critério de parada antecipada:** Para evitar o overfitting e otimizar o tempo de treinamento, foi implementado um critério de parada antecipada. O treinamento foi

monitorado utilizando a função de perda em um conjunto de validação separado. O treinamento era interrompido quando a função de perda no conjunto de validação parava de diminuir significativamente por um número predefinido de épocas (paciência de 500 épocas), indicando que o modelo começava a overfittar os dados de treinamento.

A arquitetura básica descrita acima foi encapsulada em uma função em Python, permitindo a fácil configuração e experimentação com diferentes combinações de hiperparâmetros, definidos com base nos valores especificados. A execução do treinamento foi realizada de forma independente para cada uma das bases de dados selecionadas – B1 (Brasil-NE), B2 (Brasil-AC) e D1 (Dinamarca) – permitindo avaliar o desempenho da metodologia em diferentes contextos de dados e regiões geográficas.

3.3.2 OTIMIZAÇÃO DOS HIPERPARÂMETROS

Para refinar o desempenho do modelo de rede neural, foi adotada uma estratégia de otimização de hiperparâmetros baseada em validação cruzada e busca Bayesiana, utilizando a biblioteca Optuna. A escolha da busca Bayesiana se justifica pela sua eficiência em explorar o espaço de hiperparâmetros de forma inteligente, balanceando a exploração de novas configurações com a busca de regiões promissoras já identificadas. Este processo iterativo visa encontrar a combinação de hiperparâmetros que maximize o desempenho do modelo, medido por uma métrica de avaliação específica, nos dados de validação.

O processo de otimização foi estruturado nas seguintes etapas principais:

1. **Divisão dos dados para Validação Cruzada Estratificada:** Inicialmente, cada base de dados (B1, B2 e D1) foi submetida a um processo de validação cruzada com cinco divisões estratificadas (k-fold cross-validation com $k=5$). A estratificação garante que cada dobra da validação cruzada contenha proporções semelhantes das diferentes classes ou faixas de valores da variável alvo, assegurando uma avaliação mais robusta e representativa do desempenho do modelo em diferentes subconjuntos dos dados. Esta etapa é crucial para obter estimativas de desempenho mais confiáveis e reduzir o risco de overfitting.
2. **Exploração do Espaço de Hiperparâmetros:** A busca Bayesiana foi direcionada para otimizar um conjunto chave de hiperparâmetros que exercem influência significativa sobre a arquitetura e o treinamento da rede neural. Os hiperparâmetros otimizados foram:
 - Número de neurônios por camada: Variando entre 50 e 100 neurônios, este hiperparâmetro controla a capacidade de processamento e representação da rede.

A escolha entre 50 e 100 neurônios permite explorar diferentes níveis de complexidade do modelo, buscando um equilíbrio entre capacidade de aprendizado e risco de overfitting.

- Taxa de dropout: Com valores entre 0 e 0.5 e um passo inicial de 0.05, a taxa de dropout atua como um mecanismo de regularização, prevenindo o overfitting ao desativar aleatoriamente uma proporção de neurônios durante o treinamento. A variação desta taxa permite ajustar o nível de regularização aplicada ao modelo.
- Funções de ativação para as camadas ocultas e de saída: Foram testadas diferentes funções de ativação, incluindo Linear, ReLU, Tanh, Sigmoid e Swish. A escolha da função de ativação influencia a capacidade da rede neural de aprender relações lineares e não lineares nos dados, sendo um fator determinante no desempenho do modelo para diferentes tipos de problemas.

3. **Critério de Avaliação para Otimização:** O critério utilizado para guiar a busca Bayesiana e avaliar o desempenho das diferentes configurações de hiperparâmetros foi o Erro Absoluto Médio (MAE) calculado sobre os dados de validação. O MAE foi escolhido por ser uma métrica robusta e de fácil interpretação, que quantifica a magnitude média dos erros de previsão, fornecendo uma medida direta da precisão do modelo. A minimização do MAE nos dados de validação foi o objetivo da otimização.
4. **Execução da Busca Bayesiana:** Para cada base de dados (B1, B2 e D1), foram executadas 50 rodadas de busca Bayesiana. Em cada rodada, o Optuna propôs uma nova combinação de hiperparâmetros, que foi avaliada através do treinamento e validação do modelo utilizando a validação cruzada estratificada. As informações de desempenho obtidas em cada rodada foram utilizadas pelo Optuna para refinar o espaço de busca e direcionar as próximas propostas de hiperparâmetros para regiões mais promissoras. Este processo iterativo permitiu explorar o espaço de hiperparâmetros de forma eficiente, convergindo para configurações que apresentaram bom desempenho nos dados de validação.

Com os melhores hiperparâmetros identificados, o modelo final foi treinado utilizando o conjunto de dados completo e avaliado no conjunto de teste. As métricas principais, incluindo MAE, RMSE e R^2 , foram computadas para cada execução, garantindo uma análise detalhada da capacidade preditiva do modelo.

Após 25 execuções para cada base de dados (B1, B2 e D1) obteve-se os seguintes resultados:

3.3.2.1 BASE DE DADOS B1

A Tabela 2 apresenta os melhores conjuntos de hiperparâmetros testados após as 25 execuções para B1:

Tabela 2 – Conjunto de melhores hiperparâmetros testados no modelo para B1.

neurônios	taxa dropout	ativação oculta	ativação saída	frequência	MAE mín
100	0.0	1	4	4	0.02448
100	0.0	1	2	3	0.02554
100	0.0	0	0	3	0.02698
100	0.0	0	4	3	0.02714
100	0.30	1	2	1	0.02435
100	0.1	1	1	1	0.02535
50	0.0	0	4	1	0.02538
50	0.25	0	0	1	0.02545
100	0.0	2	4	1	0.02677
100	0.1	1	4	1	0.02724
100	0.0	4	4	1	0.02730
100	0.05	1	4	1	0.02768
100	0.35	1	2	1	0.02870
50	0.30	1	1	1	0.02955
100	0.0	2	0	1	0.03098
50	0.0	0	0	1	0.03138

Observou-se que a configuração de 100 neurônios, taxa de 0.00 de "dropout", função de ativação oculta "relu" e de saída "swish" teve excelente desempenho com MAE = 0,02448, foi mais frequente entre os melhores (count = 4), indicando boa generalização, "dropout" de 0,00 sugere que overfitting não foi um problema significativo e 100 neurônios é razoável em termos de capacidade do modelo sem exagero na complexidade. Portanto, foram esses os parâmetros escolhidos para a rede do banco de dados B1.

3.3.2.2 BASE DE DADOS B2

A Tabela 3 apresenta os melhores conjuntos de hiperparâmetros testados após as 25 execuções para B2:

Tabela 3 – Conjunto de melhores hiperparâmetros testados no modelo para B2.

neurônios	dropout rate	hidden activation	output activation	count	MAE mín
100	0.0	3	3	8	0.01950
50	0.0	3	3	4	0.02377
100	0.0	2	3	3	0.02369
100	0.05	3	3	3	0.02201
100	0.0	3	2	2	0.02201
50	0.0	3	0	1	0.02644
50	0.05	2	3	1	0.02553
50	0.05	3	3	1	0.02879
50	0.15	3	3	1	0.02731
100	0.25	3	3	1	0.02684

Observou-se que a configuração de 100 neurônios, taxa de 0.00 de "drop" e funções de ativação oculta e de saída "sigmoid" alcançou o menor erro de todos ($MAE = 0,01950$), apareceu 8 vezes entre os melhores, o que indica alta consistência, usa 100 neurônios, o que é adequado para capturar padrões mais complexos, mas sem exagerar na quantidade de pesos e "dropout" 0,00 sugere que overfitting não foi um problema relevante mesmo sem regularização. Portanto, foram esses os parâmetros escolhidos para a rede do banco de dados B2.

3.3.2.3 BASE DE DADOS D1

A Tabela 4 apresenta os melhores conjuntos de hiperparâmetros testados após as 25 execuções para D1:

Tabela 4 – Conjunto de melhores hiperparâmetros testados no modelo para D1.

neurônios	dropout rate	hidden activation	output activation	count	MAE mín
100	0.0	2	0	4	0.02953
50	0.0	0	0	2	0.03238
50	0.0	2	0	2	0.03430
100	0.0	0	0	2	0.03180
50	0.0	2	3	1	0.03073
50	0.0	4	0	1	0.03035
50	0.05	0	1	1	0.03384
50	0.1	1	3	1	0.03296
50	0.2	0	3	1	0.03506
50	0.0	0	2	1	0.03854
100	0.0	0	4	1	0.02978
100	0.0	1	0	1	0.02657
100	0.0	1	3	1	0.03327
50	0.0	1	3	1	0.03383
100	0.0	2	4	1	0.03359
100	0.1	2	0	1	0.03271
100	0.3	0	3	1	0.03164
100	0.35	0	3	1	0.02954
100	0.45	0	3	1	0.03411

Observou-se que a configuração de 100 neurônios, taxa de 0.00 de "drop", função de ativação oculta "tanh" e de saída "linear" apareceu 4 vezes entre os melhores, indicando robustez, tem desempenho sólido: Excelente MAE (0.02953), quase empatado com o melhor absoluto, tem generalização aceitável: R^2 próximo de 0.42 em execuções reais e tem simplicidade: sem dropout, e arquitetura estável com 100 neurônios. Essa escolha é uma opção mais estável, ainda que não tenha o MAE mínimo absoluto — ela oferece melhor equilíbrio entre performance e robustez, o que é ideal para aplicações práticas. Portanto, foram esses os parâmetros escolhidos para a rede do banco de dados D1.

3.4 CONSTRUÇÃO E EXECUÇÃO DO MODELO DE REDE NEURAL

Após a fase de otimização dos parâmetros da rede neural (Seção 3.2.2), os conjuntos de hiperparâmetros considerados ótimos para cada base de dados (B1, B2 e D1) foram aplicados para a construção e execução final do modelo. O código implementado utilizou a arquitetura de Rede Neural Feedforward (MLP) com camadas densas totalmente conectadas, conforme detalhado na Seção 3.2.1. Para cada base de dados, a arquitetura do modelo foi instanciada com a configuração de hiperparâmetros otimizada.

O processo de execução iniciou-se com o carregamento dos conjuntos de dados (B1, B2 e D1), dividindo-os em conjuntos de treinamento e teste, conforme descrito anteriormente. Em seguida, o modelo MLP, configurado com os hiperparâmetros otimizados para a base de dados específica, foi treinado utilizando o conjunto de treinamento. O

treinamento foi realizado de forma iterativa, com os pesos da rede neural ajustados para minimizar a função de perda MSE (Erro Quadrático Médio) utilizando o otimizador Adam (Adaptive Moment Estimation), ambos detalhados na Seção 3.2.1.

Durante o treinamento, foi implementado o critério de parada antecipada, monitorando a função de perda no conjunto de validação para evitar o sobreajuste e otimizar o tempo de treinamento. O treinamento era interrompido quando não se observava melhoria significativa na função de perda no conjunto de validação por um número predefinido de épocas.

Após a conclusão do treinamento, o modelo final, com os pesos ajustados, foi utilizado para realizar previsões sobre o conjunto de dados de teste, que não foi utilizado durante o treinamento. Para avaliar o desempenho preditivo do modelo, foram calculadas as seguintes métricas no conjunto de teste: Coeficiente de Determinação (R^2), Erro Quadrático Médio (RMSE) e Coeficiente de Correlação de Pearson (R). Adicionalmente, foram gerados gráficos comparativos entre os valores reais e os valores previstos para cada base de dados e pressure head, permitindo uma análise visual do desempenho do modelo.

As métricas de desempenho e os gráficos gerados fornecem uma avaliação abrangente da capacidade preditiva do modelo de rede neural com os hiperparâmetros otimizados para cada base de dados. Os resultados detalhados dessas execuções, juntamente com a análise comparativa e discussão das implicações, serão apresentados no próximo capítulo (Capítulo 4 - Resultados e Discussões).

3.5 FLUXO METODOLÓGICO

O desenvolvimento desta pesquisa seguiu uma sequência estruturada de etapas, que visam garantir rigor metodológico na construção, calibração e avaliação dos modelos de redes neurais para predição da retenção de água no solo. A Figura 11 apresenta o fluxograma que sintetiza todo o processo metodológico adotado.

Inicialmente, procedeu-se à coleta e organização das bases de dados, contemplando três conjuntos distintos: Brasil-NE (B1), Brasil-AC (B2) e Dinamarca (D1). Esta etapa incluiu também a limpeza dos dados, com remoção de registros incompletos, e a definição das variáveis preditoras e das variáveis alvo, conforme as especificidades de cada base.

Na sequência, foi realizada a etapa de pré-processamento, que incluiu a divisão dos dados em conjuntos de treinamento e teste (70/30) e, posteriormente, a aplicação de validação cruzada (5-fold) durante o processo de otimização dos hiperparâmetros.

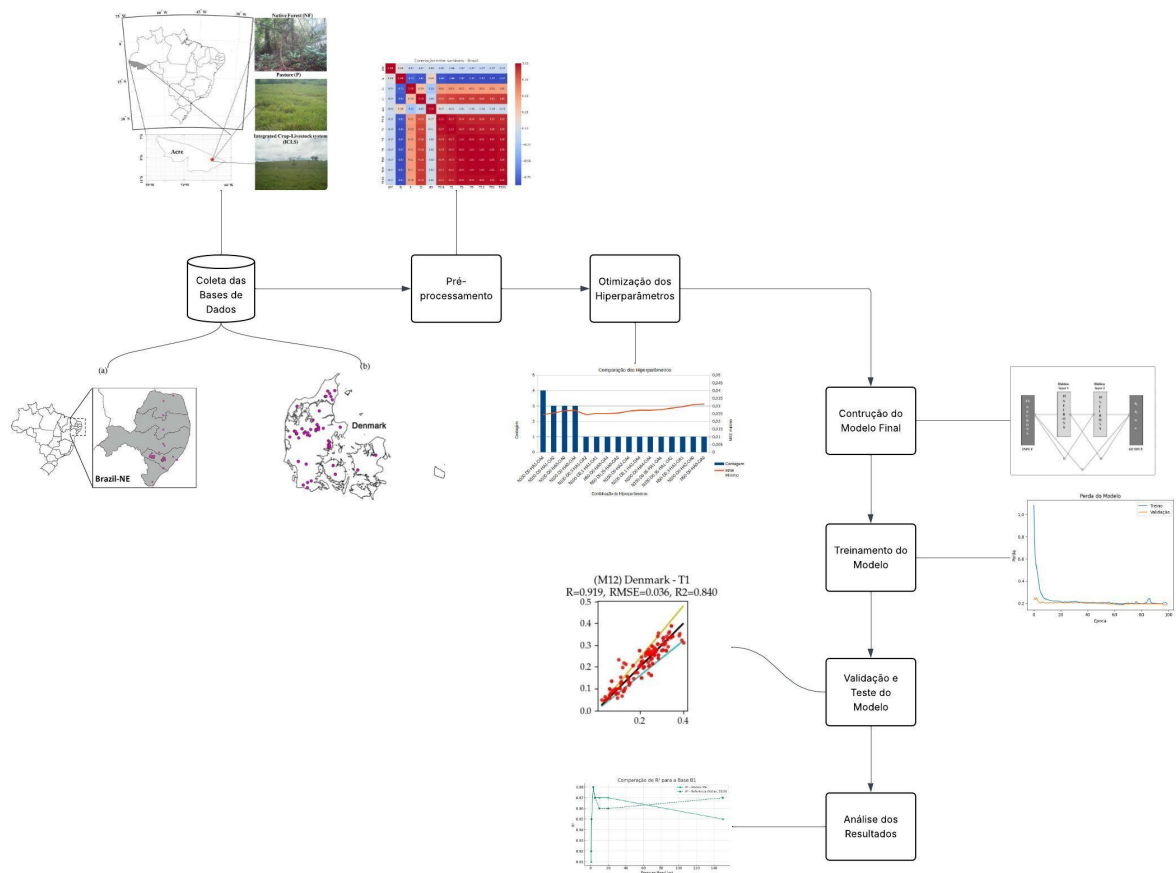
A parametrização dos modelos baseou-se na definição de variáveis como número de camadas ocultas, quantidade de neurônios por camada, taxa de dropout, e funções de ativação, tanto nas camadas ocultas quanto na camada de saída. A busca pelos melhores hiperparâmetros foi conduzida por meio de otimização bayesiana, utilizando a biblioteca

Optuna.

Após a definição da arquitetura ótima, os modelos foram treinados sobre o conjunto de treinamento e, posteriormente, avaliados no conjunto de teste, utilizando métricas como Coeficiente de Determinação (R^2), Erro Quadrático Médio (RMSE) e Erro Absoluto Médio (MAE).

Por fim, foram realizadas análises comparativas dos resultados, tanto entre as diferentes bases de dados quanto em relação à literatura, além da discussão dos pontos fortes, limitações e perspectivas de melhoria do modelo desenvolvido.

Figura 11 – Fluxograma que sintetiza o processo metodológico



Fonte: do autor

4 RESULTADOS E DISCUSSÕES

4.1 RESULTADOS

Como resultado, foram gerados gráficos para cada base de dados (B1, B2 e D1) e para cada pressure head (T), indicando pelos pontos vermelhos os valores de teste, já os pontos azuis são os valores do treino, enquanto a reta preta é a reta que melhor se adequa aos pontos do treino e as retas amarela e verde representam as retas com 20% de inclinação e declinação respectivamente. Além disso para cada gráfico foi calculado o R^2 (R-quadrado), o RMSE (Root Mean Squared Error) e o valor de R (Coeficiente de Correlação de Pearson).

4.1.1 BASE DE DADOS B1

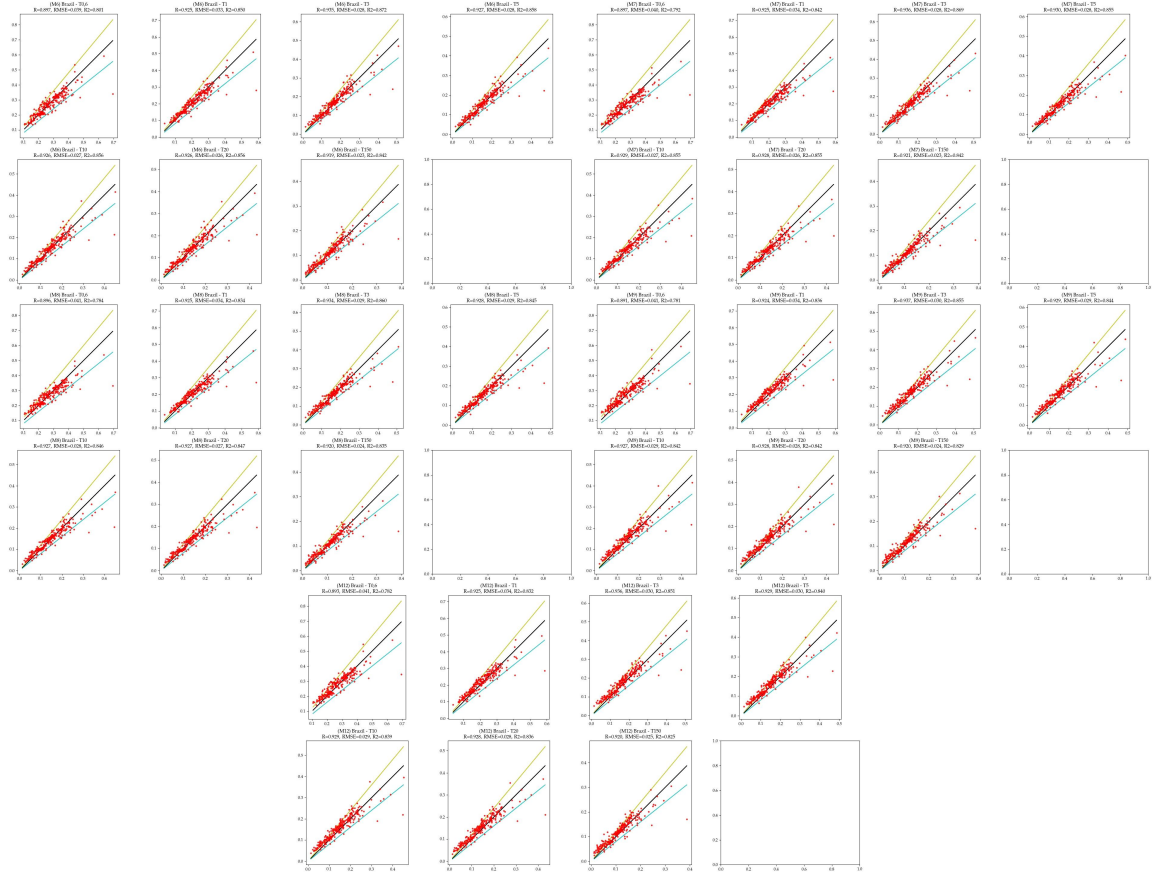
Começando pela base de dados B1, representada pelo Nordeste do Brasil, que possui as pressure heads de 0.6, 1, 3, 5, 10, 20, 150 metros. Utilizando-se os parâmetros encontrados anteriormente apresentou os seguintes gráficos:

Figura 12 – Gráficos para resultado da base de dados B1 - M1 a M5



Fonte: do autor

Figura 13 – Gráficos para resultado da base de dados B1 - M6 a M12



Fonte: do autor

Os resultados obtidos podem ser descritos na tabela abaixo, onde o modelo indica o número de camadas ocultas da rede sendo M1 equivalente a 1 camada, M2 a 2 camadas e assim sucessivamente, T indica a pressure head e R^2 o coeficiente de determinação:

Tabela 5 – Resultados obtidos para B1

Modelo	T	R	R^2	RMSE	Modelo	T	R	R^2	RMSE
M1	0.6	0.888	0.78	0.041	M6	0.6	0.897	0.80	0.039
M1	1	0.917	0.84	0.034	M6	1	0.925	0.85	0.033
M1	3	0.932	0.87	0.028	M6	3	0.935	0.87	0.028
M1	5	0.926	0.86	0.028	M6	5	0.927	0.86	0.028
M1	10	0.926	0.86	0.027	M6	10	0.926	0.86	0.027
M1	20	0.926	0.86	0.026	M6	20	0.926	0.86	0.026
M1	150	0.92	0.84	0.023	M6	150	0.919	0.84	0.023
M2	0.6	0.892	0.79	0.04	M7	0.6	0.897	0.79	0.04
M2	1	0.921	0.84	0.033	M7	1	0.925	0.84	0.034
M2	3	0.936	0.87	0.028	M7	3	0.936	0.87	0.028
M2	5	0.927	0.86	0.028	M7	5	0.93	0.86	0.028
M2	10	0.928	0.86	0.027	M7	10	0.929	0.86	0.027
M2	20	0.928	0.86	0.026	M7	20	0.928	0.86	0.026
M2	150	0.92	0.85	0.023	M7	150	0.921	0.84	0.023
M3	0.6	0.894	0.79	0.04	M8	0.6	0.896	0.78	0.041
M3	1	0.922	0.84	0.033	M8	1	0.925	0.83	0.034
M3	3	0.934	0.87	0.028	M8	3	0.934	0.86	0.029
M3	5	0.928	0.86	0.028	M8	5	0.928	0.85	0.029
M3	10	0.928	0.86	0.027	M8	10	0.927	0.85	0.028
M3	20	0.928	0.86	0.026	M8	20	0.927	0.85	0.027
M3	150	0.921	0.85	0.023	M8	150	0.92	0.84	0.024
M4	0.6	0.9	0.81	0.039	M9	0.6	0.891	0.78	0.041
M4	1	0.927	0.85	0.032	M9	1	0.924	0.84	0.034
M4	3	0.938	0.88	0.027	M9	3	0.937	0.86	0.03
M4	5	0.934	0.87	0.027	M9	5	0.929	0.84	0.029
M4	10	0.934	0.87	0.026	M9	10	0.927	0.84	0.029
M4	20	0.934	0.87	0.026	M9	20	0.928	0.84	0.028
M4	150	0.925	0.85	0.023	M9	150	0.92	0.83	0.024
M5	0.6	0.896	0.79	0.04	M12	0.6	0.893	0.78	0.041
M5	1	0.924	0.84	0.034	M12	1	0.925	0.83	0.034
M5	3	0.936	0.87	0.028	M12	3	0.936	0.85	0.03
M5	5	0.928	0.85	0.029	M12	5	0.929	0.84	0.03
M5	10	0.928	0.85	0.028	M12	10	0.929	0.84	0.029
M5	20	0.929	0.85	0.027	M12	20	0.928	0.84	0.028
M5	150	0.921	0.84	0.023	M12	150	0.92	0.83	0.025

Com base no trabalho desenvolvido por (KOTLAR et al., 2019), os valores obtidos em sua pesquisa podem ser utilizados como referência para a análise comparativa com os resultados obtidos no presente estudo. Esses valores de referência estão apresentados na Tabela 6.

Tabela 6 – Dados de referência para B1

Base de Dados	T	R^2	RMSE
B1	0.6	0.82	0.036
B1	1	0.85	0.031
B1	3	0.88	0.026
B1	5	0.87	0.026
B1	10	0.86	0.026
B1	20	0.86	0.024
B1	150	0.87	0.020

Na Tabela 7 abaixo pode-se demonstrar um comparativo de desempenho do resultado com base percentual considerando-se R^2 e RMSE.

Tabela 7 – Variação de R^2 e RMSE para os melhores resultados apresentados de B1 em comparação com os dados de referência

Modelo	T	R^2	RMSE	Variação R^2	Variação RMSE
M4	0.6	0.81	0.039	-1.22%	8.33%
M4	1	0.85	0.032	0.00%	3.23%
M4	3	0.88	0.027	0.00%	3.85%
M4	5	0.87	0.027	0.00%	3.85%
M4	10	0.87	0.026	1.16%	0.00%
M4	20	0.87	0.026	1.16%	8.33%
M4	150	0.85	0.023	-2.30%	15.00%

Realizando uma análise comparativa entre os resultados obtidos e os valores de referência da literatura, observa-se que os modelos desenvolvidos no presente estudo apresentam desempenho satisfatório, com variações pontuais a depender do valor de pressure head considerado.

Para os menores valores de pressure head ($T = 0.6$ e 1 m), os resultados demonstraram comportamentos distintos. Em $T = 1$ m, o modelo M4 obteve um desempenho equivalente ao da literatura, com R^2 de 0.85 e aumento de apenas 3,23% no RMSE, sugerindo uma boa capacidade de ajuste do modelo. Já para $T = 0.6$ m, houve uma leve redução de 1,22% no R^2 e um aumento de 8,33% no RMSE, indicando uma dificuldade um pouco maior em representar com precisão essa faixa de pressão.

Nos valores intermediários de pressão ($T = 3, 5$ e 10 m), os resultados mostraram-se bastante promissores. O modelo M4 obteve R^2 idêntico ao valor de referência para $T = 3$ e 5 m (0.88 e 0.87, respectivamente), com apenas uma pequena elevação no RMSE

(3,85%), o que evidencia um desempenho praticamente equivalente ao da literatura. Para $T = 10$ m, observou-se um ganho de 1,16% no R^2 e RMSE idêntico ao valor de referência, confirmando a eficácia da modelagem neste intervalo.

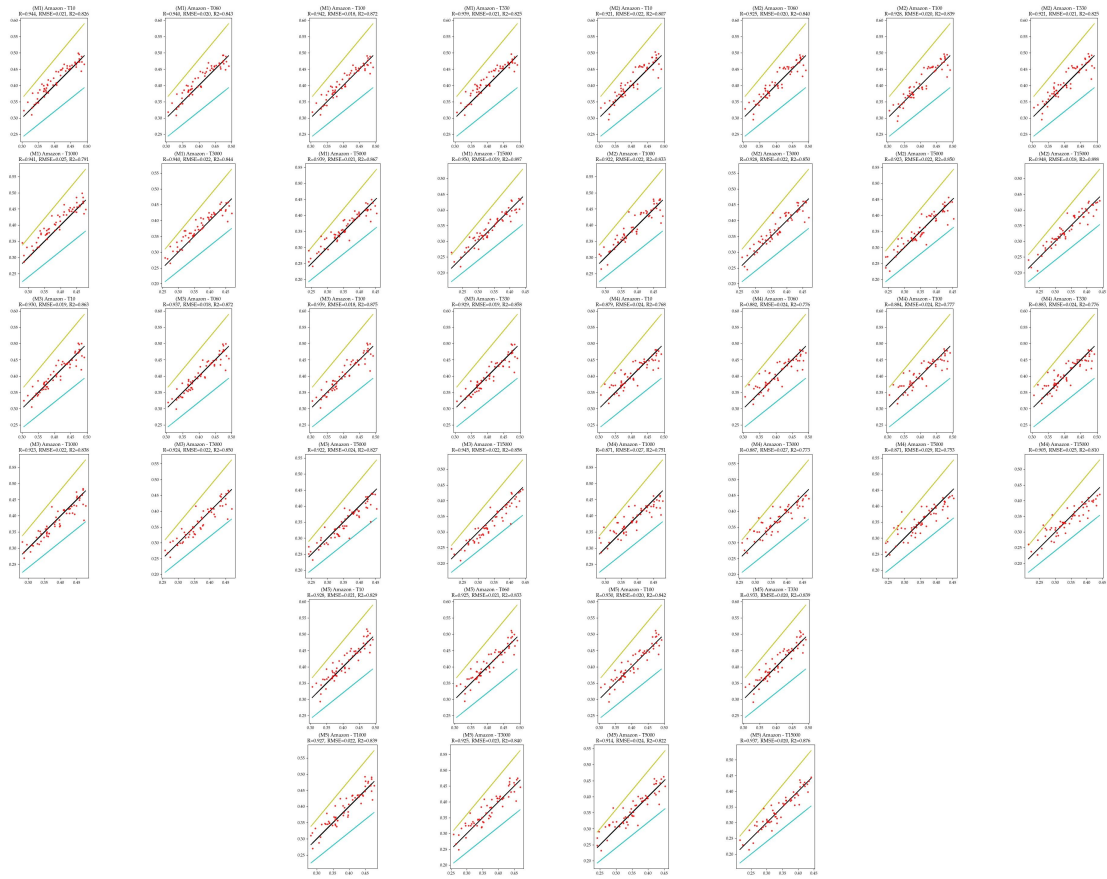
Para pressões mais elevadas, como $T = 20$ e 150 m, os modelos apresentaram desempenho ligeiramente inferior ao esperado. Em $T = 20$ m, o modelo M4 obteve um aumento de 1,16% em R^2 , mas com um acréscimo de 8,33% no RMSE. Já para $T = 150$ m, a variação foi mais significativa, com uma redução de 2,30% em R^2 e aumento de 15,00% no RMSE, indicando uma menor precisão do modelo em condições de pressão muito altas.

Dessa forma, os resultados indicam que os modelos desenvolvidos possuem um bom desempenho geral, especialmente para os valores intermediários de pressure head, onde foram observados os melhores índices de ajuste. Ainda que haja espaço para melhorias em faixas extremas de pressão — principalmente em $T = 0.6$ m e 150 m — os modelos demonstraram potencial para representar adequadamente a retenção de água no solo da região Nordeste do Brasil.

4.1.2 BASE DE DADOS B2

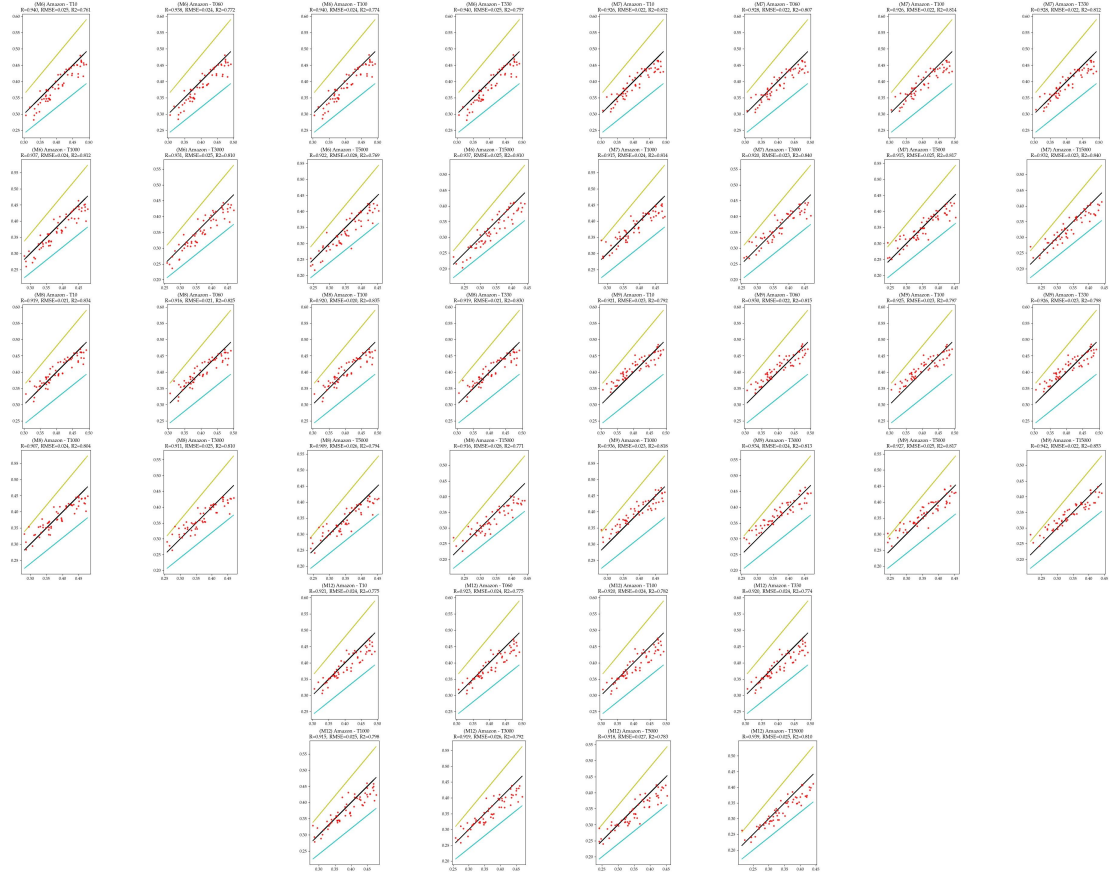
O mesmo procedimento foi aplicado à base de dados B2, que representa a região amazônica, especificamente no estado do Acre. No entanto, essa base contém valores de pressure head de 10, 60, 100, 330, 1000, 3000, 5000 e 15000 centímetros. A partir dessa base, foram gerados os seguintes gráficos:

Figura 14 – Gráficos para resultado da base de dados B2 - M1 a M5



Fonte: do autor

Figura 15 – Gráficos para resultado da base de dados B2 - M6 a M12



Fonte: do autor

Os resultados obtidos podem ser descritos na tabela abaixo, onde o modelo indica o número de camadas ocultas da rede sendo M1 equivalente a 1 camada, M2 a 2 camadas e assim sucessivamente, T indica a pressure head, R^2 o coeficiente de determinação e RMSE representa o erro quadrático médio:

Tabela 8 – Resultados obtidos para B2

Modelo	T	R	R^2	RMSE	Modelo	T	R	R^2	RMSE
M1	10	0.944	0.83	0.021	M6	10	0.94	0.76	0.025
M1	60	0.94	0.84	0.02	M6	60	0.938	0.77	0.024
M1	100	0.942	0.87	0.018	M6	100	0.94	0.77	0.024
M1	330	0.939	0.83	0.021	M6	330	0.94	0.76	0.025
M1	1000	0.941	0.79	0.025	M6	1000	0.937	0.81	0.024
M1	3000	0.94	0.84	0.022	M6	3000	0.931	0.81	0.025
M1	5000	0.939	0.87	0.021	M6	5000	0.922	0.77	0.028
M1	15000	0.95	0.90	0.019	M6	15000	0.937	0.81	0.025
M2	10	0.921	0.81	0.022	M7	10	0.926	0.81	0.022
M2	60	0.925	0.84	0.02	M7	60	0.928	0.81	0.022
M2	100	0.928	0.84	0.02	M7	100	0.926	0.81	0.022
M2	330	0.921	0.83	0.021	M7	330	0.928	0.81	0.022
M2	1000	0.922	0.83	0.022	M7	1000	0.915	0.81	0.024
M2	3000	0.928	0.85	0.022	M7	3000	0.92	0.84	0.023
M2	5000	0.923	0.85	0.022	M7	5000	0.915	0.82	0.025
M2	15000	0.948	0.90	0.018	M7	15000	0.932	0.84	0.023
M3	10	0.93	0.86	0.019	M8	10	0.919	0.83	0.021
M3	60	0.937	0.87	0.018	M8	60	0.916	0.83	0.021
M3	100	0.939	0.88	0.018	M8	100	0.92	0.84	0.02
M3	330	0.929	0.86	0.019	M8	330	0.919	0.83	0.021
M3	1000	0.923	0.84	0.022	M8	1000	0.907	0.80	0.024
M3	3000	0.924	0.85	0.022	M8	3000	0.911	0.81	0.025
M3	5000	0.922	0.83	0.024	M8	5000	0.909	0.79	0.026
M3	15000	0.945	0.86	0.022	M8	15000	0.916	0.77	0.028
M4	10	0.879	0.77	0.024	M9	10	0.921	0.79	0.023
M4	60	0.882	0.78	0.024	M9	60	0.93	0.82	0.022
M4	100	0.884	0.78	0.024	M9	100	0.925	0.80	0.023
M4	330	0.883	0.78	0.024	M9	330	0.926	0.80	0.023
M4	1000	0.871	0.75	0.027	M9	1000	0.936	0.82	0.023
M4	3000	0.887	0.77	0.027	M9	3000	0.934	0.81	0.024
M4	5000	0.871	0.75	0.029	M9	5000	0.927	0.82	0.025
M4	15000	0.905	0.81	0.025	M9	15000	0.942	0.85	0.022
M5	10	0.928	0.83	0.021	M12	10	0.921	0.78	0.024
M5	60	0.925	0.83	0.021	M12	60	0.923	0.78	0.024
M5	100	0.93	0.84	0.02	M12	100	0.92	0.78	0.024
M5	330	0.933	0.84	0.02	M12	330	0.92	0.77	0.024
M5	1000	0.927	0.84	0.022	M12	1000	0.915	0.80	0.025
M5	3000	0.925	0.84	0.023	M12	3000	0.919	0.79	0.026
M5	5000	0.914	0.82	0.024	M12	5000	0.918	0.78	0.027
M5	15000	0.937	0.88	0.02	M12	15000	0.939	0.81	0.025

Agora, usando como base o trabalho desenvolvido por (KOTLAR; de Jong van Lier; de Souza Brito, 2020), os valores obtidos em sua pesquisa podem ser utilizados como referência para a análise comparativa com os resultados obtidos no presente estudo. Esses valores de referência estão apresentados na Tabela 9.

Tabela 9 – Dados de referência para B2

Base de Dados	T	R^2	RMSE
B2	10	0.84	0.020
B2	60	0.86	0.019
B2	100	0.86	0.020
B2	330	0.88	0.019
B2	1000	0.89	0.020
B2	3000	0.89	0.020
B2	5000	0.90	0.020
B2	15000	0,87	0,023

Na Tabela 10 abaixo pode-se demonstrar um comparativo de desempenho do resultado com base percentual considerando-se R^2 e RMSE.

Tabela 10 – Variação de R^2 e RMSE para os melhores resultados apresentados de B2 em comparação com os dados de referência

Modelo	T	R^2	RMSE	Variação R^2	Variação RMSE
M3	10	0.86	0.019	2.38%	-5.00%
M3	60	0.87	0.018	1.16%	-5.26%
M3	100	0.88	0.018	2.33%	-10.00%
M3	330	0.86	0.019	-2.27%	0.00%
M5	1000	0.84	0.022	-5.62%	10.00%
M2	3000	0.85	0.022	-4.49%	10.00%
M1	5000	0.87	0.021	-3.33%	5.00%
M2	15000	0.90	0.018	3.45%	-21.74%

Realizando uma análise comparativa entre as duas propostas, observa-se que os modelos desenvolvidos no presente estudo apresentam um desempenho variável em relação aos valores de referência da literatura.

Para os menores valores de pressure head (10, 60 e 100 cm), os resultados obtidos superam ligeiramente os valores de referência, indicando que a abordagem utilizada foi capaz de capturar melhor a variabilidade dos dados nesses intervalos. Em particular, para $T = 100$ cm, o modelo M3 apresentou um aumento de 2,33% no coeficiente de determinação (R^2) e uma redução de 10,00% no RMSE, evidenciando um ajuste mais preciso para essa condição.

Por outro lado, para valores intermediários de pressure head, como 330 cm, os resultados obtidos são bastante similares aos valores de referência, com uma variação de

-2,27% em R^2 e nenhuma mudança no RMSE. Isso indica que, para esse intervalo, os modelos utilizados possuem um desempenho equivalente ao da literatura.

Já para valores mais altos de pressure head, como 1000 e 3000 cm, observa-se uma redução no desempenho dos modelos em relação aos valores de referência. Para $T = 1000$ cm, o modelo M5 apresentou uma redução de 5,62% no R^2 e um aumento de 10,00% no RMSE, sugerindo que a capacidade de generalização do modelo pode ser afetada em faixas mais elevadas de pressão. Esse comportamento também se repete para $T = 3000$ cm, onde a variação de R^2 foi de -4,49% e o RMSE aumentou 10,00%.

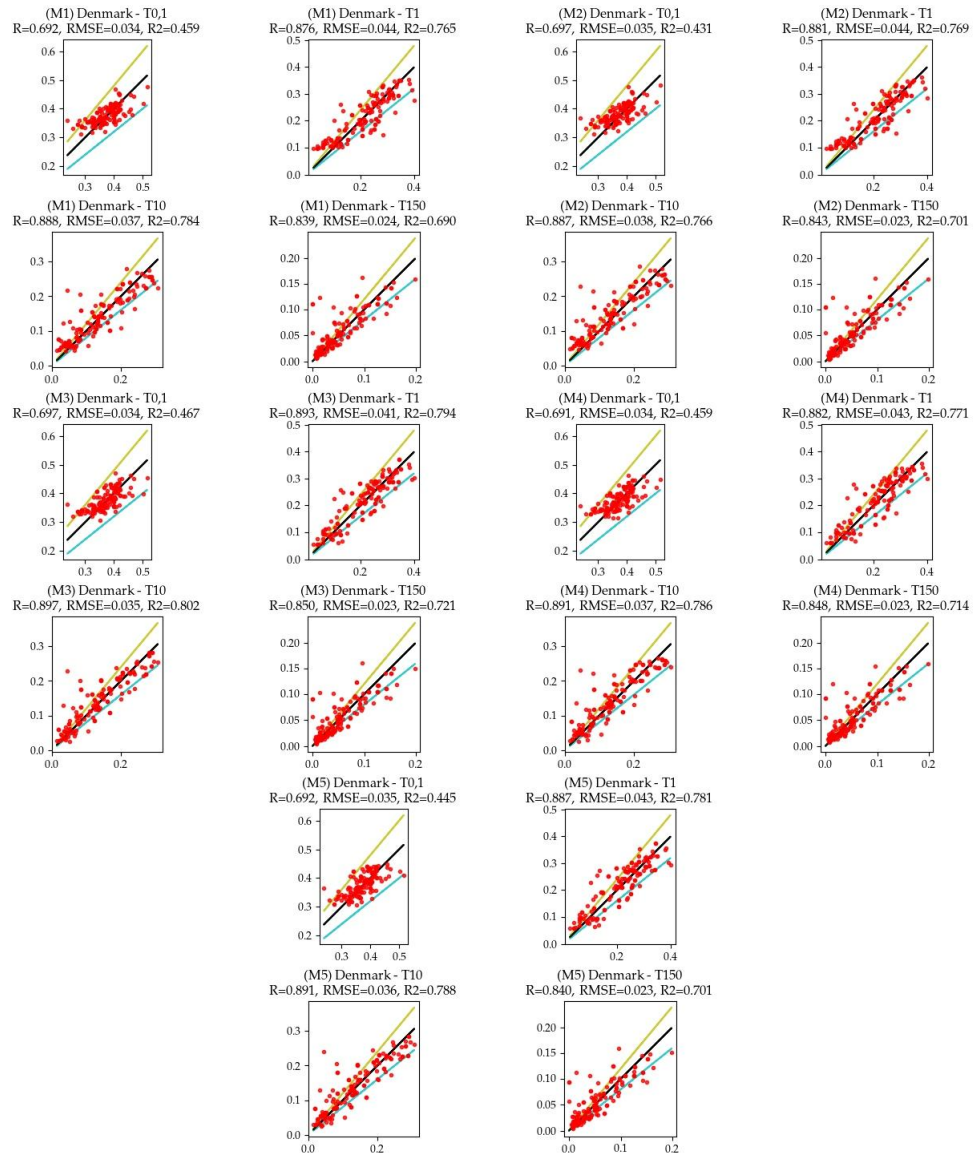
No entanto, para os maiores valores de pressure head, como 5000 e 15000 cm, os resultados obtidos voltam a se aproximar ou até mesmo superar os valores de referência. Destaca-se o caso de $T = 15000$ cm, em que o modelo M2 obteve um incremento de 3,45% no R^2 e uma redução expressiva de 21,74% no RMSE, demonstrando que a abordagem proposta pode ser mais eficaz para essas condições extremas.

Dessa forma, os resultados indicam que os modelos desenvolvidos possuem um bom desempenho geral, especialmente para os menores e maiores valores de pressure head. No entanto, para valores intermediários, ainda há espaço para aprimoramento.

4.1.3 BASE DE DADOS D1

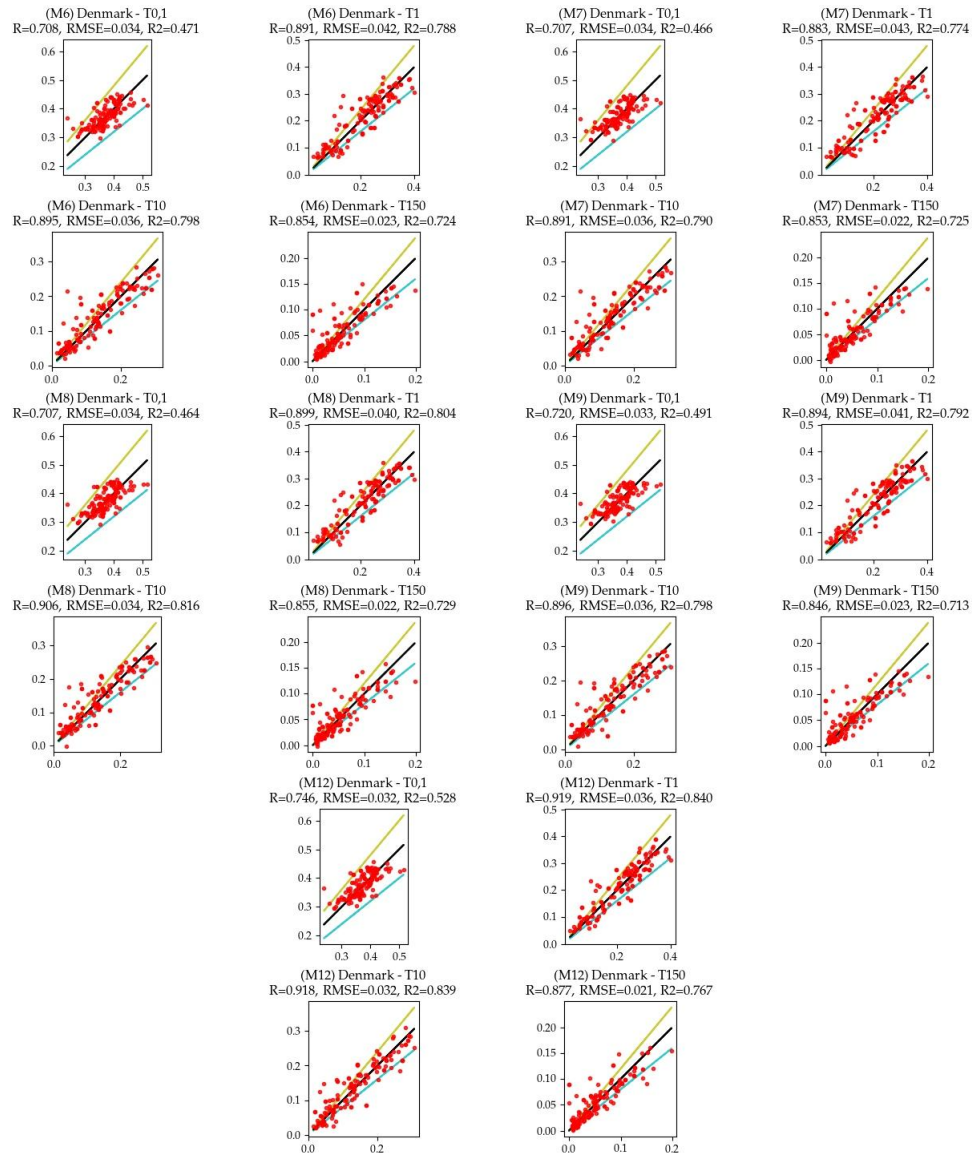
Agora, a base de dados D1, que representa a Dinamarca, possui valores de pressure head de 0.1, 1, 10 e 150 metros. A partir dessa base, foram gerados os seguintes gráficos:

Figura 16 – Gráficos para resultado da base de dados D1 - M1 a M5



Fonte: do autor

Figura 17 – Gráficos para resultado da base de dados D1 - M6 a M12



Fonte: do autor

Os resultados obtidos podem ser descritos na tabela abaixo, onde o modelo indica o número de camadas ocultas da rede sendo M1 equivalente a 1 camada, M2 a 2 camadas e assim sucessivamente, T indica a pressure head, R^2 o coeficiente de determinação e RMSE representa o erro quadrático médio:

Tabela 11 – Resultados obtidos para D1

Modelo	T	R	R^2	RMSE	Modelo	T	R	R^2	RMSE
M1	0.1	0.692	0.46	0.034	M6	0.1	0.708	0.47	0.034
M1	1	0.876	0.77	0.044	M6	1	0.891	0.79	0.042
M1	10	0.888	0.78	0.037	M6	10	0.895	0.80	0.036
M1	150	0.839	0.69	0.024	M6	150	0.854	0.72	0.023
M2	0.1	0.697	0.43	0.035	M7	0.1	0.707	0.47	0.034
M2	1	0.881	0.77	0.044	M7	1	0.883	0.77	0.043
M2	10	0.887	0.77	0.038	M7	10	0.891	0.79	0.036
M2	150	0.843	0.70	0.023	M7	150	0.853	0.73	0.022
M3	0.1	0.697	0.47	0.034	M8	0.1	0.707	0.46	0.034
M3	1	0.893	0.79	0.041	M8	1	0.899	0.80	0.040
M3	10	0.897	0.80	0.035	M8	10	0.906	0.82	0.034
M3	150	0.85	0.72	0.023	M8	150	0.855	0.73	0.022
M4	0.1	0.691	0.46	0.034	M9	0.1	0.72	0.49	0.033
M4	1	0.882	0.77	0.043	M9	1	0.894	0.79	0.041
M4	10	0.891	0.79	0.037	M9	10	0.896	0.80	0.036
M4	150	0.848	0.71	0.023	M9	150	0.846	0.71	0.023
M5	0.1	0.692	0.45	0.035	M12	0.1	0.746	0.53	0.032
M5	1	0.887	0.78	0.043	M12	1	0.919	0.84	0.036
M5	10	0.891	0.79	0.036	M12	10	0.918	0.84	0.032
M5	150	0.84	0.70	0.023	M12	150	0.877	0.77	0.021

Usando novamente como base o trabalho desenvolvido por (KOTLAR et al., 2019), os valores obtidos em sua pesquisa podem ser utilizados como referência para a análise comparativa com os resultados obtidos no presente estudo. Esses valores de referência estão apresentados na Tabela 12.

Tabela 12 – Dados de referência para D1

Base de Dados	T	R^2	RMSE
D1	0.1	0.35	0.041
D1	1	0.73	0.051
D1	10	0.73	0.041
D1	150	0.72	0.021

Na Tabela 13 abaixo pode-se demonstrar um comparativo de desempenho do resultado com base percentual considerando-se R^2 e RMSE.

Tabela 13 – Variação de R^2 e RMSE para os melhores resultados apresentados de D1 em comparação com os dados de referência

Modelo	T	R^2	RMSE	Variação R^2	Variação RMSE
M12	0.1	0.53	0.032	51.43%	-21.95%
M12	1	0.84	0.036	15.07%	-29.41%
M12	10	0.84	0.032	15.07%	-21.95%
M12	150	0.77	0.021	6.94%	0.00%

Realizando uma análise comparativa entre os resultados obtidos e os valores de referência da literatura, observa-se que os modelos desenvolvidos no presente estudo apresentaram desempenho superior em todos os níveis de pressure head, com destaque para os menores valores.

Para os menores valores de pressure head ($T = 0.1$ e 1 m), os resultados demonstraram avanços significativos em relação à literatura. Em $T = 0.1$ m, o modelo M12 apresentou um incremento de 51,43% no coeficiente de determinação (R^2) e uma redução expressiva de 21,95% no RMSE, evidenciando um ajuste muito mais preciso para essa condição. De forma semelhante, em $T = 1$ m, também se observou um aumento de 15,07% no R^2 e uma redução de 29,41% no RMSE, indicando que a abordagem utilizada conseguiu representar com maior fidelidade os dados dessa faixa.

No valor intermediário de pressure head ($T = 10$ m), os modelos mantiveram o bom desempenho observado. O modelo M12 alcançou novamente um aumento de 15,07% no R^2 , juntamente com uma redução de 21,95% no RMSE, o que reforça a consistência dos modelos propostos mesmo em faixas medianas de pressão.

Para o maior valor analisado ($T = 150$ m), o desempenho dos modelos ainda se manteve satisfatório. Embora a redução no RMSE tenha sido nula, o modelo M12 apresentou um aumento de 6,94% no R^2 em comparação com os valores de referência, demonstrando que, mesmo em condições de pressão mais elevadas, a abordagem foi capaz de manter uma boa capacidade preditiva.

Dessa forma, os resultados indicam que os modelos desenvolvidos para a base de dados D1 apresentam desempenho superior ao da literatura em todas as faixas de pressure head, com ganhos especialmente relevantes nas pressões mais baixas. Essa consistência em diferentes condições demonstra a robustez e eficácia da metodologia proposta para modelagem da retenção de água em solos da Dinamarca.

4.2 DISCUSSÕES

Nesta seção, discutimos os resultados obtidos e suas implicações. Primeiramente, os principais achados são analisados em relação às expectativas iniciais. Em seguida, comparamos os resultados com a literatura existente, destacando semelhanças e divergências.

Além disso, são apresentadas as limitações do estudo e sugestões para trabalhos futuros.

4.2.1 INTERPRETAÇÃO DOS RESULTADOS

A análise dos resultados obtidos demonstra que o desempenho do modelo de rede neural variou significativamente em função das características das bases de dados e das faixas de pressão consideradas. De forma geral, os melhores desempenhos foram observados nas pressões intermediárias, principalmente entre 1 e 10 metros, onde os modelos atingiram os maiores valores de coeficiente de determinação (R^2 até 0,90) e os menores erros médios quadráticos (RMSE até 0,018).

Este comportamento reflete a maior estabilidade física dos solos nessa faixa, onde a retenção de água é menos sensível a variações extremas na estrutura do solo. Por outro lado, o desempenho foi consistentemente inferior nas pressões extremas, como 0,6 m (representando macroporos) e 150 m (microporos e água fortemente retida). Isso sugere que o modelo tem maior dificuldade em capturar os processos associados à drenagem rápida (pressões baixas) e à retenção residual (pressões muito altas), padrão que também é relatado na literatura (ZHANG; SCHAAP, 2017; KOTLAR et al., 2019).

Ao analisar o comportamento por base de dados, nota-se que as bases B1 e B2, representativas de solos brasileiros, apresentaram desempenho superior. Esse resultado é possivelmente atribuído à maior diversidade textural, mineralógica e estrutural dessas bases, que permite à rede neural capturar relações mais robustas entre as variáveis de entrada e saída. Além disso, o maior volume de dados dessas bases contribuiu para reduzir a variância e melhorar a generalização do modelo.

Em contraste, a base D1, representativa de solos da Dinamarca, apresentou os piores resultados entre os três conjuntos analisados. Essa pior performance está associada à menor diversidade textural e ao menor tamanho da base, o que limita a capacidade do modelo de aprender padrões robustos. Além disso, as diferenças edafoclimáticas entre os solos brasileiros e dinamarqueses podem ter impactado a eficácia do modelo, que foi originalmente calibrado em bases de dados com maior heterogeneidade, especialmente em solos tropicais. Esse padrão de dificuldade na transferência entre domínios distintos também é observado nos trabalhos de (KOTLAR; de Jong van Lier; de Souza Brito, 2020) e (ZHANG; SCHAAP, 2017).

Outro aspecto relevante foi a influência da arquitetura do modelo no desempenho. As melhores configurações foram aquelas com duas ou três camadas ocultas e até 100 neurônios por camada, que apresentaram bom equilíbrio entre capacidade de generalização e desempenho preditivo. Arquiteturas mais profundas resultaram em sobreajuste, enquanto modelos excessivamente simples não foram capazes de capturar a complexidade dos dados, especialmente nas pressões intermediárias.

Por fim, a análise das métricas de desempenho reforça que o modelo proposto

apresenta robustez na predição da curva de retenção de água, especialmente quando aplicado a bases de dados representativas e nas faixas de pressão de maior relevância agrônômica. Contudo, a presença de outliers e a queda de desempenho nas tensões extremas indicam limitações, tanto na capacidade do modelo quanto na representatividade dos dados utilizados.

4.2.2 COMPARAÇÃO COM OUTROS ESTUDOS

Os resultados obtidos neste trabalho corroboram e, em alguns casos, superam os reportados na literatura especializada, especialmente quando se trata da aplicação de redes neurais artificiais (RNAs) para a modelagem da curva de retenção de água no solo.

Ao desenvolverem funções de pedotransferência (PTFs) para solos brasileiros, foram relatados coeficientes de determinação (R^2) na faixa de 0,70 a 0,82, utilizando modelos lineares simples (SLM-PTF). Embora esses modelos sejam eficazes quando as variáveis possuem alta correlação, eles apresentam limitações na captura de relações não lineares e na transferência para regiões com características edafoclimáticas distintas (KOTLAR et al., 2019). No presente trabalho, as RNAs superaram claramente esses valores, atingindo até $R^2 = 0,90$ e $RMSE = 0,018$ na base B2, com desempenho robusto também na base B1 ($R^2 = 0,88$, $RMSE = 0,027$), o que evidencia um avanço significativo no tratamento de relações não lineares.

De forma semelhante, utilizando-se o modelo Rosetta baseado em RNA simples, reportou-se R^2 entre 0,75 e 0,85, com desempenho fortemente dependente da qualidade e diversidade dos dados (SCHAAP; LEIJ; GENUCHTEN, 2001). Posteriormente, (ZHANG; SCHAAP, 2017), com a evolução para o Rosetta3, conseguiram melhorar esses números, alcançando R^2 médios de até 0,88, mas ainda com limitações claras na transferência para solos tropicais, conforme os próprios autores reconheceram. O presente trabalho, ao trabalhar com solos brasileiros, não só alcança valores semelhantes ou superiores, como demonstra que modelos de RNA, quando bem parametrizados e aplicados em bases diversas, podem superar as limitações de modelos anteriores.

Em outro estudo, utilizando-se machine learning combinado com regressões não lineares para estimar propriedades hidráulicas, também reportou-se R^2 na faixa de 0,80 a 0,90, com $RMSE$ variando conforme a base (ZHAO et al., 2016). Embora seus resultados sejam robustos, eles destacaram que os desempenhos caem significativamente em solos tropicais ou quando há elevada heterogeneidade textural. Este mesmo padrão foi observado em outro estudo, que identificou que as PTFs tradicionais não conseguem capturar adequadamente a variabilidade dos solos tropicais, especialmente devido à presença de materiais altamente intemperizados, elevada concentração de oxihidróxidos e estruturas muito variáveis (BOTULA; RANST; CORNELIS, 2014).

Neste sentido, os resultados obtidos neste trabalho reforçam a superioridade dos

modelos baseados em RNA, que demonstraram maior acurácia e menor erro em contextos de alta variabilidade, como os solos brasileiros. Nas bases B1 e B2, o modelo obteve $R\hat{s} \geq 0,88$ e $RMSE \leq 0,027$, enquanto que na base dinamarquesa (D1), esses valores foram mais modestos ($R^2 = 0,84$, $RMSE = 0,021$). Essa diferença é atribuída, principalmente, à menor diversidade da base D1, ao seu tamanho reduzido e às diferenças edafoclimáticas — como textura mais homogênea, mineralogia menos variável e ausência de características típicas de solos tropicais, como elevada porosidade estrutural e grande presença de agregados biológicos e minerais secundários.

Além disso, a ausência de análise formal de incerteza, fortemente presente em modelos como o Rosetta3 (ZHANG; SCHAAP, 2017) e no trabalho de (KOTLAR; de Jong van Lier; de Souza Brito, 2020), aparece como um ponto de melhoria do presente estudo. No entanto, os resultados aqui obtidos são competitivos, mesmo sem essa etapa, indicando que o modelo é estatisticamente robusto.

Entre os principais avanços em relação à literatura destacam-se:

- Maior acurácia, especialmente nas bases B1 e B2, superando os limites reportados por Kotlar et al. (2019, 2020) e Schaap et al. (2001).
- Baixo custo operacional, visto que elimina a necessidade de ensaios laboratoriais demorados e onerosos, alinhado ao que foi demonstrado por (ARSHAD et al., 2013).
- Redução significativa no tempo para obtenção da curva de retenção, o que tem impacto direto em levantamentos pedológicos, diagnósticos de uso agrícola e planejamento hidrológico.
- Aplicabilidade prática comprovada, especialmente em solos tropicais, que são tradicionalmente um desafio para modelos empíricos e PTFs convencionais.
- Desempenho expressivo em solos tropicais, o que reforça o papel das RNAs como ferramenta adequada para ambientes altamente heterogêneos, como observado nas bases brasileiras.

Em contrapartida, a queda de desempenho na base D1 revela que, assim como apontado na literatura (KOTLAR; de Jong van Lier; de Souza Brito, 2020; ZHANG; SCHAAP, 2017), ainda persistem desafios na construção de modelos realmente transferíveis entre diferentes condições edafoclimáticas. Esse fato sugere que abordagens como aprendizado por transferência, análise probabilística ou o desenvolvimento de modelos híbridos podem ser caminhos promissores para superar essas limitações.

Tabela 14 – Comparação dos resultados deste trabalho com estudos da literatura

Estudo	R ²	RMSE	Observações
Schaap et al. (2001) – Rosetta	0,75 – 0,85	–	RNA simples, dependente da qualidade dos dados, dificuldade em solos tropicais.
Zhang & Schaap (2017) – Rosetta3	até 0,88	–	Melhoria com análise de incerteza, mas ainda com limitações em transferência.
Kotlar et al. (2019, 2020)	0,70 – 0,82	–	Modelos lineares, baixa performance em solos tropicais.
Zhao et al. (2016)	0,80 – 0,90	variável	ML robusto, mas com queda em alta heterogeneidade.
Este trabalho – Base B1 (Brasil)	0,88	0,027	Alto desempenho em solos tropicais.
Este trabalho – Base B2 (Brasil)	0,90	0,018	Melhor desempenho geral.
Este trabalho – Base D1 (Dinamarca)	0,84	0,021	Desempenho inferior, devido à baixa diversidade textural da base.

4.2.3 LIMITAÇÕES E MELHORIAS

O modelo desenvolvido demonstrou excelente desempenho na captura de relações não lineares e na previsão de propriedades hidráulicas do solo. No entanto, algumas limitações foram identificadas e merecem atenção. A principal delas reside na forte dependência da diversidade e qualidade dos dados de treinamento, fator que impacta diretamente na capacidade de generalização dos modelos, especialmente quando aplicados a contextos diferentes daqueles para os quais foram treinados.

Adicionalmente, observou-se tendência ao sobreajuste em arquiteturas mais profundas, o que compromete o desempenho preditivo em bases menores ou menos representativas, como foi o caso da base D1. Outro ponto crítico é a limitada interpretabilidade dos modelos de RNA, característica inerente à sua estrutura de "caixa-preta", dificultando a compreensão do impacto de cada variável na predição.

Por fim, destaca-se a ausência de uma análise formal de incerteza associada às predições, aspecto essencial para aplicações em engenharia e gestão de recursos naturais.

Como melhorias, recomenda-se a incorporação de técnicas de regularização mais robustas, como dropout dinâmico ou penalizações L1/L2 ajustadas, além da aplicação de estratégias de aumento de dados e validação cruzada mais rigorosa. Também se sugere explorar ferramentas de interpretabilidade e incorporar métodos de quantificação de incertezas para fortalecer a aplicabilidade prática do modelo.

4.2.4 PERSPECTIVAS FUTURAS

Para aprimorar a abordagem desenvolvida, futuras pesquisas devem priorizar a análise e incorporação de incertezas nos modelos preditivos, utilizando técnicas estatísticas e computacionais adequadas. Além disso, é recomendada a ampliação dos bancos de dados, incluindo diferentes tipos de solos, condições climáticas e propriedades físico-químicas, de modo a fortalecer a generalização e robustez das soluções propostas.

Outros caminhos promissores incluem a integração de redes neurais com métodos estatísticos e de aprendizado de máquina mais interpretáveis, bem como a aplicação de estratégias de aprendizado por transferência para aumentar a eficiência na adaptação dos modelos a novos cenários.

O desenvolvimento de ferramentas computacionais de fácil acesso, como aplicativos web ou plugins para sistemas de informação geográfica (SIG), também representa uma possibilidade concreta para tornar os modelos desenvolvidos aplicáveis de forma prática por profissionais das áreas agrícola, ambiental e de engenharia. Por fim, a realização de análises econômicas pode esclarecer os benefícios da adoção de PTFs baseadas em aprendizado de máquina frente aos métodos tradicionais.

5 CONCLUSÃO

Este trabalho teve como objetivo geral avaliar a aplicação de funções de pedotransferência (PTFs) por meio de modelos baseados em redes neurais artificiais (RNAs) para estimar propriedades hidráulicas do solo. Utilizando três bases de dados distintas — representando diferentes condições geográficas e texturais —, foi possível analisar a robustez e a capacidade de generalização dos modelos.

Os resultados obtidos demonstraram que as RNAs são ferramentas eficazes na modelagem de PTFs, especialmente pela capacidade de capturar relações não lineares entre variáveis, superando, em diversos cenários, métodos tradicionais. A otimização dos hiperparâmetros, realizada via busca bayesiana, foi determinante para alcançar um equilíbrio adequado entre desempenho preditivo e generalização.

As métricas de desempenho — com coeficientes de determinação (R^2) de 0,88 na base B1, 0,90 na base B2 e 0,84 na base D1, e RMSE variando entre 0,018 e 0,027 nas melhores configurações — reforçam a viabilidade da proposta, embora tenham sido observadas variações relevantes entre as bases de dados. Modelos aplicados às bases B1 e B2 apresentaram desempenho robusto, enquanto na base D1 foram identificadas limitações associadas à heterogeneidade dos dados e possível sobreajuste.

Comparativamente à literatura, os resultados alinham-se com estudos como Kotlar et al. (2019), reforçando o potencial das RNAs na modelagem de propriedades do solo. Contudo, também foram observadas divergências, particularmente relacionadas à incerteza preditiva e à capacidade de transferência dos modelos para diferentes contextos edafoclimáticos.

As principais limitações deste estudo incluem a dependência de dados de qualidade e diversidade para treinamento e a sensibilidade dos modelos a cenários com menor representatividade. Como perspectivas futuras, recomenda-se expandir o banco de dados, aplicar técnicas avançadas de regularização, explorar abordagens híbridas (como aprendizado por transferência) e, sobretudo, realizar análises de incerteza mais robustas.

Este estudo contribui, portanto, para o avanço da modelagem de funções de pedotransferência, oferecendo uma metodologia replicável, robusta e de potencial aplicação prática em contextos como planejamento agrícola, modelagem hidrológica e gestão ambiental.

REFERÊNCIAS

- ARFEEN, N.; KHAN, T. A. Evaluation of psd models for the estimation of hydraulic conductivity for different soil textural classes. *International Journal of Engineering Works*, 2020. 31
- ARSHAD, R. R. et al. Predicting saturated hydraulic conductivity by artificial intelligence and regression models. *Isrn Soil Science*, 2013. 33, 40, 43, 83
- BAGNALL, D. K. et al. Carbon-sensitive pedotransfer functions for plant available water. *Soil Science Society of America Journal*, 2022. 35, 37, 45, 47
- BARROS, A. d. et al. Pedotransfer functions to estimate water retention parameters of soils in northeastern brazil. *Revista Brasileira de Ciência do Solo*, v. 37, p. 379–391, 2013. 52
- BENKE, K. K. et al. Development of pedotransfer functions by machine learning for prediction of soil electrical conductivity and organic carbon content. *Geoderma*, 2020. 32, 33, 39
- BOHNE, K.; RENGGER, M.; WESSOLEK, G. New pedotransfer function (“crc”) for the prediction of unsaturated soil hydraulic conductivity using soil water retention data. *International Agrophysics*, 2019. 30, 31, 35, 39, 40, 41, 44
- BORMANN, H. Towards a hydrologically motivated soil texture classification. *Geoderma*, 2010. 30
- BOTULA, Y.; RANST, E.; CORNELIS, W. Pedotransfer functions to predict water retention for soils of the humid tropics: a review. *Revista Brasileira de Ciência do Solo*, SciELO, v. 38, n. 3, 2014. 15, 82
- CHIRICO, G. B.; MEDINA, H.; ROMANO, N. Functional evaluation of ptf prediction uncertainty: An application at hillslope scale. *Geoderma*, 2010. 32, 33, 39, 41
- CIORUTA, B.-V.; COMAN, M. A theoretical perspective on the role and functions of soil resources in the information and knowledge society. *Asian Soil Research Journal*, 2022. 25
- COOPER, E. et al. *Using Data Assimilation to Optimize Pedotransfer Functions Using Large-Scale In-Situ Soil Moisture Observations*. 2020. 38
- COOPER, E. et al. Using data assimilation to optimize pedotransfer functions using field-scale in situ soil moisture observations. *Hydrology and Earth System Sciences*, 2021. 38
- D’EMILIO, A. et al. Artificial neural networks for predicting the water retention curve of sicilian agricultural soils. *Water*, MDPI, v. 10, n. 10, p. 1431, 2018. 13, 42
- DRAYE, X. et al. Model-assisted integration of physiological and environmental constraints affecting the dynamic and spatial patterns of root water uptake from soils. *Journal of Experimental Botany*, 2010. 30
- ELGELDAWI, E. et al. Hyperparameter tuning for machine learning algorithms used for arabic sentiment analysis. *Informatics*, v. 8, p. 79, 2021. 50

- ELSHEWEY, A. et al. Bayesian optimization with support vector machine model for parkinson disease classification. *Sensors*, v. 23, p. 2085, 2023. 50
- ESMAEELNEJAD, L. et al. Selection of a suitable model for the prediction of soil water content in north of iran. *Spanish Journal of Agricultural Research*, v. 13, n. 1, p. e1202, Feb. 2015. Disponível em: <<https://sjar.revistas.csic.es/index.php/sjar/article/view/6111>>. 18
- FIDALSKI, J.; HELBEL, C. Available water content for the management of irrigated crops in the northwestern region of parana state. *Revista Brasileira De Agricultura Irrigada*, 2020. 35, 40, 43
- FU, X. et al. Evaluation of permeable brick pavement on the reduction of stormwater runoff using a coupled hydrological model. *Water*, v. 12, n. 10, 2020. ISSN 2073-4441. Disponível em: <<https://www.mdpi.com/2073-4441/12/10/2821>>. 19, 20
- GAMIE, R.; SMEDT, F. D. Experimental and statistical study of saturated hydraulic conductivity and relations with other soil properties of a desert soil. *European Journal of Soil Science*, 2017. 31
- GHANBARIAN, B. et al. Sample dimensions effect on prediction of soil water retention curve and saturated hydraulic conductivity. *Journal of Hydrology*, 2015. 27, 28, 29
- GHANBARIAN-ALAVIJEH, B.; HUNT, A. G. Estimation of soil-water retention from particle-size distribution. *Soil Science*, 2012. 27
- GHAWI, R.; PFEFFER, J. Efficient hyperparameter tuning with grid search for text categorization using knn approach with bm25 similarity. *Open Computer Science*, v. 9, p. 160–180, 2019. 48
- GU, B. et al. Optimizing large-scale hyperparameters via automated learning algorithm. 2021. 48
- GUNARATHNA, M. H. J. P. et al. *Machine Learning Approaches to Develop Pedotransfer Functions for Tropical Sri Lankan Soils*. 2019. 35, 40, 44
- GUZMÁN, G. et al. Water related properties to assess soil quality in two olive orchards of south spain under different management strategies. *Water*, 2019. 29
- GÜLSER, C.; CANDEMIR, F. Using soil moisture constants and physical properties to predict saturated hydraulic conductivity. *Eurasian Journal of Soil Science (Ejss)*, 2014. 32, 39
- HARDIE, M. et al. Influence of antecedent soil moisture on hydraulic conductivity in a series of texture-contrast soils. *Hydrological Processes*, 2012. 29
- HE, Q. et al. Global optimization of soil texture maps from satellite-observed soil moisture drydowns and its implementation in noah-mp land surface model. *Journal of Advances in Modeling Earth Systems*, 2024. 30
- HENGL, T. et al. Soilgrids250m: Global gridded soil information based on machine learning. *Plos One*, 2017. 35, 40, 44, 45

- HOQUE, K.; ALJAMAAN, H. Impact of hyperparameter tuning on machine learning models in stock price forecasting. *Ieee Access*, v. 9, p. 163815–163830, 2021. 48, 49, 50
- HORNIK, K.; STINCHCOMBE, M.; WHITE, H. Multilayer feedforward networks are universal approximators. *Neural Networks*, v. 2, n. 5, p. 359–366, 1989. 21
- HOUBORG, R.; MCCABE, M. F. A hybrid training approach for leaf area index estimation via cubist and random forests machine-learning. *ISPRS Journal of Photogrammetry and Remote Sensing*, v. 135, p. 173–188, 2018. ISSN 0924-2716. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0924271617303209>>. 16, 17
- IBRAHIM, H. et al. Patterns of soil nitrogen sequestration in drylands explored with pedotransfer functions and bayesian analysis. *Research Journal of Applied Sciences Engineering and Technology*, 2014. 38
- IRESON, A. et al. Using observed soil moisture to constrain the uncertainty of simulated hydrological fluxes. *Hydrological Processes*, 2022. 38
- IVERSEN, B. et al. Risk predicting of macropore flow using pedotransfer functions, textural maps, and modeling. *Vadose Zone Journal*, v. 10, p. 1185–1195, 2011. 53
- JENA, S. K. et al. Toward developing a generalizable pedotransfer function for saturated hydraulic conductivity using transfer learning and predictor selector algorithm. *Water Resources Research*, 2021. 35, 40, 44, 45
- JIA, L. et al. Effects of sponge city development on soil moisture and water quality in a typical city in the loess plateau in china. *Frontiers in Earth Science*, v. 8, 2020. ISSN 2296-6463. Disponível em: <<https://www.frontiersin.org/articles/10.3389/feart.2020.00125>>. 20
- JIMÉNEZ, B. H.; FLORES, F. I. D. I. C.; DELGADO-INIESTA, M. J. Edaphosphere: A perspective of soil inside the biosphere. *Earth*, 2023. 24
- JIN, H. Hyperparameter importance for machine learning algorithms. 2022. 48
- JORDA, H. et al. *Using Boosted Regression Trees to Explore Key Factors Controlling Saturated and Near-saturated Hydraulic Conductivity*. 2015. 35, 38, 40, 45
- KALUMBA, M. et al. Development and functional evaluation of pedotransfer functions for soil hydraulic properties for the <scp>zambezi</scp> river basin. *European Journal of Soil Science*, 2020. 33, 39, 43
- KANG, S. et al. Effects of muddy water infiltration on the hydraulic conductivity of soils. *Agronomy*, 2024. 30
- KATEB, Z. et al. Hydrological modeling using the swat model based on two types of data from the watershed of beni haroun dam, algeria. *Journal of Water and Land Development*, 2019. 29
- KESKIN, H.; GRUNWALD, S.; HARRIS, W. G. Digital mapping of soil carbon fractions with machine learning. *Geoderma*, v. 339, p. 40–58, 2019. ISSN 0016-7061. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S001670611732030X>>. 17

- KOCH, A. et al. Soil security: Solving the global soil crisis. *Global Policy*, 2013. 24, 25
- KOESTEL, J. et al. *Estimating the Permeability of Naturally Structured Soil From Percolation Theory and Pore Space Characteristics Imaged by X-Ray*. 2018. 36, 45
- KOTLAR, A. M.; de Jong van Lier, Q.; de Souza Brito, E. Pedotransfer functions for water contents at specific pressure heads of silty soils from amazon rainforest. *Geoderma*, v. 361, p. 114098, 2020. ISSN 0016-7061. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0016706119315368>>. 41, 53, 54, 75, 81, 83
- KOTLAR, A. M. et al. Development and uncertainty assessment of pedotransfer functions for predicting water contents at specific pressure heads. *Vadose Zone Journal*, v. 18, n. 1, p. 190063, 2019. Disponível em: <<https://doi.org/10.2136/vzj2019.06.0063>>. 21, 52, 53, 56, 70, 79, 81, 82
- KOZAITI, M.; KOSTOPOULOU, S. The effect of compaction on water retention in the vineyard's root zone. *Agrofor*, 2019. 28
- KUANG, X.; JIAO, J. J. A new equation for the soil water retention curve. *European Journal of Soil Science*, 2014. 27, 28, 29
- KUANG, X. et al. A modification to the van genuchten model for improved prediction of relative hydraulic conductivity of unsaturated soils. *European Journal of Soil Science*, 2020. 27, 28, 29
- KUBBEN, P.; DUMONTIER, M.; DEKKER, A. (Ed.). *Fundamentals of Clinical Data Science*. Cham: Springer, 2019. Figura 8.8, disponível sob licença Creative Commons Attribution 4.0. ISBN 978-3-319-99712-4. Disponível em: <<https://doi.org/10.1007/978-3-319-99713-1>>. 50
- LI, Y.; PAN, L.; XU, G. On quantifying hydraulic conductance at the soil-root interface. *Hydrological Processes*, 2013. 30
- LIAO, K. et al. Comparison of seven water retention functions used for modelling soil hydraulic conductivity due to film flow. *Soil Use and Management*, 2018. 30
- LIN, N. et al. A comparative study of machine learning models with hyperparameter optimization algorithm for mapping mineral prospectivity. *Minerals*, v. 11, p. 159, 2021. 48
- LOOY, K. V. et al. Pedotransfer functions in earth system science: Challenges and perspectives. *Reviews of Geophysics*, 2017. 13, 32, 34, 37, 38, 41, 42, 45, 47
- MCBRATNEY, A. B.; MINASNY, B.; TRANTER, G. *Necessary Meta-Data for Pedotransfer Functions*. 2011. 36, 38, 45, 47
- MOEYS, J.; BERGHEAUD, V.; COQUET, Y. Pedotransfer functions for isoproturon sorption on soils and vadose zone materials. *Pest Management Science*, 2011. 37, 47
- MOHANTY, M.; SINHA, N. K.; PAINULI, D. K. e. a. Modelling soil water contents at field capacity and permanent wilting point using artificial neural network for indian soils. *Natl. Acad. Sci. Lett.*, v. 38, p. 373–377, 2015. 17

- MONTZKA, C. et al. A global data set of soil hydraulic properties and sub-grid variability of soil water retention and hydraulic conductivity curves. *Earth System Science Data*, v. 9, n. 2, p. 529–543, 2017. 21
- MUGHNI, M.; FAHRUDIN, T.; KAMISUTARA, M. Classification of toddler nutritional status based on anthropometric index and feature discrimination using support vector machine hyperparameter tuning. *Ijconsist Journals*, v. 2, p. 60–65, 2021. 48, 49
- OBIERO, J. P. O. et al. Development of pedotransfer functions for saturated hydraulic conductivity. *Open Journal of Modern Hydrology*, 2013. 32, 35, 39, 40, 44
- OLIVEIRA VANESSA SARI, N. M. d. R. C. Marquis Henrique Campos de; PEDROLLO, O. C. Estimation of soil water content in watershed using artificial neural networks. *Hydrological Sciences Journal*, Taylor Francis, v. 62, n. 13, p. 2120–2138, 2017. Disponível em: <<https://doi.org/10.1080/02626667.2017.1364844>>. 17
- OSTOVARI, Y.; ASGARI, K.; CORNELIS, W. Performance evaluation of pedotransfer functions to predict field capacity and permanent wilting point using unsoda and hypres datasets. *Arid Land Research and Management*, 2015. 31, 33, 39, 41
- PACHEPSKY, Y. et al. Scaling and pedotransfer in numerical simulations of flow and transport in soils. *Vadose Zone Journal*, 2014. 35, 36, 40, 44
- PADARIAN, J.; MINASNY, B.; MCBRATNEY, A. Machine learning and soil sciences: a review aided by machine learning tools. *Soil*, EGU, v. 6, n. 1, p. 35–52, 2020. 15, 19, 41
- PATIL, N.; SINGH, S. Pedotransfer functions for estimating soil hydraulic properties: A review. *Pedosphere*, Elsevier, v. 26, n. 4, p. 417–430, 2016. 14
- PINHEIRO, A. G. et al. Characterization of alluvial soil hydrodynamics in the upper ipanema river basin using the beerkan method. *Dyna*, 2021. 30
- PINNINGTON, E. et al. Improving soil moisture prediction of a high-resolution land surface model by parameterising pedotransfer functions through assimilation of smap satellite data. *Hydrology and Earth System Sciences*, 2021. 38
- PINNINGTON, E.; QUAIFE, T.; BLACK, E. Impact of remotely sensed soil moisture and precipitation on soil moisture prediction in a data assimilation system with the jules land surface model. *Hydrology and Earth System Sciences*, 2018. 31
- QIAN, D. Analysis of the hyperparameter selection in machine learning. *Applied and Computational Engineering*, v. 104, p. 122–128, 2024. 48, 49
- QIAO, J. et al. Development of pedotransfer functions for soil hydraulic properties in the critical zone on the loess plateau, china. *Hydrological Processes*, 2018. 31, 32, 33, 39, 41
- RAFRAF, S. et al. Quantification of hysteresis effects on a soil subjected to drying and wetting cycles. *International Agrophysics*, 2016. 28
- ROSE, S. et al. Machine learning and statistical approaches used in estimating parameters that affect the soil fertility status: A survey. In: *2018 Second International Conference on Green Computing and Internet of Things (ICGCIoT)*. [S.l.]: IEEE, 2018. 13

- SAFFARI, R. et al. Effects of biological stabilization on the water retention properties of unsaturated soils. *Journal of Geotechnical and Geoenvironmental Engineering*, 2019. 29
- SCHAAP, M.; LEIJ, F.; GENUCHTEN, M. van. Rosetta: A computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions. *Journal of Hydrology*, v. 251, n. 3–4, p. 163–176, 2001. 82
- SCHARNAGL, B. et al. Inverse modelling of in situ soil water dynamics: Investigating the effect of different prior distributions of the soil hydraulic parameters. *Hydrology and Earth System Sciences*, 2011. 37, 47
- SCHOONOVER, J. E.; CRIM, J. F. An introduction to soil concepts and the role of soils in watershed management. *Journal of Contemporary Water Research & Education*, 2015. 24
- SHAHRIARI, B. et al. Taking the human out of the loop: A review of bayesian optimization. *Proceedings of the IEEE*, IEEE, v. 104, n. 1, p. 148–175, 2016. 51
- SHUSTER, W. D. et al. κ in an urban world: New contexts for hydraulic conductivity. *Jawra Journal of the American Water Resources Association*, 2021. 38
- SONG, X.-D. et al. Can pedotransfer functions based on environmental variables improve soil total nutrient mapping at a regional scale? *Soil and Tillage Research*, v. 202, p. 104672, 2020. ISSN 0167-1987. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0167198720304542>>. 17
- SPOKAS, K. A. et al. Biomass or biochar – which is better at improving soil hydraulic properties? *Acta Horticulturae*, 2016. 31
- STEENHUIS, T. S. et al. A saturated excess runoff pedotransfer function for vegetated watersheds. *Vadose Zone Journal*, 2013. 38, 47
- STUKE, A.; RINKE, P.; TODOROVIĆ, M. Efficient hyperparameter tuning for kernel ridge regression with bayesian optimization. 2020. 50
- SUHAILA, M. R. N. et al. Soil data definition for hydrologic response unit analysis in swat model of langkawi island, malaysia. *Eurasian Journal of Soil Science (Ejss)*, 2021. 27
- TAO, G. et al. A fractal approach for predicting unsaturated hydraulic conductivity of deformable clay. *Geofluids*, 2019. 29
- TOBELLA, A. B. et al. Determinants of field-saturated soil hydraulic conductivity across sub-saharan africa: Texture and beyond. *Water Resources Research*, 2024. 32, 33, 39, 41
- TUTOVA, G. F. et al. Response of earthworms to changes in the aggregate structure of floodplain soils. *Iop Conference Series Earth and Environmental Science*, 2022. 26
- TWARAKAVI, N. K. C.; ŠIMŮNEK, J.; SCHAAP, M. G. Can texture-based classification optimally classify soils with respect to soil hydraulics? *Water Resources Research*, 2010. 25, 26, 29
- TÓTH, B. et al. New generation of hydraulic pedotransfer functions for europe. *European Journal of Soil Science*, 2014. 33, 38, 40, 43

- WAINER, J.; CAWLEY, G. Nested cross-validation when selecting classifiers is overzealous for most practical applications. 2018. 49
- WANG, Y. et al. Effects of aggregate size on water retention capacity and microstructure of lime-treated silty soil. *Géotechnique Letters*, 2015. 27, 29
- WANG, Y. et al. Aggregate size effect on the water retention properties of a lime-treated compacted silt during curing. *E3s Web of Conferences*, 2016. 27
- WILPISZESKI, R. L. et al. Soil aggregate microbial communities: Towards understanding microbiome interactions at biologically relevant scales. *Applied and Environmental Microbiology*, 2019. 25
- WÖSTEN, J.; PACHEPSKY, Y.; RAWLS, W. Pedotransfer functions: bridging the gap between available basic soil data and missing soil hydraulic characteristics. *Journal of Hydrology*, v. 251, n. 3–4, p. 123–150, 2001. 56
- XU, C. et al. Enhancing pedotransfer functions (ptfs) using soil spectral reflectance data for estimating saturated hydraulic conductivity in southwestern china. *Catena*, Elsevier, v. 158, p. 350–356, 2017. 15, 16
- YAWSON, D. O. et al. Putting soil security on the policy agenda: Need for a familiar framework. *Challenges*, 2016. 25
- YIN, K. et al. A review of sand–clay mixture and soil–structure interface direct shear test. *Geotechnics*, 2021. 27
- ZEINELDIN, F. I.; BIRO, K.; ALGHANNAM, A. O. Influence of natural and artificial soil conditioners on water holding capacity and hydraulic conductivity of sandy soils. *International Journal of Food Science and Agriculture*, 2021. 30
- ZHANG, Y. et al. Changes in soil aggregate fractions, stability, and associated organic carbon and nitrogen in different land use types in the loess plateau, china. *Sustainability*, 2022. 25, 26
- ZHANG, Y.; SCHAAP, M. G. Weighted recalibration of the rosetta pedotransfer model with improved estimates of hydraulic parameter distributions and summary statistics (rosetta3). *Journal of Hydrology*, 2017. 35, 37, 40, 44, 47, 81, 82, 83
- ZHANG, Y.; SCHAAP, M. G. Estimation of saturated hydraulic conductivity with pedotransfer functions: A review. *Journal of Hydrology*, v. 575, p. 1011–1030, 2019. 21, 41
- ZHAO, C. et al. Using pedotransfer functions to estimate soil hydraulic conductivity in the loess plateau of china. *Catena*, Elsevier, v. 143, p. 1–6, 2016. 15, 82
- ZHOU, M. et al. Soil aggregates stability and storage of soil organic carbon respond to cropping systems on black soils of northeast china. *Scientific Reports*, 2020. 26