

UNIVERSIDADE DO ESTADO DE SANTA CATARINA – UDESC
CENTRO DE CIÊNCIAS TECNOLÓGICAS – CCT
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA CIVIL

NION MARON DRANSFELD

**APLICAÇÃO DE MÉTODOS DE APRENDIZADO DE MÁQUINA PARA
MAPEAMENTO DE SUSCETIBILIDADE À INUNDAÇÃO**

JOINVILLE
2025

NION MARON DRANSFELD

**APLICAÇÃO DE MÉTODOS DE APRENDIZADO DE MÁQUINA PARA
MAPEAMENTO DE SUSCETIBILIDADE À INUNDAÇÃO**

Dissertação apresentada como requisito parcial
para obtenção do título de mestre em
Engenharia Civil pelo Programa de Pós-
Graduação em Engenharia Civil da
Universidade do Estado de Santa Catarina –
UDESC.

Orientadora: Prof.^a Dra. Elisa Henning

**JOINVILLE
2025**

**Ficha catalográfica elaborada pelo programa de geração automática da
Biblioteca Universitária Udesc,
com os dados fornecidos pelo(a) autor(a)**

Dransfeld, Nion Maron

Aplicação de métodos de aprendizado de máquina para
mapeamento de suscetibilidade à inundação / Nion Maron
Dransfeld. -- 2025.

219 p.

Orientadora: Elisa Henning

Dissertação (mestrado) -- Universidade do Estado de Santa
Catarina, Centro de Ciências Tecnológicas, Programa de
Pós-Graduação , Joinville, 2025.

1. Mapas de suscetibilidade. 2. Inundação. 3. Floresta aleatória.
4. Máquina de vetores de suporte. I. Henning, Elisa . II.
Universidade do Estado de Santa Catarina, Centro de Ciências
Tecnológicas, Programa de Pós-Graduação . III. Título.

NION MARON DRANSFELD

**APLICAÇÃO DE MÉTODOS DE APRENDIZADO DE MÁQUINA PARA
MAPEAMENTO DE SUSCETIBILIDADE À INUNDAÇÃO**

Dissertação apresentada como requisito parcial
para obtenção do título de mestre em
Engenharia Civil pelo Programa de Pós-
Graduação em Engenharia Civil da
Universidade do Estado de Santa Catarina –
UDESC.

BANCA EXAMINADORA

Orientadora:

Prof.^a Dra. Elisa Henning
Universidade do Estado de Santa Catarina - UDESC

Membros:

Prof. Dr. Leonardo Romero Monteiro
Universidade do Estado de Santa Catarina - UDESC

Prof.^a Dra. Amanara Potykytã de Sousa Dias Vieira
Universidade Federal de Santa Catarina - UFSC

Joinville, 06 de maio de 2025

AGRADECIMENTOS

Agradeço à minha orientadora, Elisa Henning, por sua orientação, pelas horas de dedicação ao trabalho, pela paciência e pela compreensão das minhas dificuldades e limitações.

Sou grato à Universidade do Estado de Santa Catarina, pela oportunidade de participar do programa de mestrado, e aos professores, por todos os conhecimentos compartilhados.

Agradeço profundamente aos meus pais, que sempre me incentivaram na busca pelo conhecimento, e à minha família, cujo apoio constante foi essencial em todas as etapas da minha vida.

Estendo ainda minha gratidão aos amigos que, de maneira direta ou indireta, contribuíram para a realização desta pesquisa.

"Human beings - not nature - are the cause of disaster losses, which stem from choices about where and how human development will proceed." (Mileti, 1999, p. 12)

RESUMO

Este estudo investiga o desenvolvimento de aprendizado de máquina para o monitoramento de áreas suscetíveis às inundações e aplica os algoritmos floresta aleatória (FA) e máquina de vetores de suporte (MVS) para prever a suscetibilidade à inundação na bacia hidrográfica do Rio Cachoeira, localizada no município de Joinville (Santa Catarina, Brasil). Foram feitas avaliações comparativas da eficácia preditiva entre os modelos de FA e MVS, analisando seu desempenho na captura de relações não lineares entre as variáveis condicionantes identificadas na área de estudo. Foram utilizadas métricas de avaliação de classificação e técnicas de otimização hiperparamétrica, visando estabelecer um procedimento rigoroso para seleção dos melhores modelos. Para o modelo final foram integradas quatorze variáveis geoambientais, abrangendo parâmetros topográficos, geológicos, hidrológicos e meteorológicos, selecionadas mediante revisão sistemática de fatores condicionantes de inundação. A implementação computacional, desenvolvida em linguagem R, empregou pacotes como *caret*, otimização hiperparamétrica, validação cruzada *k-fold*, *raster* (processamento de dados georreferenciados) e *rgeos* (análises geoespaciais), garantindo reprodutibilidade e escalabilidade metodológica. Os modelos demonstraram desempenho elevado, com acurácias de 99,60% (FA) e 99,40% (MSV-R), atribuídas à robustez na seleção dos fatores, a geomorfologia da bacia e representatividade estatística da amostragem. A saída cartográfica gerada, além de identificar zonas críticas com alta resolução (10x10m), permite atualizações dinâmicas mediante inclusão de novos dados ambientais ou cenários climáticos, consolidando-se como ferramenta estratégica para gestão de riscos. Os resultados evidenciam a boa acurácia de técnicas de aprendizado supervisionado na assimilação de relações não lineares em sistemas naturais, reforçando seu potencial para políticas públicas de mitigação de desastres em regiões urbanas e com geomorfologias complexas.

Palavras-chave: Mapas de suscetibilidade. Inundação. Floresta aleatória. Máquina de vetores de suporte.

ABSTRACT

This study investigates the development of machine learning for monitoring flood-susceptible areas and applies the random forest (RF) and support vector machine (SVM) algorithms to predict flood susceptibility in the Cachoeira River watershed, located in the municipality of Joinville (Santa Catarina, Brazil). Comparative evaluations of the predictive efficacy between the RF and SVM models were performed, analyzing their performance in capturing non-linear relationships among the conditioning variables identified in the study area. Classification evaluation metrics and hyperparameter optimization techniques were utilized to establish a rigorous procedure for selecting the best models. For the final model, fourteen geo-environmental variables were integrated, encompassing topographic, geological, hydrological, and meteorological parameters, selected through a systematic review of flood conditioning factors. The computational implementation, developed in the R language, employed packages such as caret for hyperparameter optimization and *k*-fold cross-validation, raster for processing georeferenced data, and rgeos for geospatial analyses, ensuring methodological reproducibility and scalability. The models demonstrated elevated performance, with accuracies of 99.60% (RF) and 99.40% (SVM-R), attributed to the robustness in the selection of factors, the geomorphology of the basin, and the statistical representativeness of the sampling. The generated cartographic output, in addition to identifying critical zones with high resolution (10x10m), allows for dynamic updates through the inclusion of new environmental data or climatic scenarios, consolidating it as a strategic tool for risk management. The results show the high accuracy of supervised learning techniques in assimilating non-linear relationships in natural systems, reinforcing their potential for public policies on disaster mitigation in urban regions and with complex geomorphologies.

Keywords: Flood Susceptibility. Machine Learning Models. Support Vector Machine. Random Forest.

LISTA DE FIGURAS

Figura 1 - Representação da propensão da suscetibilidade à inundação gradual	24
Figura 2 - Linha do tempo <i>machine learning</i>	39
Figura 3 - Tipos de aprendizado de máquina	40
Figura 4 - Exemplo de matriz de confusão para área inundável	45
Figura 5 - Exemplos de curvas ROC ilustrando níveis de acurácia de modelos preditivos conforme a classificação de Swets (1988).....	47
Figura 6 - Fluxo de desenvolvimento de mapas de suscetibilidade à inundação por aprendizado de máquina	50
Figura 7 - Esquema simplificado do algoritmo floresta aleatória	53
Figura 8 - Esquema de classificação por meio do SVM	56
Figura 9 - Exemplos de funções do Kernel	56
Figura 10 - Hiperparâmetros Custo e Gama	57
Figura 11 - Diagrama da metodologia da pesquisa bibliométrica	59
Figura 12 - Mapa de localização do município de Joinville.....	61
Figura 13 - Hipsometria do Município de Joinville	62
Figura 14 - Bacias Hidrográficas de Joinville	63
Figura 15 - Localização da bacia hidrográfica do Rio Cachoeira	64
Figura 16 - Fluxograma para obtenção de mapas de suscetibilidade à inundação por aprendizado de máquina	66
Figura 17 - Esquema do processo de validação cruzada <i>K-Fold</i> aplicado ao conjunto de treinamento	68
Figura 18 - Diagrama de coleta de dados de áreas de inundações	71
Figura 19- Processamento do fator de declividade do terreno	76
Figura 20 - Sequência de processamento fator DCH	77
Figura 21 - Sequência de processamento fator DMV.....	77
Figura 22- Sequência de processamento do fator NDVI.....	78
Figura 23 - Processamento do fator rugosidade do terreno	79
Figura 24 - Processamento do fator aspecto do terreno.....	80
Figura 25 - Processamento do fator curvatura.....	81
Figura 26 - Precipitação média anual de Joinville.....	84
Figura 27 - Classificação dos níveis de suscetibilidade	85

Figura 28 - Análise das publicações por países: a) rede de colaboração, b) total de citações e c) distribuição por país	87
Figura 29 - Principais autores e artigos citados em estudos sobre suscetibilidade à inundação com aprendizado de máquina: a) autores; b) artigos	89
Figura 30 - Frequência de uso das variáveis em estudos de mapeamento de suscetibilidade à inundação (MSI)	94
Figura 31- Divisão espacial dos dados de treino e teste	108
Figura 32 - Mapas temáticos das variáveis independentes: a) MDE, b) DCH, c) DMV, d) IVDN, e) DEC, f) RUG, g) IUT, h) IPE, i) ASP, j) PREC, k) CUR_PER, l) CUR_PLAN, m) TSO, n) MARE, o) UOS	110
Figura 33 - Matriz de correlação	112
Figura 34 - Diagramas de caixa (<i>box plot</i>) de sensibilidade e especificidade dos modelos de treinamento	115
Figura 35 – Análise das importâncias das variáveis através da função ‘ <i>varImp</i> ’	117
Figura 36 - Matrizes de confusão da amostra de teste: a) FA, b) MVS-R, c) MVS-P e d) MVS-L.....	120
Figura 37 - Curva ROC dados de validação	122
Figura 38 - Mapa de suscetibilidade à inundação algoritmo floresta aleatória	123
Figura 39 - Mapa de suscetibilidade à inundação - algoritmo máquina de vetores de suporte	124
Figura 40 - Diagramas aluviais das variáveis: a) MDE, b) DEC, c) IPE e d) IUT	127
Figura 41 - Diagramas aluviais das variáveis: a) DCH, b) UOS, c) DMV e d) IVDN	129
Figura 42 - Comparação entre os mapas de suscetibilidade à inundação: a) CPRM, b) FA, c) sobreposição CPRM e FA	131

LISTA DE QUADROS

Quadro 1 - Ferramentas para estudo de mapas de suscetibilidade principais vantagens e desvantagens	34
Quadro 2 - Principais algoritmos de aprendizado de máquina supervisionado.....	41
Quadro 3 - Principais métricas de validação aplicadas a modelos de classificação por aprendizado de máquina	44
Quadro 4 - Principais hiperparâmetros do algoritmo floresta aleatória.....	54
Quadro 5 - Lista de artigos analisados na bibliometria	183

LISTA DE TABELAS

Tabela 1 - Detalhamento das fontes dos dados utilizados	74
Tabela 2 - Variáveis para modelagem de mapas de suscetibilidade à inundação encontrados na literatura.....	91
Tabela 3 - Descrição dos principais algoritmos encontrados	95
Tabela 4 - Algoritmos e programas utilizados.....	103
Tabela 5 - Distribuição aleatória das amostras (pontos).....	107
Tabela 6 - Análise da multicolinearidade	114
Tabela 7 - Comparação de áreas e níveis de suscetibilidade	125
Tabela 8 - Inventário de pontos de inundação e não inundação	196
Tabela 9 - Tabela de contingência	202

LISTA DE ABREVIATURAS E SIGLAS

AHP	<i>Analytical Hierarchical Process</i> (Análise Hierárquica de Processos)
ANA	Agência Nacional de Águas
AUC	<i>Area under the ROC Curve</i> (Área sob a curva ROC)
CEMADEN	Centro Nacional de Monitoramento e Alertas de Desastres Naturais
COBRADE	Classificação e Codificação Brasileira de Desastres
CPRM	Serviço Geológico do Brasil (Companhia de Pesquisa de Recursos Minerais)
EPSG	<i>European Petroleum Survey Group</i>
FA	Floresta Aleatória
GPT	<i>Generative Pre-Trained Transformer</i>
IA	Inteligência Artificial
IBGE	Instituto Brasileiro de Geografia e Estatística
IDW	<i>Inverse Distance Weighting</i>
INPE	Instituto Nacional de Pesquisas Espaciais
IPPUJ	Instituto de Pesquisa e Planejamento Urbano de Joinville
LAT	Latitude
LiDAR	<i>Light Detection and Ranging</i>
LON	Longitude
MDE	Modelo digital de elevação
ML	<i>Machine Learning</i> (Aprendizado de máquina)
MSI	Mapa de suscetibilidade à inundação
MVS	Máquina de Vetores de Suporte
netCDF	<i>Network Common Data Form</i>
NDVI	<i>Normalized Difference Vegetation Index</i> (Índice de vegetação da diferença normalizada)
PIB	Produto interno bruto
PMJ	Prefeitura Municipal de Joinville
RF	<i>Random forest</i> (Floresta aleatória)
ROC	<i>Receiver Operating Characteristic</i> (Curva característica de operação)
SAGA	<i>System for Automated Geoscientific Analyses</i>

SEPUD	Secretaria de Planejamento Urbano e Desenvolvimento Sustentável (Joinville)
SIG	Sistema de Informações Geográficas
SIMGeo	Sistema de informações municipais georreferenciadas (Joinville)
SPI	<i>Stream power index</i> (Índice de potência do fluxo)
SRC	Sistemas de Referência de Coordenadas
SVM	<i>Support Vector Machine</i> (Máquina de vetores de suporte)
VIF	<i>Variance inflation factor</i> (fator de inflação da variância)
TWI	<i>Topographic Wetness Index</i> (Índice de Saturação do Solo)
UFSC	Universidade Federal de Santa Catarina
IVDN	Índice de Vegetação por Diferença Normalizada

SUMÁRIO

1.	INTRODUÇÃO.....	18
1.1	JUSTIFICATIVA.....	18
1.2	OBJETIVOS.....	20
1.2.1	Objetivo Geral.....	20
1.2.2	Objetivos Específicos	20
1.3	SELEÇÃO DA ÁREA DE ESTUDO	20
1.4	ESTRUTURA DO TRABALHO	21
2.	REFERENCIAL TEÓRICO.....	23
2.1	SUSCETIBILIDADE À INUNDAÇÃO	23
2.1.1	Conceitos e definições	23
2.1.2	Fatores geoespaciais de influência	25
<i>2.1.2.1</i>	<i>Fatores topográficos</i>	<i>26</i>
<i>2.1.2.1.1</i>	<i>Elevação</i>	<i>26</i>
<i>2.1.2.1.2</i>	<i>Declividade</i>	<i>27</i>
<i>2.1.2.1.3</i>	<i>Orientação da vertente</i>	<i>27</i>
<i>2.1.2.1.4</i>	<i>Índice de umidade topográfica</i>	<i>27</i>
<i>2.1.2.1.5</i>	<i>Índice de potência de fluxo</i>	<i>28</i>
<i>2.1.2.1.6</i>	<i>Curvatura.....</i>	<i>28</i>
<i>2.1.2.1.7</i>	<i>Índice de posição topográfica</i>	<i>28</i>
<i>2.1.2.2</i>	<i>Fatores hidrológicos e ambientais.....</i>	<i>28</i>
<i>2.1.2.2.1</i>	<i>Precipitação pluviométrica.....</i>	<i>29</i>
<i>2.1.2.2.2</i>	<i>Distância à rede de drenagem</i>	<i>29</i>
<i>2.1.2.2.3</i>	<i>Litologia.....</i>	<i>30</i>
<i>2.1.2.2.4</i>	<i>Índice de vegetação por diferença normalizada</i>	<i>30</i>
<i>2.1.2.2.5</i>	<i>Grupos hidrológicos do solo</i>	<i>30</i>
<i>2.1.2.3</i>	<i>Fatores antrópicos</i>	<i>31</i>
<i>2.1.2.3.1</i>	<i>Uso e ocupação do solo.....</i>	<i>31</i>
<i>2.1.2.3.2</i>	<i>Distância a vias de transporte.....</i>	<i>32</i>
2.1.3	Instrumentos para mapeamento.....	32
<i>2.1.3.1</i>	<i>Plataformas de análise espacial</i>	<i>35</i>
<i>2.1.3.2</i>	<i>Métodos preditivos</i>	<i>36</i>
2.2	FUNDAMENTOS DE APRENDIZADO DE MÁQUINA	37

2.2.1	Tipos de aprendizado de máquina.....	39
2.2.2	Hiperparâmetros	42
2.2.3	Métricas de validação	43
2.2.3.1	<i>Matriz de confusão.....</i>	45
2.2.3.2	<i>Curva de característica de operação do receptor</i>	46
2.3	INTEGRAÇÃO DE TÉCNICAS DE APRENDIZADO DE MÁQUINA PARA MAPEAMENTOS DE SUSCETIBILIDADE À INUNDAÇÃO	48
2.3.1	Fluxo de desenvolvimento de modelos	49
2.3.2	Algoritmos de aprendizado de máquina	52
2.3.2.1	<i>Floresta Aleatória.....</i>	53
2.3.2.2	<i>Máquina de vetores de suporte</i>	55
3.	METODOLOGIA	59
3.1	ANÁLISE BIBLIOMÉTRICA	59
3.2	MODELAGEM DE MAPA DE SUSCETIBILIDADE À INUNDAÇÃO	60
3.2.1	Área de estudo	60
3.2.1.1	<i>Área de estudo: bacia hidrográfica do Rio Cachoeira.....</i>	63
3.2.2	Fluxograma da modelagem	64
3.2.3	Pré-processamento e empilhamento de dados.....	66
3.2.4	Multicolinearidade	67
3.2.5	Validação cruzada e <i>tuning</i>	68
3.2.6	Avaliação e validação dos modelos	69
3.2.7	Coleta da variável dependente	70
3.2.7.1	<i>Inventário de áreas de inundação.....</i>	70
3.2.7.2	<i>Inventário de áreas de não inundação.....</i>	72
3.2.8	Coleta das variáveis independentes	74
3.2.8.1	<i>Modelo digital de elevação</i>	75
3.2.8.2	<i>Declividade</i>	76
3.2.8.3	<i>Distância de corpos hídricos</i>	76
3.2.8.4	<i>Distanciamento da malha viária.....</i>	77
3.2.8.5	<i>Índice de vegetação por diferença normalizada.....</i>	78
3.2.8.6	<i>Índice de rugosidade do terreno</i>	78
3.2.8.7	<i>Índice de umidade topográfico</i>	79
3.2.8.8	<i>Índice de potência de escoamento</i>	79
3.2.8.9	<i>Aspecto</i>	80

3.2.8.10	<i>Curvatura</i>	81
3.2.8.11	<i>Tipo de solo</i>	81
3.2.8.12	<i>Área de influência da maré</i>	82
3.2.8.13	<i>Uso e ocupação do solo</i>	82
3.2.8.14	<i>Precipitação</i>	83
3.2.9	Níveis de suscetibilidade	84
3.2.10	Bibliotecas utilizadas para manipulação dos dados	85
4.	RESULTADOS E DISCUSSÃO DA ANÁLISE BIBLIOMÉTRICA	86
4.1	PUBLICAÇÕES POR PAÍSES	86
4.2	PRINCIPAIS AUTORES E PUBLICAÇÕES	88
4.3	PRINCIPAIS FATORES	90
4.4	PRINCIPAIS ALGORITMOS E PROGRAMAS	95
5.	RESULTADOS E DISCUSSÃO DOS MAPAS DE SUSCETIBILIDADE	106
5.1	INVENTÁRIO DE INUNDAÇÃO E NÃO INUNDAÇÃO	106
5.2	MAPAS TEMÁTICOS	109
5.3	RESULTADO DA CORRELAÇÃO	112
5.4	RESULTADO DA MULTICOLINEARIDADE	113
5.5	PROCESSAMENTO DOS MODELOS	114
5.6	RESULTADOS DOS MODELOS DE TREINAMENTO	115
5.7	HIPERPARÂMETROS DOS MODELOS FINAIS	116
5.8	ANÁLISE DA IMPORTÂNCIA DAS VARIÁVEIS	117
5.9	RESULTADOS DA AMOSTRA DE TESTE	120
5.10	MAPAS DE SUSCETIBILIDADE À INUNDAÇÃO	122
5.11	ANÁLISE DA SUSCETIBILIDADE À INUNDAÇÃO EM FUNÇÃO DAS VARIÁVEIS	126
5.12	COMPARAÇÃO COM OUTROS ESTUDOS	130
5.13	LIMITAÇÕES	133
5.13.1	Avaliação da quantidade adequada de fatores	133
5.13.2	Limitações decorrentes das fontes de dados e sua qualidade	134
5.13.3	Limitações da base de dados do inventário de inundações	135
5.13.4	Limitações da base de dados do inventário de não inundações	137
5.13.5	Limitações metodológicas dos modelos de aprendizado de máquina	138
5.13.6	Limitações quanto ao tipo de inundação modelada	139
6.	CONCLUSÃO	141

REFERÊNCIAS	145
APÊNDICE A – LISTA DE ARTIGOS ANALISADOS NA BIBLIOMETRIA.....	183
APÊNDICE B – REGISTROS E LOCALIZAÇÃO DO INVENTÁRIO DE PONTOS DE INUNDAÇÃO E NÃO INUNDAÇÃO	196
APÊNDICE C – TABELAS CRUZADAS DE SUSCETIBILIDADE E FATORES .	202
APÊNDICE D – CÓDIGO DE PROCESSAMENTO DO MODELO	212

1. INTRODUÇÃO

Devido ao crescimento da população e intervenções antrópicas ocorridas nos últimos anos, observa-se uma maior importância dada pela sociedade ao gerenciamento de desastres (Jha; Lamond, 2011). De acordo com o banco de dados de eventos de emergência (EM-DAT, 2025), no período de 1900 a 2025, as inundações representaram o tipo de desastre mais frequente, causando danos econômicos, sociais e ambientais nas últimas décadas. Portanto, identificar áreas suscetíveis às inundações é essencial para o planejamento urbano e para mitigar os impactos nessas áreas.

Nos últimos 20 anos, o Brasil esteve entre os dez países com maior número de registros de inundações, segundo o EM-DAT (2025). Devido à recorrência de desastres naturais no território brasileiro, o inciso I do art. 7º da Lei Federal nº 12.608, de 10 de abril de 2012, determina que compete aos Estados identificar e mapear áreas de risco, além de realizar estudos de identificação de ameaças, suscetibilidades e vulnerabilidades, em articulação com a União e os Municípios. Adicionalmente, nos incisos I e IV do parágrafo 2º do artigo 3º-A da mesma lei, o Governo Federal estabelece que os municípios inscritos no cadastro nacional de municípios com áreas suscetíveis a deslizamentos de grande impacto, inundações bruscas ou processos geológicos ou hidrológicos correlatos a elaborar mapeamentos dessas áreas suscetíveis.

Segundo Capraria (2017), os principais instrumentos presentes na literatura para auxiliar ou gerar modelos para identificação de áreas de alagamentos e inundações são referentes à modelagem hidrológica, modelagem hidráulica, modelo de análise hierárquica de processos (AHP), sistema de informações geográficas (SIG) e sensoriamento remoto. Porém, nos últimos anos com a popularização da computação, métodos de aprendizado de máquina (*Machine Learning* em inglês) vêm ganhando destaque como instrumento para elaboração de análises espaciais para reconhecer e explorar padrões de um grande volume de dados, tendo aplicações diretas na elaboração de mapas de suscetibilidades às inundações (Tehrany *et al.*, 2013; Lee *et al.*, 2017; Zhao *et al.*, 2019).

1.1 JUSTIFICATIVA

Os eventos e impactos gerados pelas inundações têm se intensificado em Joinville/SC (Silveira *et al.*, 2008; Smagalla *et al.*, 2020). Uma previsão futura para a bacia hidrográfica do rio Cachoeira, que possui aproximadamente 82 km², realizada pelo PDDU (2011), alerta para

um aumento da área inundável de uma média atual de 16,17 km² para 20,00 km², representando cerca de 25% da área total. Esse aumento nos eventos de inundação é atribuído ao crescimento populacional, aumento do nível da maré e à expansão urbana em áreas de risco. Nesse contexto, a análise espacial oferece uma maneira eficaz de identificar áreas de suscetibilidade, auxiliar no planejamento urbano e implementar medidas preventivas, contribuindo assim para a redução e prevenção de desastres (Nachappa, *et al.*, 2020; Vojtek, *et al.*, 2021).

Em geral, os métodos mais comuns usados para simular e prever a distribuição espacial das inundações são os modelos hidrodinâmicos e hidrológicos, na qual simulam a extensão da zona de inundação e a profundidade da água na região inundada (Tanga *et al.*, 2020). No entanto, esses modelos requerem dados de entrada complexos e que são difíceis de se obter, o que limita sua aplicação (Chen *et al.*, 2012).

De acordo com Lamovec (2013), as técnicas de aprendizado de máquina aceleram o mapeamento de inundações, reduzindo o tempo necessário para analisar e identificar áreas de risco. Além disso, uma série de estudos conduzidos por Tehrany *et al.* (2013, 2018, 2020) reforçam a eficácia desses métodos, demonstrando que diversas abordagens de aprendizado de máquina são capazes de determinar com precisão áreas suscetíveis à inundação.

Existem diversos algoritmos de aprendizado de máquina que podem ser testados para resolver problemas espaciais, cuja eficiência varia conforme a natureza do problema, a disponibilidade de dados e as características da região. Entre os algoritmos mais populares para o mapeamento da suscetibilidade à inundação estão a floresta aleatória, redes neurais artificiais e máquinas de vetores de suporte (Falah *et al.*, 2019; Costache *et al.*, 2020; Vafakhah *et al.*, 2020). Entre esses métodos, o algoritmo MVS vem ganhando crescente reconhecimento para problemas de inundação (Zhao *et al.*, 2018).

Além disso, a aplicação do aprendizado de máquina na criação de mapas de suscetibilidade à inundação é frequentemente limitada a regiões específicas, ressaltando a necessidade de ampliar estudos que avaliem a precisão desses métodos. Esse aprofundamento auxilia na análise da confiabilidade dos modelos e pode estimular seu uso. Neste contexto, o presente estudo é pioneiro ao explorar novas abordagens para o mapeamento de suscetibilidade na bacia hidrográfica do Rio Cachoeira, em Joinville/SC, utilizando técnicas de aprendizado de máquina. A pesquisa desenvolverá mapas de suscetibilidade, comparando-os com os mapas de manchas de inundações existentes e avaliando seu desempenho. Essa análise fornecerá uma contribuição significativa para a melhoria das estratégias de gestão de riscos de desastres naturais na área de estudo.

1.2 OBJETIVOS

1.2.1 Objetivo Geral

O objetivo principal desta pesquisa é modelar espacialmente a suscetibilidade à inundação na bacia hidrográfica do Rio Cachoeira (BHRC), localizada no perímetro urbano de Joinville (Santa Catarina, Brasil), por meio da integração de algoritmos de aprendizado de máquina (AM).

1.2.2 Objetivos Específicos

Para atingir o objetivo geral foram definidos os seguintes objetivos específicos:

- a) realizar uma revisão bibliométrica para identificar os principais autores, contribuições, tendências e abordagens relacionadas à aplicação de aprendizado de máquina na modelagem de suscetibilidade à inundação;
- b) levantar as principais variáveis e algoritmos de aprendizado de máquina utilizados na elaboração de mapas de suscetibilidade à inundação, com base na literatura técnica e científica;
- c) modelar, testar e comparar algoritmos de aprendizado de máquina, exclusivamente os algoritmos floresta aleatória e máquina de vetores de suporte, para avaliar seu desempenho na predição de áreas suscetíveis à inundação na BHRC;
- d) identificar as variáveis mais relevantes para a área de estudo, analisando os graus de importância atribuídos pelos algoritmos de aprendizado de máquina nos modelos com melhor desempenho;
- e) gerar um mapa de suscetibilidade à inundação para toda a área de estudo, classificando as regiões em cinco níveis de suscetibilidade, com o objetivo de representar espacialmente os diferentes graus de suscetibilidade.

1.3 SELEÇÃO DA ÁREA DE ESTUDO

A seleção da bacia hidrográfica do Rio Cachoeira (BHRC) em Joinville/SC como área de estudo para o mapeamento de suscetibilidade à inundação é justificada por vários fatores críticos. Joinville é o maior município de Santa Catarina em termos populacionais, com aproximadamente 661.600 habitantes (IBGE, 2022), sendo um importante polo econômico que

enfrenta significativos desafios devido às frequentes inundações (Silveira *et al.*, 2008; PDDU, 2011). A área em estudo (BHRC), localiza-se no centro urbano de Joinville. Situada a apenas dois metros acima do nível do mar, esta região é extremamente suscetível a eventos de inundações, tanto fluviais quanto marítimas. As condições são intensificadas por fatores como marés astronômicas e tempestades, além do impacto da urbanização intensa, conforme documentado por Silveira *et al.* (2008). Essa confluência de fatores faz da bacia do Rio Cachoeira um cenário crítico para verificar os níveis de suscetibilidades à inundação.

1.4 ESTRUTURA DO TRABALHO

O presente trabalho está organizado em seis capítulos principais, seguindo uma estrutura lógica que integra fundamentação teórica, metodologia, análise de resultados e conclusões. Cada capítulo foi delineado para abordar aspectos específicos do mapeamento de suscetibilidade à inundação por meio de técnicas de aprendizado de máquina, garantindo coerência e progressão sistemática do conhecimento. A seguir, descreve-se a estrutura do trabalho por capítulos.

No primeiro capítulo apresenta-se a introdução, abordando a contextualização do problema estudado, incluindo uma análise sobre a relevância das inundações como um problema socioambiental significativo no Brasil e no mundo. São apresentados ainda a justificativa para a realização do estudo, os objetivos gerais e específicos, detalhando o propósito principal da pesquisa e as metas específicas a serem atingidas.

No segundo capítulo discute-se o referencial teórico que fundamenta este trabalho. São abordados conceitos relacionados à suscetibilidade à inundação, destacando suas características e fatores condicionantes, agrupados em fatores topográficos, hidrológicos, ambientais e antrópicos. O capítulo também investiga as bases do aprendizado de máquina, delineando os paradigmas de aprendizagem supervisionada, a otimização de hiperparâmetros, validação cruzada, e métricas de validação de modelos, como matriz de confusão, e curva ROC para avaliação de desempenho preditivo. Por fim, articula-se a integração desses elementos metodológicos, com destaque para os algoritmos floresta aleatória (FA) e máquinas de vetores de suporte (MVS), em fluxos de trabalho geoespaciais, visando à geração de mapas probabilísticos de suscetibilidade.

No terceiro capítulo apresenta-se a metodologia adotada neste estudo. Inicialmente é descrita a análise bibliométrica conduzida para contextualizar a temática estudada no cenário científico atual. Em seguida, detalha-se a metodologia utilizada para a modelagem do mapa de

suscetibilidade à inundação, incluindo a seleção e caracterização da área de estudo (BHRC). Também é apresentado um fluxograma que sintetiza o processo metodológico proposto, incluindo a definição das variáveis independentes e dependentes, procedimentos para validação cruzada, *tuning* dos algoritmos e avaliação da qualidade dos modelos gerados. Neste capítulo, ainda são detalhados os fatores utilizados na análise, assim como as bibliotecas empregadas na manipulação dos dados.

No quarto e quinto capítulo são apresentados os resultados da pesquisa e sua discussão detalhada. Inicialmente é discutida a análise bibliométrica realizada, destacando as publicações por países, principais fatores condicionantes identificados e os algoritmos mais utilizados para estudos semelhantes. Na sequência, são detalhados os resultados obtidos através da modelagem de aprendizado de máquina, incluindo os inventários das áreas de inundação e não inundação, mapas temáticos produzidos, análises de desempenho dos algoritmos empregados (FA e MSV), bem como os resultados obtidos nas etapas de treinamento, validação e avaliação dos modelos. São também apresentados os mapas finais de suscetibilidade à inundação, seguidos de uma comparação crítica com resultados encontrados na literatura especializada.

Por fim, o sexto capítulo apresenta as considerações finais, destacando as conclusões principais obtidas no estudo, limitações identificadas e recomendações para futuras pesquisas.

2. REFERENCIAL TEÓRICO

Este capítulo estabelece os fundamentos teóricos que subsidiam a análise da suscetibilidade à inundação mediante técnicas de aprendizado de máquina (AM), contextualizando sua evolução histórica e aplicabilidade na gestão de riscos naturais.

2.1 SUSCETIBILIDADE À INUNDAÇÃO

No âmbito da análise de riscos naturais, a suscetibilidade à inundação pode ser entendida como a predisposição ou propensão de um ambiente físico a sofrer os efeitos de processos naturais, como inundações, independentemente da presença humana direta (CPRM, 2014; Lima e Amorim, 2023). Essa suscetibilidade do meio físico diante de certos fenômenos é influenciada por fatores como relevo, uso do solo, impermeabilização e declividade, entre outros (CPRM, 2014; Houdijk 2012).

A suscetibilidade distingue-se conceitualmente de outros elementos centrais da análise de risco, como perigo, exposição e vulnerabilidade, por focar exclusivamente em condições intrínsecas do ambiente, sem incorporar fatores dinâmicos como a probabilidade de ocorrência dos eventos, o período de retorno ou a intensidade dos efeitos causados (Corominas *et al.*, 2014; Domínguez-Cuesta, 2013).

2.1.1 Conceitos e definições

Para delimitar precisamente o conceito de suscetibilidade em análises espaciais, é importante distinguir termos correlatos, como risco, perigo, vulnerabilidade e exposição. Na literatura, o risco (*risk* em inglês) é definido como produto de um perigo (*hazard* em inglês) e de suas consequências (Kron, 2005; Houdijk, 2012; Macedo e Bressani, 2015). Regiões na qual não existem pessoas ou valores que podem ser afetados por um fenômeno natural, não há risco (Kron, 2005). Conforme descrito por Olimpio e Zanella (2017, p. 105) “se o risco é uma situação percebida enquanto probabilidade de perdas, o desastre refere-se à própria materialização desta condição”.

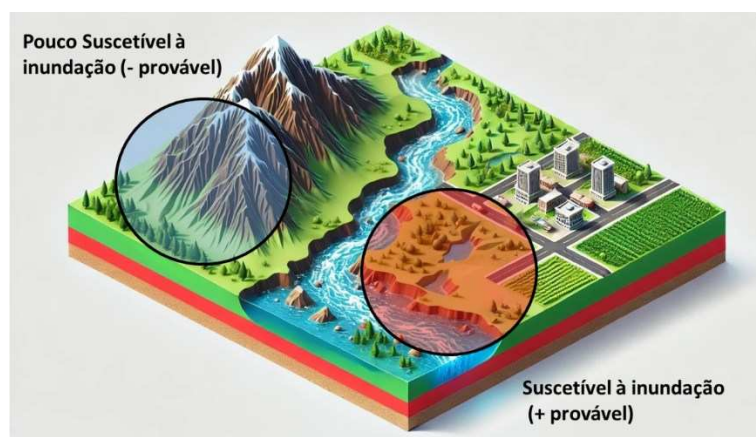
Segundo Bressani e Costa (2015, p. 17) o perigo é “uma condição com potencial para causar consequência indesejável, ou a probabilidade que uma ameaça específica (portanto, com danos) ocorrer em um dado período de tempo”.

Houdijk (2012) considera a vulnerabilidade (*Vulnerability*) como um conceito composto, que consiste em exposição (*exposure* em inglês) e suscetibilidade (*susceptibility* em inglês), na qual a exposição são os valores em risco, como as perdas materiais ou humanas. Já a suscetibilidade pode ser definida como a propensão ao desenvolvimento de um fenômeno ou processo em uma dada área (CPRM, 2014).

Nas abordagens do CPRM (2014) e Houdijk (2012), um mapa de suscetibilidade à inundação (MSI) pode ser definido como uma representação espacial das áreas propensas ao desenvolvimento de processos de inundação. Esses mapas identificam regiões que, devido a características específicas, são mais suscetíveis às inundações. Importante ressaltar que um mapa de suscetibilidade à inundação (MSI) não inclui análises de perigo, risco ou vulnerabilidade, mas sim podem complementar essas análises (Domínguez-Cuesta, 2013; Houdijk, 2012).

Com base na Figura 1, que representa de forma genérica áreas com maior ou menor suscetibilidade à inundação, observa-se que regiões montanhosas ou de maior altitude apresentam menor propensão a eventos de inundação gradual devido à redução da acumulação de água e ao aumento do escoamento superficial. Por outro lado, áreas localizadas nas proximidades de corpos hídricos ou com gradientes topográficos suaves tendem a exibir maior suscetibilidade, uma vez que a baixa declividade favorece a estagnação de água, enquanto a adjacência a rios amplia a exposição a transbordamentos e alagamentos. Essa distribuição espacial reflete a influência direta de fatores geomorfológicos e hidrológicos na dinâmica de inundações, conforme evidenciado em estudos de suscetibilidade (Kazakis *et al.*, 2015; Choubin *et al.*, 2019; Costache *et al.*, 2020).

Figura 1 - Representação da propensão da suscetibilidade à inundação gradual



Fonte: Figura gerada por inteligência artificial (modelo GPT-4o) e adaptada pelo autor (2024).

Para este estudo, a suscetibilidade à inundação será definida como a propensão de uma região ser afetada por eventos de inundação, determinada pela interação de variáveis naturais e antrópicas (Fang *et al.*, 2020). Fatores como altitude, declividade, distância de redes de drenagem, padrões pluviométricos e alterações no uso do solo desempenham papel crítico nessa avaliação (Alia *et al.*, 2020). Por exemplo, áreas planas e de baixa altitude, especialmente em zonas ripárias, são mais vulneráveis devido à menor capacidade de infiltração e à convergência de fluxos hídricos (Pham *et al.*, 2021). Adicionalmente, intervenções humanas, como impermeabilização do solo e desmatamento, podem exacerbar riscos ao modificar processos hidrológicos naturais (Vojtek, *et al.*, 2021). Estudos desenvolvidos por Tehrany *et al.* (2013) e Kazakis *et al.* (2015) evidenciam a necessidade de integrar abordagens espaciais e análises multivariadas para o mapeamento preciso da suscetibilidade à inundação, ao mesmo tempo que destacam a complexidade inerente à modelagem desse fenômeno.

2.1.2 Fatores geoespaciais de influência

A suscetibilidade à inundação é uma função complexa de múltiplos fatores inter-relacionados, que podem ser agrupados em categorias topográficas (elevação, declividade, orientação da vertente, índice de umidade topográfica, índice de potência de fluxo, etc), hidrológicas (precipitação, distância à rede de drenagem, etc.), ambientais (litologia, cobertura vegetal, índice de vegetação por diferença normalizada, etc.) e antrópicas (proximidade às vias de comunicação, uso e ocupação do solo, etc.) (Nachappa *et al.*, 2020; Vafakhah *et al.*, 2020). A seleção criteriosa destes fatores condicionantes é crucial para a elaboração de mapas de suscetibilidade à inundação (MSI), garantindo maior acurácia nos resultados (Costache *et al.*, 2020). A ponderação relativa de cada fator é dependente das características intrínsecas da área da bacia hidrográfica (Tehrany *et al.*, 2013). Variáveis como distância a cursos d'água, precipitação acumulada, declividade, pedologia, índice de umidade topográfico, curvatura da vertente, uso e ocupação do solo e litologia são frequentemente incorporadas em análises de suscetibilidade (Costache, 2019).

A elucidação dos determinantes intrínsecos e extrínsecos da propensão dos eventos de inundação constitui um desafio metodológico para a formulação de estratégias de mitigação de riscos baseadas em evidências e para a otimização de políticas de ordenamento territorial em zonas de vulnerabilidade. Mediante uma análise crítica sistemática da literatura especializada, são sistematizados, a seguir, os principais fatores condicionantes que, conforme demonstram

estudos multidisciplinares (hidrologia, geografia e física), exercem influência estatisticamente significativa na formação e intensificação de episódios de inundação.

2.1.2.1 Fatores topográficos

A topografia de uma área, representada por fatores como elevação, declive e curvatura, desempenha um papel crucial na determinação da suscetibilidade à inundação. Estes fatores influenciam a forma como a água se move e se acumula numa determinada região, tornando algumas áreas mais propensas às inundações do que outras (Zhao *et al.*, 2019; Costache; Bui, 2019).

A relação entre a elevação e a suscetibilidade à inundação é complexa e heterogênea. Em áreas de baixa altitude, o escoamento superficial tende a concentrar-se em superfícies planas adjacentes aos rios, tornando estas áreas mais vulneráveis à inundação gradual (Kazakis *et al.*, 2015). Além disso, áreas com baixa elevação geralmente têm lençóis freáticos mais altos, o que as torna mais propensas às inundações (Pham *et al.*, 2021). A topografia também influencia a distribuição da vegetação, que por sua vez afeta a suscetibilidade à inundação (Nachappa *et al.*, 2020). Abaixo os principais fatores relacionados à topografia em estudos relacionados a suscetibilidade.

2.1.2.1.1 Elevação

A elevação topográfica, quantificada como a altitude absoluta de coordenadas geodésicas em relação ao plano de referência vertical adotado, constitui uma variável morfométrica crítica para a parametrização de modelos preditivos em geociências (Janizadeh *et al.*, 2019). A obtenção desses dados, realizada por meio de modelos digitais de elevação (MDE's) de alta resolução espacial, permite a espacialização de variáveis hidrológicas intrínsecas, como gradientes de declividade, índices de convergência de fluxo e redes de drenagem, essenciais para a calibração de sistemas de simulação hidrodinâmica (Pham *et al.*, 2021).

Em contextos hidrológicos, MDE's subsidiam a implementação de algoritmos de direcionamento de fluxo (D8, D- ∞) e a estimativa de parâmetros-chave para equações de Manning e modelos de onda cinemática, fundamentais na previsão de tempos de concentração e vazões máximas (O'Callaghan e Mark, 1984). Na hidráulica computacional, a topografia de alta precisão, obtida via LIDAR ou interferometria SAR, viabiliza a modelagem de cenários

extremos por meio de soluções numéricas de equações de Saint-Venant 2D, permitindo a simulação de padrões de inundação com resolução malha (Sander, 2007; Hartnett e Nash, 2017).

2.1.2.1.2 Declividade

A declividade representa a inclinação do terreno, sendo calculada com base em MDEs e expressa em graus ou em porcentagem (Pham *et al.*, 2021). Terrenos com declives suaves e planos tendem a concentrar o escoamento superficial, o que pode elevar o risco de inundações graduais. Em contrapartida, áreas com maior declividade favorecem o aumento da velocidade do escoamento superficial, reduzindo o tempo de concentração da água e propiciando inundações bruscas (Pham *et al.*, 2021; Liu *et al.*, 2021).

2.1.2.1.3 Orientação da vertente

A orientação da vertente (aspecto) indica a direção para a qual uma encosta está voltada (norte, sul, leste, oeste, etc.). Derivada do MDE, essa variável influencia a quantidade de radiação solar recebida, o que afeta a evapotranspiração, a umidade do solo e, consequentemente, a magnitude do escoamento superficial (Vafakhah *et al.*, 2020; Pham *et al.*, 2021). Apesar desse fator afetar às inundações apenas indiretamente, muitos pesquisadores o incluem como um dos fatores condicionantes no mapeamento da suscetibilidade à inundação (Sachdeva *et al.*, 2017; Bui *et al.*, 2019; Choubin *et al.*, 2019; Mosavi *et al.*, 2020).

2.1.2.1.4 Índice de umidade topográfica

O índice de umidade topográfica (em inglês, *Topographic Wetness Index* - TWI) é um indicador do potencial de acúmulo de água em um determinado ponto da paisagem, sendo calculado com base na área de contribuição a montante (área que drena para um determinado ponto) e na declividade local (Beven *et al.*, 1991; Sørensen *et al.*, 2006). Valores elevados de TWI indicam maior probabilidade de saturação do solo e, portanto, maior suscetibilidade à inundação (Liu *et al.*, 2021).

2.1.2.1.5 Índice de potência de fluxo

O índice de potência de fluxo (em inglês, *Stream Power Index* - SPI) quantifica o potencial erosivo do fluxo de água concentrado. É calculado utilizando a área de contribuição e a declividade (Jain e Kothyari, 2000). Valores altos de SPI indicam maior capacidade de erosão e transporte de sedimentos, o que pode agravar os processos de inundação e assoreamento. Já um SPI baixo indica que o rio tem menos força para erodir e transportar sedimentos, o que pode levar ao acúmulo desses materiais no próprio leito, assoreando o rio e também contribuindo para inundações (Turoğlu; Dölek, 2011; Pham *et al.*, 2021).

2.1.2.1.6 Curvatura

A curvatura do terreno, derivada do modelo digital de elevação (MDE), quantifica a forma da superfície terrestre, classificando-a como côncava, convexa ou plana. Essa característica é fundamental para entender processos hidrológicos e erosivos (Wilson e Gallant, 2000). Superfícies côncavas tendem a concentrar o fluxo (aumentando a suscetibilidade à inundação), enquanto superfícies convexas tendem a dispersá-lo (Khosravia *et al.*, 2019).

2.1.2.1.7 Índice de posição topográfica

O índice de posição topográfica (em inglês, *Topographic Position Index* - TPI) classifica o relevo em diferentes formas (cristas, encostas, vales), comparando a elevação de um ponto com a elevação média de sua vizinhança. Essa classificação ajuda a identificar áreas mais propensas à inundação gradual (vales) ou a deslizamentos e enxurradas (encostas íngremes) (Weiss, 2001).

2.1.2.2 Fatores hidrológicos e ambientais

A susceptibilidade à inundação é influenciada por diversos fatores, sendo os fatores hidrológicos um dos componentes importantes a serem considerados. Estes fatores estão relacionados com as características da água e do terreno, e o seu estudo é fundamental para a compreensão e modelagem do risco de ocorrência de inundações (Almeida, 2015). Vários estudos têm demonstrado a importância dos fatores hidrológicos na modelagem da susceptibilidade à inundação (Tehrany *et al.*, 2013; Almeida, 2015; Khosravia *et al.*, 2019). A

seleção de fatores hidrológicos relevantes e a sua correta ponderação são passos essenciais para a criação de mapas de suscetibilidade à inundação precisos (Zhao *et al.*, 2018; Alia *et al.*, 2020; Dodangeh *et al.*, 2020; Islam *et al.*, 2021).

A dinâmica das inundações também é governada por uma complexa interação de fatores ambientais. A litologia influencia a permeabilidade do substrato, onde formações porosas facilitam a infiltração, em contraste com litologias compactas, que promovem o escoamento superficial. As propriedades edáficas, incluindo textura, estrutura e granulometria, modulam as taxas de infiltração e a capacidade de retenção hídrica (Çelik *et al.*, 2012). A lista a seguir apresenta os principais fatores hidrológicos e ambientais que, com base em estudos recentes, demonstraram eficácia na previsão da suscetibilidade à inundação.

2.1.2.2.1 Precipitação pluviométrica

A precipitação, medida em milímetros (mm), representa a quantidade de chuva que ocorre em uma determinada área e período. Dados de precipitação podem ser obtidos de estações meteorológicas (pluviômetros), radares meteorológicos ou modelos de previsão do tempo (Zhao *et al.*, 2018). Eventos de precipitação intensa ou prolongada podem exceder a capacidade de infiltração do solo e a capacidade de vazão dos sistemas de drenagem, resultando em inundações (Vafakhah *et al.*, 2018; Costache; 2019; Khosravia *et al.*, 2019).

2.1.2.2.2 Distância à rede de drenagem

A suscetibilidade à inundação é fortemente influenciada pela proximidade à rede de drenagem (Nachappa *et al.*, 2020). Áreas situadas próximas aos rios, córregos e outros corpos d'água estão mais expostas às inundações devido ao transbordamento. Essa proximidade é frequentemente analisada através da criação de zonas de influência (*buffers*) ao redor dos cursos d'água, definindo áreas de risco decrescente com o aumento da distância (Costache *et al.*, 2020; Vafakhah *et al.*, 2020). Além disso, a capacidade da bacia hidrográfica em remover o excesso de água, representada pela densidade de drenagem e pela densidade da rede fluvial, influencia a rapidez e magnitude da resposta aos eventos de precipitações (Yaseen *et al.*, 2022). Altas densidades indicam um sistema potencialmente mais eficiente, mas essa eficiência pode ser superada por eventos extremos, levando às inundações mesmo em áreas com boa capacidade de drenagem (Vafakhah *et al.*, 2020; Pham *et al.*, 2021). A combinação destes fatores

(proximidade e densidade) permite uma avaliação mais precisa do mapeamento de inundação (Nachappa *et al.*, 2020; Liu *et al.*, 2021).

2.1.2.2.3 Litologia

A análise litológica, baseada em mapas geológicos, é crucial para entender o risco de inundações, pois a permeabilidade e resistência das rochas determinam a capacidade de infiltração da água (Shafapour *et al.*, 2013). Materiais permeáveis reduzem o escoamento e o risco, enquanto sedimentos inconsolidados (areias, argilas) em planícies de inundação, deltas e zonas costeiras, por serem menos resistentes e mais propensos à erosão, intensificam a ameaça de inundações (Smith e Ward, 1998; Rahman *et al.*, 2019; Pham *et al.*, 2021).

2.1.2.2.4 Índice de vegetação por diferença normalizada

Originalmente proposto por Rouse *et al.* (1974), o índice de vegetação por diferença normalizada (IVDN), ou NDVI (do inglês *Normalized Difference Vegetation Index*), consolidou-se como um parâmetro biofísico amplamente utilizado para a mensuração da densidade, distribuição e vigor da cobertura vegetal. Obtido por meio de dados de sensoriamento remoto multiespectral, o IVDN, que varia em uma escala de -1 a +1, fundamenta-se na propriedade espectral da vegetação saudável de apresentar alta reflectância na banda do infravermelho próximo e elevada absorção na banda do vermelho (Tehrany *et al.*, 2013; Nachappaa *et al.*, 2020; Yaseen *et al.*, 2022).

A relação entre IVDN e inundações podem variar dependendo do contexto (Khosravi *et al.*, 2019; Liu *et al.*, 2021; Yaseen *et al.*, 2022). Em algumas situações, um alto IVDN pode indicar menor risco de inundações, pois a vegetação intercepta a chuva e aumenta a infiltração da água no solo. No entanto, em regiões na qual a intensidade da precipitação excede a capacidade de interceptação e absorção da vegetação, o efeito mitigador proporcionado pela cobertura vegetal pode ser reduzido, de modo que o IVDN pode apresentar pouca ou nenhuma influência sobre a suscetibilidade à inundação.

2.1.2.2.5 Grupos hidrológicos do solo

A suscetibilidade à inundação está intrinsecamente ligada às características do solo, que pode ser categorizada pelos grupos hidrológicos (A, B, C e D), conforme proposto pelo *Soil*

Conservation Service - SCS (United States, 1986). Solos com alta taxa de infiltração (grupo A), geralmente arenosos e bem drenados, promovem maior absorção da água, reduzindo o escoamento superficial e o risco de enchentes. Em contrapartida, solos argilosos e com baixa infiltração (grupo D) dificultam a penetração da água, aumentando o volume que escoam superficialmente e, conseqüentemente, elevando a probabilidade e a gravidade das inundações (Costache¹, 2019; Rahman *et al.*, 2019; Costache, *et al.*, 2020).

2.1.2.3 Fatores antrópicos

Fatores antrópicos representam o conjunto de influências e modificações impostas ao meio ambiente como resultado direto das atividades humanas (Vafakhah, *et al.*, 2020). Tais ações, abrangendo desde a urbanização e o desmatamento até a implementação de infraestruturas (estradas, edificações, etc.), promovem alterações significativas na paisagem natural e nos processos hidrológicos associados. Em particular, a dinâmica de drenagem hídrica e a infiltração no solo são frequentemente impactadas. A modificação dos padrões naturais, juntamente com a criação de barreiras físicas, pode elevar a suscetibilidade de determinadas áreas a desastres naturais, como inundações. Este aumento decorre, em parte, da intensificação da concentração e acúmulo de água em locais inadequados, devido à modificação dos fluxos hídricos originais.

2.1.2.3.1 Uso e ocupação do solo

A análise do uso e ocupação do solo abrange uma investigação da distribuição de características antrópicas e naturais da superfície terrestre, contemplando, por exemplo, áreas florestais, urbanas, agrícolas e agropastoris (Costache, 2019; Costache *et al.*, 2020; Yaseen *et al.*, 2022). A quantificação e a classificação dessas coberturas são comumente realizadas por meio de técnicas de sensoriamento remoto, que envolvem o processamento de imagens de satélite, utilização de fotografias aéreas e a validação em campo (Jensen, 2015; Costache *et al.*, 2020).

A classificação do uso e ocupação do solo é um fator determinante na suscetibilidade à inundação, pois diferentes categorias de uso (florestal, urbano, agrícola, entre outros) possuem características próprias que influenciam a absorção e o escoamento da água (Rahman *et al.*, 2019). No contexto das áreas urbanas, identificadas e classificadas por meio dessas análises, a impermeabilização do solo emerge como um fator crítico. A predominância de superfícies

construídas (como asfalto e concreto), altera significativamente o ciclo hidrológico, reduzindo a infiltração, aumentando o escoamento superficial e, conseqüentemente, elevando a probabilidade de ocorrência de inundações (Tehrany *et al.*, 2013; Zhao *et al.*, 2019).

2.1.2.3.2 Distância a vias de transporte

A relação entre a suscetibilidade à inundação e a distância das vias de comunicação pode ser observada sob diferentes perspectivas. Em áreas próximas a grandes rodovias ou ferrovias, a infraestrutura pode, por um lado, atuar como barreira física, alterando o fluxo natural das águas e, conseqüentemente, contribuindo para o acúmulo de água em determinados pontos (Nachappa, *et al.*, 2020). Por outro lado, a proximidade dessas vias muitas vezes implica em áreas urbanizadas, onde o solo apresenta alto grau de impermeabilização, aumentando o risco de enchentes devido à menor capacidade de infiltração da água (Zhao *et al.*, 2019).

Conforme estudos realizados por Zhao *et al.* (2019), a análise da relação espacial entre estradas e áreas de inundação constitui um campo de pesquisa importante. Nesse contexto, a métrica de distância da estrada (do inglês, *distance to road* – DRO) apresenta-se como uma ferramenta quantitativa fundamental para avaliar de que forma a proximidade das vias influencia a suscetibilidade à inundação.

2.1.3 Instrumentos para mapeamento

Para a elaboração de mapas de suscetibilidade, torna-se imprescindível integrar instrumentos tecnológicos e metodologias analíticas de natureza multidisciplinar. Nesse contexto, é fundamental combinar técnicas de coleta, processamento e análise de dados espaciais e ambientais, de modo a garantir elevado grau de precisão e confiabilidade. Para esta seção, será adotada a definição proposta por Caprario (2017), segundo a qual o conjunto de ferramentas, metodologias, técnicas, métodos e procedimentos é designado pelo termo genérico “instrumento para mapeamento”.

As principais ferramentas para mapeamento de áreas de alagamentos e inundações incluem a modelagem hidrológica, modelagem hidráulica, análise hierárquica de processos (AHP), sistemas de informações geográficas (SIG) e sensoriamento remoto (Caprario, 2017).

A modelagem hidrodinâmica é uma das ferramentas mais utilizadas para representar processos e elaborar prognósticos do estudo do movimento dos fluidos (Almeida e Serra, 2017; Santos, 2015). Esta categoria de modelo permite analisar cenários, avaliar alternativas, prever

fenômenos, entre outras possibilidades. Nos estudos para identificar áreas de inundação, esse método pode ter grande precisão, pois consegue simular manchas de inundação com eficiência e confiabilidade (Santos, 2015; Caprario, 2017).

Segundo Almeida e Serra (2017, p.136), a modelagem hidrológica é “uma ferramenta científica desenvolvida com o intuito de se compreender e representar o funcionamento das bacias hidrográficas e simular diferentes cenários”. Nesse contexto, diversos autores têm explorado essa metodologia em seus estudos, como exemplo Bagstevold (2015) e Joo *et al.* (2014) que utilizaram a modelagem hidrológica como instrumento para mapear áreas suscetíveis a alagamentos, enchentes e inundações através do *software* HEC-HMS.

O método de análise hierárquica de processos (*analytical hierarchical process* - AHP) é comum na literatura para gerar mapas de suscetibilidade à inundação (MSI) devido a sua facilidade e simplicidade (Caprario, 2015; Nachappa *et al.*, 2020). Este método foi proposto por Thomas Lorie Saaty na década de setenta (Saaty, 1977). Para esta categoria de problema a aplicação desse método é representada por uma matriz de comparação de pares, na qual vários elementos são comparados e classificados em uma escala. A aplicação deste método pode ser dividida em três etapas principais: 1) criar uma hierarquia; 2) gerar prioridade de critérios individuais (fatores condicionantes de inundação); e 3) avaliar a qualidade da matriz de comparação par a par (Costache *et al.*, 2020). Estudos que utilizaram esse método como Nachappa *et al.* (2020), Malik *et al.* (2020) e Rahman *et al.* (2019) obtiveram modelos com acurácia de 85,9%, 86,90% e 70,60%, respectivamente para estudos de suscetibilidade à inundação.

Com a evolução das técnicas de geoprocessamento, manipulações de dados espaciais podem ser executadas por sistemas específicos, mais comumente chamados de sistemas de informação geográfica (SIG). O termo SIG pode ser definido como um sistema que processa dados gráficos e não gráficos, com ênfase em análises espaciais e modelagens de superfícies (Burrough *et al.*, 2015). Como exemplo, pode se destacar os estudos de Lawal *et al.* (2014) na qual devolveu um modelo de extração das extensões das inundações ocorridas no Estado de Perlis, Malásia.

Outro instrumento utilizado para mapeamento é o sensoriamento remoto, que consiste na obtenção de dados da superfície terrestre por meio de sensores à distância, os quais podem estar em satélites, plataformas terrestres, aviões ou outros meios. Esta ferramenta pode ser empregada como suporte a modelos de monitoramento de inundações em áreas urbanas, com a vantagem de oferecer rapidez e baixo custo operacional frente aos métodos convencionais (Freddo *et al.*, 2018).

O Quadro 1 apresenta um resumo dos pontos positivos e limitações das ferramentas de mapeamento utilizadas em diversos estudos científicos, conforme o levantamento realizado por Caprario (2017).

Quadro 1 - Ferramentas para estudo de mapas de suscetibilidade principais vantagens e desvantagens

(continua)

Ferramentas	Vantagens	Desvantagens
Modelagem hidrológica	Realização de prognósticos através de cenários; Baixo custo operacional; Tempo de simulação curto.	Demanda quantidade e qualidade dos dados; Dificuldade de modelar matematicamente alguns processos hidrológicos; Simplificação da espacialização das variáveis e dos fenômenos envolvidos.
Modelagem hidráulica ou modelagem hidrodinâmica	Destaque devido ao alcance da sua modelagem, detalhando além dos processos hidrológicos, todos os processos hidráulicos ocorridos no interior da bacia hidrográfica; Tem grande precisão e tem a melhor representação física do escoamento.	Exige um complexo código computacional; Necessidade de levantamento de uma grande quantidade de dados de campo, como exemplo as informações espaciais do canal: geometria, declividade, coeficiente de rugosidade de Manning e outros.
Método AHP	Simplicidade, clareza e facilidade de uso visto à diversidade de materiais que explicam o passo a passo da metodologia; Interação entre o analista do método matemático e o especialista que fez a avaliação;	Necessidade de consenso no processo de priorização dos critérios e alternativas; Independência ou diferença relevante entre os critérios de cada nível; Risco de decisões subjetivas ou baseadas em opiniões pessoais e preconceitos; Necessidade de procedimento para estruturar o questionário de perguntas e preferências; Limite máximo de nove alternativas para uso do método;
Sistema de Informações Geográficas (SIG)	A velocidade e a quantidade de informações processadas, supera muitas vezes os resultados alcançados manualmente; Capacidade dos SIGs em manipular dados gráficos e não-gráficos de forma integrada;	Necessidade de uma equipe técnica especializada, com total domínio da ferramenta; Demanda quantidade e qualidade das informações utilizadas para a composição do banco de dados;

FONTE: adaptado de Caprario (2017).

Quadro 1- Ferramentas para estudo de mapas de suscetibilidade principais vantagens e desvantagens

(conclusão)

Ferramentas	Vantagens	Desvantagens
Sensoriamento remoto	Possibilidade de programação de coleta sistemática de dados sobre um determinado alvo;	Dificuldade de obter imagens com alta resolução temporal e espacial;
	Imageamento do terreno por varredura linear;	Necessidade de calibração dos dados antes e após o lançamento do sensor;
	Capacidade de realizar a varredura de toda a superfície do planeta em um curto espaço de tempo;	Possibilidade de a radiação emitida por alguns sensores afetar o alvo a ser imageado.
	Capacidade de obter imagens em várias faixas do espectro eletromagnético.	

FONTE: adaptado de Caprario (2017).

O Quadro 1 apresenta um panorama das técnicas e métodos empregados na elaboração de mapas de suscetibilidade à inundação (MSI), abrangendo desde a fase de coleta de dados até a etapa de modelagem. Estudos como Caprario (2017) e trabalhos mais recentes (Costache, 2019; Askar *et al.*, 2022; Balogun *et al.*, 2022) vêm demonstrando a evolução desses instrumentos, que podem ser organizados em três categorias principais: ferramentas de coleta de dados, plataformas de análise espacial e métodos preditivos.

2.1.3.1 Plataformas de análise espacial

No âmbito do geoprocessamento, os sistemas de informação geográfica (SIG) consolidam-se como ferramentas principais para a integração, manipulação e análise de dados espaciais, oriundos de diversas fontes e representados em diferentes formatos, como matricial (raster) e vetorial (Viana *et al.*, 2019). Softwares amplamente utilizados, a exemplo do QGIS e do ArcGIS, viabilizam a combinação de múltiplas camadas de informação, abrangendo variáveis como declividade do terreno, uso e ocupação do solo e proximidade aos corpos hídricos. Essa capacidade de integração de dados heterogêneos é fundamental para a execução de análises espaciais avançadas. Dentre essas análises, destacam-se a sobreposição de camadas

(*overlay*), aplicação de álgebra, e a delimitação de áreas de influência por meio de técnicas como a geração de *buffers* e a segmentação de imagens (Viana *et al.*, 2019).

Adicionalmente, a integração de técnicas de ponderação multicritério aos SIGs amplia significativamente a capacidade de análise de suscetibilidade à inundação. O *Analytic Hierarchy Process* (AHP), por exemplo, possibilita a atribuição de pesos a diferentes critérios, refletindo sua importância relativa no processo de previsão de inundação (Bera *et al.*, 2022). Paralelamente, a modelagem hidrológica e hidráulica, por meio de *softwares* especializados como HEC-HMS e HEC-RAS, possibilita a simulação do comportamento do escoamento superficial e da propagação de ondas de cheia. Esses modelos incorporam variáveis chave, incluindo intensidade e distribuição da precipitação, coeficiente de escoamento e capacidade de infiltração do solo (Caprario, 2017). Essa integração de SIG, com as metodologias de AHP e modelagem hidrológica-hidráulica, pode proporcionar cenários de inundação detalhados (Caprario, 2017).

Conforme abordado em Pebesma e Bivand (2023) e Dorman *et al.* (2025), em um contexto em que se intensificam as demandas por análises espaciais cada vez mais complexas e customizadas, a sinergia entre sistemas de informação geográfica (SIG) e linguagens de programação, como R e Python, revela-se como um recurso importante. Por meio de suas bibliotecas de código aberto, essas linguagens permitem a interação direta com softwares SIG reconhecidos, a exemplo de QGIS e ArcGIS, ao mesmo tempo em que possibilitam o acesso e a manipulação de uma vasta gama de formatos de dados geoespaciais, tais como *shapefiles*, GeoTIFFs e netCDF.

Em síntese, a implementação de um fluxo de trabalho integrado, no qual os dados geoespaciais podem ser importados de um ambiente SIG, processados e analisados em ambientes computacionais (R/Python) mediante técnicas de estatística espacial, aprendizado de máquina, análise hierárquica de processos (AHP) e modelagem hidrológico-hidráulica, estabelece um fluxo analítico completo para estudos de suscetibilidades (Liu *et al.*, 2021; Rana e Mahanta, 2022).

2.1.3.2 Métodos preditivos

Para converter dados em previsões robustas, fazem-se necessários modelos preditivos e técnicas estatísticas avançadas. Modelos de simulação hidrológica e hidráulica, como o HEC-HMS, HEC-RAS entre outros, podem simular os processos de chuva-vazão e o escoamento em canais, possibilitando a geração de hidrogramas de cheias e mapas de inundações para diversos

cenários (Almeida, 2015; Caprario, 2017). De forma complementar, abordagens estatísticas, como as regressões logísticas, e análises de frequência também permitem estabelecer relações quantitativas entre variáveis preditoras e a suscetibilidade à inundação (Zhao *et al.*, 2019; Rahman *et al.*, 2019).

Com aumento da capacidade de processamento, a inteligência artificial (IA), por meio de algoritmos de aprendizado de máquina (AM), como redes neurais artificiais, florestas aleatórias e máquinas de vetor de suporte, expande a capacidade de previsão (Tehrany *et al.*, 2014; Dodangeh *et al.*, 2020; Vafakhah *et al.*, 2020). Esses algoritmos podem detectar padrões complexos e não lineares em grandes bases de dados, podendo superar limitações de métodos mais tradicionais como o AHP (Nachappa, 2020; Vojtek *et al.*, 2021). Entretanto, Liu *et al.* (2021) ressalta que embora o AM apresente bom desempenho preditivo, nenhum método ou técnica é universalmente superior. As deficiências inerentes de cada abordagem justificam a crescente adoção de estratégias híbridas, que combinam o aprendizado de máquina com métodos estatísticos e tradicionais.

No contexto da geração de mapas de suscetibilidade à inundação (MSI), Tehrany *et al.* (2014), Mosavi *et al.* (2020), Islam *et al.* (2021), Bui *et al.* (2019) e Samanta *et al.* (2018) demonstram a existência de uma variedade de métodos. A escolha da abordagem ótima depende de fatores como a disponibilidade e precisão dos dados, a dimensão da área de estudo e a disponibilidade de tempo. Métodos que exigem grande volume de dados precisos ou que demandam tempo computacional elevado podem ser inviáveis em situações com recursos limitados (Burrough *et al.*, 2015; Zhao *et al.*, 2019). Portanto, a seleção da metodologia mais adequada requer expertise do pesquisador, que deve considerar as restrições e potencialidades de cada abordagem para o contexto específico que pretende estudar.

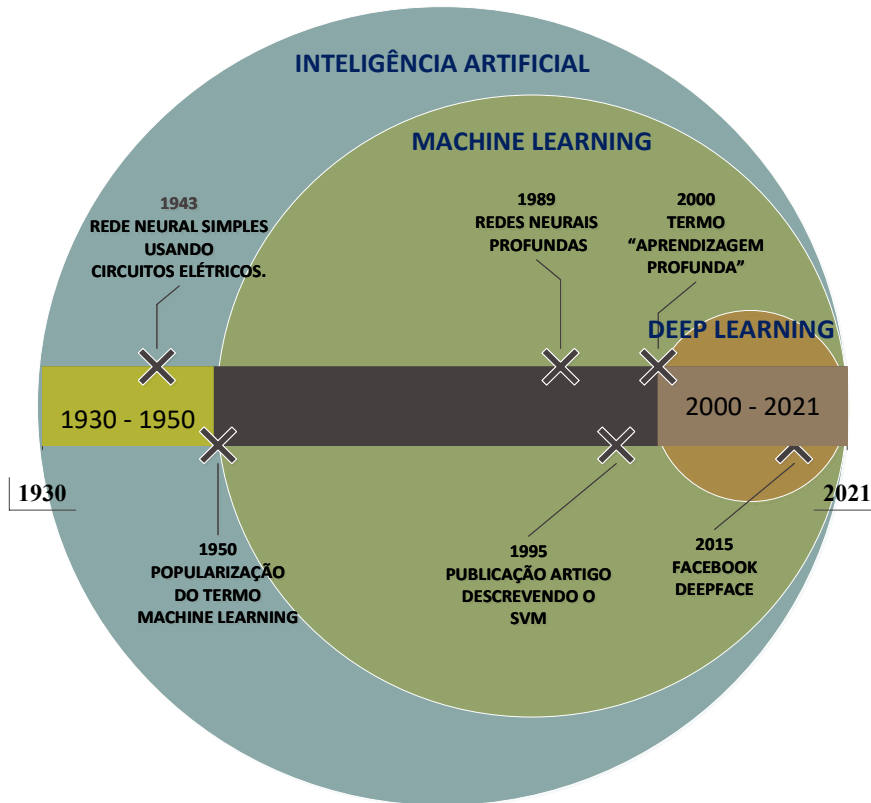
2.2 FUNDAMENTOS DE APRENDIZADO DE MÁQUINA

O aprendizado de máquina (do inglês *machine learning*) é um sub-ramo da inteligência artificial (IA) e que se concentra no aspecto de aprendizagem de computadores (Tappert, 2019). Arthur Lee Samuel foi um grande estudioso da inteligência artificial e popularizou o termo *Machine Learning* em 1959 através da publicação de seu estudo intitulado “*Some Studies in Machine Learning using the game of Checkers*”, ele definiu ML como uma parte da ciência da computação que dá aos computadores a capacidade de aprender sem serem explicitamente programados (Samuel, 1959). Desde então técnicas de aprendizado de máquina tem evoluído

com diferentes algoritmos para automatizar processos e decisões para diversas áreas e com diferentes aplicações.

A Figura 2 mostra a evolução de alguns eventos do surgimento do *Machine Learning*. Ninguém sabe ao certo o ponto inicial do surgimento da inteligência artificial, as raízes da IA podem ser rastreadas até a década de 1940 (Haenlein e Kaplan, 2019). Segundo Masson (1990) em 1943 foi construído a primeira rede neural (simples) pelos neurofisiologistas Warren McCulloch e o matemático Walter Pitts. Em 1995, Cortes e Vapnik (1995) introduziram o algoritmo máquina de vetores de suporte. Após os anos 2000 começou a se popularizar o termo “aprendizagem profunda” conhecido como *Deep Learning* através dos estudos de Geoffrey Hinton e de outros autores (Tappert, 2019).

As redes Transformer têm representado um avanço significativo na área de aprendizado profundo, revolucionando a forma como as máquinas processam e compreendem a linguagem e outros tipos de dados sequenciais. Apresentado em 2017 no artigo “*Attention Is All You Need*”, por Vaswani *et al.* (2017), essas redes têm se destacado por sua arquitetura inovadora, baseada no mecanismo de atenção (*self-attention*), que permite capturar dependências de longo alcance em sequências de dados. Ao contrário das redes neurais recorrentes (RNNs), que processam dados sequencialmente, os Transformers operam de forma paralela, tornando o treinamento mais eficiente e escalável (Vaswani *et al.*, 2017; Tay *et al.*, 2022).

Figura 2 - Linha do tempo *machine learning*

Fonte: Elaborada pelo autor (2025).

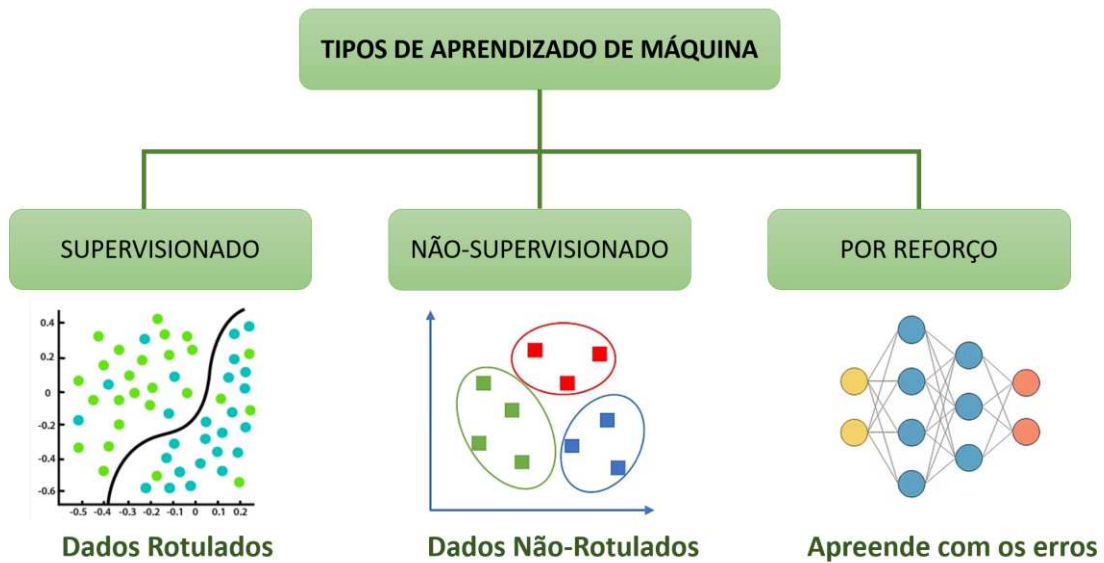
O período pós-2020 tem sido caracterizado por avanços notáveis no campo do aprendizado de máquina, especialmente em relação aos modelos generativos de grande porte. Arquiteturas como as do GPT (Brown *et al.*, 2020) e *Diffusion Models* (Ho *et al.*, 2020) destacam-se pela capacidade de sintetizar conteúdo com resultados expressivos em múltiplas tarefas. Entretanto, a aplicação dessas técnicas em contextos científicos, que exigem a incorporação de modelos físicos, matemáticos, químicos ou operacionais, ainda representam um obstáculo a ser superado (Bengio *et al.*, 2021; Karniadakis *et al.*, 2021). Este cenário tem motivado uma disputa global, tanto na indústria quanto na academia, visando ao desenvolvimento dos melhores modelos capazes de unir eficiência computacional e precisão científica (Conroy e Mallapaty, 2025).

2.2.1 Tipos de aprendizado de máquina

O aprendizado de máquina engloba uma ampla variedade de algoritmos e métodos que permitem a extração de padrões e a tomada de decisões automatizadas a partir de dados (LeCun *et al.*, 2025). Entre os principais métodos utilizados estão redes neurais artificiais, árvores de

decisão, máquinas de vetores de suporte e modelos probabilísticos, cada um adequado para diferentes tipos de problemas e contextos (Breiman, 2001; Jordan, 2004). Esses algoritmos podem ser aplicados em tarefas como reconhecimento de padrões, previsão, agrupamento e otimização. Conforme Muphy (2012), os métodos de aprendizado de máquina podem ser categorizados em três tipos principais (Figura 3) dependendo do objetivo: aprendizagem supervisionada, aprendizagem não supervisionada e aprendizagem por reforço.

Figura 3 - Tipos de aprendizado de máquina



Fonte: Elaborada pelo autor (2025).

Conforme ilustrado na Figura 3, o aprendizado de máquina supervisionado pode ser formalmente descrito como um processo de modelagem de uma função $y(x)$, em que x representa um vetor de dados de entrada e y corresponde a um vetor de saída codificado em classes ou valores contínuos (Murphy, 2012). A função $y(x)$ é inferida durante a etapa de treinamento, utilizando um conjunto de dados rotulados $\{(x_i, y_i)\}_{i=1}^N$, onde N denota o número de amostras de treino. A sequência de operações matemáticas e regras estatísticas que parametrizam $y(x)$ são denominados algoritmos de aprendizado (Bishop, 2006). Após a modelagem de $y(x)$, a função é validada em um conjunto de teste, composto por dados não vistos durante o treinamento, para avaliar sua capacidade preditiva. Nesse contexto, segundo Bishop (2006), a capacidade do algoritmo de categorizar corretamente novos exemplos que diferem daqueles usados para treinamento é conhecida como “generalização”. Essa característica está intrinsecamente ligada à variabilidade dos vetores de entrada x , uma vez que

o espaço de possíveis entradas pode ser infinito ou de alta dimensionalidade, enquanto o conjunto de treinamento abrange apenas uma subamostra limitada (Stone, 1974; Bishop, 2006).

Algoritmos de aprendizado de máquina supervisionado podem ser usados para problemas de regressão quanto para problemas de classificação. Nos problemas de classificação, a variável de saída (y) é categórica, assumindo valores discretos. Por outro lado, nos problemas de regressão, a variável de saída assume valores numéricos contínuos (Brownlee, 2016). O Quadro 2 apresenta exemplos representativos de alguns tipos de algoritmos de aprendizado de máquina supervisionado.

Quadro 2 - Principais algoritmos de aprendizado de máquina supervisionado

Algoritmo	Tipo(s) de Problema(s)	Descrição
Floresta Aleatória (<i>Random Forest</i>)	classificação, regressão	Combina múltiplas árvores de decisão para melhorar a precisão e generalização (Breiman, 2001).
Máquina de Vetores de Suporte (<i>Support Vector Machine</i>)	classificação, regressão	Encontra o hiperplano ótimo para separação de classes (classificação) ou ajuste de dados (regressão) (Cortes e Vapnik, 1995).
Redes Neurais (<i>Neural Networks</i>)	classificação, regressão	Modelos inspirados no cérebro, com camadas de nós interconectados, capazes de aprender relações complexas (LeCun <i>et al</i> , 2015).
k-Vizinhos Mais Próximos (<i>k-Nearest Neighbors</i>)	classificação, regressão	Classifica/prevê com base nos k vizinhos mais próximos (Cover e Hart, 1967).
Árvores de Decisão Impulsionadas (<i>Boosted Trees</i>)	classificação, regressão	Ex: Gradient Boosting, XGBoost, LightGBM, CatBoost. Combina sequencialmente árvores de decisão "fracas" (Friedman, 2001).
Regressão Logística (<i>Logistic Regression</i>)	classificação (principalmente), regressão (ordinal)	Modela a probabilidade de uma classe. Pode ser adaptada para regressão ordinal (Hosmer e Lemeshow, 2000).
Naive Bayes (<i>Naive Bayes</i>)	classificação	Algoritmo probabilístico baseado no teorema de Bayes (Domingos e Pazzani, 1997).

Fonte: Elaborada pelo autor (2024).

O aprendizado não supervisionado é um ramo da aprendizagem de máquina que se distingue pela análise de conjuntos de dados desprovidos de rótulos ou variáveis de saída pré-definidas (Brownlee, 2016). Em contraste com o aprendizado supervisionado, que busca mapear entradas para saídas conhecidas, o aprendizado não supervisionado concentra-se em inferir a estrutura subjacente, a distribuição de probabilidade ou os padrões intrínsecos aos dados (Brownlee, 2016). Essa abordagem é particularmente eficaz em tarefas como *clustering*

(agrupamento), onde o objetivo é particionar os dados em grupos homogêneos, e em mineração de regras de associação, que visa identificar relações estatisticamente significativas entre variáveis (Johnson, 1967; MacQueen, 1967; Agrawal; Srikant, 1994). Exemplos de algoritmos incluem o *k-means* para *clustering*, que busca minimizar a variância *intra-cluster*, e o algoritmo *Apriori*, amplamente utilizado na descoberta de regras de associação em bancos de dados transacionais (MacQueen, 1967; Agrawal; Srikant, 1994).

O aprendizado por reforço (do inglês, *reinforcement learning*) é um tipo de aprendizado de máquina no qual agentes autônomos aprendem a tomar decisões sequenciais em um ambiente dinâmico, visando maximizar uma recompensa cumulativa ao longo do tempo (Sutton e Barto, 2015). Conforme Arulkumaran *et al.* (2017), o aprendizado por reforço é modelado como um processo de decisão de Markov.

No contexto do aprendizado por reforço, existem diversos algoritmos, cada um diferindo principalmente em suas estratégias de exploração do ambiente e na forma como atualizam as estimativas do valor de estados e/ou ações (Sutton; Barto, 2015). Além disso, o conceito de “política” se refere à estratégia que o agente adota para tomar decisões e, essencialmente, corresponde a uma função que mapeia cada estado do ambiente para a ação que deve ser executada (Silver *et al.*, 2014). É essa “política” que equilibra a exploração de novas ações e a exploração do conhecimento adquirido, influenciando de maneira direta a eficiência e a convergência do processo de aprendizado (Thrun, 1992).

2.2.2 Hiperparâmetros

A definição de hiperparâmetros constitui uma etapa crítica no desenvolvimento de modelos de aprendizado de máquina. Diferentemente dos parâmetros internos do modelo (aprendidos durante o treinamento), os hiperparâmetros são configurações pré-definidas pelo pesquisador que governam o comportamento do algoritmo, influenciando diretamente sua capacidade de generalização, eficiência computacional e equilíbrio entre *overfitting* e *underfitting* (Hutter *et al.*, 2019).

Em algoritmos como floresta aleatória, os hiperparâmetros como o número de árvores (*n_estimators*), a profundidade máxima das árvores (*max_depth*) e o critério de divisão (*criterion*) determinam a complexidade do modelo e sua robustez a ruídos (Breiman, 2001). Já para os algoritmos de máquinas de vetores de suporte, a escolha do *kernel* (linear, polinomial, RBF) e o parâmetro de regularização *C* controlam a flexibilidade na separação de classes (Cortes e Vapnik, 1995). Em redes neurais profundas, hiperparâmetros como taxa de

aprendizado (*learning rate*), tamanho do *batch* e arquitetura de camadas são decisivos para a convergência do treinamento (Goodfellow *et al.*, 2016).

A otimização de algoritmos frequentemente exige a combinação de seus hiperparâmetros específicos. Para isso, a busca por configurações ideais se baseia em métodos sistemáticos que exploram o espaço de hiperparâmetros, visando obter o melhor desempenho possível (Box; Wilson, 1951; Stone, 1974; Bergstra; Bengio, 2012). Para isto podem ser aplicados os seguintes métodos:

- a) validação cruzada (*cross-validation*): técnica de reamostragem para avaliar desempenho em múltiplos subconjuntos de dados, reduzindo o risco de *overfitting* (Stone, 1974).
- b) busca em grade (*grid search*): varredura exaustiva de combinações pré-definidas, adequada para espaços de busca pequenos e discretos (Box; Wilson, 1951; Hutter *et al.*, 2019).
- c) busca aleatória (*random search*): amostragem estocástica mais eficiente que *grid search* em espaços de alta dimensionalidade (Bergstra; Bengio, 2012).
- d) otimização bayesiana: uso de modelos probabilísticos (ex., processos gaussianos) para direcionar a busca, minimizando avaliações dispendiosas (Snoek *et al.*, 2012).

A escolha ótima dos hiperparâmetros e métodos de otimização depende criticamente de três fatores inter-relacionados: a natureza dos dados, incluindo volume, dimensionalidade e distribuição estatística (Hutter *et al.*, 2019); a complexidade da tarefa, como classificação, regressão ou aprendizado não supervisionado, que define os requisitos de flexibilidade do modelo (Hutter *et al.*, 2019); e as restrições operacionais, como recursos computacionais (CPU/GPU) e tempo disponível, que limitam a viabilidade de métodos exaustivos (Bergstra; Bengio, 2012). A interação entre esses componentes impacta diretamente na eficiência e eficácia do processo de ajuste, influenciando diretamente na busca do melhor modelo de aprendizado (Snoek *et al.*, 2012).

2.2.3 Métricas de validação

Ao avaliar modelos de aprendizado de máquina, é essencial selecionar as métricas de validação mais apropriadas para o contexto específico do problema. Para problemas de classificação, métricas como precisão, sensibilidade, especificidade e a área sob a curva ROC são frequentemente utilizadas para quantificar a capacidade do modelo de distinguir entre as

classes (Powers, 2011). Já problemas de regressão geralmente empregam métricas como o erro quadrático médio (MSE) e o coeficiente de determinação (R^2) para avaliar a precisão das previsões dos modelos em relação aos valores reais (Murphy, 2012). A escolha adequada dessas métricas permite uma avaliação mais robusta e significativa do desempenho do modelo, abaixo o Quadro 3 apresenta algumas métricas e casos de uso.

Quadro 3 - Principais métricas de validação aplicadas a modelos de classificação por aprendizado de máquina

Métrica	Definição	Casos de Uso
Acurácia (<i>Accuracy</i>)	Proporção de previsões corretas em relação ao total.	Classificação de imagens, reconhecimento de objetos.
Precisão (<i>Precision</i>)	Proporção de verdadeiros positivos entre os positivos previstos.	Recomendação de sistema, quando falso positivos são custosos.
Sensibilidade (<i>Recall</i>)	Proporção de positivos verdadeiros identificados corretamente.	Diagnóstico médico, detecção de fraudes.
Especificidade (<i>Specificity</i>)	Proporção de negativos verdadeiros identificados corretamente.	Testes de triagem, sistemas de alarme.
Área sob a Curva ROC (AUC-ROC)	Capacidade do modelo de distinguir entre classes em diferentes limiares.	Previsão de risco, análise de crédito.
Erro Quadrático Médio (MSE)	Média dos quadrados dos erros entre previsões e valores reais.	Previsão de preços, previsão de demanda.
F1-Score	Média harmônica de precisão e sensibilidade	Quando falso positivos e falsos negativos são relevantes.

Fonte: Elaborada pelo autor (2024), com base em Bishop (2006) e Buckland e Davis (1992).

Através das métricas apresentadas no Quadro 3, os modelos de aprendizado de máquina de classificação podem ser avaliados por meio da matriz de confusão e de gráficos que relacionam a sensibilidade e especificidade (Bishop, 2006). A matriz de confusão oferece uma análise detalhada dos resultados do classificador, permitindo a identificação dos diferentes tipos de erros (Duda *et al.*, 2001). Em paralelo, a curva ROC fornece uma representação gráfica da capacidade discriminante do modelo em diferentes limiares, possibilitando uma avaliação global de sua sensibilidade e especificidade (Fawcett, 2006).

2.2.3.1 Matriz de confusão

A matriz de confusão (do inglês, *confusion matrix*) foi introduzida no contexto de avaliação de modelos de classificação por Ron Kohavi e Foster Provost com base em sua publicação ‘*Glossary of Terms*’ em 1998, na qual apresentou terminologias específicas para a área de aprendizado de máquina (Kohavi; Provost, 1998). Entretanto, sua estrutura básica tem raízes em conceitos estatísticos clássicos, como tabelas de contingência (Pearson, 1904) e métricas de erro em testes de hipóteses (Neyman; Pearson, 1933).

A matriz de confusão constitui uma ferramenta essencial para a avaliação do desempenho de modelos de aprendizado de máquina, conforme apresentado por Duda *et al.* (2001). Essa abordagem organiza os resultados da classificação em quatro categorias fundamentais: verdadeiros negativos (VN), verdadeiros positivos (VP), falsos negativos (FN) e falsos positivos (FP), conforme discutido por Basgalupp (2010, p. 20-21). Os FP e FN são interpretados como classificações incorretas, associadas, respectivamente, aos erros de tipo I e tipo II, enquanto os VP e VN correspondem às classificações corretas, conforme demonstrado na Figura 4. Essa estrutura conceitual oferece uma base teórica robusta para a análise crítica dos resultados em estudos que empregam técnicas de classificação.

Figura 4 - Exemplo de matriz de confusão para área inundável

Classe Real	Área inundável	VP Verdadeiro positivo	FN Falso negativo Erro tipo II
	Área não inundável	FP Falso positivo Erro tipo I	VN Verdadeiro negativo
		Classe predita	
		Área inundável	Área não inundável

Fonte: Elaborada pelo autor (2025) com base em Basgalupp (2010).

A acurácia (A), também denominada eficiência, é uma métrica fundamental na avaliação do desempenho de modelos de aprendizado de máquina, pois indica a proporção de instâncias corretamente classificadas, conforme descrito na Equação 3 (Hastie, 2009; Seydi *et al.*, 2022). No entanto, essa métrica, quando considerada isoladamente, não diferencia adequadamente erros decorrentes de falsos positivos (erro tipo I) e de falsos negativos (erro tipo II). Por essa razão, é recomendável complementar com análises de sensibilidade e a especificidade (Powers, 2011). A sensibilidade do modelo é representada pela taxa de verdadeiros positivos (TVP), conforme a Equação 1, enquanto a especificidade é dada pela taxa de falsos positivos (TFP), calculada pela Equação 2 (Hastie, 2009; Powers, 2011).

$$\text{Sensibilidade} = TVP = \frac{VP}{VP + FN} \quad (1)$$

$$\text{Especificidade} = TFP = \frac{VN}{VN + FP} \quad (2)$$

$$\text{Acurácia} = A = \frac{VP + VN}{VP + VN + FP + FN} \quad (3)$$

com TVP (taxa de verdadeiro positivo); TFP (taxa de falso positivo); VP (verdadeiro positivo); VN (verdadeiro negativo); FN (falso negativo) e FP (falso positivo).

A avaliação de uma matriz de confusão requer uma análise dos custos associados aos erros de classificação, uma vez que tais custos são dependentes do domínio específico da aplicação (Sokolova e Lapalme, 2009). Em contextos como diagnóstico médico, onde a subnotificação de casos positivos (FN) pode levar a tratamentos tardios ou óbitos, a minimização de FN é prioritária, mesmo que isso aumente FP (Chawla *et al.*, 2002; Caruana, *et al.*, 2015). Por outro lado, FP em sistemas de detecção de spam, podem ser críticos, pois pode comprometer a confiabilidade do usuário, enquanto FN pode ter um impacto tolerável (Qazi *et al.*, 2024). Essa contraposição mostra que a escolha de métricas de avaliação e a própria otimização do modelo devem ser cuidadosamente adaptadas às consequências específicas dos erros em cada contexto.

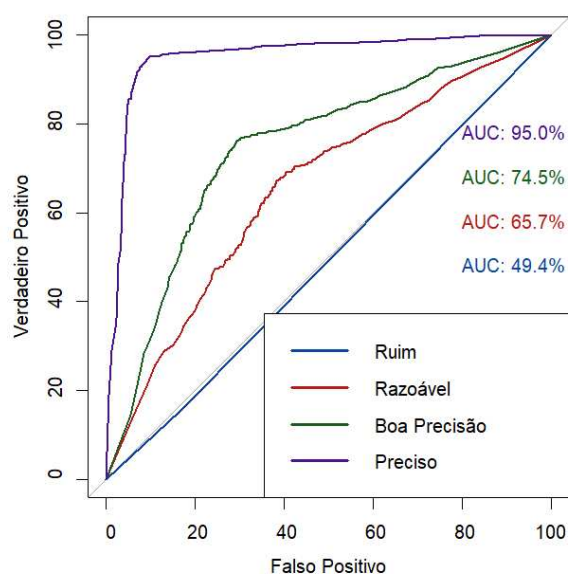
2.2.3.2 Curva de característica de operação do receptor

A curva característica de operação do receptor (do inglês, *receiver operating characteristic* - ROC), foi desenvolvida originalmente no contexto da teoria de detecção de sinais (*signal detection theory* - SDT), durante a segunda guerra mundial para análise de

sistemas de radar e comunicação (Green e Swets, 1966; Duda *et al.*, 2000). Sua formalização matemática e aplicação científica foram estabelecidas nas décadas de 1950 e 1960, principalmente por pesquisadores como Peterson e Birdsall (1954) e Green e Swets (1966).

Em problemas de classificação binária, a curva ROC (Figura 5) constitui um instrumento essencial para avaliar o desempenho de modelos ao distinguir entre duas classes (Brownlee, 2016). Essa curva estabelece uma relação entre a taxa de verdadeiros positivos (sensibilidade) e a taxa de falsos positivos (especificidade), permitindo analisar como o modelo equilibra a correta identificação dos casos pertencentes à classe de interesse e a minimização de classificações equivocadas dos demais casos. Dessa forma, a curva ROC possibilita a visualização da capacidade do modelo em separar efetivamente as instâncias de cada classe conforme varia o critério de decisão (Bradley, 1997).

Figura 5 - Exemplos de curvas ROC ilustrando níveis de acurácia de modelos preditivos conforme a classificação de Swets (1988)



Fonte: Elaborada pelo autor (2025).

A área sob a curva ROC da Figura 5 quantifica a capacidade discriminatória de um modelo de classificação, ao relacionar a taxa de verdadeiro positivo (sensibilidade) e a taxa de falso positivo (especificidade) para diferentes limiares de decisão (Hanley; Mcneil, 1982; Bradley, 1997). Essa métrica não depende de um ponto de corte fixo, o que a torna robusta para comparar modelos distintos (Bradley, 1997; Brownlee, 2016). Em termos interpretativos, um AUC de 1,0 indica um desempenho perfeito (TVP = 1 e TFP = 0 para algum limiar), ao passo

que $AUC = 0,5$ reflete um modelo aleatório, sem poder discriminativo (Bradley, 1997). Valores intermediários ($0,5 < AUC < 1,0$) sugerem graus variáveis de eficácia, onde curvas mais arqueadas acima da diagonal apontam maior precisão (Hanley; Mcneil, 1982). Dessa forma, a Figura 5 (exemplo hipotético) ilustra três cenários distintos para a curva ROC: (1) uma curva de alto desempenho ($AUC=0,950$), que se aproxima do limite superior esquerdo do gráfico; (2) curvas intermediárias, posicionada acima da linha de referência ($AUC=0,745$ e $AUC=0,657$); e (3) a linha de base, cuja AUC é igual ou inferior a 0,5, representando a ausência de poder preditivo.

De acordo com os critérios de interpretação da área sob a curva ROC (AUC) propostos por Swets (1988), modelos com $AUC \geq 0,5$ são considerados aceitáveis, enquanto valores inferiores a 0,5 indicam desempenho abaixo do esperado, tornando o modelo inaceitável. Além disso, o valor de AUC pode ser classificada em diferentes níveis de precisão: valores entre 0,5 e 0,7 indicam uma precisão suficiente, enquanto um intervalo de 0,7 a 0,9 representa boa precisão. Por fim, modelos com $AUC \geq 0,9$ são considerados altamente precisos, demonstrando excelente capacidade discriminativa (Figura 5).

2.3 INTEGRAÇÃO DE TÉCNICAS DE APRENDIZADO DE MÁQUINA PARA MAPEAMENTOS DE SUSCETIBILIDADE À INUNDAÇÃO

Os estudos sobre mapeamentos de suscetibilidade à inundação usando técnicas de aprendizado de máquina têm sua origem nas publicações no campo de deslizamentos de terra (Brabb *et al.*, 1972; Carrara, 1983; Lee *et al.*, 2003; Manso *et al.*, 2007). Burton *et al.* (1978) introduziu conceitos-chave como "suscetibilidade" e "resiliência" no contexto da análise de perigos naturais, fornecendo uma base teórica importante para a avaliação de áreas de risco. Carrara (1983), avançou significativamente ao apresentar modelos estatísticos multivariados para o mapeamento de áreas suscetíveis a deslizamentos, estabelecendo uma abordagem quantitativa. Posteriormente, Varnes (1984) consolidou o conhecimento existente ao propor uma metodologia sistemática para o mapeamento de suscetibilidade à deslizamento, contribuindo para a padronização e disseminação dessas técnicas.

A partir da década de 1990, a difusão de sistemas de informação geográfica (SIG) revolucionou o campo de estudos de mapeamentos de suscetibilidade. Westen (1993) e Gökeceoglu e Aksoy (1996) integraram o geoprocessamento e análise espacial, desenvolvendo estruturas metodológicas que associavam variáveis como declividade, tipo de solo e cobertura vegetal a algoritmos de sobreposição matricial. Esses estudos não apenas aumentaram a

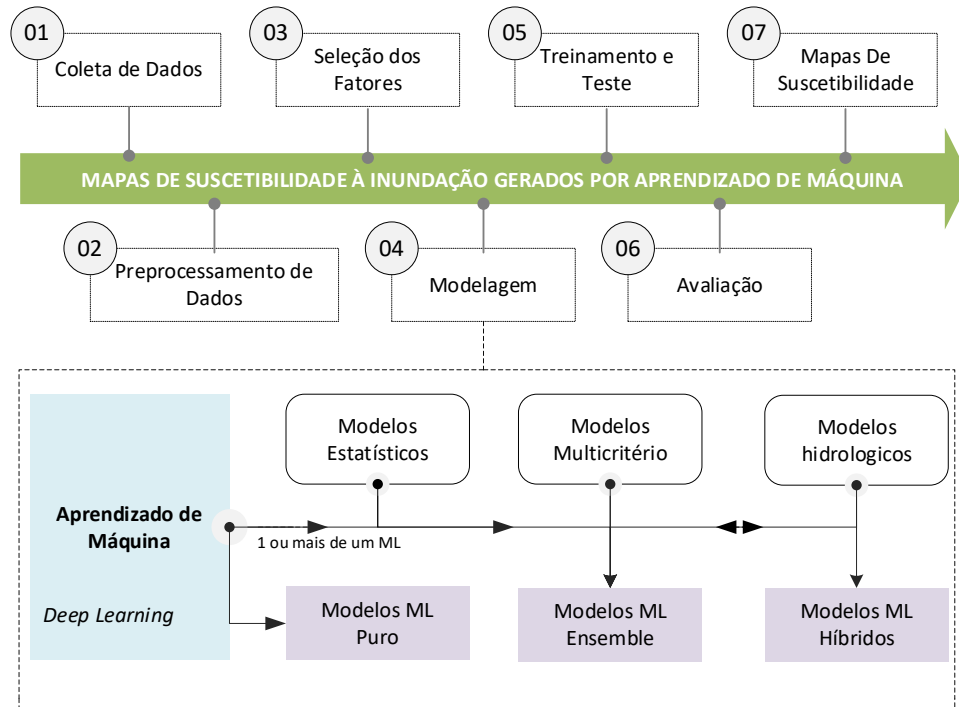
acurácia dos mapas, mas também permitiram a análise de grandes áreas (Lee *et al.*, 2003). Paralelamente, avanços em inteligência artificial (IA), como redes neurais, máquinas de vetores de suporte (Vapnik, 1995) e algoritmos de aprendizagem profunda (Bentivoglio *et al.*, 2022), começaram a ser adaptados para otimizar a predição de suscetibilidade. A capacidade de processar volumes massivos de dados heterogêneos (imagens de satélite, séries históricas) transformou as abordagens convencionais e permitiu o desenvolvimento de novos modelos preditivos (Bentivoglio *et al.*, 2022).

Os primeiros estudos a aplicar métodos de aprendizado de máquina (AM) no mapeamento de suscetibilidade à inundação emergiram nos anos 2000, impulsionados pela crescente disponibilidade de dados geoespaciais e avanços em computação. Trabalhos realizados pelos pesquisadores Biswajeet Pradhan e Mahyat Shafapour Tehrany, introduziram abordagens híbridas, integrando técnicas estatísticas com algoritmos de AM para estudos de suscetibilidades à inundação (Pradhan, 2008, Pradhan, 2010; Pradhan; Youssef, 2011; Kia *et al.*, 2012; Tehrany 2013; Tehrany 2014). Apesar de algumas limitações, esses estudos demonstraram o potencial do AM para capturar relações complexas entre variáveis ambientais e eventos de inundação, estabelecendo bases de referências para novas pesquisas.

2.3.1 Fluxo de desenvolvimento de modelos

A construção de modelos de AM para a previsão da suscetibilidade à inundação segue uma sequência metodológica bem definida, abrangendo desde a aquisição e tratamento de dados até a geração de mapas preditivos. Conforme Hastie *et al.* (2009), o processo de AM inicia-se com a coleta de dados espaciais e ambientais relevantes (etapa 1), seguida do pré-processamento para garantir a qualidade e padronização das informações (etapa 2). Em seguida, ocorre a seleção dos fatores preditivos com maior potencial explicativo para a modelagem (etapa 3), que subsidiam a construção do modelo computacional (etapa 4). A fase subsequente envolve o treinamento e teste dos algoritmos de AM (etapa 5), permitindo ajustes para otimização do desempenho preditivo. Após essa etapa, a avaliação dos modelos (etapa 6) verifica a robustez e a capacidade de generalização das previsões. Por fim, os modelos validados são utilizados para a geração dos mapas de suscetibilidade (etapa 7), os quais representam espacialmente as áreas mais propensas às inundações (Figura 6).

Figura 6 - Fluxo de desenvolvimento de mapas de suscetibilidade à inundação por aprendizado de máquina



Fonte: Elaborada pelo autor (2025).

A modelagem de suscetibilidade à inundação baseada em aprendizado de máquina pode ser estruturada por meio do uso de diversas abordagens metodológicas (Figura 6), incluindo modelos puros (do inglês, *single model* ou *base classifier*), modelos *ensemble* (“comitê”) e modelos híbridos (do inglês, *hybrid model*), cada qual com uma arquitetura distinta e com critérios específicos de implementação (Hastie *et al*, 2009; Zhou, 2012; Liu *et al.*, 2021; Saini *et al.*, 2023).

Os modelos puros (ou modelos simples) de aprendizado de máquina referem-se ao uso de um único algoritmo ou classificador de forma isolada para realizar a predição (Hastie *et al*, 2009). Ou seja, é quando se aplica um único modelo de AM diretamente aos dados de entrada para estimar a suscetibilidade à inundação, sem combinações com outros métodos ou algoritmos. Como exemplo, os estudos de Tehrany *et al.* (2015) e Nachappa *et al.* (2020), na qual foram empregados modelos puros de aprendizagem de máquina, utilizando redes neurais artificiais e máquinas de vetor de suporte (MVS) para gerar mapas de suscetibilidade.

Os modelos *ensemble* são abordagens que combinam múltiplos modelos de aprendizado de máquina com objetivo de produzir uma predição final mais precisa (Zhou, 2012). Nesse

contexto, em vez de confiar em um único algoritmo (como nos modelos puros), utilizam-se vários classificadores ou instâncias de modelos simultaneamente, integrando seus resultados (Rana e Mahanta, 2021). A ideia central é que diferentes algoritmos podem complementar-se, corrigindo erros uns dos outros e, assim, melhorar o desempenho preditivo geral (Breiman, 2001). Diversos estudos tem testados essa metodologia com o intuito de aprimorar os modelos utilizados (Tehrany *et al.*, 2015; Tehrany *et al.*, 2019; Pham *et al.*, 2021).

Os modelos híbridos caracterizam-se pela integração sinérgica de abordagens metodológicas heterogêneas em uma estrutura preditiva integrada, combinando algoritmos de aprendizado de máquina (AM) com métodos complementares (modelos físicos, estatísticos ou de outras naturezas) para otimizar a generalização e a precisão dos resultados (Plataridis; Mallios, 2023; Saini *et al.*, 2023). Diferentemente dos *ensembles*, que combinam vários modelos, os híbridos costumam mesclar diferentes paradigmas de modelagem. Em estudos de suscetibilidade à inundação, isso pode ocorrer de diversas formas, como nos exemplos:

- a) combinação com modelos físicos: integrar um modelo hidrológico (físico) de chuva-vazão com um modelo de aprendizado de máquina. A literatura traz casos de uso de saídas de modelos como HEC-HMS/HEC-RAS combinadas com redes neurais profundas, formando um sistema híbrido que considera tanto princípios físicos quanto padrões extraídos dos dados (Tamiru; Wagari, 2021; Riche *et al.* 2024). Esses modelos híbridos podem demonstrar desempenho superior na detecção de inundações, justamente por unirem conhecimento teórico e aprendizado de dados (Riche *et al.* 2024).
- b) combinação com métodos estatísticos: um exemplo comum é usar técnicas bivariadas ou multicritério (como *frequency ratio* (FR), *weight of evidence* (WoE)) para pré-processamento ou geração de variáveis, alimentando em seguida um classificador de AM (Chapi *et al.*, 2017; Wang *et al.*, 2021). Nesse caso, o modelo resultante é híbrido, pois incorpora pesos ou índices estatísticos calculados a priori em conjunto com a capacidade de aprendizado não linear de um algoritmo de AM (Wang *et al.*, 2021). Por exemplo, pode-se calcular a suscetibilidade inicial por um modelo estatístico e então refinar a predição usando um algoritmo de AM, visando melhorar a acurácia.
- c) utilizando técnicas de otimização: outra abordagem híbrida é empregar algoritmos de otimização (genéticos, colônia artificial de abelhas, etc.) para ajustar os hiperparâmetros ou selecionar variáveis de um modelo de AM (Sachdeva *et al.*, 2017; Plataridis e Mallios, 2023). Assim, o *framework* final não é puramente um aprendizado de máquina, mas sim um híbrido entre otimização matemática e AM.

Estudos como o de Plataridis e Mallios (2023), demonstram que a integração dessas diferentes abordagens pode estar em conjunto em um único modelo, combinando métodos estatísticos, modelo *ensemble* e metaheurística de otimização. A fusão dessas técnicas evidencia que a modelagem híbrida do AM não possui um limite fixo, permitindo uma ampla gama de combinações entre modelos físico, estatístico e de otimização (Chapi *et al.*, 2017; Termeh *et al.*, 2018; Darabi *et al.*, 2021).

2.3.2 Algoritmos de aprendizado de máquina

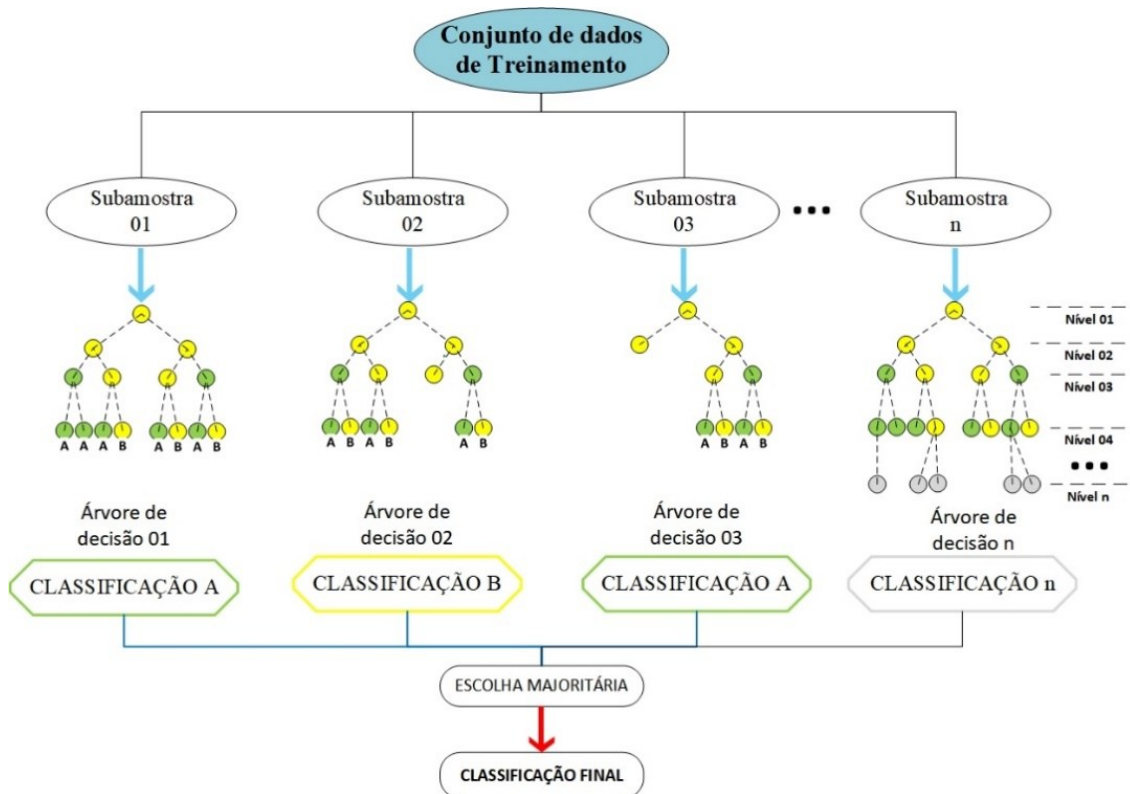
No contexto da avaliação da suscetibilidade à inundação, diversos algoritmos de aprendizado de máquina têm emergido como ferramentas poderosas para gerar mapas preditivos (Vojtek *et al.*, 2021). Estes modelos, que incluem redes neurais artificiais (RNA), como a perceptron multicamadas (MLP) e RNA, máquinas de vetores de suporte (MVS) com diferentes tipos de kernel e abordagens híbridas, e árvores de decisão como árvores de classificação e regressão (CART), florestas aleatórias (FA), árvores de decisão alternadas (ADT) e árvores de decisão impulsionadas (BCT/BRT), têm demonstrado significativa capacidade em modelar relações complexas entre fatores condicionantes e a ocorrência de inundações (Costachea *et al.*, 2020; Nachappa *et al.*, 2020; Vojtek *et al.*, 2021; Yaseen *et al.*, 2022). Adicionalmente, modelos de aprendizado profundo como redes de crenças profundas (DBN), redes neurais convolucionais (CNN), máquinas de aprendizado extremo (ELM), e classificadores como naïve bayes (NB) e árvores naïve bayes (NBT), têm sido aplicados com sucesso (Khosravi *et al.*, 2019; Liu *et al.*, 2021; Pham *et al.*, 2021).

Contudo, diversos estudos comparativos confirmam o excelente desempenho preditivo dos algoritmos de floresta aleatória (FA) e máquinas de vetores de suporte (MVS) no mapeamento de suscetibilidade (Tehrany *et al.*, 2015; Chen *et al.*, 2020; Tanga *et al.*, 2020; Hitouri *et al.*, 2024). Em alguns trabalhos, esses algoritmos apresentaram acurácia superior ou equivalente à de outros métodos tradicionais, consolidando-os como algoritmos para gerar modelos de referência para comparações e identificação de suscetibilidade à inundação (Tehrany *et al.*, 2015; Sachdeva *et al.*, 2017; Lee *et al.*, 2017; Chen *et al.*, 2020; Tanga *et al.*, 2020).

2.3.2.1 Floresta Aleatória

Floresta aleatória (do inglês, *Random Forest* - RF) é um tipo de algoritmo de aprendizado de máquina, muito utilizado para problemas de classificação e regressão (Murphy, 2012). Ele pode ser definido como uma coleção de árvores de decisão que são geradas usando um subconjunto aleatório de dados, conforme ilustra a Figura 7.

Figura 7 - Esquema simplificado do algoritmo floresta aleatória



Fonte: Elaborado pelo autor (2025).

O nome floresta aleatório vem da combinação da aleatoriedade que é usada para selecionar o subconjunto de dados com um grupo de árvores de decisão, portanto, uma floresta (Hartshorn, 2016). O algoritmo de floresta aleatória é amplamente utilizado devido à sua adaptabilidade a diversos tipos de problemas. No entanto, uma limitação conhecida deste algoritmo é o risco de sobreajuste, ou *overfitting*. Esse fenômeno ocorre quando o modelo se ajusta excessivamente aos dados de treino, resultando em um desempenho inferior quando exposto a novos dados de teste (Hartshorn, 2016).

O algoritmo floresta aleatória possui diversos hiperparâmetros (Quadro 4) que influenciam o desempenho de predição. Um dos principais parâmetros é conhecido como *mtry*

(*number of drawn candidate variables in each split*) que pode ser definido como o número de variáveis selecionadas aleatoriamente para cada divisão. O valor de *mtry* pode variar de 1 até *n* (número total de variáveis). De acordo com Probst *et al.* (2019), valores mais baixos para o parâmetro '*mtry*' resultam em árvores mais diversificadas e menos correlacionadas entre si, o que contribui significativamente para a estabilidade dos modelos.

Quadro 4 - Principais hiperparâmetros do algoritmo floresta aleatória

Hiperparâmetro	Descrição
<i>mtry</i>	Número de variáveis selecionadas aleatoriamente para cada divisão de nó. Valores menores reduzem a correlação entre árvores, mas podem prejudicar a acurácia.
Tamanho mínimo do nó	Número mínimo de observações em um nó terminal. Valores menores permitem árvores mais complexas (risco de sobreajuste); valores maiores simplificam o modelo.
Número de árvores (B)	Quantidade de árvores na floresta. Mais árvores reduzem a variância, mas aumentam custo computacional. Valores típicos: 100 a 1000.
Critério de divisão	Métrica para selecionar divisões (ex: impureza de Gini ou entropia para classificação; MSE ou MAE para regressão).
Amostragem <i>bootstrap</i>	Determina se cada árvore do modelo será treinada com amostras geradas por <i>bootstrap</i> .

Elaborado pelo autor (2024), com base em Breiman (2001) e Hastie *et al.* (2009).

O número de árvores é um hiperparâmetro do modelo de floresta aleatória, sendo que em geral o aumento dessa quantidade contribui para reduzir a variância e estabilizar o erro de predição, melhorando a acurácia do modelo. No entanto, acima de um determinado valor, o incremento no número de árvores não gera melhorias adicionais relevantes no desempenho e apenas aumenta o tempo e os recursos computacionais necessários para construir o modelo (Santana, 2020).

O parâmetro tamanho do nó (*nodesize*) especifica o número mínimo de observações em um nó terminal. Esse parâmetro determina a quantidade mínima de amostras que restará em cada um dos nós, sendo os valores padrão 5 para regressão e 1 para classificação (Probst *et al.*, 2019; Santana, 2020)

Os hiperparâmetros do algoritmo floresta aleatória (Quadro 4) podem variar significativamente dependendo da biblioteca ou ferramenta utilizada para sua implementação. Isso ocorre porque cada biblioteca adapta o algoritmo original (Breiman, 2001) para otimizar funcionalidades, desempenho ou usabilidade.

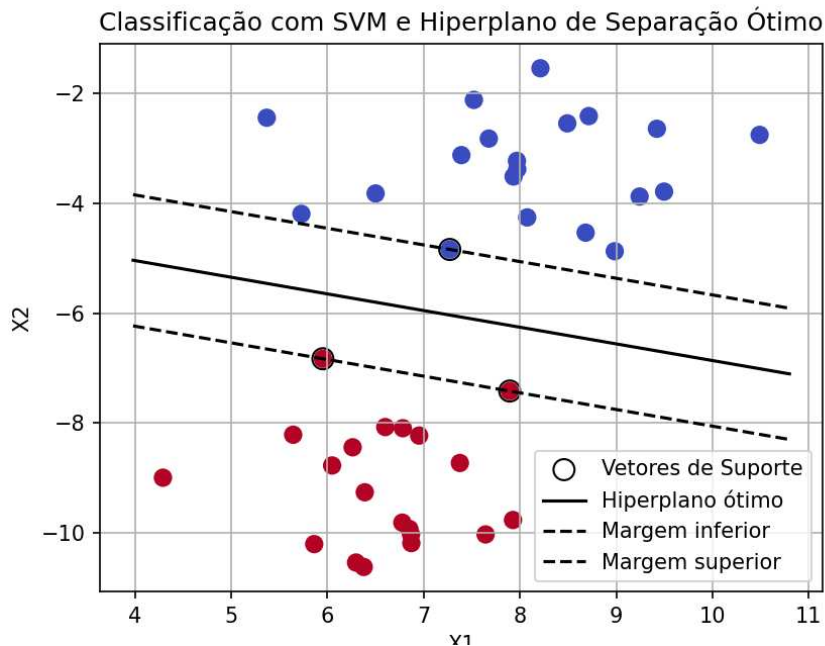
O algoritmo FA tem se mostrado eficaz na geração de mapas de suscetibilidade à inundação, frequentemente superando outros algoritmos de aprendizado de máquina (Youssef *et al.*, 2022; Vu *et al.*, 2023; Hitouri *et al.*, 2023). Estudos realizados no Vietnã (Vu *et al.*, 2023) e no noroeste do Marrocos (Hitouri *et al.*, 2023) evidenciam que o algoritmo FA obteve elevados valores de AUC (superiores a 0,80), indicando bom desempenho na identificação de áreas suscetíveis à inundação. Mesmo quando comparado a outros algoritmos amplamente utilizados, como CART, SVM e XGBoost, o FA apresentou desempenho superior, resultado atribuído à sua capacidade de reduzir o sobreajuste por meio de amostragem *bootstrap* e de modelar eficientemente relações complexas entre variáveis preditoras (Youssef *et al.*, 2022; Hitouri *et al.*, 2023).

2.3.2.2 Máquina de vetores de suporte

Máquina de vetores de suporte (do inglês, *Support Vector Machine* - SVM) foi desenvolvido pela AT&T (*Bell Laboratories*) por Vladimir Vapnik com seus colegas (Cortes e Vapnik, 1995). Cortes e Vapnik (1995) apresentaram o algoritmo *Support Vector Machine* (SVM) por meio da publicação *Support-Vector Networks*. Em seu estudo eles conceituaram que os vetores de entrada (dados) são mapeados não linearmente para um espaço de características de dimensões mais altas. Neste espaço de recursos, uma superfície de decisão linear é construída. Propriedades especiais da superfície de decisão (hiperplano) garante a alta capacidade de generalização do aprendizado de máquina (Figura 8).

Conforme Cortes e Vapnik (1995), o hiperplano no algoritmo atua como a superfície discriminante que separa os dados em diferentes classes no espaço de características. Esse hiperplano é determinado com base na maximização da distância entre os grupos, denominada *margem*, a qual consiste na faixa entre duas superfícies paralelas ao hiperplano, livre de quaisquer pontos de dados, exceto aqueles que tangenciam diretamente (Figura 8). Os pontos que tangenciam as margens, são chamados vetores de suporte, pois influenciam a posição e a orientação do hiperplano (Bishop, 2006). Dessa forma, o objetivo fundamental do SVM é identificar o hiperplano que maximize essa margem, fazendo dos vetores de suporte os elementos-chave na definição e no desempenho do classificador.

Figura 8 - Esquema de classificação por meio do SVM

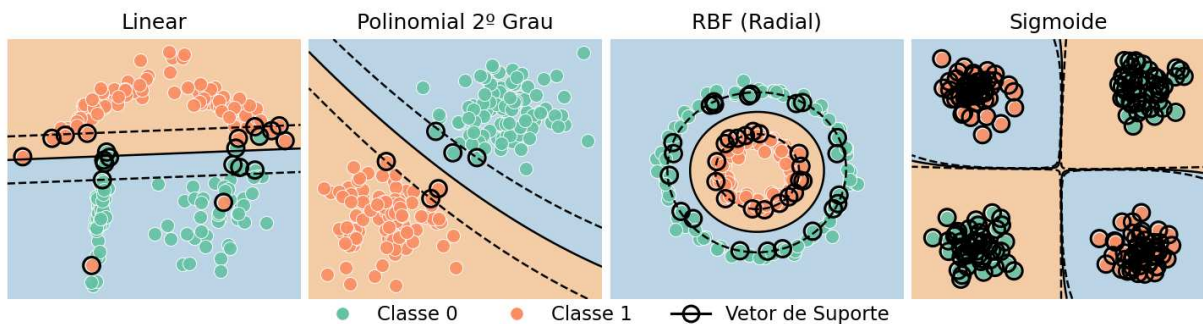


Fonte: Elaborado pelo autor (2025).

Os algoritmos SVMs precisam de ajustes cujos valores são usados para controlar o processo de aprendizado de máquina. O algoritmo SVM possui alguns hiperparâmetros para ajustes, dentre os principais podem-se citar: Kernel, Custo (C) e Gamma.

Em aprendizado de máquina as funções de Kernel são usadas como um classificador para resolver problemas não lineares. Existem muitas funções de Kernel, porém os mais comuns são: linear, polinomial, radial e sigmoide (Merghadi *et al.*, 2020). A Figura 9 mostra exemplos de situações em que algumas funções de Kernel podem separar dados não linearmente.

Figura 9 - Exemplos de funções do Kernel

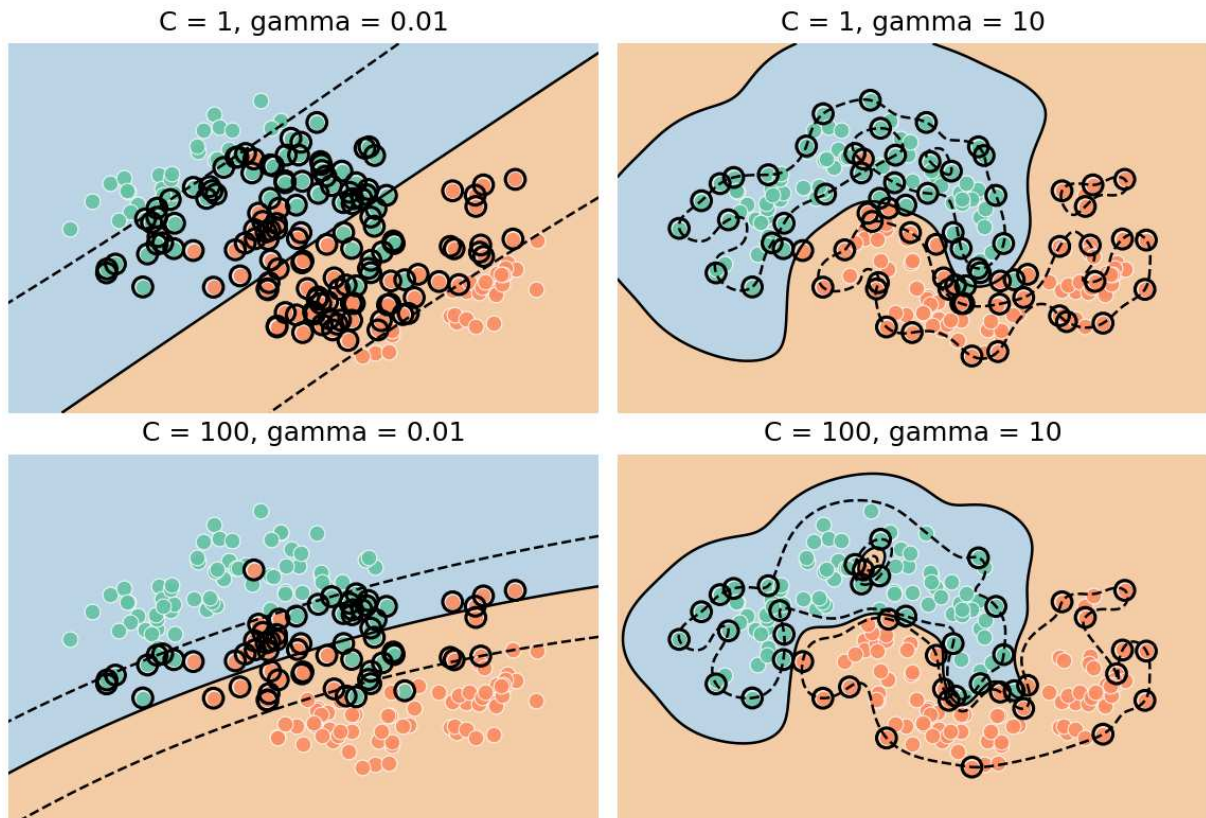


Fonte: Elaborado pelo autor (2025).

O hiperparâmetro Gamma controla a influência que cada valor tem na posição do hiperplano e é usado por todas as funções do kernel, exceto o kernel linear. Quanto mais alto

for o valor de gama, mais próximo será o ajuste dos dados de treinamento (Merghadi *et al.*, 2020). O hiperparâmetro Custo otimiza um hiperplano da classificação incorreta, se o valor de C for grande a separação do hiperplano pode ser boa, mas pode ajustar-se demais aos dados (Merghadi *et al.*, 2020). A Figura 12 ilustra os hiperparâmetros Custo e Gamma.

Figura 10 - Hiperparâmetros Custo e Gama



Fonte: Elaborado pelo autor (2025).

Em um contexto de estudos de suscetibilidade, em que os dados podem ser separados por uma reta ou plano simples, o kernel linear pode ser uma opção rápida e fácil de interpretar, como mostram alguns estudos (Cortes e Vapnik, 1995; Tehrany *et al.*, 2015). Já em casos complexos, onde os padrões são "embaralhados", o kernel RBF pode se destacar por identificar relações não lineares (Tehrany *et al.*, 2015). Outra questão é quando há muitos dados ruidosos ou valores extremos, sendo que o parâmetro C define quanto o algoritmo "tolera" erros, enquanto o gamma (no RBF) controla a sensibilidade a detalhes na qual os valores altos podem fazer o modelo "decorar" os dados em vez de aprender (Tehrany *et al.*, 2019; Hitouri *et al.*, 2024).

Em resumo, os estudos recentes (Tehrany *et al.*, 2015; Sachdeva *et al.*, 2017; Hitouri *et al.*, 2024) indicam que o algoritmo de MVS (com kernel adequado, tipicamente RBF) alcançam

bom desempenho preditivo em casos de suscetibilidade à inundação. Os modelos de MVS costumam apresentar valores elevados de AUC (frequentemente $>0,80$), indicando alta capacidade discriminativa. Além disso, exibem precisão e *recall* elevados, sinalizando que conseguem prever áreas propensas à inundação com grande acerto e consistência.

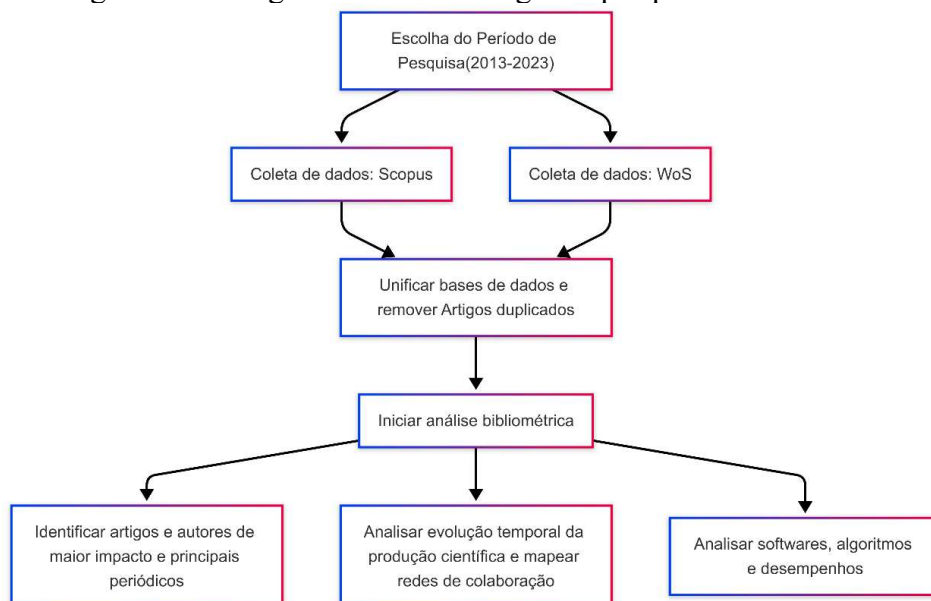
3. METODOLOGIA

Para alcançar o objetivo desta pesquisa, a metodologia foi organizada em duas etapas principais. Na primeira, descreve-se a abordagem adotada para a análise das produções científicas relacionadas à aplicação de métodos de aprendizado de máquina na área de suscetibilidade às inundações, com o intuito de identificar tendências, lacunas e contribuições relevantes ao tema. Na segunda etapa, apresenta-se a metodologia empregada na modelagem de mapas de suscetibilidade à inundação na bacia hidrográfica do Rio Cachoeira (BHRC), situada na região urbana de Joinville, evidenciando as etapas e os procedimentos adotados para a aplicação prática das técnicas de aprendizado de máquina.

3.1 ANÁLISE BIBLIOMÉTRICA

Para avaliar e fazer inferências sobre a produção científica referente às pesquisas de mapas de suscetibilidade à inundação (MSI) gerados por aprendizado de máquina, foram aplicadas técnicas de coleta e tratamento de dados bibliográficos nas bases *Scopus* e *Web of Science* (WoS), para publicações entre o período de 2013 a 2023, conforme representado no diagrama da Figura 11. Para isto, foi utilizado a bibliometria, um método quantitativo para mapeamento e avaliação da área de pesquisa a partir dos dados bibliográficos de sua produção científica (Broadus, 1987; Zupic; Čater, 2015).

Figura 11 - Diagrama da metodologia da pesquisa bibliométrica



Fonte: Elaborado pelo autor (2024).

Para estabelecer os critérios de inclusão e exclusão dos artigos a serem estudados, definiram-se concomitantemente as palavras-chave em inglês “*machine learning*” e “*flood susceptibility*” e, na sequência, procedeu-se à análise dos resultados recuperados no banco de dados, excluindo-se os estudos que não abordavam a geração de mapas de suscetibilidade à inundação por meio de aprendizado de máquina (AM). Em seguida, um subconjunto de estudos foi selecionado para análise mais aprofundada, com o objetivo de identificar os principais fatores considerados e os programas computacionais empregados. Por fim, realizou-se a análise bibliométrica com o uso do pacote *bibliometrix* (Aria; Cuccurullo, 2023). Seguindo o fluxo descrito na Figura 11, foi possível identificar os artigos e autores de maior impacto, os periódicos mais relevantes, a evolução da produção científica ao longo do tempo, as redes de colaboração existentes, bem como os principais softwares, algoritmos aplicados e seus respectivos desempenhos.

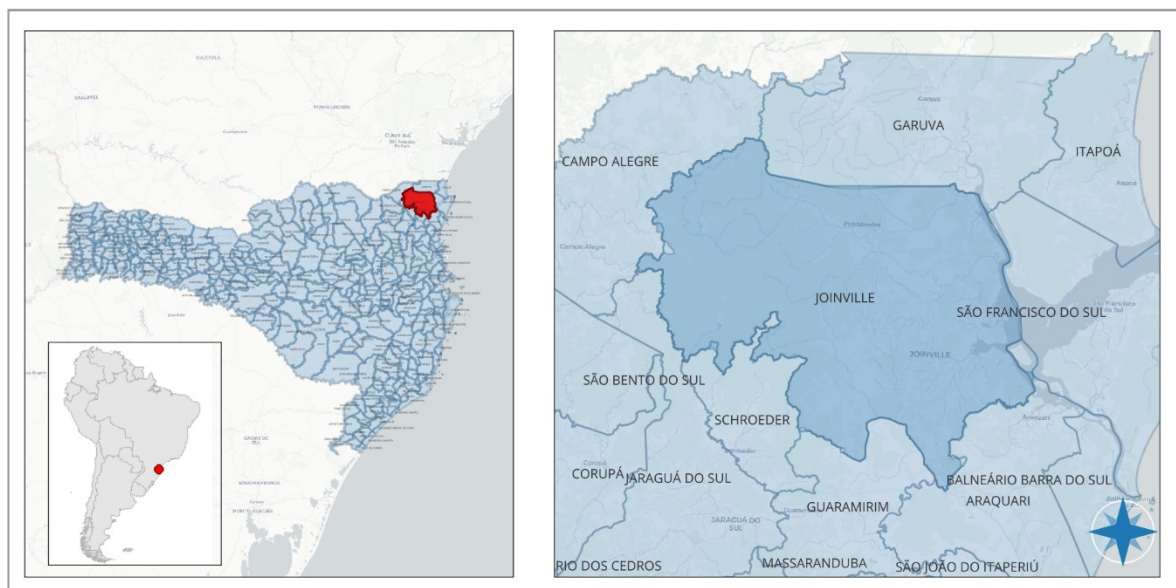
3.2 MODELAGEM DE MAPA DE SUSCETIBILIDADE À INUNDAÇÃO

A seguir, descreve-se a metodologia empregada para a criação de modelos de aprendizado de máquina, aplicados à previsão e elaboração de mapas de suscetibilidade à inundação na bacia hidrográfica do Rio Cachoeira (BHRC), situada na região urbana de Joinville, Santa Catarina, Brasil.

3.2.1 Área de estudo

O município de Joinville/SC (Figura 12) cobre uma área de 1.127,95 km², dos quais 210,4 km² correspondem ao perímetro urbano. Segundo o censo do IBGE de 2022, a população é de aproximadamente 661.600 habitantes. Localizada no hemisfério ocidental, abaixo da linha do Capricórnio, Joinville situa-se no norte litorâneo do Estado de Santa Catarina, na região Sul do Brasil. A leste, o município é limitado pela Baía da Babitonga, enquanto a oeste faz fronteira com um segmento da Serra do Mar, destacando sua posição geográfica estratégica e sua diversidade topográfica.

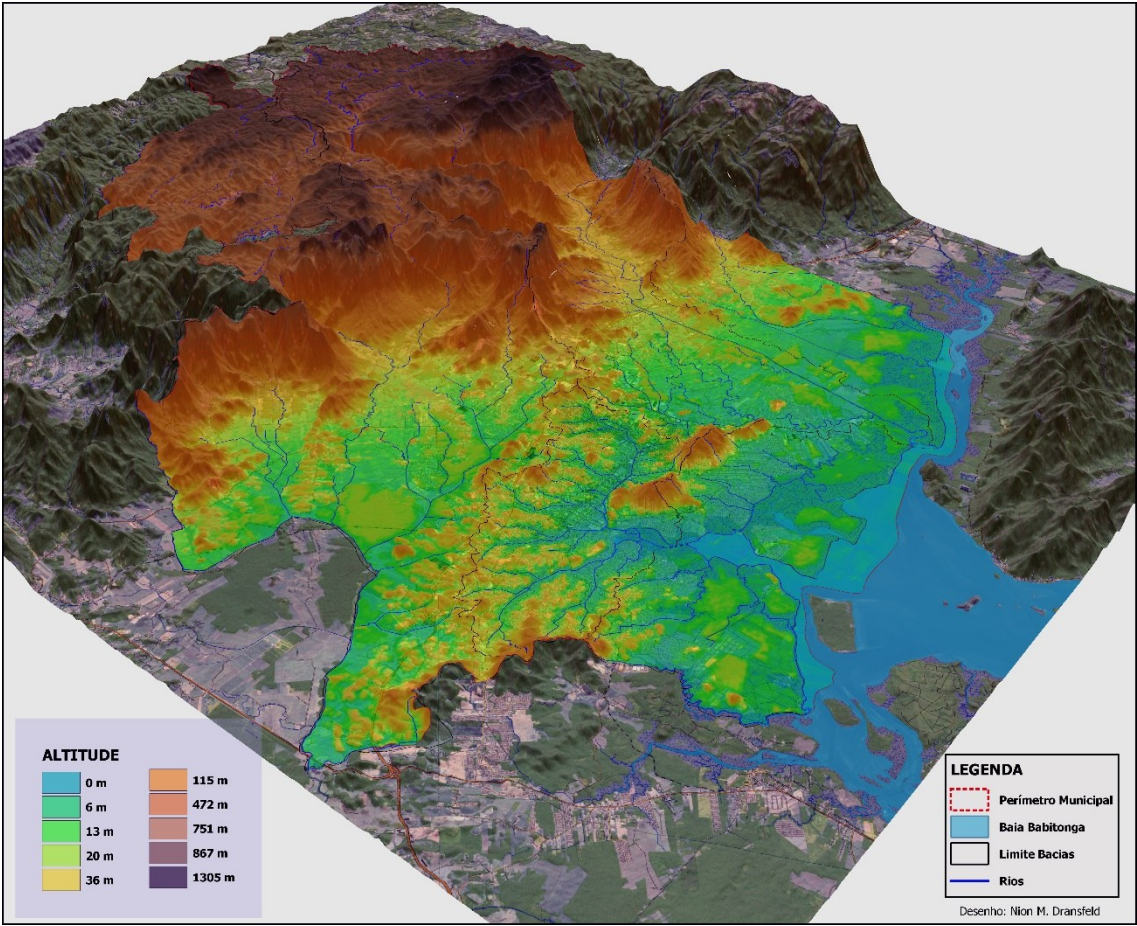
Figura 12 - Mapa de localização do município de Joinville



Fonte: Elaborada pelo autor (2025).

No relevo do município encontram-se as escarpas da serra, com vertentes inclinadas (mais de 60°), vales profundos e encaixados, sendo que o ponto mais alto atinge 1.325 metros. (Figura 14). Na parte leste ocorre uma região de planícies, resultado de processos sedimentares aluviais nas partes mais interioranas e marinhas na linha da costa, onde ocorrem os mangues (Joinville Cidade em Dados, 2020). Justamente nas áreas próximas de mangues se desenvolve a ocupação humana (área agricultável e urbana), com altitude que varia de 0 a 20 metros. Inseridos na região da planície ocorrem morros isolados, constituídos de formas de relevo arredondados, conhecidas como mar de morros, sendo o Morro da Boa Vista o mais alto da área urbana, com 220 metros (Joinville Cidade em Dados, 2020).

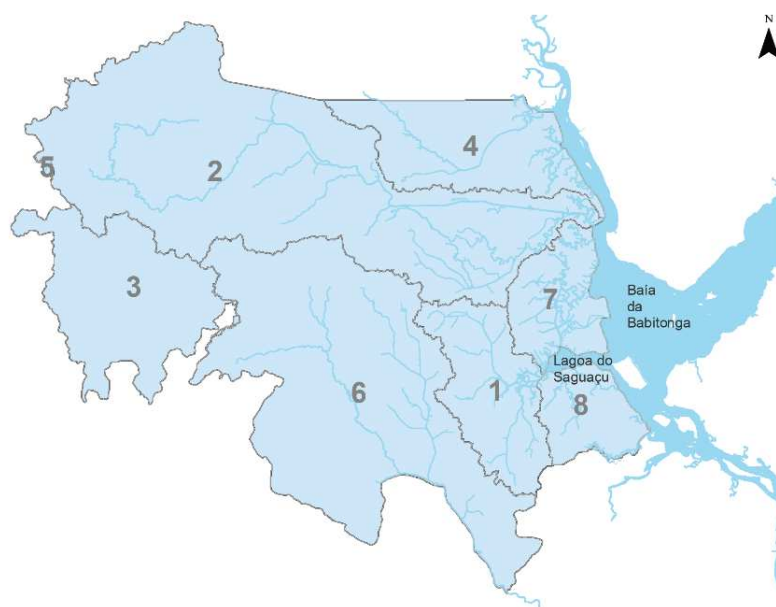
Figura 13 - Hipsometria do Município de Joinville



Fonte: Elaborada pelo autor (2024).

Joinville localiza-se na Vertente Atlântica da Serra do Mar, que é formada por um conjunto de bacias isoladas sendo que o perímetro do município é composto por 8 bacias conforme mostra a Figura 14. Parte da rede hidrográfica de Joinville faz parte do complexo hídrico da Baía da Babitonga, composto pelas bacias hidrográficas do Rio Cubatão (2), Rio Palmital (4), Rio Cachoeira (1), Bacias Hidrográficas Independentes da Vertente Leste (7) e da Vertente Sul (8). Já os rios da bacia hidrográfica do Rio Pirai (6) e do Rio Itapocuzinho (3) fazem parte da Bacia Hidrográfica do Rio Itapocu e estes desaguam na Lagoa do Itapocu, no município de Barra Velha. E em uma pequena parte (lado oeste) Joinville tem parte da bacia hidrográfica do Rio Paraná (5) (SEPUD: Joinville Cidade em Dados 2020 – ambiente natural).

Figura 14 - Bacias Hidrográficas de Joinville



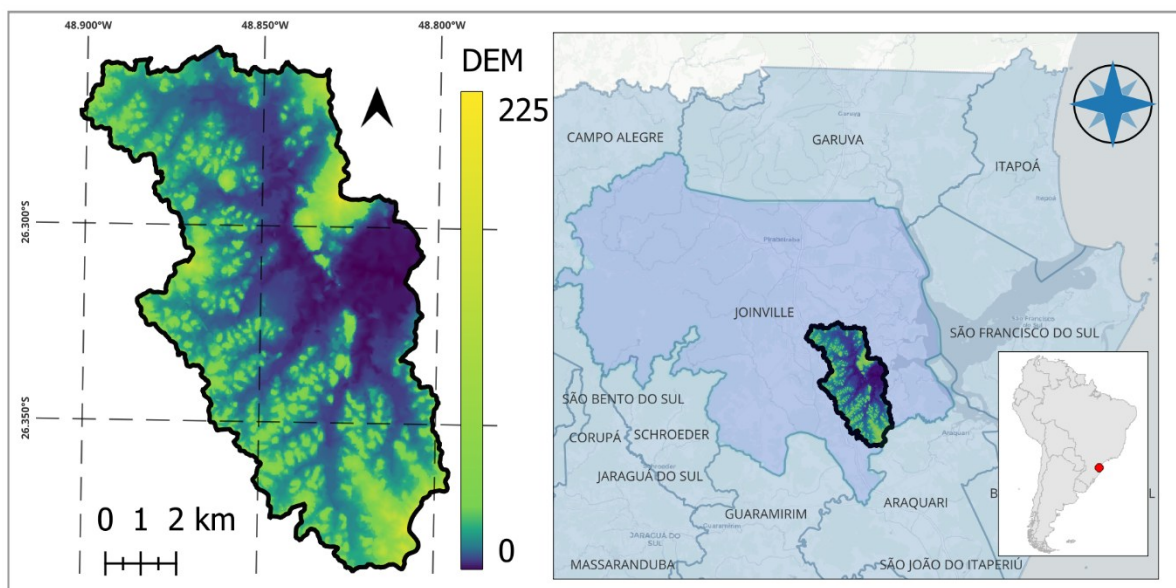
Fonte: SEPUD: Joinville Cidade em Dados 2020/Ambiente Natural.

São abrangidas pela área urbana consolidada do município de Joinville as seguintes bacias hidrográficas: Cubatão, Piraí, Cachoeira, Independentes da Vertente Leste e Sul e do Palmital.

3.2.1.1 Área de estudo: *bacia hidrográfica do Rio Cachoeira*

A bacia hidrográfica do Rio Cachoeira (Figura 15), situada na região central de Joinville, abrange uma área de 83,12 km², correspondendo a 7,3% do território do município. A nascente do rio é encontrada no bairro Costa e Silva, a uma altitude de 40 metros acima do nível do mar. Historicamente, desde a chegada dos primeiros colonizadores, o leito do Rio Cachoeira sofreu múltiplas intervenções para mitigar enchentes, resultando em um curso majoritariamente artificial (IBGE BIBLIOTECA; Ribeiro; Oliveira, 2014)

Figura 15 - Localização da bacia hidrográfica do Rio Cachoeira



Fonte: Elaborada pelo autor (2025).

De acordo com o Plano Diretor de Drenagem Urbana da Bacia Hidrográfica do Rio Cachoeira (PDDU, 2011), a bacia enfrenta desafios ambientais significativos, principalmente devido às ocupações urbanas que se estendem por todas as várzeas até as margens dos rios, incluindo construções sobre o leito dos rios. A ocupação nas áreas de planície foi realizada sobre aterros de cerca de 1,60 metros acima do nível médio da Baía da Babitonga, o que corresponde à altura da maré máxima bianual. Eventos de maré astronômica combinados com condições meteorológicas adversas têm provocado cheias maiores a cada década, com inundações que atingem uma média de 4,66 km² da área urbanizada. A frequência desses eventos está relacionada ao aumento do grau de impermeabilização do solo nas sub-bacias, e a tendência é que a área média inundada aumente significativamente se a expansão urbana continuar sem critérios adequados de manejo do solo. Estima-se que para um evento com probabilidade de ocorrência de uma vez a cada 50 anos, a área inundada poderia aumentar em 25%, superando 20 km² (PDDU, 2011).

3.2.2 Fluxograma da modelagem

Esta pesquisa adotou como método o aprendizado de máquina supervisionado de classificação para criar os mapas de suscetibilidade à inundação (MSI). Para desenvolver a programação para gerar os mapas, se optou em utilizar a linguagem de programação R (R Core

Team, 2024), junto com o *framework* do Rstudio (Rstudio Team, 2024). Essa escolha se deve à licença gratuita, aos manuais de uso abrangente e à ampla disponibilidade de bibliotecas para algoritmos de aprendizado de máquina.

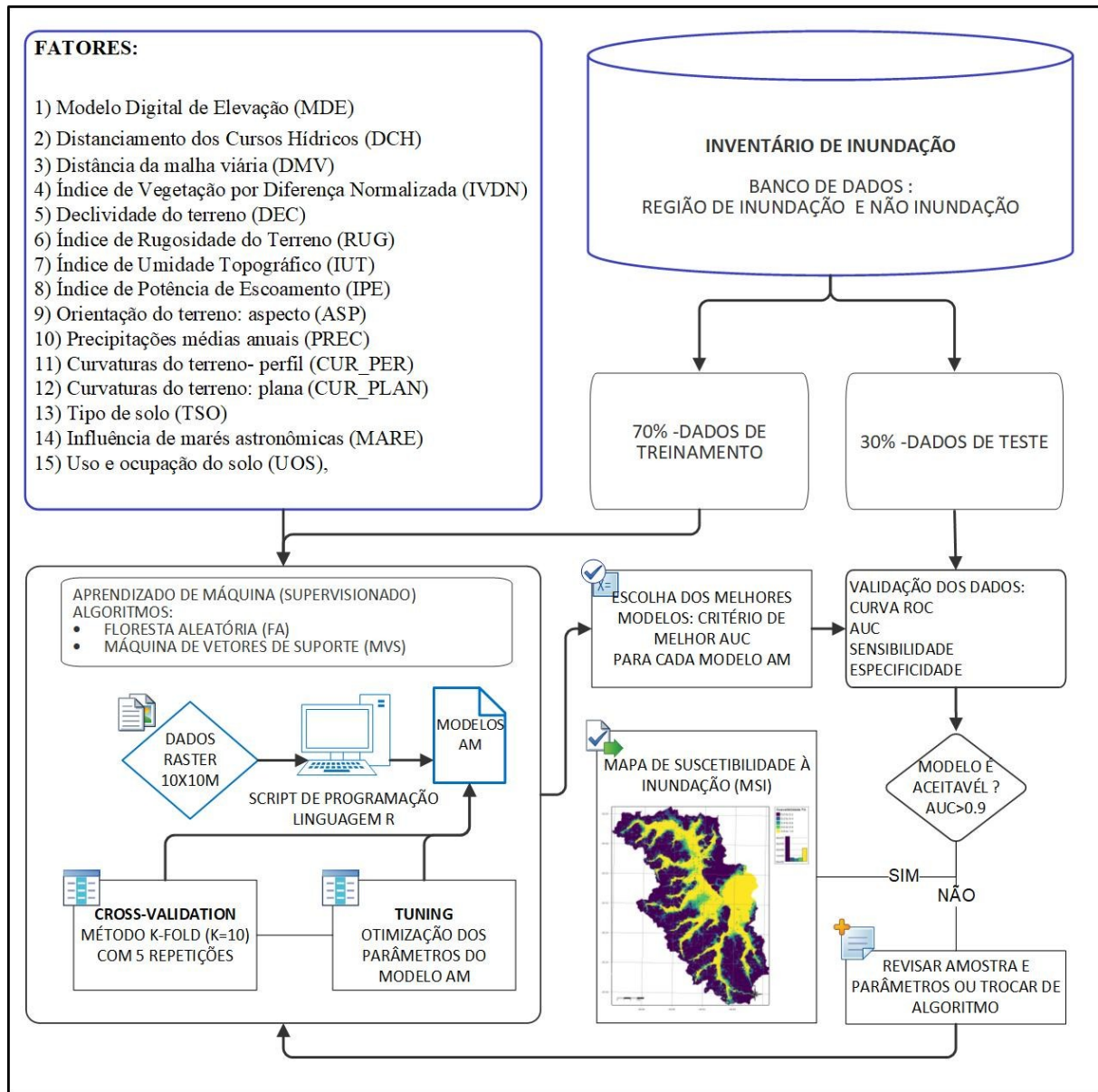
A metodologia aplicada para a construção dos modelos de aprendizado de máquina (AM) visando à previsão da suscetibilidade à inundação seguiu uma sequência bem definida, que abrangeu desde a aquisição e tratamento de dados até a geração dos mapas preditivos (Figura 16). Inicialmente, realizou-se a coleta de dados relevantes por meio de imagens de satélite, estações meteorológicas, levantamentos topográficos, mapeamentos governamentais e bases de dados geoespaciais. Posteriormente, procedeu-se ao pré-processamento dos dados para assegurar sua padronização e qualidade, utilizando sistemas de informação geográfica (SIG), na qual foram aplicadas técnicas de interpolação, reamostragem e normalização.

Em seguida, efetuou-se a seleção dos fatores preditivos, considerando a disponibilidade e relevância dos dados, o que possibilitou a inclusão de variáveis como declividade, uso do solo, precipitação, proximidade a corpos hídricos e características do solo. Na etapa subsequente, realizou-se a modelagem computacional por meio da linguagem de programação em R.

Conforme ilustrado na Figura 16, foi realizado a separação dos dados na proporção de 70-30 (treinamento-teste). Essa escolha justifica-se pela maior demanda de dados para a etapa de treinamento do modelo de aprendizado de máquina, além de refletir o padrão adotado em grande parte dos estudos científicos dessa natureza (Chowdhuri *et al.*, 2019; Tehrany *et al.*, 2014; Choubin *et al.*, 2019).

A avaliação do desempenho preditivo foi realizada com base em métricas estatísticas, como AUC-ROC e matriz de confusão, as quais quantificaram a capacidade do modelo de classificar corretamente as áreas suscetíveis à inundação. Por fim, os modelos validados foram empregados para a geração dos mapas de suscetibilidade, nos quais a probabilidade de inundação foi representada em uma escala contínua de 0 a 1, permitindo a visualização espacial das áreas de maior e menor propensão.

Figura 16 - Fluxograma para obtenção de mapas de suscetibilidade à inundação por aprendizado de máquina



Fonte: Elaborada pelo autor (2024).

3.2.3 Pré-processamento e empilhamento de dados

Para iniciar a modelagem de aprendizado de máquina espacial, todas as informações coletadas foram integradas por meio de uma ferramenta de sistemas de informação geográfica (SIG). Cada fator ou variável foi convertido em imagens georreferenciadas compatibilizadas, garantindo a uniformidade dos dados para análises subsequentes. Para tanto, foram gerados mapas temáticos em formato *raster* com resolução espacial de 10x10 metros, compatível com

a resolução do MDE e das imagens do satélite Sentinel-2 utilizadas neste estudo. Todos os dados foram referenciados ao sistema geodésico SIRGAS 2000, com projeção UTM Zona 22S (EPSG 31982). Essa padronização espacial foi fundamental para assegurar que os diferentes conjuntos de dados pudessem ser comparados e integrados de maneira precisa e consistente.

Em seguida, as imagens individuais foram organizadas e combinadas para formar uma única imagem com múltiplas bandas, que serviu como base para a implementação do modelo de aprendizado de máquina. O empilhamento possibilitou a integração harmoniosa de todas as variáveis, facilitando a identificação de padrões e a análise conjunta dos fatores que influenciam o fenômeno estudado. Para a geração dos mapas temáticos, utilizou-se o software QGIS versão 3.4 Madeira (QGIS Development Team, 2020), enquanto a compilação e execução dos modelos de aprendizado de máquina foram realizadas com a linguagem R, garantindo a consistência metodológica e a qualidade técnica dos resultados obtidos.

3.2.4 Multicolinearidade

Para a seleção adequada de variáveis em modelos estatísticos e de aprendizado de máquina, é essencial empregar métodos de avaliação de multicolinearidade, como o cálculo do coeficiente de correlação de Pearson, o fator de inflação da variância (*variance inflation factor* - VIF) ou a análise de tolerância (Khosravi *et al.*, 2019; Chen, *et al.*, 2020; Tang *et al.*, 2020). Essas técnicas permitem diagnosticar interdependências lineares entre preditores, identificando níveis elevados de correlação que podem comprometer a estabilidade e a interpretação dos modelos (Chan *et al.*, 2022). Para este estudo se aplicou, o fator de inflação da variância (VIF), que é uma métrica para diagnosticar multicolinearidade, particularmente em estudos espaciais e ambientais que integram múltiplas variáveis preditoras derivadas de dados rasterizados (Dormann *et al.*, 2013).

Embora o algoritmo FA possa apresentar menor sensibilidade à multicolinearidade entre variáveis preditoras (Kaveh *et al.*, 2023), tal característica não se aplica ao algoritmo MVS. Portanto, neste estudo, foi adotado a análise da correlação e do fator de inflação da variância (*Variance Inflation Factor* - VIF) como critério para avaliar e mitigar possíveis efeitos de multicolinearidade (Dormann *et al.*, 2013; Montgomery *et al.*, 2012).

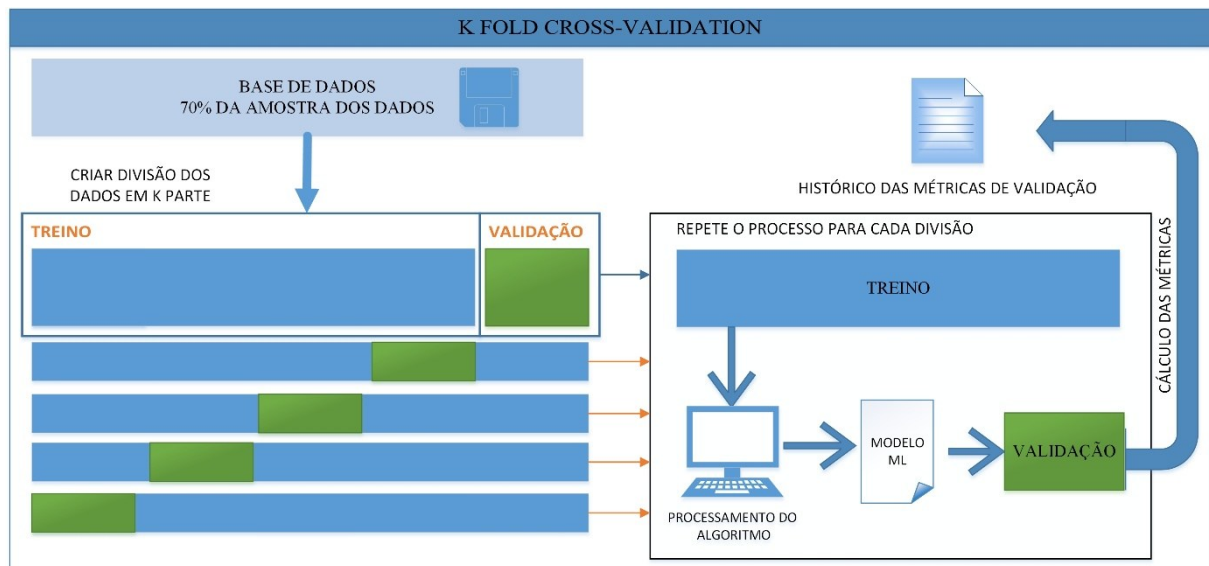
O VIF quantifica o aumento proporcional na variância dos coeficientes de regressão devido à correlação entre variáveis explicativas, comprometendo a confiabilidade das estimativas (Belsley, *et al.*, 1980). Valores elevados de VIF (≥ 5) sugerem multicolinearidade moderada, enquanto valores ≥ 10 indicam colinearidade severa, distorcendo a significância

estatística, prejudicando a identificação de preditores relevantes e reduzindo a estabilidade do modelo (Belsley *et al.*, 1980; James *et al.*, 2013). Para análises espaciais, recomenda-se um limiar conservador de $VIF < 10$, embora limiares mais rigorosos ($VIF < 5$) sejam adotados em contextos com alta sensibilidade a variáveis correlacionadas, como em modelos ecológicos ou previsões baseadas em sensoriamento remoto (Dormann *et al.*, 2013; Montgomery *et al.*, 2012).

3.2.5 Validação cruzada e *tuning*

A técnica de validação cruzada (*cross-validation* - CV) é utilizada para seleção e avaliação de modelos de aprendizado de máquina. O método CV consiste na divisão do conjunto de dados em duas ou mais partes, com objetivo de estimar o desempenho do algoritmo. No particionamento dos dados, um conjunto de dados recebe o nome de treinamento e o segundo de validação (Figura 17). A estimativa do desempenho do algoritmo é realizada com o conjunto de dados de validação (Junior, 2018).

Figura 17 - Esquema do processo de validação cruzada K-Fold aplicado ao conjunto de treinamento



Fonte: Elaborada pelo autor (2024).

O método CV apresenta diferentes formas de divisão dos dados, para este projeto será utilizado o método de validação cruzada por k-partições (*k-fold*). Este método cria uma divisão dos dados em k partes, em tamanhos iguais. Durante a divisão, uma parte é utilizada para treino e outra como validação (Figura 17). Após o processamento de todas as divisões, uma estimativa de precisão geral é fornecida (Brownlee, 2016). Com o resultado do método de validação

cruzada pode se escolher os melhores modelos que se adequaram aos dados, para subsequentemente testar com os dados selecionados para teste.

Para implementar essa metodologia no R, foi utilizado o pacote *caret* (Kuhn, 2020), que disponibiliza um conjunto abrangente de funções para treinamento, validação e avaliação de modelos de aprendizado de máquina. Dentro deste pacote, a função *trainControl* foi fundamental para definir a estratégia de validação cruzada, na qual foi utilizado o parâmetro *method = "repeatedcv"*. No código executado, o *tuning* (ou ajuste fino) foi conduzido pela função *train* do pacote *caret* em conjunto com o *trainControl*, onde a validação cruzada repetida (*repeatedcv*) e a métrica de avaliação, a área sob a curva ROC, foi utilizada para comparar diversas configurações de hiperparâmetros. Durante a execução, cada conjunto de parâmetros foi testado e avaliado em múltiplos subconjuntos (*folds*) dos dados, de modo a obter uma estimativa consistente da performance de cada configuração, possibilitando a seleção do ajuste que produz os melhores modelos.

3.2.6 Avaliação e validação dos modelos

Após a etapa de treinamento dos modelos preditivos baseados nos algoritmos de floresta aleatória (FA) e máquina de vetores de suporte (MSV), otimizados mediante técnicas de validação cruzada e ajuste fino (*tuning*), procedeu-se à seleção dos modelos com melhor desempenho para validação externa. Nesta fase, utilizou-se um conjunto de teste independente, correspondente a 30% da amostra total, previamente segregado e não utilizado durante o treinamento, com o objetivo de quantificar a capacidade de generalização dos modelos em cenários não vistos.

A validação consistiu na comparação sistemática entre as predições geradas pelos modelos e os dados observacionais de referência, representados espacialmente por um mapa georreferenciado de inundações históricas. Inicialmente, derivou-se uma matriz de confusão binária (classe positiva: áreas suscetíveis; classe negativa: áreas não suscetíveis), a partir da qual foram calculados métricas de desempenho discriminatório:

- a) sensibilidade (*recall*): proporção de áreas verdadeiramente suscetíveis corretamente identificadas;
- b) especificidade: proporção de áreas não suscetíveis adequadamente classificadas;
- c) precisão (*precision*): acurácia nas predições positivas.

Para completar esta etapa, foi realizada a avaliação da curva ROC, por meio da razão entre a taxa de verdadeiros positivos (TVP) e a taxa de falsos positivos (TFP). A área sob a curva (AUC-ROC) foi adotada como métrica para quantificar o poder discriminativo dos modelos, expressando o equilíbrio entre sensibilidade (detecção de áreas suscetíveis) e especificidade (exclusão de áreas não suscetíveis) ao longo do gradiente de probabilidades.

Considerando que mapas de suscetibilidade à inundação são aplicados na gestão de riscos de desastres e a consequente necessidade de alta confiabilidade em decisões críticas, estabeleceu-se, como critério final de validação dos modelos, uma área sob a curva ROC (AUC) superior a 0,9. Conforme Swets (1988), este limiar representa um nível de alta precisão, caracterizado por um desempenho discriminativo excelente e, portanto, apropriado para o contexto em questão.

3.2.7 Coleta da variável dependente

A geração de mapas de suscetibilidade à inundação por meio de modelos de aprendizado supervisionado requer a definição clara de classes binárias: áreas inundáveis (classe 1) e não inundáveis (classe 0). Essa abordagem binária foi essencial para treinar os algoritmos FA e MVS, que dependem de dados rotulados para identificar padrões espaciais associados (Tehrany *et al.*, 2013). No entanto, desafios significativos incluem a delimitação precisa das áreas de treinamento e o equilíbrio das amostras para evitar viés no modelo.

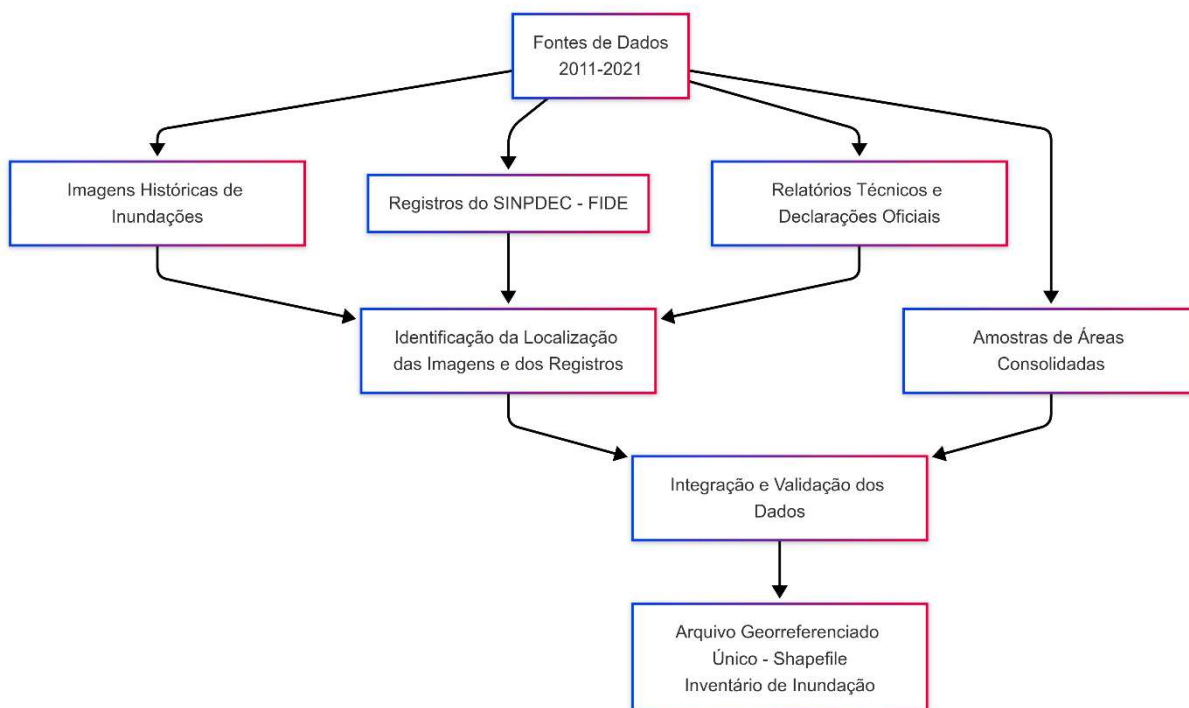
Em problemas binários, o desequilíbrio entre classes (*class imbalance*) pode comprometer o desempenho dos modelos, especialmente quando uma classe domina os dados. Isso pode levar a modelos com alta acurácia aparente, mas baixa sensibilidade para a classe minoritária (He e Garcia, 2009). Para mitigar esse efeito e garantir a representatividade de ambas as classes, este estudo adotou uma estratégia de balanceamento amostral visando uma distribuição equitativa. Inicialmente, um inventário detalhado de eventos de inundação foi compilado, mapeando as áreas historicamente afetadas.

3.2.7.1 Inventário de áreas de inundação

A construção do inventário de áreas inundáveis fundamentou-se na integração dos dados em um formato georreferenciados em formato *shapefile*, consolidando múltiplas fontes para assegurar a validação espacial e temporal das informações (Figura 18). Foram utilizadas

imagens históricas de inundações, associadas aos registros sistematizados pelo Sistema Nacional de Proteção e Defesa Civil (SINPDEC), obtidos por meio do Formulário de Informações do Desastre (FIDE), o qual padroniza a coleta de dados sobre situações de emergência. Paralelamente, relatórios técnicos da Defesa Civil de Joinville e declarações oficiais de emergência ou calamidade pública (conforme Lei Federal nº 10.878/2004) forneceram informações validadas quanto à localização, extensão e severidade dos eventos.

Figura 18 - Diagrama de coleta de dados de áreas de inundações



Fonte: Elaborada pelo autor (2024).

O período de coleta dos dados estendeu-se de 2011 a 2021, definido com base na disponibilidade e na confiabilidade dos registros institucionais, garantindo maior consistência na representação dos padrões de inundação. Períodos anteriores a 2011 apresentavam lacunas na documentação do SINPDEC e inconsistências de localização, dificultando a análise e prejudicando a precisão dos registros.

3.2.7.2 Inventário de áreas de não inundação

Diversos estudos tratam o inventário de áreas de não inundação por meio de diferentes metodologias, refletindo a diversidade de abordagens na literatura. Vafakhah *et al.* (2020) empregaram pontos aleatórios dentro da área de estudo e os classificaram como de não inundação na ausência de registros históricos de eventos de inundação nesses locais. De maneira semelhante, Pham *et al.* (2021), ao analisar a região de Quang Nam, no Vietnã, utilizaram a ferramenta "*create random point tool*" no ArcGIS para selecionar pontos de não inundação aleatoriamente, garantindo que o número de locais escolhidos fosse equivalente ao das áreas historicamente inundadas.

A seleção de locais de não inundação pode envolver tanto conhecimento especializado quanto interpretação visual. No estudo de Lamovec *et al.* (2013), a definição dos pontos de treinamento de inundação e não inundação, fundamentou-se na interpretação visual de imagens de satélite, medições hidráulicas e observações de campo realizadas imediatamente após episódios de chuvas torrenciais. Em Bangladesh, Mandal *et al.* (2022) definiram as áreas de não inundação como aquelas que permaneceram livres de água em todas as imagens de satélite analisadas para eventos de inundação específicos. É importante notar que alguns estudos reconhecem a falta de um consenso geral sobre como os pontos de não inundação devem ser selecionados (Costache; Bui, 2019).

Uma abordagem mais comum é a seleção de áreas com características topográficas que tornam menos propensas à inundação. No estudo realizado na bacia do rio Buzău, na Romênia, Costache *et al.* (2020) geraram amostras representativas de áreas não suscetíveis às inundações, selecionando locais com elevada declividade ou situados a distâncias consideráveis dos cursos d'água, onde eventos de inundação apresentam baixa probabilidade de ocorrência (Tehrany *et al.*, 2014). De maneira análoga, no estudo da bacia do rio Prahova, Costache (2019) descreveu a geração aleatória de 158 locais de não inundação nas zonas interfluviais ou em encostas muito íngremes, onde o fenômeno da inundação é considerado quase impossível. Na região metropolitana de Pequim, Zhao *et al.* (2019) selecionaram um número igual de locais não inundados aos locais de inundação, baseando-se em áreas de alta elevação e alta capacidade de drenagem, que foram então selecionados aleatoriamente para treinamento e teste.

Para delimitar as características topográficas para compor o inventário de não inundação, foram utilizadas as informações referentes à bacia hidrográfica do Rio Cachoeira (Mello *et al.*, 2017) e os resultados da literatura científica na qual demonstra uma forte correlação entre as características topográficas e a ocorrência de inundações:

- a) Mello *et al.* (2017): a BHRC apresenta altitudes entre 0 e 224 m, com predominância de altitudes entre 0 e 20 m (66,4% da área). A declividade dominante é de 0 a 3%, abrangendo 69,8% da área.
- b) Vojtek *et al.* (2021): atribuem maior importância à inclinação do terreno, classificando áreas entre 0 e 2 graus como de alta suscetibilidade às inundações fluviais. Inclinações maiores (35,1 a 57,2 graus) recebem a menor classificação de suscetibilidade.
- c) Costache *et al.* (2020): observam que a classe de inclinação entre 0 e 3 graus concentra 65% dos pixels de inundação na bacia do rio Buzău, indicando forte correlação entre baixa inclinação e inundação.
- d) Vafakhah *et al.* (2020): destaca que às inundações ocorrem mais frequentemente em áreas de menor elevação e inclinação.

Embora a literatura aponte que áreas com menor inclinação, baixa altitude e proximidade a corpos d'água apresentem maior suscetibilidade à inundação, a construção de um modelo preditivo requer uma base de dados que contemple diferentes condições. Assim, para a condução deste estudo, a seleção das áreas de não inundação foi norteada por duas abordagens metodológicas:

- a) zonas topográficas de baixa probabilidade de inundação: identificaram-se regiões com características topográficas que indicam risco mínimo de inundação, considerado hidrologicamente improvável e catastrófico, a partir da análise de parâmetros específicos que devem ocorrer concomitantemente. Tais parâmetros foram definidos com base em análise geomorfológica conduzida por este autor, considerando princípios de dinâmica hidrológica e morfometria fluvial. A escolha dos limiares (altitude > 150 m, declividade > 8%, e distância > 100 m do canal principal) fundamenta-se na relação entre relevo elevado, maior capacidade de escoamento superficial e afastamento de zonas de alagamento, reduzindo substancialmente a probabilidade de ocorrência de inundações;
- b) rotas de evacuação seguras da defesa civil¹: foram mapeadas as rotas de evacuação previamente definidas pela Defesa Civil de Joinville, e amostras dessas rotas foram geradas de forma aleatória para compor o inventário de áreas não inundáveis. A

¹ JOINVILLE. Mapas de rotas seguranças em casos de inundação. Disponível em: <https://www.joinville.sc.gov.br/publicacoes/mapas-de-rotas-seguranças-em-casos-de-inundacao/>.

incorporação dessas áreas justifica-se pela sua relevância estratégica para a segurança pública e pelo histórico de confiabilidade.

Para garantir a confiabilidade e evitar viés no modelo de aprendizado de máquina, as diferentes fontes das amostras de áreas de não inundação foram equilibradas. Durante a seleção, as áreas consideradas seguras, seja por características topográficas como elevação e inclinação ou por integrarem rotas de evacuação definidas pela Defesa Civil, receberam atenção especial para manter esse equilíbrio, visto que as rotas se sobrepõem às vias de acesso para veículos. Dessa maneira, as medidas adotadas buscaram mitigar a geração de padrões equivocados ao longo do processo de treinamento e validação do modelo, contribuindo para a confiabilidade dos resultados preditivos.

3.2.8 Coleta das variáveis independentes

A determinação de áreas suscetíveis à inundação é influenciada por um conjunto multifatorial que integra aspectos naturais, ambientais e antrópicos (Vafakhah *et al.*, 2020; Rana e Mahanta, 2022). Na metodologia adotada, foram escolhidas quinze variáveis preditoras com base em uma revisão sistemática da literatura especializada e na análise das particularidades fisiográficas da região de estudo. Os critérios de seleção consideraram a representatividade dos fatores nos processos hidrológicos de geração de cheias e na disponibilidade de dados em escala espacial. A Tabela 1, apresenta os quinze fatores selecionados, detalhando o nome do dado, o tipo de arquivo, a data de geração dos dados e as fontes utilizadas.

Tabela 1 - Detalhamento das fontes dos dados utilizados

(continua)

Dados	Tipo de arquivo gerado	Data dos dados (fonte)	Fonte geradora
Modelo digital de elevação	<i>Raster</i> 10x10m	Restituição aerofotogramétrica de 2010	Prefeitura de Joinville
Distância do rio (distância euclidiana)	<i>Raster</i> 10x10m	Ano 2011	Mapa Hidrográfico / Prefeitura de Joinville
Distância malha viária (distância euclidiana)	<i>Raster</i> 10x10m	Atualizado 11/03/2020	Mapa Viário formato shapefile Prefeitura de Joinville

Fonte: Elaborada pelo autor (2021).

Tabela 1- Detalhamento das fontes dos dados utilizados

(conclusão)

Dados	Tipo de arquivo gerado	Data dos dados (fonte)	Fonte geradora
Índice de Vegetação por Diferença Normalizada	<i>Raster</i> 10x10m	24/05/2020 (outono) 22/08/2020 (inverno) 01/10/2020 (primavera) 18/02/2021 (verão)	Satélite Sentinel 2: bandas 4 e 8 com correção radiométrica.
Inclinação do terreno	<i>Raster</i> 10x10m	Ano 2010	Calculado através do MDE
Índice de Rugosidade do terreno	<i>Raster</i> 10x10m	Ano 2010	Calculado através do MDE
Índice de Saturação do Solo	<i>Raster</i> 10x10m	Ano 2010	Calculado através do MDE
Índice de potência de fluxo	<i>Raster</i> 10x10m	Ano 2010	Calculado através do MDE
Aspecto	<i>Raster</i> 10x10m	Ano 2010	Calculado através do MDE
Curvatura (plana e de perfil)	<i>Raster</i> 10x10m	Ano 2010	Calculado através do MDE
Tipo de solo	<i>Raster</i> 10x10m	Ano 2012 Atualizado em 07/02/2019	Mapa de pedologia / Prefeitura de Joinville
Uso e ocupação do solo	<i>Raster</i> 10x10m	2020/04/29 2020/05/24	Classificação através das imagens retirada do Satélite Sentinel-2
Área de influência da Maré	<i>Raster</i> 10x10m	2008- 2021	Plano Diretor de Drenagem Urbana (PDDU) do Município de Joinville e dados dos marégrafos
Precipitação	<i>Raster</i> 10x10m	2015	CPRM, 2015. Carta de suscetibilidade a movimentos gravitacionais de massa e inundações: município de Joinville - SC

Fonte: Elaborada pelo autor (2024).

3.2.8.1 Modelo digital de elevação

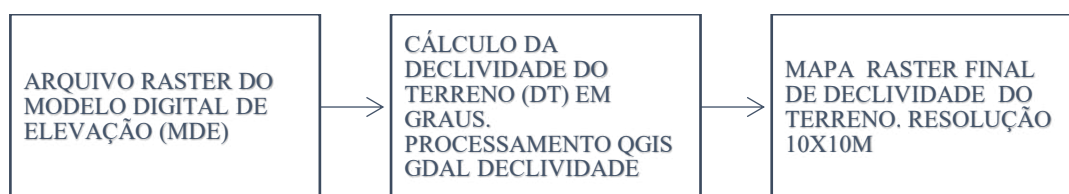
A altitude/elevação do terreno tem grande influência no acúmulo de água. Diversos estudos que utilizaram AM para gerar mapas de suscetibilidade apontaram a elevação/altitude

como um dos fatores mais relevantes para a acurácia dos modelos (Bui *et al.*, 2016; Choubin *et al.*, 2019; Khosravia *et al.*, 2019). Para este estudo foi utilizado o modelo digital de elevação do terreno (MDE)² fornecido pela prefeitura municipal de Joinville (PMJ).

3.2.8.2 Declividade

A inclinação/declividade de um terreno pode ser obtida pelo modelo digital de elevação (MDE). Para muitos autores que estudaram MSI com algoritmos AM concluíram ser um fator de extrema importância, pois o fluxo de água se relaciona à inclinação nas áreas de inundações (Costache *et al.*, 2020; Choubin *et al.*, 2019). A Figura 19 representa o fluxo de processamento para obter o mapa temático de declividade.

Figura 19- Processamento do fator de declividade do terreno



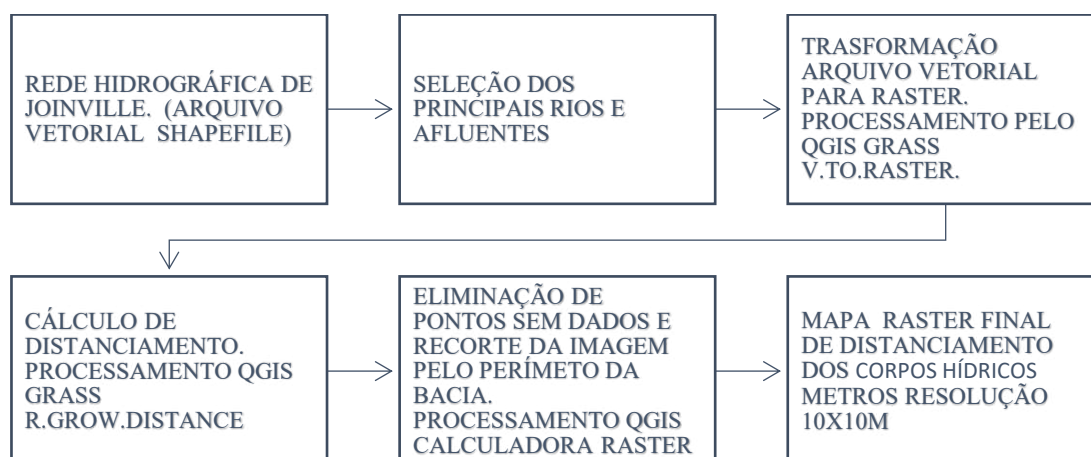
Fonte: Elaborada pelo autor (2024).

3.2.8.3 Distância de corpos hídricos

A distância em relação aos rios, calculada por meio da distância euclidiana, atua como um fator importante em modelos de aprendizado de máquina voltados à previsão de suscetibilidade à inundação. Por meio dessa variável, pode-se identificar a intensidade do impacto exercido pela proximidade dos corpos hídricos sobre o risco de alagamentos, considerando que as áreas mais próximas tendem a ser mais suscetíveis à ocorrência de enchentes. Dessa forma, a inclusão do mapa de distância de corpos hídricos (DCH) nos modelos de aprendizado de máquina contribui para a melhoria da precisão e confiabilidade das previsões de inundação, servindo como base para estratégias de planejamento territorial e mitigação de desastres naturais. A Figura 20 mostra as etapas de processamento para gerar o mapa de distanciamento dos corpos hídricos (DCH).

² Secretaria de Planejamento, Orçamento e Gestão/Prefeitura Municipal de Joinville. 2010. Base Cartográfica do Município de Joinville Escala 1:10.000 / 1:5.000. Executado Por: Aeroimagem Engenharia e Aerolevantamento, ano de 2010

Figura 20 - Sequência de processamento fator DCH

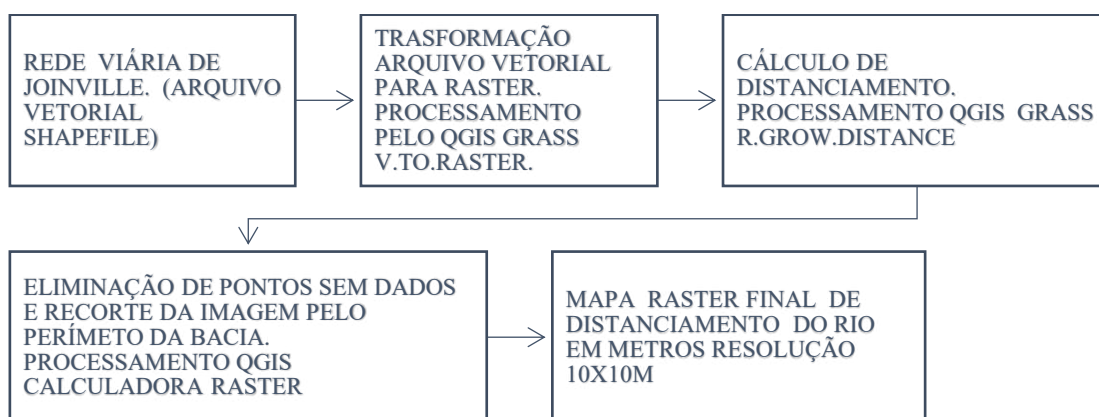


Fonte: Elaborada pelo autor (2024).

3.2.8.4 Distanciamento da malha viária

Esta condicionante não é muito utilizada na literatura, porém está intimamente relacionada ao uso e à ocupação do solo, pois regiões mais afastadas das vias de transporte tendem a sofrer menor influência antrópica. Estudos como Tehrany *et al.* (2019) e Arora *et al.* (2021) demonstraram a relevância de considerar a malha viária em modelos de suscetibilidade à inundação, uma vez que essas vias podem dividir regiões e redirecionar fluxos de água, alterando a dinâmica de drenagem. Nesse contexto, as distâncias das áreas de estudo em relação às vias de transporte foram calculadas pelo método de distância euclidiana, permitindo identificar zonas potencialmente mais vulneráveis. A Figura 21 ilustra o processo de obtenção desse mapa.

Figura 21 - Sequência de processamento fator DMV



Fonte: Elaborada pelo autor (2024).

3.2.8.5 Índice de vegetação por diferença normalizada

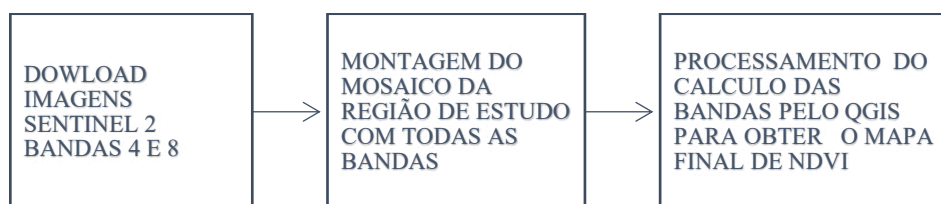
O índice de vegetação por diferença normalizada (IVDN, do inglês *normalized difference vegetation index – NDVI*) foi desenvolvido por Rouse *et al.* (1973) e se constitui um elemento importante para o estudo de mudanças ambientais, utilizando a vegetação como índice de cálculo, atribuindo à área estudada um determinado valor que dependerá do estado em que se encontra essa vegetação. Conforme Shimabukuro e Kuplich (2012): valores negativos de NDVI correspondem a água; valores muito próximos de 0 correspondem a superfície não vegetada; e quanto mais próximo de 1, mais densa é a vegetação. A Equação 4 mostra como calcular o NDVI.

$$NDVI = \frac{(NIR - RED)}{(NIR + RED)} \quad (4)$$

em que, NIR = banda do infravermelho próximo e RED = banda vermelho visível.

Para elaboração desse índice de vegetação foi utilizado as bandas B4 no vermelho visível e B8 no infravermelho próximo obtidas pelas imagens do Sentinel-2. A Figura 22 mostra o fluxo das etapas para obter o mapa final de IVDN.

Figura 22- Sequência de processamento do fator NDVI

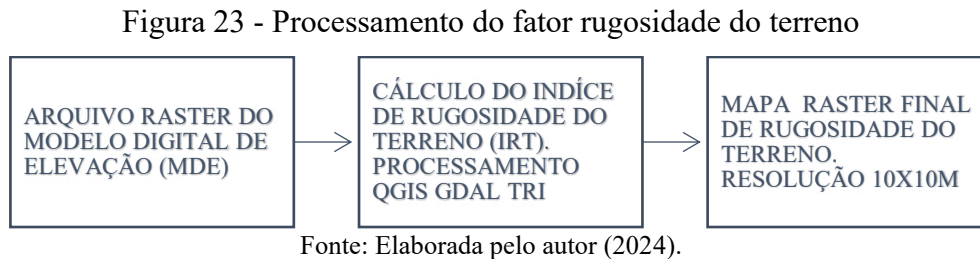


Fonte: Elaborada pelo autor (2024).

3.2.8.6 Índice de rugosidade do terreno

O índice de rugosidade do terreno adotado neste estudo representa a macro rugosidade, calculada com base no MDE com resolução espacial de 10 metros (Fig. 23). Esse índice quantifica a variação altimétrica da superfície terrestre em escala métrica, refletindo as irregularidades topográficas em uma malha regular. O decaimento dessa rugosidade é fortemente influenciado pelo volume e intensidade das chuvas e enxurradas (Bertol *et al.*, 2006). Em termos gerais, valores baixos de rugosidade indicam terrenos mais nivelados,

enquanto valores elevados apontam para superfícies extremamente acidentadas (Alia *et al.*, 2020). A Figura 23 ilustra as etapas de processamento adotadas para a obtenção do fator de rugosidade (RUG).



3.2.8.7 Índice de umidade topográfico

O índice de umidade topográfica (IUT, do inglês *topographic wetness index* – TWI) é um parâmetro hidrogeomorfológico proposto por Beven e Kirkby (1979). O TWI é um fator topográfico utilizado em modelo de escoamento de bacias onde α e β são a área de captação específica ($\text{m}^2 \text{m}^{-1}$) e o gradiente de inclinação (em radianos), respectivamente. O cálculo de TWI é representado pela Equação 5 (Choubin *et al.*, 2019).

$$TWI = \ln \left(\frac{\alpha}{tg(\beta)} \right) \quad (5)$$

com α = área de captação específica e β = gradiente de inclinação.

Manfreda *et al.* (2011) mostraram que o TWI pode representar fortemente as áreas dentro de uma bacia hidrográfica que estão expostas a inundações. O TWI foi calculado utilizando QGIS com o *plugin* SAGA.

3.2.8.8 Índice de potência de escoamento

O índice de potência de escoamento (IPE, do inglês *stream power index* - SPI) é um atributo topográfico secundário que descreve o potencial da topografia para concentrar grandes quantidades de água superficial, promovendo o entalhe. SPI é derivado da área de contribuição específica (A_s) do fluxo acumulado e da declividade (β), a partir de um modelo digital de elevação, sendo representado pela Equação 6 (Wilson; Gallant, 2000).

$$SPI = As \tan \beta \quad (6)$$

em que As é derivado da área de contribuição específica do fluxo acumulado e β é a declividade.

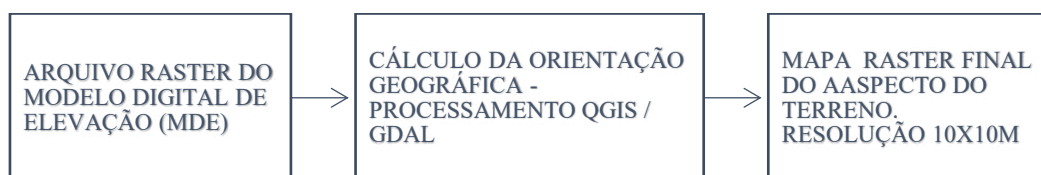
Conforme Turoglu e Dolke (2011), o SPI mostra o poder abrasivo de um rio e foi avaliado como um parâmetro de importância intermediária no risco de inundação. Regiões com valores baixos de SPI indicam regiões na qual a água e materiais podem se acumular.

3.2.8.9 Aspecto

O aspecto do terreno é derivado do modelo digital de elevação (MDE) e indica a orientação geográfica das vertentes, ou seja, a direção para a qual uma encosta está voltada (ex.: norte, sul, leste, oeste). Conforme destacado por Nachappa *et al.* (2020), essa variável é fundamental na análise de suscetibilidade à inundação, uma vez que influencia diretamente a insolação e a dinâmica de absorção de água no solo. Vertentes com orientação predominante para norte (no Hemisfério Sul) recebem maior incidência solar, o que acelera a evaporação e reduz a umidade do solo. Por outro lado, encostas voltadas para sul tendem a reter mais água devido à menor exposição ao sol, aumentando a saturação do solo e, conseqüentemente, o risco de escoamento superficial e eventos inundáveis. Assim, o aspecto atua como um indicador indireto da capacidade de infiltração do terreno, contribuindo para a identificação de áreas críticas em estudos hidrológicos.

Para calcular o aspecto, algoritmos de processamento geográfico analisam a variação altimétrica entre cada célula e suas vizinhas, determinando a direção da máxima inclinação do terreno em cada ponto. Essa direção é expressa em graus (de 0° a 360°), correspondendo à orientação geográfica (ex.: 0° ou 360° = norte; 90° = leste). Áreas planas, sem inclinação definida, recebem valores nulos ou específicos. A Figura 24 ilustra as etapas de processamento adotadas para a obtenção da variável aspecto do terreno.

Figura 24 - Processamento do fator aspecto do terreno



Fonte: Elaborada pelo autor (2024).

3.2.8.10 Curvatura

A curvatura é outro fator determinante da suscetibilidade à inundação e pode ser obtida através do modelo digital de elevação. Curvatura é a segunda derivada de uma superfície, ou seja, a inclinação da inclinação. Conforme Beven e Kirby (1979) existem duas classificações de curvatura do terreno: a curvatura no plano e a curvatura no perfil.

A curvatura do perfil é paralela à direção da inclinação máxima do terreno, na qual um valor negativo indica que a superfície é convexa, enquanto um valor positivo sinaliza uma superfície côncava e, um valor zero indica uma superfície linear. A curvatura do perfil afeta a aceleração ou desaceleração do fluxo na superfície (Minella; Merten, 2012).

A curvatura do plano é perpendicular à direção da inclinação máxima do terreno. De acordo com a curvatura no plano, uma encosta pode ser classificada como convergente, divergente ou reta. (Minella; Merten, 2012).

Para este estudo, foram utilizados os dois tipos de curvatura, a Figura 25 representa as etapas de processamento.

Figura 25 - Processamento do fator curvatura



Fonte: Elaborada pelo autor (2024).

3.2.8.11 Tipo de solo

A geologia desempenha um papel importante na suscetibilidade às inundações, pois diferentes tipos de solo possuem variações significativas na capacidade de infiltração e retenção de água (Nachappa *et al.*, 2020). Solos com alta capacidade de infiltração, como solos arenosos, podem reduzir a probabilidade de inundações superficiais, enquanto solos com baixa capacidade de infiltração, como solos argilosos, podem aumentar essa suscetibilidade (Bedient *et al.*, 2015).

Para compor este fator no modelo de aprendizado de máquina, foi utilizado o mapa pedológico de Joinville (PMJ, 2012). Este mapa é uma classe de feição vetorial do tipo poligonal que categoriza as classes de solo do município, permitindo uma análise detalhada das variações pedológicas na região.

As classes de solo identificadas foram tratadas como variáveis categóricas no modelo de AM. A inclusão desta variável é fundamental, pois permite ao modelo considerar as diferenças nas características dos solos, podendo melhorar a precisão na previsão das áreas suscetíveis às inundações. Estudos têm demonstrado que a incorporação de variáveis referente as características físicas do solo podem aprimorar a acurácia dos modelos de suscetibilidade (Tehrany *et al.*, 2015; Samanta *et al.*, 2018; Sahana *et al.*, 2019.).

3.2.8.12 *Área de influência da maré*

Na revisão da literatura não foram encontrados estudos que utilizaram de forma direta a maré como condicionante em modelos de aprendizado de máquina (AM) para gerar mapas de suscetibilidade à inundação, mesmo em áreas litorâneas (Tehrany *et al.*, 2015; Tehrany *et al.*, 2019). Contudo, devido à particularidade de uma parte significativa da área de estudo sofrer inundações constantes causadas pelas influências da maré provenientes da Baía da Babitonga, este fator foi incluído no modelo proposto.

Para o desenvolvimento deste fator foram considerados estudos da influência da mancha de inundação devido à maré, publicados na região em 2011, especificamente para o plano diretor de drenagem urbana (PDDU), que considerou manchas de maré de 2,5m para a bacia do rio Cachoeira (PDDU, 2011). Esse fator, disponibilizado pela prefeitura de Joinville em forma de arquivo *shapefile*, foi incorporado ao modelo de aprendizado de máquina como uma variável categórica binária, representando áreas com e sem influência de maré. A categorização foi realizada da seguinte forma: 1 para áreas sem influência de maré e 2 para áreas com influência de maré.

3.2.8.13 *Uso e ocupação do solo*

A metodologia adotada para a elaboração do mapa de uso e ocupação do solo (UOS) utilizou o QGIS com técnicas padrão da plataforma, como o *plugin semi-automatic classification plugin* (SCP). Em um processo integrado, as imagens do Sentinel-2, já corrigidas, foram importadas e processadas para segmentação e classificação, utilizando as bandas azul (B2), verde (B3), vermelha (B4) e infravermelho próximo (B8) para realçar as características dos diferentes tipos de cobertura. Nesse procedimento, as classes definidas foram: (1) corpos d'água, (2) floresta densa, (3) campos, (4) vegetação agrícola, (5) solo exposto e (6) áreas de

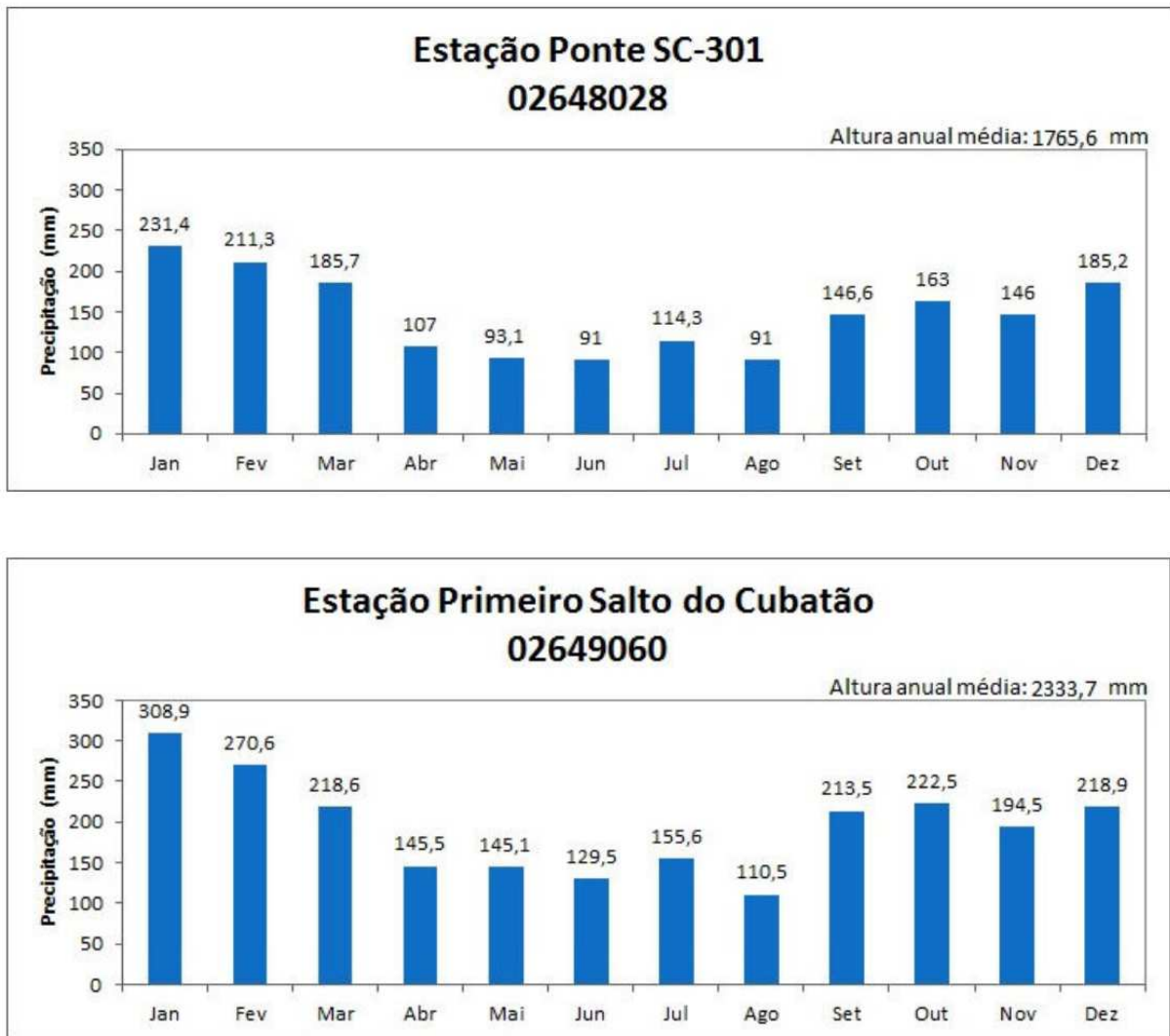
construções e edificações. Essa abordagem possibilitou a geração de um mapa de uso e ocupação do solo.

3.2.8.14 *Precipitação*

Regiões que apresentam chuvas intensas e contínuas demonstram maior propensão à inundação (Alia *et al.*, 2020; Zhao *et al.*, 2018). Para caracterizar essa condição, diversas pesquisas têm utilizado dados de precipitação representados por isoietas (Nachappa *et al.*, 2020; Zhao *et al.*, 2020), permitindo uma análise espacial detalhada dos eventos pluviométricos. Nesse contexto, Zhao *et al.* (2019) empregaram dois fatores de precipitação – a precipitação máxima diária anual e a frequência de chuvas intensas – para uma área de estudo de 2266 km², ressaltando a relevância desses parâmetros na avaliação dos riscos de inundação. Complementarmente, Costache (2019) utilizou a interpolação da distribuição espacial das chuvas por meio do método IDW, em uma região de 2600 km² com alta variação de altitude (128 a 2505 m), evidenciando a influência da topografia na dinâmica da precipitação.

Para a representação espacial do fator precipitação optou-se pela construção de um mapa temático de precipitações médias anuais. Inicialmente, buscou-se coletar dados hidrológicos de nove estações localizadas na área de estudo, por meio do Sistema HidroWEB da Agência Nacional de Águas (ANA). Entretanto, identificou-se que as séries temporais disponíveis apresentavam inconsistências, incluindo lacunas operacionais, discrepâncias instrumentais e predominância de registros recentes (a partir de 2015), inviabilizando a composição de uma série histórica consistente de 10 anos. Diante desse viés metodológico, adotaram-se dados consolidados pela Companhia de Pesquisa de Recursos Minerais (CPRM), referenciados no estudo de 2015 que publicou um estudo de suscetibilidade para a região de Joinville. Essa base, abrangeu o período de 1973 a 2012 e integrou os registros pluviométricos das estações 02648028 (Estação Ponte SC-301) e 02649060 (Estação Primeiro Salto do Cubatão), garantindo uma representação pluviométrica para área de estudo (Figura 26).

Figura 26 - Precipitação média anual de Joinville



Fonte: CPRM (2015).

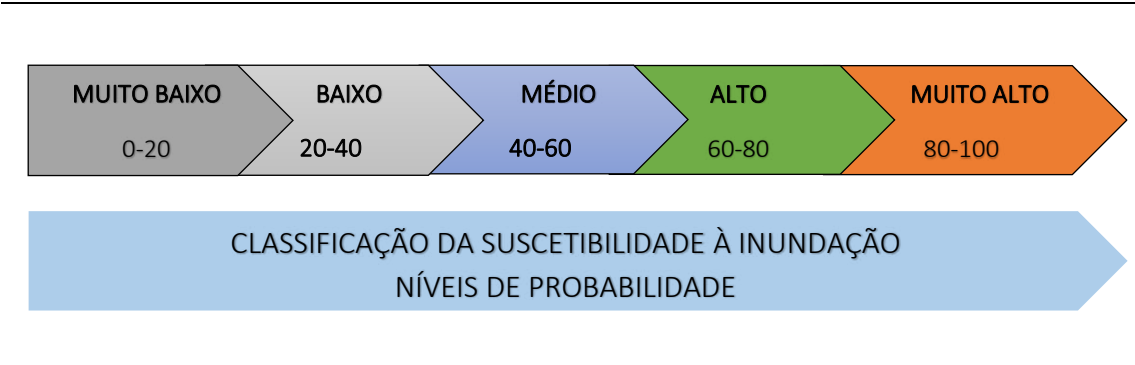
3.2.9 Níveis de suscetibilidade

Nos algoritmos de classificação floresta aleatória e máquina de vetores de suporte empregados no método proposto, os resultados são binários, representados pelos valores 0 (área não inundável) e 1 (área inundável). Considerando a natureza binária do problema, é possível estimar a probabilidade associada a cada decisão do algoritmo. Tais probabilidades, que representam a chance de determinada área pertencer à classe 0 ou 1, possibilitam a geração de uma classificação contínua ou graduada dos níveis de suscetibilidade à inundação.

Para este estudo, será utilizada uma classificação em cinco níveis de suscetibilidade à inundação, adotando uma abordagem baseada em estudos científicos que empregaram metodologias semelhantes (Tehrany *et al.*, 2015; Fang *et al.*, 2020; Vafakhah *et al.*, 2020). A

aplicação dessa abordagem permite uma análise mais detalhada e estratificada, facilitando a identificação de áreas de maior e menor suscetibilidade (Figura 27).

Figura 27 - Classificação dos níveis de suscetibilidade



Fonte: Elaborada pelo autor (2024).

3.2.10 Bibliotecas utilizadas para manipulação dos dados

Para a manipulação dos dados, foram utilizados diversos pacotes na linguagem de programação R, cada um com funções específicas que contribuíram para as diferentes etapas do processo analítico. O pacote *raster* foi empregado para a leitura e manipulação de imagens (Hijmans, 2021), enquanto os pacotes *rgeos* (Bivand; Rundel, 2021) e *rgdal* (Bivand *et al.*, 2021) realizaram operações com dados espaciais. Para a leitura de arquivos no formato Excel, utilizou-se o pacote *readxl* (Wickham; Bryan, 2019). A geração de mapas interativos foi efetuada por meio dos pacotes *tmap* (Tennekes, 2018) e *leaflet* (Cheng *et al.*, 2021), e o pacote *tidyverse* forneceu uma coleção de funções para a manipulação e análise de dados (Wickham *et al.*, 2019).

Para a criação de modelos de aprendizado de máquina multi-modelos, empregou-se o pacote *caretEnsemble* (Deane-Mayer; Knowles, 2019). A construção de grids de plotagens foi realizada com o pacote *cowplot* (Wilke, 2020), e a análise de correlação foi facilitada pelos pacotes *corrplot* (Wei; Simko, 2021) e *corr* (Kuhn *et al.*, 2020). A criação de múltiplos gráficos com *ggplot* foi efetuada por meio do pacote *ggpubr* (Kassambara, 2020), a geração de curvas ROC utilizou o pacote *pROC* (Robin *et al.*, 2011), e a criação de grids com mapas foi facilitada pelo pacote *rasterVis* (Perpiñán; Hijmans, 2021).

4. RESULTADOS E DISCUSSÃO DA ANÁLISE BIBLIOMÉTRICA

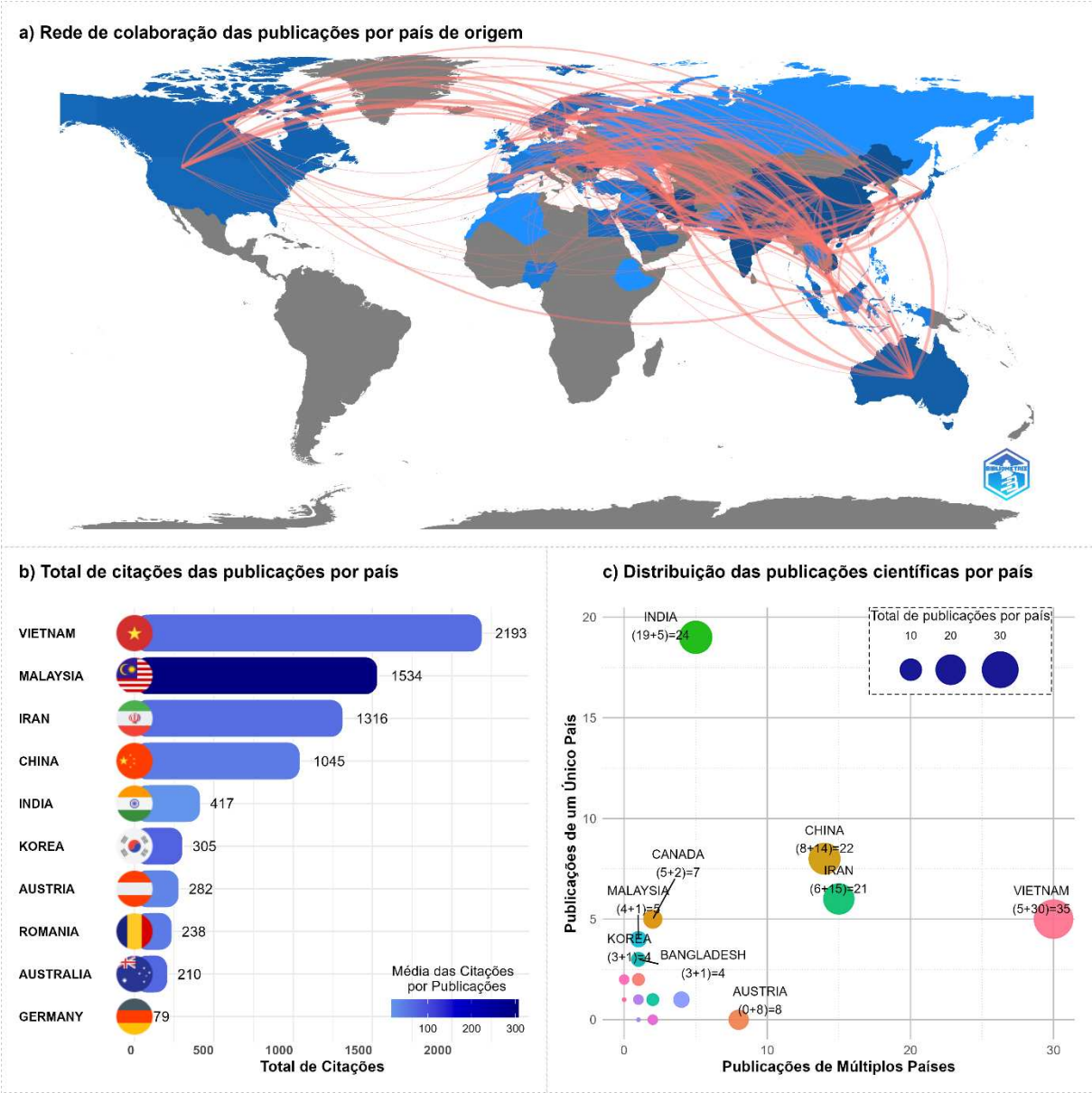
A investigação bibliométrica, conduzida em 12 de abril de 2023 e detalhada na seção metodológica deste estudo, resultou na identificação inicial de 350 registros na base de dados *Scopus* e 500 registros na base de dados do *Web of Science* (WoS), dos quais alguns apresentavam um número de publicações repetidas. Mediante a aplicação de critérios de exclusão para identificação e remoção das publicações duplicatas, realizou-se a triagem dos documentos. Após esse processo, 169 artigos foram selecionados para compor a análise bibliográfica final. Todos os artigos selecionados estão listados no Apêndice A.

4.1 PUBLICAÇÕES POR PAÍSES

A análise da produção científica por país de origem constitui um eixo central em estudos bibliométricos, uma vez que desvenda a geopolítica do conhecimento e a contribuição estratégica de nações para o avanço de domínios específicos (Dordi *et al.*, 2022; Rotimi *et al.*, 2024). No contexto da modelagem de suscetibilidade à inundação com aprendizado de máquina, essa análise pode fornecer dados sobre a distribuição geográfica, revelando investimentos e prioridades nacionais (Badareu *et al.*, 2025). Essas informações são importantes para mapear a capacidade de inovação global, destacando países que lideram a geração de evidências técnicas e a formulação de abordagens computacionais aplicadas a desastres hidrológicos (Dordi *et al.*, 2022). Adicionalmente, a distribuição geográfica dos artigos reflete disparidades em investimentos em ciência e tecnologia, bem como prioridades políticas vinculadas a gestão de riscos ambientais (Dordi *et al.*, 2022; Rotimi *et al.*, 2024).

A Figura 28 apresenta uma análise do país de origem dos artigos selecionados, destacando padrões de distribuição geográfica, redes de colaboração internacional e impacto científico na pesquisa sobre modelos avançados (AM) aplicados em MSI. Os resultados evidenciam uma concentração de publicações no continente asiático, com destaque para Vietnã, Índia, China e Irã (Figura 28c). O Vietnã apresentou o maior volume de publicações e cooperações internacionais, sugerindo a consolidação de uma rede global de pesquisa. Por outro lado, a Índia também se destacou em número de estudos, embora com menor incidência de parcerias internacionais e média de 17 citações por publicação. (Figura 28b). Um achado relevante refere-se à Malásia, que, apesar do baixo volume de publicações, registrou artigos com impacto alto, alcançando até 306 citações por publicação (Figura 28b).

Figura 28 - Análise das publicações por países: a) rede de colaboração, b) total de citações e c) distribuição por país



Fonte: Elaborada pelo autor (2024).

A predominância asiática na produção científica pode estar associada a exposição frequente da população à eventos extremos de inundações (Rentschler *et al.*, 2022), segundo o EM-DAT (2020), a Ásia concentrou o maior número de registros de inundações nas últimas duas décadas. Essa correlação, entretanto, não implica causalidade direta. Variáveis como capacidade econômica, infraestrutura de pesquisa e políticas públicas de ciência e tecnologia também influenciam a produção acadêmica (Rotimi *et al.*, 2024). Países como a China e Índia, além de sua vulnerabilidade climática, possuem sistemas abrangentes de pesquisa e

desenvolvimento, o que amplia sua representatividade (Zhao *et al.*, 2018; Madushani *et al.*, 2024).

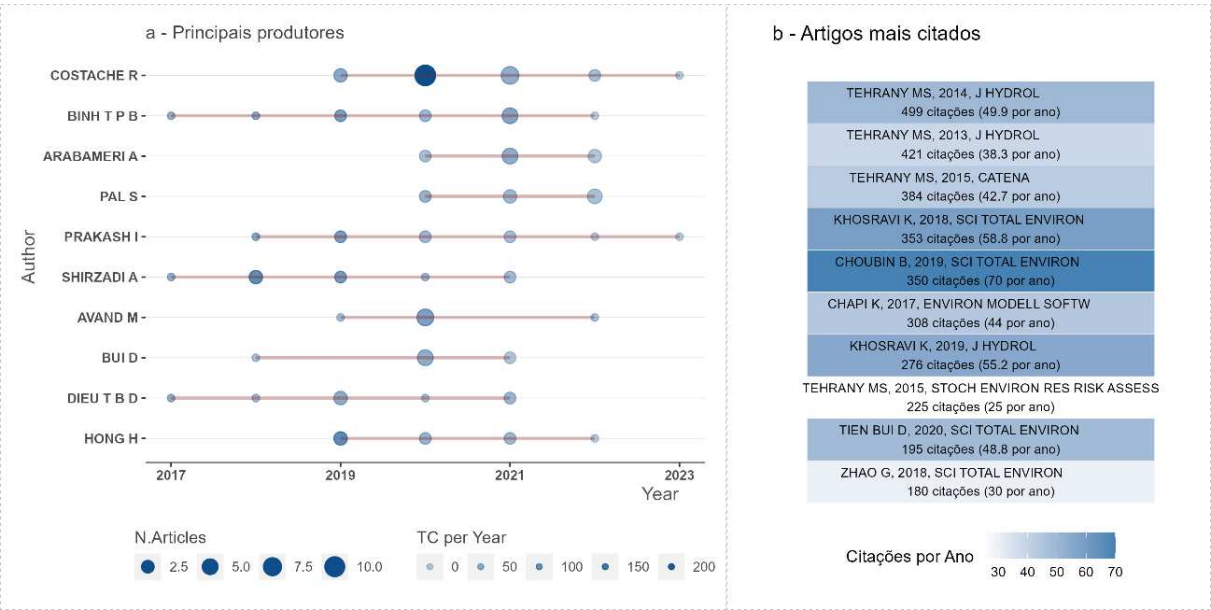
Apesar do avanço dos estudos em MSI com uso de aprendizado de máquina, ainda é evidente a escassez de pesquisas que contemplem regiões como a América Latina e a África Austral (Figura 28a), revelando uma lacuna significativa na representatividade espacial desses modelos. Essa sub-representação pode indicar possíveis limitações estruturais, como financiamento insuficiente, falta de colaborações internacionais ou priorização insuficiente de temas relacionados à resiliência climática (Tandon, 2021; Virgüez *et al.*, 2024). A assimetria geográfica identificada (Figura 28) reforça a necessidade de iniciativas multilaterais para capacitação técnica e fomento a redes de cooperação, especialmente em regiões marginalizadas. Embora existam estudos de MSI na América Latina e África, a escassez de abordagens baseadas em AM nessas regiões pode limitar a adoção de estratégias mais robustas e adaptativas de mitigação de riscos, ressaltando a importância de fomentar pesquisas com métodos inovadores em contextos locais.

4.2 PRINCIPAIS AUTORES E PUBLICAÇÕES

A análise bibliométrica desempenha um papel fundamental ao mapear lideranças intelectuais, identificar contribuições científicas influentes e detectar padrões emergentes em áreas de pesquisa específicas (Rathnasiri, 2023). Nesse contexto, o levantamento evidenciou um total de 607 autores, com média de 5,71 coautores por documento, evidenciando a estrutura colaborativa e a densidade intelectual do campo. Além disso, as métricas de citação revelam tanto a produtividade acadêmica quanto o impacto das publicações, permitindo não apenas apontar os pesquisadores mais ativos, mas também destacar aqueles cujos trabalhos servem como destaque para avanços subsequentes, refletindo a dinâmica de consolidação e evolução do conhecimento científico na área (Rentschler *et al.*, 2022; Rathnasiri, 2023).

A Figura 29-b sintetiza a contribuição dos autores mais produtivos, evidenciando aqueles cujas publicações registraram o maior número de citações. A relação observada entre o número de citações e o impacto científico aponta para as principais linhas de investigação priorizadas pela comunidade acadêmica. Além disso, a expressiva taxa de colaboração (média de 5,71 autores por documento) aponta uma tendência à interdisciplinaridade. Esses achados não apenas delineiam o cenário atual da produção científica, mas também orientam novas pesquisas ao evidenciar lacunas de conhecimento e oportunidades de sinergia entre diferentes grupos.

Figura 29 - Principais autores e artigos citados em estudos sobre suscetibilidade à inundação com aprendizado de máquina: a) autores; b) artigos



Fonte: Elaborada pelo autor (2024).

A pesquisadora Mahyat Shafapour Tehrany destacou-se como referência na modelagem de suscetibilidade à inundação utilizando AM, com publicações influentes nos anos de 2013 a 2023. Seu trabalho pioneiro, particularmente o artigo *"Flood susceptibility mapping using a novel ensemble weights-of-evidence and support vector machine models in GIS"* (Tehrany et al., 2014), introduziu um modelo *ensemble* inovador que integrou o método *weights of evidence* (WofE) e *support vector machine* (SVM). Essa abordagem mista, caracterizada por alta acurácia preditiva (AUC = 0,96), representou um marco metodológico ao superar limitações de modelos clássicos, consolidando-se como referência em um período marcado pela escassez de estudos aplicando técnicas de aprendizado de máquina (AM) para suscetibilidade à inundação por aprendizado de máquina (Apêndice A). Por ter sido uma das primeiras a integrar algoritmos de aprendizado de máquina em análises geoespaciais, sua contribuição passou a ser amplamente referenciada em estudos subsequentes.

Romulus Costache emerge como o autor mais produtivo do campo, com 24 artigos publicados e índice de autoria fracionada de 4,99 (Figura 28-a). Sua produção científica concentra-se na evolução de modelos *ensemble* e híbridos, incorporando avanços como *deep learning* para aprimorar a precisão de previsões (Costache e Bui, 2019; Costache et al., 2020, 2021). Seus estudos demonstram sinergia entre o progresso tecnológico em sensoriamento remoto e a disponibilização de dados geoespaciais de alta resolução, fatores que catalisaram a testagem de arquiteturas computacionais complexas. Paralelamente, Bahram Choubin sobressai

com impacto temporal alto: seu artigo de 2019, que aplicou um modelo *ensemble* combinando técnicas estatísticas, algoritmos de classificação e SVM (Choubin *et al.*, 2019), atingiu um índice de citação anual de ≈ 70 (Figura 28-b).

Coletivamente, esses pesquisadores parecem refletir diferentes trajetórias de contribuição: Tehrany com possíveis avanços metodológicos, Costache associado a uma maior produtividade técnica e Choubin frequentemente citado em contextos de aplicação translacional. Suas contribuições fortalecem um paradigma em que a hibridização de algoritmos clássicos e modernos, aliada à disponibilidade de *big data* geoespacial, potencializa a confiabilidade de modelos preditivos. Por meio das contribuições desses pesquisadores, observa-se indícios de uma possível tendência à adoção de abordagens interdisciplinares, integrando ciência de dados e geotecnologias como suporte à mitigação de incertezas na gestão de riscos hidrológicos.

4.3 PRINCIPAIS FATORES

A identificação das variáveis preditivas mais relevantes em modelos de suscetibilidade à inundação por aprendizado de máquina (AM) é essencial para compreender os mecanismos hidrológicos e geomorfológicos que influenciam a ocorrência de eventos extremos. Entre os 169 artigos analisados, 47 foram selecionados (Tabela 2 e 3) pela clareza metodológica e pela descrição explícita das variáveis independentes. A análise evidenciou um conjunto recorrente de fatores predisponentes, agrupados em cinco categorias principais: (i) parâmetros topográficos (p. ex., altitude, declividade, índice topográfico de umidade); (ii) proximidade a corpos hídricos (distância de rios, densidade de drenagem); (iii) características pedológicas e geológicas (tipo de solo, permeabilidade, litologia); (iv) uso e ocupação do solo; e (v) padrões pluviométricos (precipitação média, frequência de chuvas intensas).

Tabela 2 - Variáveis para modelagem de mapas de suscetibilidade à inundação encontrados na literatura

(continua)

Publicação	Principais variáveis	Quantidade de variáveis utilizadas
LAMOVEC <i>et al.</i> , 2013.	rios (distância); elevação; declividade; NDVI;	7
TEHRANY <i>et al.</i> , 2013.	rios (distância); elevação; TWI; declividade; SPI; curvatura; uso e ocupação do solo; precipitação; tipo de solo; geologia;	10
TEHRANY <i>et al.</i> , 2014.	rios (distância); elevação; declividade; SPI; curvatura; precipitação; geologia;	15
TEHRANY <i>et al.</i> , 2014.	rios (distância); elevação; TWI; declividade; SPI; curvatura; uso e ocupação do solo; precipitação; litologia; geologia;	10
TEHRANY <i>et al.</i> , 2015.	rios (distância); elevação; TWI; declividade; SPI; curvatura; uso e ocupação do solo; tipo de solo; geologia;	11
BUI <i>et al.</i> , 2016.	rios (distância); elevação; TWI; declividade; NDVI; SPI; curvatura; densidade de drenagem; precipitação; litologia;	10
LEE <i>et al.</i> , 2017.	rios (distância); elevação; TWI; declividade; SPI; curvatura; uso e ocupação do solo; geologia;	12
SACHDEVA <i>et al.</i> , 2017.	rios (distância); elevação; TWI; declividade; NDVI; SPI; curvatura; uso e ocupação do solo; precipitação; aspecto; geologia;	11
BUI <i>et al.</i> , 2018.	rios (distância); elevação; TWI; declividade; SPI; curvatura; uso e ocupação do solo; densidade de drenagem; precipitação; litologia;	10
LEE <i>et al.</i> , 2018.	rios (distância); elevação; TWI; declividade; SPI; uso e ocupação do solo; geologia;	10
LOLLO <i>et al.</i> , 2018.	elevação; uso e ocupação do solo;	2
SAMANTA <i>et al.</i> , 2018.	elevação; TWI; declividade; densidade de drenagem; precipitação; geologia;	8
SAMANTA <i>et al.</i> , 2018.	rios (distância); elevação; TWI; declividade; uso e ocupação do solo; litologia; tipo de solo;	8
SHAFIZADEH-MOGHADAM <i>et al.</i> , 2018.	rios (distância); elevação; TWI; declividade; NDVI; SPI; curvatura; uso e ocupação do solo; densidade de drenagem; precipitação; litologia;	11
WANG <i>et al.</i> , 2018.	rios (distância); elevação; TWI; declividade; NDVI; curvatura; uso e ocupação do solo; precipitação; litologia; geologia;	11
ZHAO <i>et al.</i> , 2018.	elevação; declividade; NDVI; curvatura; densidade de drenagem; precipitação;	12
ARABAMERI <i>et al.</i> , 2019.	rios (distância); elevação; TWI; declividade; NDVI; SPI; curvatura; uso e ocupação do solo; densidade de drenagem; aspecto; litologia; tipo de solo;	12

Fonte: Elaborada pelo autor (2024).

Tabela 2 - Variáveis para modelagem de mapas de suscetibilidade à inundação encontrados na literatura

(continuação)

Publicação	Principais variáveis	Quantidade de variáveis utilizadas
BUI <i>et al.</i> , 2019.	elevação; TWI; declividade; NDVI; SPI; curvatura; densidade de drenagem; precipitação; aspecto; litologia; tipo de solo;	12
CHOUBIN <i>et al.</i> , 2019.	rios (distância); elevação; TWI; declividade; curvatura; uso e ocupação do solo; densidade de drenagem; aspecto; litologia;	11
CHOWDHURI <i>et al.</i> , 2019.	rios (distância); elevação; TWI; NDVI; SPI; uso e ocupação do solo; precipitação; geologia;	8
FALAH <i>et al.</i> , 2019.	rios (distância); elevação; declividade; uso e ocupação do solo; densidade de drenagem;	5
KANANI-SADATA <i>et al.</i> , 2019.	rios (distância); elevação; TWI; declividade; NDVI; curvatura; precipitação;	10
KHOSRAVIA <i>et al.</i> , 2019.	rios (distância); elevação; TWI; declividade; NDVI; STI; SPI; curvatura; uso e ocupação do solo; precipitação; litologia; tipo de solo;	12
RAHMAN <i>et al.</i> , 2019.	elevação; declividade; uso e ocupação do solo; densidade de drenagem; precipitação; geologia;	9
SAHANA <i>et al.</i> , 2019.	rios (distância); elevação; TWI; declividade; curvatura; densidade de drenagem; precipitação; aspecto; tipo de solo; geologia;	12
TEHRANY <i>et al.</i> , 2019.	rios (distância); elevação; TWI; declividade; SPI; curvatura; uso e ocupação do solo; precipitação; aspecto; vias de transporte (distância); tipo de solo; geologia;	12
WANG <i>et al.</i> , 2019.	rios (distância); elevação; TWI; declividade; NDVI; STI; SPI; curvatura; uso e ocupação do solo; precipitação; aspecto; litologia; tipo de solo;	13
ZHAO <i>et al.</i> , 2019.	rios (distância); elevação; TWI; declividade; precipitação; vias de transporte (distância);	6
ALIA <i>et al.</i> , 2020.	rios (distância); elevação; TWI; declividade; NDVI; SPI; curvatura; uso e ocupação do solo; densidade de drenagem; precipitação; litologia;	14
CHEN <i>et al.</i> , 2020.	rios (distância); elevação; TWI; declividade; NDVI; SPI; curvatura; uso e ocupação do solo; precipitação; aspecto; geologia;	13
COSTACHE <i>et al.</i> , 2020.	rios (distância); elevação; TWI; declividade; curvatura; uso e ocupação do solo; precipitação; aspecto; litologia;	12
DODANGEH <i>et al.</i> , 2020.	rios (distância); elevação; declividade; NDVI; curvatura; uso e ocupação do solo; precipitação; aspecto; litologia;	9
FANG <i>et al.</i> , 2020.	rios (distância); elevação; TWI; declividade; NDVI; STI; SPI; curvatura; uso e ocupação do solo; precipitação; aspecto; litologia; geologia;	13
FANG <i>et al.</i> , 2020.	rios (distância); elevação; TWI; declividade; NDVI; STI; SPI; curvatura; uso e ocupação do solo; precipitação; aspecto; litologia; geologia;	13

Fonte: Elaborada pelo autor (2024).

Tabela 2 - Variáveis para modelagem de mapas de suscetibilidade à inundação encontrados na literatura

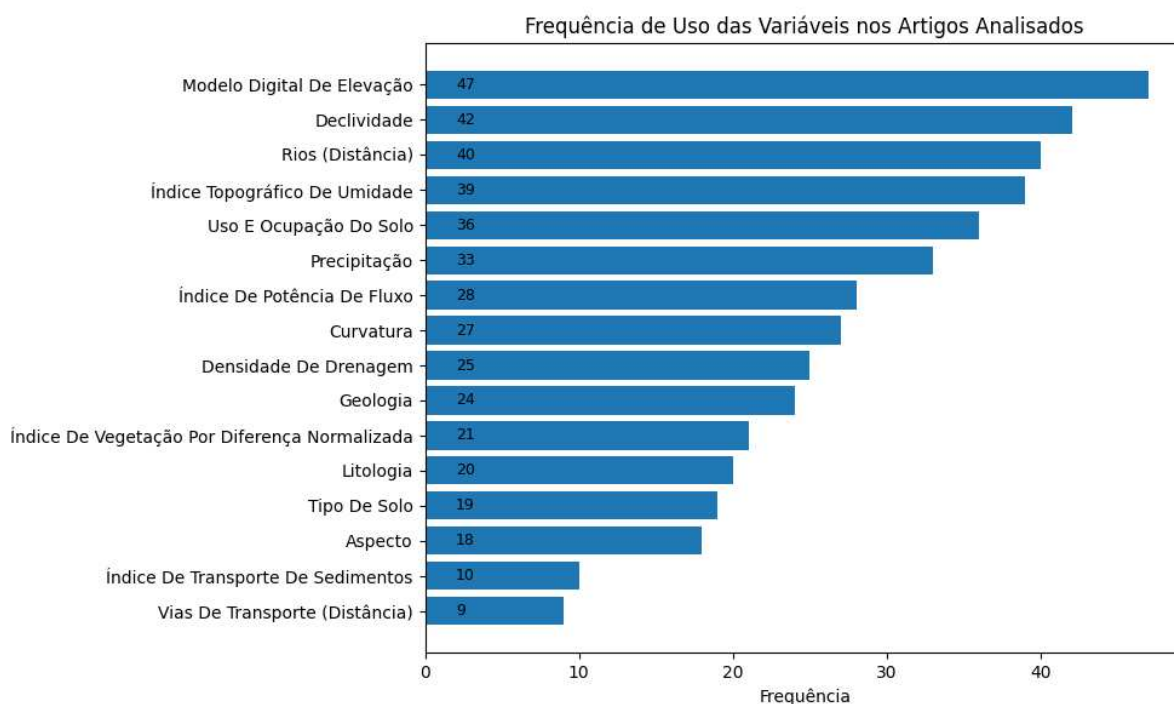
Publicação	Principais variáveis	Quantidade de variáveis utilizadas (conclusão)
MALIK <i>et al.</i> , 2020.	rios (distância); elevação; TWI; declividade; NDVI; SPI; curvatura; uso e ocupação do solo; densidade de drenagem; precipitação; aspecto; geologia;	14
MOSAVI <i>et al.</i> , 2020.	rios (distância); elevação; TWI; declividade; curvatura; uso e ocupação do solo; densidade de drenagem; aspecto; litologia; geologia;	10
NACHAPPA <i>et al.</i> , 2020.	rios (distância); elevação; TWI; declividade; NDVI; SPI; uso e ocupação do solo; precipitação; aspecto; vias de transporte (distância); geologia;	11
SOUISSI <i>et al.</i> , 2020.	rios (distância); elevação; TWI; declividade; uso e ocupação do solo; densidade de drenagem; precipitação; litologia;	8
TALUKDAR <i>et al.</i> , 2020.	rios (distância); elevação; TWI; declividade; STI; SPI; curvatura; uso e ocupação do solo; precipitação; aspecto; geologia;	12
TANGA <i>et al.</i> , 2020.	rios (distância); elevação; TWI; declividade; NDVI; STI; curvatura; uso e ocupação do solo; precipitação; aspecto; litologia; geologia;	12
VAFAKHAH <i>et al.</i> , 2020.	rios (distância); elevação; TWI; declividade; SPI; curvatura; uso e ocupação do solo; densidade de drenagem; precipitação; aspecto; litologia;	11
WANG <i>et al.</i> , 2020.	rios (distância); elevação; TWI; declividade; NDVI; STI; SPI; curvatura; uso e ocupação do solo; precipitação; aspecto; litologia; geologia;	13
ZHAO <i>et al.</i> , 2020.	rios (distância); elevação; TWI; declividade; densidade de drenagem; precipitação; vias de transporte (distância);	9
ARORA <i>et al.</i> , 2021.	rios (distância); elevação; TWI; declividade; curvatura; uso e ocupação do solo; densidade de drenagem; aspecto; vias de transporte (distância); tipo de solo;	12
ISLAM <i>et al.</i> , 2021.	rios (distância); elevação; TWI; declividade; STI; SPI; curvatura; uso e ocupação do solo; precipitação; aspecto; geologia;	12
PHAM <i>et al.</i> , 2021.	rios (distância); elevação; declividade; curvatura; uso e ocupação do solo; densidade de drenagem; aspecto; geologia;	10

Fonte: Elaborada pelo autor (2024).

Nos mapeamentos de suscetibilidade à inundação (Tabela 2), o número de variáveis utilizadas variou entre os estudos analisados, oscilando de apenas dois fatores (conforme LOLLO *et al.*, 2018) até quinze (TEHRANY *et al.*, 2014), com uma média de aproximadamente onze variáveis por artigo. Essa variação reflete as diferentes abordagens metodológicas, os objetivos específicos de cada pesquisa e as características singulares das bacias estudadas, que podem demandar a inclusão ou exclusão de determinados fatores preditivos. Entre os fatores mais recorrentes (Figura 30), destacam-se a elevação (modelo digital de elevação), declividade, distanciamento dos rios, índice topográfico de umidade e uso e ocupação do solo, sendo que a

elevação/altitude aparece em todos os trabalhos, sublinhando sua importância como variável fundamental. Essa prevalência evidencia que a altitude é considerada um elemento importante na determinação da suscetibilidade à inundação, influenciando diretamente a dinâmica hidrológica e a distribuição dos demais fatores ambientais. Além disso, a diversidade na seleção das variáveis aponta para a complexidade dos sistemas naturais e reforça a necessidade de abordagens adaptativas que integrem tanto aspectos geomorfológicos quanto climáticos e de uso do solo, possibilitando a construção de modelos preditivos mais robustos e representativos dos variados cenários de risco hidrogeomorfológico.

Figura 30 - Frequência de uso das variáveis em estudos de mapeamento de suscetibilidade à inundação (MSI)



Fonte: Elaborada pelo autor (2025).

A análise das frequências das principais variáveis de cada artigo revela um padrão consistente na utilização para estudos de suscetibilidade à inundação. A variável “elevação” destacou-se como a mais recorrente, com 47 ocorrências, seguida de “declividade” (42) e “rios (distância)” (40), evidenciando a ênfase dos pesquisadores em aspectos geomorfológicos e hidrológicos para a definição de áreas propensas à inundação. O índice topográfico de umidade com 39 registros, reforça a importância dos parâmetros de umidade do solo na predição de inundações, enquanto “uso e ocupação do solo” (36) e “precipitação” (33) sublinham a

relevância dos fatores relacionados à cobertura do solo e às condições climáticas. Em uma escala mais baixa, variáveis como SPI (28), curvatura (27) e densidade de drenagem (25) complementam os modelos, apontando para uma abordagem integrada que combina múltiplos atributos ambientais. Outras variáveis, incluindo geologia (24), NDVI (21), litologia (20), tipo de solo (19) e aspecto (18), também aparecem com regularidade, embora em menor escala. Por fim, parâmetros como o STI (10) e a distância a vias de transporte (9) demonstram uma aplicação mais restrita, sugerindo que esses indicadores podem funcionar de forma complementar, reforçando a complexidade e a multidimensionalidade dos estudos de mapeamento de inundações.

4.4 PRINCIPAIS ALGORITMOS E PROGRAMAS

Após a etapa de seleção, 43 artigos de um total inicial de 169 foram escolhidos para uma análise detalhada dos algoritmos e softwares empregados em suas respectivas pesquisas. Os resultados dessa análise estão sumarizados nas Tabelas 3 e 4, as quais apresentam um panorama dos principais algoritmos utilizados para os estudos de MSI. Estes algoritmos abrangem diversas áreas, incluindo aprendizado de máquina (supervisionado e não supervisionado), algoritmos de otimização, modelos estatísticos, modelos multicritério (MCDM) e lógica *fuzzy*, demonstrando a diversidade de abordagens metodológicas aplicadas.

Tabela 3 - Descrição dos principais algoritmos encontrados (continua)

Sigla	Descrição do algoritmo
ADT (1)	<i>Alternating Decision Tree</i> (Árvore de Decisão Alternada): combina árvores de decisão com o método de <i>boosting</i> . Cada nó da árvore é um <i>preditor</i> , e a árvore cresce alternando entre nós de <i>decisão</i> e nós de <i>predição</i> (Janizadeh <i>et al.</i> , 2019; Pham <i>et al.</i> , 2021).
AdaBoost (1)	<i>Adaptive Boosting</i> (Boosting Adaptativo): método <i>ensemble</i> que combina múltiplos classificadores fracos (como árvores de decisão) sequencialmente. Cada classificador subsequente dá mais peso aos exemplos que foram classificados incorretamente pelos classificadores anteriores, melhorando a precisão em dados difíceis. (Abu-Salih <i>et al.</i> , 2023; Saravanan <i>et al.</i> , 2023)

Fonte: Elaborada pelo autor (2024).

Tabela 3 - Descrição dos principais algoritmos encontrados

(continuação)

Sigla	Descrição do algoritmo
ANFIS (1/6)	<i>Adaptive Neuro-Fuzzy Inference System</i> (Sistema de Inferência Neuro-Fuzzy Adaptativo): combina redes neurais artificiais com sistemas de inferência fuzzy. A rede neural aprende a estrutura das regras fuzzy a partir dos dados, adaptando-se a relações não lineares e complexas. <i>Híbrido entre aprendizado de máquina e lógica fuzzy</i> (Bui <i>et al.</i> , 2018; Costache <i>et al.</i> , 2019; Vojtek <i>et al.</i> , 2021; Wang <i>et al.</i> , 2019).
ANN (1)	<i>Artificial Neural Network</i> (Rede Neural Artificial): modelo computacional inspirado na estrutura e função do cérebro biológico. Consiste em camadas de nós interconectados ("neurônios") que processam informações e aprendem padrões a partir dos dados (Sachdeva <i>et al.</i> , 2017; Shafizadeh-Moghadam <i>et al.</i> , 2018; Zhao <i>et al.</i> , 2018; Falah <i>et al.</i> , 2019; Rahman <i>et al.</i> , 2019; Rana and Mahanta, 2022; Singh <i>et al.</i> , 2020; Vafakhah <i>et al.</i> , 2020; Vojtek <i>et al.</i> , 2021; Zhao <i>et al.</i> , 2019; Nachappa <i>et al.</i> , 2020).
ANNRBF (1)	<i>Artificial Neural Network based on Radial Basis Function</i> (Rede Neural Artificial baseada em Função de Base Radial): tipo de rede neural que usa funções de base radial (RBFs) como funções de ativação na camada oculta. As RBFs medem a distância entre um ponto de entrada e um centro, tornando-as adequadas para interpolação e aproximação de funções (Pham <i>et al.</i> , 2021).
Bag-REPTree (1)	<i>Bagging with Reduced-Error Pruning Trees</i> (Bagging com Árvores de Poda com Erro Reduzido): método <i>ensemble</i> que combina <i>bagging</i> (<i>Bootstrap</i> Aggregating) com árvores de decisão que utilizam poda para reduzir o erro (Reduced-Error Pruning). O <i>bagging</i> cria múltiplos modelos treinando em subamostras aleatórias dos dados (Vafakhah <i>et al.</i> , 2020).
BCT (1)	<i>Boosted Classification Tree</i> (Árvore de Classificação Impulsionada): método <i>ensemble</i> que combina várias árvores de decisão "fracas" (com desempenho ligeiramente melhor que o aleatório) de forma sequencial. Cada árvore é treinada dando mais peso aos exemplos que foram classificados incorretamente pelas árvores anteriores, aumentando a precisão (Vojtek <i>et al.</i> , 2021).
BRT (1)	<i>Boosted Regression Tree</i> (Árvore de Regressão Impulsionada): similar ao BCT, mas aplicado a problemas de regressão (previsão de valores contínuos) em vez de classificação (Vafakhah <i>et al.</i> , 2020; Vojtek <i>et al.</i> , 2021).
BTR (1)	<i>Boosted Tree Regression</i> (Regressão por Árvores Impulsionadas): uma variação do BRT focada em regressão, otimizando a função de perda para previsões contínuas (Lee <i>et al.</i> , 2017; Shafizadeh-Moghadam <i>et al.</i> , 2018).
CART (1)	<i>Classification and Regression Tree</i> (Árvore de Classificação e Regressão): algoritmo que constrói árvores de decisão de forma recursiva, dividindo os dados em subconjuntos mais homogêneos com base em limiares de variáveis. Pode ser usado para classificação (variáveis categóricas) ou regressão (variáveis contínuas) (Costache and Bui, 2019; Vafakhah <i>et al.</i> , 2020).

Fonte: Elaborada pelo autor (2024).

Tabela 3 - Descrição dos principais algoritmos encontrados

(continuação)

Sigla	Descrição do algoritmo
CatBoost (1)	<i>Category Boosting</i> (Boosting de Categoria): algoritmo de <i>gradient boosting</i> projetado para lidar eficientemente com variáveis categóricas. Ele usa técnicas de codificação especializadas para evitar vazamento de dados e <i>overfitting</i> (Saravanan <i>et al.</i> , 2023).
CDT (1)	<i>Credal Decision Tree</i> (Árvore de Decisão Credal): uma generalização da árvore de decisão que usa conjuntos credais (conjuntos de distribuições de probabilidade) em vez de probabilidades únicas para representar a incerteza. Isso a torna mais robusta a dados imprecisos ou incompletos (Ngo <i>et al.</i> , 2023).
DEBP (1)	<i>Deep Belief Network based on Extreme Learning Machine with backpropagation and Particle Swarm Optimization</i> (Rede de Crença Profunda baseada em Máquina de Aprendizado Extremo com retropropagação e Otimização por Enxame de Partículas): arquitetura de aprendizado profundo que combina uma Rede de Crença Profunda (DBN) para extração de características com uma Máquina de Aprendizado Extremo (ELM) para classificação rápida. Utiliza retropropagação para ajuste fino e Otimização por Enxame de Partículas (PSO) para otimizar os pesos da rede (Pham <i>et al.</i> , 2021).
DT (1)	<i>Decision Tree</i> (Árvore de Decisão): estrutura de árvore hierárquica que representa decisões e seus possíveis resultados. Cada nó interno representa um teste em um atributo, cada ramo representa o resultado do teste, e cada nó folha representa uma classe ou valor de previsão (Tehrany <i>et al.</i> , 2013; Nachappa <i>et al.</i> , 2020; Vafakhah <i>et al.</i> , 2020; Yaseen <i>et al.</i> , 2022).
FT/FTree (1)	<i>Functional Tree</i> (Árvore Funcional): tipo de árvore de decisão onde os nós de decisão (internos) não contêm testes simples em atributos, mas sim <i>funções lineares</i> dos atributos. Isso permite modelar relações mais complexas nos dados (Janizadeh <i>et al.</i> , 2019; Pham <i>et al.</i> , 2021).
FSVM (1/6)	<i>Fuzzy Support Vector Machine</i> (Máquina de Vetores de Suporte Fuzzy): extensão da SVM que incorpora conceitos de lógica fuzzy. Atribui graus de pertinência fuzzy aos dados de treinamento, permitindo lidar com incerteza e sobreposição entre classes. Híbrido entre aprendizado de máquina e lógica fuzzy (Costache <i>et al.</i> , 2019).
GBoost (1)	<i>Gradient Boosting</i> (Boosting de Gradiente): método <i>ensemble</i> que combina múltiplos modelos fracos (geralmente árvores de decisão) sequencialmente. Cada modelo subsequente é treinado para corrigir os erros dos modelos anteriores, minimizando uma função de perda por meio do gradiente descendente (Abu-Salih <i>et al.</i> , 2023; Saravanan <i>et al.</i> , 2023).
KLR (1)	<i>Kernel Logistic Regression</i> (Regressão Logística Kernel): aplica o "truque do kernel" à regressão logística. O truque do kernel mapeia os dados para um espaço de dimensão superior (implicitamente), onde a separação linear se torna mais fácil, permitindo modelar relações não lineares. (Janizadeh <i>et al.</i> , 2019).
KNN (1)	<i>K-Nearest Neighbors</i> (K-Vizinhos Mais Próximos): algoritmo simples que classifica um novo ponto de dados com base na classe da maioria dos seus k vizinhos mais próximos no conjunto de treinamento. A distância é geralmente medida usando a distância euclidiana ou outra métrica de distância. Pode ser usado para classificação (supervisionado) ou agrupamento (não supervisionado) (Rana and Mahanta, 2022; Hasan <i>et al.</i> , 2023).

Fonte: Elaborada pelo autor (2024).

Tabela 3 - Descrição dos principais algoritmos encontrados

(continuação)

Sigla	Descrição do algoritmo
LMTTree (1)	<i>Logistic Model Tree</i> (Árvore de Modelo Logístico): combina árvores de decisão com modelos de regressão logística. As folhas da árvore contêm modelos logísticos em vez de rótulos de classe simples, permitindo previsões probabilísticas (Pham <i>et al.</i> , 2021).
LR (1/3)	<i>Logistic Regression</i> (Regressão Logística): modelo estatístico para prever a probabilidade de uma variável dependente categórica (geralmente binária). Usa uma função logística (sigmoide) para mapear uma combinação linear de variáveis independentes para um valor entre 0 e 1 (Tehrany <i>et al.</i> , 2013; Tehrany <i>et al.</i> , 2014; Sachdeva <i>et al.</i> , 2017; Lee <i>et al.</i> , 2018; Chowdhuri <i>et al.</i> , 2019; Rahman <i>et al.</i> , 2019; Zhao <i>et al.</i> , 2019; Vafakhah <i>et al.</i> , 2020; Pham <i>et al.</i> , 2021; Yaseen <i>et al.</i> , 2022, Chen <i>et al.</i> , 2023).
MLP (1)	<i>Multilayer Perceptron</i> (Perceptron Multicamadas): tipo de rede neural artificial com uma ou mais camadas ocultas entre a entrada e a saída. As camadas ocultas permitem que o MLP aprenda representações não lineares complexas dos dados (Janizadeh <i>et al.</i> , 2019; Costache and Bui, 2019; Chen <i>et al.</i> , 2023).
NB (1/3)	<i>Naive Bayes</i> (Naive Bayes): classificador probabilístico baseado no teorema de Bayes, com a suposição "ingênua" de que as variáveis preditoras são independentes entre si, dada a classe. Simples e eficiente, mas pode sofrer se a suposição de independência for violada (Khosravia <i>et al.</i> , 2019; Khosravi <i>et al.</i> , 2019).
NBT (1)	<i>Naive Bayes Tree</i> (Árvore de Bayes Naive): combina árvores de decisão com classificadores Naive Bayes. A árvore é construída de forma a particionar os dados, e um classificador Naive Bayes é treinado em cada nó folha (Khosravia <i>et al.</i> , 2019; Khosravi <i>et al.</i> , 2019).
QDA (1)	<i>Quadratic Discriminant Analysis</i> (Análise Discriminante Quadrática): classificador que assume que cada classe segue uma distribuição normal, mas, ao contrário da Análise Discriminante Linear (LDA), permite que cada classe tenha sua própria matriz de covariância. Isso permite modelar superfícies de decisão quadráticas (Janizadeh <i>et al.</i> , 2019).
QUEST (1)	<i>Quick, Unbiased, Efficient Statistical Tree</i> (Árvore Estatística Rápida, Não Viciada e Eficiente): um algoritmo de árvore de decisão projetado para ser rápido, ter menos viés de seleção de variáveis e ser computacionalmente eficiente (Vafakhah <i>et al.</i> , 2020).
RF (1)	<i>Random Forest</i> (Floresta Aleatória): método <i>ensemble</i> que constrói múltiplas árvores de decisão durante o treinamento. Cada árvore é treinada em um subconjunto aleatório dos dados e um subconjunto aleatório das variáveis preditoras. A previsão final é feita agregando (por exemplo, por votação majoritária ou média) as previsões de todas as árvores (Lamovec <i>et al.</i> , 2013; Lee <i>et al.</i> , 2017; Sachdeva <i>et al.</i> , 2017; Zhao <i>et al.</i> , 2018; Costache and Bui, 2019; Vafakhah <i>et al.</i> , 2020; Chen <i>et al.</i> , 2023; Hasan <i>et al.</i> , 2023; Nachappa <i>et al.</i> , 2020).
RS-REPTree (1)	<i>Random Subspace Reduced-Error Pruning Trees</i> (Árvores de Poda de Erro Reduzido com Subespaço Aleatório): similar ao Random Forest, mas em vez de amostrar aleatoriamente as amostras para cada árvore, amostra aleatoriamente as variáveis preditoras (o "subespaço"). Também usa poda para reduzir o erro. (Vafakhah <i>et al.</i> , 2020).
SGB (1)	<i>Stochastic Gradient Boosting</i> (Boosting de Gradiente Estocástico): variação do <i>gradient boosting</i> que introduz aleatoriedade no processo de treinamento. Em cada iteração, um subconjunto aleatório dos dados de treinamento é usado para ajustar o modelo fraco, o que pode melhorar a generalização e a velocidade. (Saravanan <i>et al.</i> , 2023).

Fonte: Elaborada pelo autor (2024).

Tabela 3 - Descrição dos principais algoritmos encontrados

(continuação)

Sigla	Descrição do algoritmo
SVM (1)	<i>Support Vector Machine</i> (Máquina de Vetores de Suporte): algoritmo que encontra o hiperplano que melhor separa as classes em um espaço de alta dimensão. O hiperplano é escolhido de forma a maximizar a margem entre as classes, o que geralmente leva a uma boa generalização. (Tehrany <i>et al.</i> , 2014; Tehrany <i>et al.</i> , 2015; Sachdeva <i>et al.</i> , 2017; Zhao <i>et al.</i> , 2018; Choubin <i>et al.</i> , 2019; Zhao <i>et al.</i> , 2019; Vafakhah <i>et al.</i> , 2020; Liu <i>et al.</i> , 2021; Vojtek <i>et al.</i> , 2021; Bera <i>et al.</i> , 2022; Rana and Mahanta, 2022; Chen <i>et al.</i> , 2023; Nachappa <i>et al.</i> , 2020; Costache <i>et al.</i> , 2020).
WELLSVM (1)	<i>Weakly-Labeled Support Vector Machine</i> (Máquina de Vetores de Suporte com Rótulos Fracos): variação do SVM projetada para lidar com dados onde os rótulos (classes) são imperfeitos, incompletos ou ruidosos (Zhao <i>et al.</i> , 2019).
XGB/XGBoost (1)	<i>Extreme Gradient Boosting</i> (Boosting de Gradiente Extremo): implementação otimizada do <i>gradient boosting</i> , conhecida por sua alta eficiência, precisão e flexibilidade. Inclui regularização para evitar <i>overfitting</i> , lida com dados esparsos e permite paralelização (Hasan <i>et al.</i> , 2023; Abu-Salih <i>et al.</i> , 2023; Saravanan <i>et al.</i> , 2023).
ANFIS-BBO (2)	<i>ANFIS with Biogeography-Based Optimization</i> (ANFIS com Otimização Baseada em Biogeografia): combina o ANFIS com o algoritmo de otimização BBO (Biogeography-Based Optimization). O BBO é inspirado na migração de espécies entre ilhas e é usado para otimizar os parâmetros do ANFIS (Wang <i>et al.</i> , 2019).
ANFIS-ICA (2)	<i>ANFIS with Imperialist Competitive Algorithm</i> (ANFIS com Algoritmo Competitivo Imperialista): combina o ANFIS com o algoritmo de otimização ICA (Imperialist Competitive Algorithm). O ICA é inspirado na competição sociopolítica entre impérios e é usado para ajustar os parâmetros do ANFIS (Wang <i>et al.</i> , 2019).
GARP (2)	<i>Genetic Algorithm for Rule-set Production</i> (Algoritmo Genético para Produção de Regras): algoritmo evolutivo que usa os princípios da seleção natural (seleção, cruzamento, mutação) para gerar um conjunto de regras de classificação. Cada regra é representada como um cromossomo, e a aptidão de cada regra é avaliada com base em sua capacidade de classificar corretamente os dados (Vafakhah <i>et al.</i> , 2020).
QPSO (2)	<i>Quantum Particle Swarm Optimization</i> (Otimização Quântica por Enxame de Partículas): uma variação do algoritmo de Otimização por Enxame de Partículas (PSO) que usa princípios da mecânica quântica para melhorar a capacidade de busca e evitar mínimos locais.
WOE (3)	<i>Weights of Evidence</i> (Pesos de Evidência): uma transformação de dados que converte variáveis categóricas (ou variáveis contínuas discretizadas) em valores numéricos que refletem a força preditiva de cada categoria em relação a uma variável dependente binária (Vafakhah <i>et al.</i> , 2020; Yaseen <i>et al.</i> , 2022; Costache and Bui, 2019).
FR (3)	<i>Frequency Ratio</i> (Razão de Frequência): mede a razão entre a probabilidade de ocorrência de um evento em uma determinada classe de uma variável preditora e a probabilidade de ocorrência do evento em toda a população. (Tehrany <i>et al.</i> , 2013; Tehrany <i>et al.</i> , 2015; Lee <i>et al.</i> , 2018; Samanta <i>et al.</i> , 2018; Arabameri <i>et al.</i> , 2019; Rahman <i>et al.</i> , 2019; Costache <i>et al.</i> , 2019; Vafakhah <i>et al.</i> , 2020; Nachappa <i>et al.</i> , 2020).

Fonte: Elaborada pelo autor (2024).

Tabela 3 - Descrição dos principais algoritmos encontrados

(continuação)

Sigla	Descrição do algoritmo
FR-AHP (3/4)	<i>Frequency Ratio - Analytic Hierarchy Process</i> (Razão de Frequência - Processo Analítico Hierárquico): combinação de FR e AHP. A FR avalia a importância de cada <i>categoria</i> dentro de uma variável, enquanto o AHP avalia a importância relativa das <i>variáveis</i> em si. (Vafakhah <i>et al.</i> , 2020).
FR-SVM (1/3)	<i>Frequency Ratio - Support Vector Machine</i> (Razão de Frequência - Máquina de Vetores de Suporte): combina FR e SVM, o FR é utilizado como pré-processamento. (Vafakhah <i>et al.</i> , 2020).
GAM (3)	<i>Generalized Additive Model</i> (Modelo Aditivo Generalizado): extensão da regressão linear que permite que a relação entre a variável dependente e as variáveis independentes seja modelada por meio de funções suaves (não necessariamente lineares). A variável dependente pode seguir diferentes distribuições. (Shafizadeh-Moghadam <i>et al.</i> , 2018).
GLM (3)	<i>Generalized Linear Model</i> (Modelo Linear Generalizado): generalização da regressão linear que permite que a variável dependente tenha uma distribuição diferente da distribuição normal (por exemplo, binomial, Poisson). A relação entre a média da variável dependente e as variáveis independentes é modelada por meio de uma função de ligação. (Vafakhah <i>et al.</i> , 2020).
IoE (3)	<i>Index of Entropy</i> (Índice de Entropia): medida de incerteza ou impureza em um conjunto de dados. Valores mais altos de entropia indicam maior desordem ou incerteza. Pode ser usado para seleção de variáveis ou para avaliar a qualidade de um modelo. (Costache <i>et al.</i> , 2020)
SE (3)	<i>Shannon Entropy</i> (Entropia de Shannon): medida da quantidade de informação ou incerteza em uma variável aleatória ou em um conjunto de dados. Amplamente utilizada em teoria da informação e aprendizado de máquina. (Vafakhah <i>et al.</i> , 2020)
SI (3)	<i>Statistical Index</i> (Índice Estatístico): termo genérico que abrange uma ampla variedade de índices estatísticos usados em diferentes contextos. A descrição específica dependeria do índice em questão. (Costache <i>et al.</i> , 2019)
AHP (4)	<i>Analytic Hierarchy Process</i> (Processo Hierárquico Analítico): método de tomada de decisão multicritério que estrutura um problema complexo em uma hierarquia de objetivos, critérios e alternativas. Usa comparações par a par para derivar pesos de importância para os critérios e pontuações para as alternativas. (Vafakhah <i>et al.</i> , 2020; Nachappa <i>et al.</i> , 2020; Rahman <i>et al.</i> , 2019; Bera <i>et al.</i> , 2022)
ANP (4)	<i>Analytic Network Process</i> (Processo de Rede Analítica): generalização do AHP que permite <i>dependências</i> e <i>feedback</i> entre os elementos da hierarquia (objetivos, critérios, alternativas). Modela problemas mais complexos onde os elementos não são independentes. (Nachappa <i>et al.</i> , 2020).
FAHP (4/6)	<i>Fuzzy Analytic Hierarchy Process</i> (Processo de Hierarquia Analítica Fuzzy): extensão do AHP que usa números fuzzy (e lógica fuzzy) para representar as comparações par a par. Isso permite lidar com a imprecisão e a subjetividade inerentes ao julgamento humano. (Costache <i>et al.</i> , 2020).
MCDM (4)	<i>Multi-Criteria Decision Making</i> (Tomada de Decisão Multicritério): campo de estudo que lida com a tomada de decisões em situações onde há múltiplos critérios, muitas vezes conflitantes, a serem considerados. (Khosravi <i>et al.</i> , 2019).

Fonte: Elaborada pelo autor (2024).

Tabela 3 - Descrição dos principais algoritmos encontrados

(conclusão)

Sigla	Descrição do algoritmo
VIKOR (4)	<i>VIsekriterijumska Optimizacija I Kompromisno Resenje</i> (Otimização Multicritério e Solução de Compromisso): um método MCDM que busca uma solução de compromisso que esteja mais próxima do ideal e mais distante do anti-ideal. (Khosravi <i>et al.</i> , 2019).
TOPSIS (4)	<i>Technique for Order Preference by Similarity to Ideal Solution</i> (Técnica para Ordenação de Preferências por Similaridade com a Solução Ideal): um método MCDM que escolhe a alternativa que é mais próxima da solução ideal (melhor desempenho em todos os critérios) e mais distante da solução anti-ideal (pior desempenho em todos os critérios). (Khosravi <i>et al.</i> , 2019).
SAW (4)	<i>Simple Additive Weighting</i> (Ponderação Aditiva Simples): um dos métodos MCDM mais simples. Atribui pesos aos critérios e calcula uma pontuação geral para cada alternativa somando os produtos dos pesos pelos valores normalizados dos critérios. (Khosravi <i>et al.</i> , 2019).
FL (5)	<i>Fuzzy Logic</i> (Lógica Fuzzy): extensão da lógica clássica (booleana) que permite representar e manipular graus de verdade, em vez de apenas valores binários (verdadeiro/falso). Lida com imprecisão e incerteza. (Kanani-Sadata <i>et al.</i> , 2019).
Lógica Neuro-Fuzzy (1/6)	<i>Neuro-Fuzzy Logic</i> (Lógica Neuro-Fuzzy): É uma combinação de técnicas de redes neurais artificiais e lógica fuzzy.(Yaseen <i>et al.</i> , 2022).
D-S (6)	<i>Dempster-Shafer theory</i> (Teoria de Dempster-Shafer): teoria matemática das evidências, usada para combinar informações de diferentes fontes e lidar com incertezas. Generaliza a teoria bayesiana da probabilidade. (Nachappa <i>et al.</i> , 2020).
EBF (6)	<i>Evidential Belief Function</i> (Função de Crença Evidencial): usado na teoria de Dempster-Shafer, representa o grau de crença em uma proposição com base em evidências. (Nachappa <i>et al.</i> , 2020).
MaxEnt (6)	<i>Maximum Entropy</i> (Entropia Máxima): princípio usado em modelagem, que busca a distribuição de probabilidade que melhor representa o estado atual do conhecimento, dado um conjunto de restrições (por exemplo, valores esperados de certas variáveis). Frequentemente usado em modelagem de nicho ecológico e processamento de linguagem natural. (Vafakhah <i>et al.</i> , 2020).

(1) aprendizado de máquina (supervisionado): inclui algoritmos que aprendem a partir de dados rotulados. A maioria dos algoritmos nesta categoria é usada para tarefas de classificação ou regressão. Alguns algoritmos podem se encaixar em outras categorias, dependendo do uso.

(2) algoritmos de otimização: inclui algoritmos projetados para encontrar a melhor solução para um problema, geralmente minimizando ou maximizando uma função objetivo.

(3) estatística / modelagem estatística: inclui métodos estatísticos tradicionais e modelos usados para analisar dados, encontrar relações entre variáveis e fazer previsões.

(4) modelos multicritério (MCDM): inclui métodos para tomar decisões em situações com múltiplos critérios, muitas vezes conflitantes.

(5) lógica *fuzzy*: refere-se a algoritmos/técnicas que utilizam a lógica *fuzzy*, que lida com graus de verdade em vez de valores binários (verdadeiro/falso).

(6) outros / híbridos: inclui algoritmos que não se encaixam perfeitamente nas outras categorias ou que combinam elementos de diferentes áreas.

Fonte: Elaborada pelo autor (2024).

Um dos algoritmos mais recorrentemente utilizados foi floresta aleatória (FA), um método *ensemble* que combina múltiplas árvores de decisão para alcançar maior precisão e robustez, como evidenciado nos trabalhos de Lamovec *et al.* (2013), Lee *et al.* (2017), Sachdeva

et al. (2017), Zhao *et al.* (2018), Costache e Bui (2019), Vafakhah *et al.* (2020), Chen *et al.* (2023), Hasan *et al.* (2023) e Nachappa *et al.* (2020). Sua capacidade de lidar com dados de alta dimensionalidade o tornam uma escolha atrativa em diversas aplicações.

O algoritmo máquina de vetores de suporte (MVS) também aparece com grande frequência, sendo empregado em estudos como os de Tehrany *et al.* (2014, 2015), Sachdeva *et al.* (2017), Zhao *et al.* (2018), Choubin *et al.* (2019), Zhao *et al.* (2019), Vafakhah *et al.* (2020), Liu *et al.* (2021), Wojtek *et al.* (2021), Bera *et al.* (2022), Rana e Mahanta (2022), Chen *et al.* (2023), Nachappa *et al.* (2020) e Costache *et al.* (2020). O algoritmo MVS se destaca por sua capacidade de encontrar o hiperplano ótimo que separa diferentes classes de dados, mesmo em espaços de alta dimensão, através do uso de *kernels*.

Algoritmos de *boosting*, que combinam múltiplos modelos fracos para criar um modelo forte, também são proeminentes (Janizadeh *et al.*, 2029). O Extreme Gradient Boosting (XGBoost), uma implementação otimizada do *gradient boosting*, é utilizado por Hasan *et al.* (2023), Abu-Salih *et al.* (2023) e Saravanan *et al.* (2023). Outras variações do *boosting*, como AdaBoost, *Gradient Boosting* e *Stochastic Gradient Boosting* (SGB), também foram aplicados, como visto nos trabalhos de Abu-Salih *et al.* (2023) e Saravanan *et al.* (2023).

Além destes, outros algoritmos de aprendizado de máquina supervisionado aparecem com menor frequência, mas ainda são relevantes, como *Decision Trees* (DT) (Tehrany *et al.*, 2013; Nachappa *et al.*, 2020; Vafakhah *et al.*, 2020; Yaseen *et al.*, 2022), *Logistic Regression* (LR) (Tehrany *et al.*, 2013, 2014; Sachdeva *et al.*, 2017; Lee *et al.*, 2018; Chowdhuri *et al.*, 2019; Rahman *et al.*, 2019; Zhao *et al.*, 2019; Vafakhah *et al.*, 2020; Pham *et al.*, 2021; Yaseen *et al.*, 2022; Chen *et al.*, 2023), *Functional Trees* (Janizadeh *et al.* 2019, Pham *et al.* 2021) e Naive Bayes (Khosravia *et al.*, 2019; Tang *et al.*, 2020).

A análise dos softwares empregados nos estudos (Tabela 4) revela uma predominância de ferramentas voltadas para sistemas de informação geográfica (SIG), com destaque para o ArcGIS, utilizado em diversas versões e extensões (Tehrany *et al.*, 2014; Tehrany *et al.*, 2014; Tehrany *et al.*, 2015; Lee *et al.*, 2017; Lee *et al.*, 2018; Samanta *et al.*, 2018; Shafizadeh-Moghadam *et al.*, 2018; Arabameri *et al.*, 2019; Bui *et al.*, 2019; Choubin *et al.*, 2019; Costache e Bui, 2019; Khosravia *et al.*, 2019; Costache *et al.*, 2019; Costache *et al.*, 2020; Nachappa *et al.*, 2020; Wojtek *et al.*, 2021; Bera *et al.*, 2022; Chen *et al.*, 2023). A frequência com que o ArcGIS é adotado pode estar associada à sua capacidade de gerenciar dados espaciais, realizar análises geoespaciais complexas e gerar mapas de alta qualidade, funcionalidades consideradas importantes em estudos de suscetibilidade à inundação.

Tabela 4 - Algoritmos e programas utilizados

(continua)

Publicação	Algoritmos Utilizados	Programas Utilizados
Tehrany <i>et al.</i> (2013)	LR, FR, DT	-
Lamovec <i>et al.</i> (2013)	RF	Weka, Clus
Tehrany <i>et al.</i> (2014)	LR	ArcGIS, SPSS V.19
Tehrany <i>et al.</i> (2014)	SVM	SPSS Clementine V.14.2, ArcGIS
Tehrany <i>et al.</i> (2015)	SVM, FR	ArcGIS, SPSS Modeler
Lee <i>et al.</i> (2017)	RF, BTR	STATISTICA, ArcGIS
Sachdeva <i>et al.</i> (2017)	RF, SVM, ANN, LR	-
Bui <i>et al.</i> (2018)	ANFIS	MATLAB, HEC-RAS
Lee <i>et al.</i> (2018)	LR, FR	ArcGIS, SPSS
Samanta <i>et al.</i> (2018)	FR	SPSS
Samanta <i>et al.</i> (2018)	FR	ArcGIS, Erdas Imagine
Shafizadeh-Moghadam <i>et al.</i> (2018)	ANN, GAM, BTR, MARS	ArcGIS; R
Zhao <i>et al.</i> (2018)	RF, SVM, ANN	MATLAB, R
Arabameri <i>et al.</i> (2019)	FR	ArcGIS, SPSS
Bui <i>et al.</i> (2019)	MARS	MATLAB, ArcGIS, IDRISI Selva
Choubin <i>et al.</i> (2019)	SVM	R, ArcGIS
Chowdhuri <i>et al.</i> (2019)	LR	SPSS
Costache e Bui (2019)	MLP, RF, CART (combinados com FR e WOE)	ArcCatalog; ArcGIS; MATLAB
Falah <i>et al.</i> (2019)	ANN	R
Janizadeh <i>et al.</i> (2019)	ADT, FT, KLR, MLP, QDA	-
Kanani-Sadata <i>et al.</i> (2019)	FL	SAGA; HEC-GeoHMS
Khosravia <i>et al.</i> (2019)	NBT, NB	ArcGIS
Khosravi <i>et al.</i> (2019)	MCDM (VIKOR, TOPSIS, SAW), NBT, NB	MATLAB, R software
Rahman <i>et al.</i> (2019)	ANN, LR, FR	-
Rahman <i>et al.</i> (2019)	ANN, AHP, LR, FR	-
Wang <i>et al.</i> (2019)	ANFIS-BBO, ANFIS-ICA, ANFIS	MATLAB
Zhao <i>et al.</i> (2019)	WELLSVM, LR, ANN, SVM	MATLAB, R software
Costache <i>et al.</i> (2019)	SI, FR, FSVM, ANFIS	Weka, MATLAB, ArcGIS
Costache <i>et al.</i> (2020)	FAHP, IoE, SVM, ...	Microsoft Excel, ArcGIS, Weka
Nachappa <i>et al.</i> (2020)	FR, AHP, ANP, SVM, RF, EBF, D-S, DT, LR, ANN	MATLAB, Google Earth Engine, ArcGIS

Fonte: Elaborada pelo autor (2024).

Tabela 4 - Algoritmos e programas utilizados

(conclusão)

Publicação	Algoritmos Utilizados	Programas Utilizados
Singh <i>et al.</i> (2020)	ANN	-
Vafakhah <i>et al.</i> (2020)	DT, FR, LR, SVM, WOE, AHP, FR-AHP, FR-SVM, SE, MaxEnt, BRT, GLM, GARP, QUEST, ANN, Bag-REPTree, RS-REPTree, Combinações de FR/WOE com ANN/RF/CART	MATLAB
Liu <i>et al.</i> (2021)	SVM	R software
Pham <i>et al.</i> (2021)	DEBP, ANNRBF, LR, LMTree, FTree, ADTree	MATLAB, WEKA
Vojtek <i>et al.</i> (2021)	ANN, ANFIS, SVM, BCT, BRT	ArcGIS
Bera <i>et al.</i> (2022)	AHP, SVM	ArcGIS
Rana and Mahanta (2022)	ANN, SVM, KNN	-
Yaseen <i>et al.</i> (2022)	WOE, LR, Lógica Neuro-Fuzzy, DT	-
Chen <i>et al.</i> (2023)	MLP, LR, SVM, RF	ArcGIS 10.2
Ngo <i>et al.</i> (2023)	QPSO, CDT	Geospatial data processing tools, Sentinel-1 C-band SAR images
Hasan <i>et al.</i> (2023)	RF, XGBoost, KNN	GIS techniques
Abu-Salih <i>et al.</i> (2023)	Extremely Randomized Trees, AdaBoost, GBoost, XGBoost	AutoML tools
Saravanan <i>et al.</i> (2023)	Adaboost, Gradient Boosting, XGB, CatBoost, SGB	-

Fonte: Elaborada pelo autor (2024).

O ambiente de programação MATLAB também se destaca como uma ferramenta importante, frequentemente utilizado para a implementação de algoritmos de aprendizado de máquina e modelagem estatística (Bui *et al.*, 2018; Zhao *et al.*, 2018; Bui *et al.*, 2019; Khosravi *et al.*, 2019; Wang *et al.*, 2019; Zhao *et al.*, 2019; Costache *et al.*, 2019; Nachappa *et al.*, 2020; Vafakhah *et al.*, 2020; Pham *et al.*, 2021). Sua flexibilidade e a disponibilidade de bibliotecas especializadas o tornam uma escolha popular para pesquisadores que buscam personalizar seus modelos e análises (Warren, 2015). Já a linguagem de programação R, com seu foco em computação estatística e gráficos, também aparece em vários estudos (Shafizadeh-Moghadam *et al.*, 2018; Zhao *et al.*, 2018; Choubin *et al.*, 2019; Falah *et al.*, 2019; Khosravi *et al.*, 2019; Zhao *et al.*, 2019; Liu *et al.*, 2021), evidenciando sua relevância na comunidade científica para análise de dados e desenvolvimento de modelos estatísticos (Aria; Cuccurullo, 2017).

Outras ferramentas mencionadas (Tabela 4) incluem o SPSS (Tehrany *et al.*, 2014; Tehrany *et al.*, 2014; Tehrany *et al.*, 2015; Lee *et al.*, 2018; Samanta *et al.*, 2018; Arabameri *et al.*

al., 2019; Chowdhuri *et al.*, 2019), amplamente utilizado para análises estatísticas, e o Weka (Lamovec *et al.*, 2013; Costache *et al.*, 2019; Costache *et al.*, 2020; Pham *et al.*, 2021), uma plataforma de aprendizado de máquina de código aberto. Além disso, ferramentas que automatizam a construção de modelos de aprendizado de máquina (*AutoML*), estão presentes no trabalho de Abu-Salih *et al.* (2023). Softwares menos frequentes, mas relevantes em contextos específicos, incluem STATISTICA (Lee *et al.*, 2017), HEC-RAS e HEC-GeoHMS (Bui *et al.*, 2018; Kanani-Sadata *et al.*, 2019), Erdas Imagine (Samanta *et al.*, 2018), IDRISI Selva (Bui *et al.*, 2019), SAGA (Kanani-Sadata *et al.*, 2019), e Google Earth Engine (Nachappa *et al.*, 2020).

5. RESULTADOS E DISCUSSÃO DOS MAPAS DE SUSCETIBILIDADE

Este capítulo apresenta os resultados obtidos mediante a aplicação da metodologia proposta, com ênfase na construção, avaliação e comparação de modelos preditivos para mapeamento de suscetibilidade às inundações na bacia hidrográfica do Rio Cachoeira. Foram empregados quatro algoritmos de aprendizado de máquina: floresta aleatória (FA) e três variantes do algoritmo máquina de vetores de suporte (MVS), diferenciadas por suas funções kernel (linear, polinomial e radial). A seleção desses métodos baseou-se em sua eficácia consolidada na literatura para interpretação de grandes volumes de dados, além de sua capacidade de lidar com complexidade e alta dimensionalidade, características intrínsecas nos estudos geoambientais que demandam precisão na classificação de padrões espaciais.

A avaliação dos modelos foi realizada com base em métricas consolidadas de desempenho, incluindo precisão, sensibilidade, especificidade e a área sob a curva ROC (*Receiver Operating Characteristic*), as quais permitiram quantificar a acurácia na predição de áreas suscetíveis às inundações. O processo metodológico abrangeu etapas de pré-processamento, como normalização dos dados e balanceamento de variáveis, seguidos de validação, por meio de divisão estratificada entre conjuntos de treinamento e teste. Essa abordagem garantiu a avaliação da generalização dos modelos, evitando *overfitting* e identificando as técnicas mais robustas para as particularidades do conjunto de dados em análise. Posteriormente, os modelos selecionados foram integrados às variáveis preditoras para gerar os mapas finais de suscetibilidade à inundação, considerando a resolução espacial original dessas variáveis, resultando na representação final dos níveis de suscetibilidade para a região estudada.

5.1 INVENTÁRIO DE INUNDAÇÃO E NÃO INUNDAÇÃO

A delimitação das áreas suscetíveis à inundação na BHRC foi conduzida seguindo a abordagem metodológica descrita no capítulo de metodologia. Esta abordagem combinou registros históricos de eventos de inundação, documentados pela Defesa Civil de Joinville e pelo FIDE, com a delimitação de áreas sob influência de marés elevadas (definidas como cota de 2,5 metros, em concordância com o Plano Diretor de Drenagem Urbana) e zonas correspondentes aos leitos permanentes de rios. Para assegurar a representatividade espacial da amostra de áreas inundáveis e manter o balanço estatístico, um algoritmo de amostragem

aleatória estratificada foi implementado para gerar pontos nas áreas consolidadas de inundação (leitos fluviais e zonas de influência de maré). Este procedimento visou preservar a proporção amostral entre as classes de áreas inundáveis e não inundáveis, mitigando o risco de viés decorrente da sobreposição desproporcional de uma determinada fonte de dados.

Em contrapartida, as áreas não inundáveis, conforme explicado na metodologia, foram identificadas com base em critérios geomorfológicos objetivos (conforme explicado na metodologia), como topos de morros (altitude $\geq 150\text{m}$), encostas com alta inclinação ($> 8\%$, derivadas de modelos digitais de elevação) e rotas seguras homologadas pela Defesa Civil, mantendo distância mínima de 100 metros de cursos hídricos e ausência de planícies aluviais. Um algoritmo de seleção aleatória foi aplicado para evitar aglomeração de pontos em micro-regiões. O resultado final é um inventário binário (Apêndice B) com 250 pontos de inundação e 237 de não inundação, unificados e validados, combinando técnicas estatísticas e de geoprocessamento para garantir a precisão e representação espacial nas amostras de teste e treino. A Tabela 5 mostra a listagem dos identificadores espaciais (registros) dos pontos classificados como "Inundação" e "Não-Inundação", distribuídos aleatoriamente entre os conjuntos de treino (70%) e teste (30%). Essa divisão foi empregada para treinar e validar os modelos de suscetibilidade à inundação, assegurando representatividade e equilíbrio entre as classes.

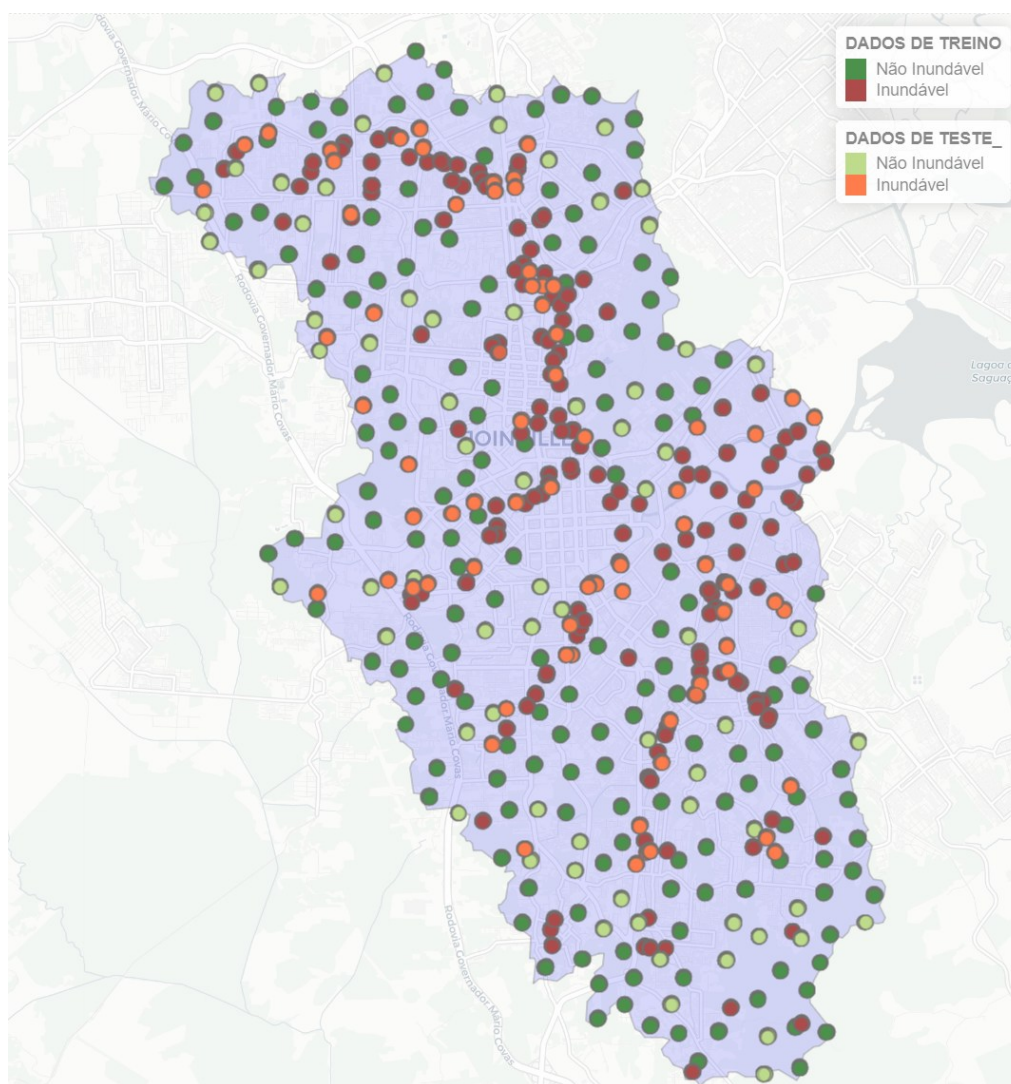
Tabela 5 - Distribuição aleatória das amostras (pontos)

Tipo de amostra	Registros
Treino Inundação (175)	95, 46, 224, 36, 328, 74, 330, 239, 35, 117, 115, 212, 125, 33, 20, 246, 345, 257, 15, 305, 219, 45, 253, 222, 249, 66, 17, 27, 108, 314, 57, 333, 26, 118, 240, 10, 41, 34, 311, 237, 23, 67, 209, 22, 324, 43, 232, 301, 98, 16, 116, 30, 331, 64, 90, 78, 220, 9, 6, 96, 319, 124, 215, 109, 225, 91, 309, 14, 332, 44, 315, 275, 51, 113, 338, 215, 24, 245, 79, 336, 208, 344, 72, 244, 28, 251, 313, 241, 201, 200, 256, 102, 341, 1, 77, 73, 221, 230, 141, 337, 103, 217, 56, 63, 306, 75, 310, 300, 87, 128, 228, 233, 223, 53, 202, 68, 94, 132, 243, 93, 307, 235, 11, 308, 302, 59, 123, 317, 8, 92, 329, 60, 238, 4, 107, 47, 234, 70, 343, 127, 76, 215, 312, 323, 100, 3, 226, 52, 18, 29, 318, 110, 25, 250, 82, 114, 304, 252, 86, 65, 99, 131, 227, 7, 19, 58, 130, 84, 122, 320, 178, 119, 171, 204, 203
Treino Não-Inundação (166)	564, 666, 733, 654, 504, 715, 633, 672, 573, 660, 642, 590, 519, 673, 699, 521, 687, 629, 698, 720, 696, 525, 529, 722, 622, 526, 718, 663, 639, 583, 580, 569, 625, 510, 728, 635, 561, 516, 554, 647, 727, 582, 684, 512, 734, 596, 678, 661, 714, 543, 540, 614, 688, 681, 559, 545, 607, 620, 655, 693, 601, 719, 644, 691, 735, 542, 662, 604, 706, 707, 501, 605, 528, 613, 736, 612, 649, 552, 697, 616, 566, 626, 703, 572, 585, 668, 539, 731, 577, 535, 730, 670, 643, 638, 671, 524, 518, 679, 646, 594, 592, 507, 683, 705, 544, 575, 606, 634, 621, 597, 568, 579, 556, 717, 513, 675, 576, 624, 598, 686, 557, 610, 711, 724, 560, 656, 503, 716, 530, 689, 664, 632, 640, 701, 581, 608, 534, 617, 713, 586, 690, 658, 600, 641, 665, 574, 505, 726, 712, 593, 602, 570, 558, 531, 657, 509, 619, 702, 567, 550, 609, 551, 615, 587, 515, 584
Teste Inundação (75)	129, 13, 335, 346, 321, 210, 81, 48, 50, 342, 205, 85, 106, 71, 327, 347, 334, 255, 130, 31, 61, 206, 231, 207, 303, 225, 254, 55, 38, 121, 89, 322, 325, 21, 247, 69, 40, 326, 101, 154, 39, 80, 112, 214, 242, 213, 2, 211, 340, 126, 248, 236, 32, 218, 88, 217, 96, 54, 316, 111, 105, 12, 37, 120, 229, 83, 42, 339, 130, 62, 216, 104, 5, 97, 49
Teste Não Inundação (71)	599, 682, 650, 536, 677, 527, 652, 667, 523, 506, 532, 571, 578, 636, 725, 546, 700, 537, 549, 721, 588, 692, 522, 659, 514, 603, 653, 708, 508, 729, 541, 631, 695, 669, 548, 547, 704, 630, 611, 618, 591, 555, 553, 651, 511, 502, 562, 637, 723, 676, 674, 623, 627, 538, 732, 710, 709, 500, 517, 563, 565, 648, 533, 645, 595, 694, 589, 685, 628, 680, 520

Fonte: Elaborada pelo autor (2025)

A partir dos 487 pontos georreferenciados (Figura 31), cada registro foi convertido em um polígono quadrado, permitindo a geração de 13 pixels (10x10m) por unidade geométrica. Essa abordagem metodológica foi necessária para padronizar os mapas temáticos em uma única base *raster* com resolução espacial homogênea.

Figura 31- Divisão espacial dos dados de treino e teste



Fonte: Elaborada pelo autor (2024).

Para garantir a estratificação adequada dos pixels nas amostras de treinamento (70%) e teste (30%), aplicou-se uma divisão por registro pontual. Esse procedimento evita o viés de autocorrelação do modelo final, assegurando que os 13 pixels derivados de um mesmo ponto original fossem alocados integralmente em apenas uma das divisões (treino ou teste). Dessa forma, preservou-se a independência estatística entre os conjuntos, uma vez que os *pixels* de mesma origem geométrica compartilham dependência espacial.

5.2 MAPAS TEMÁTICOS

Conforme os procedimentos estabelecidos na seção ‘Coleta de Variáveis Independentes’, o modelo de aprendizado de máquina aplicado neste estudo incorporou um conjunto multifatorial para avaliação da suscetibilidade à inundação na bacia hidrográfica do Rio Cachoeira (BHRC). Os parâmetros preditivos selecionados abrangeram variáveis geomorfológicas, hidrológicas, climáticas e antrópicas, detalhadas a seguir:

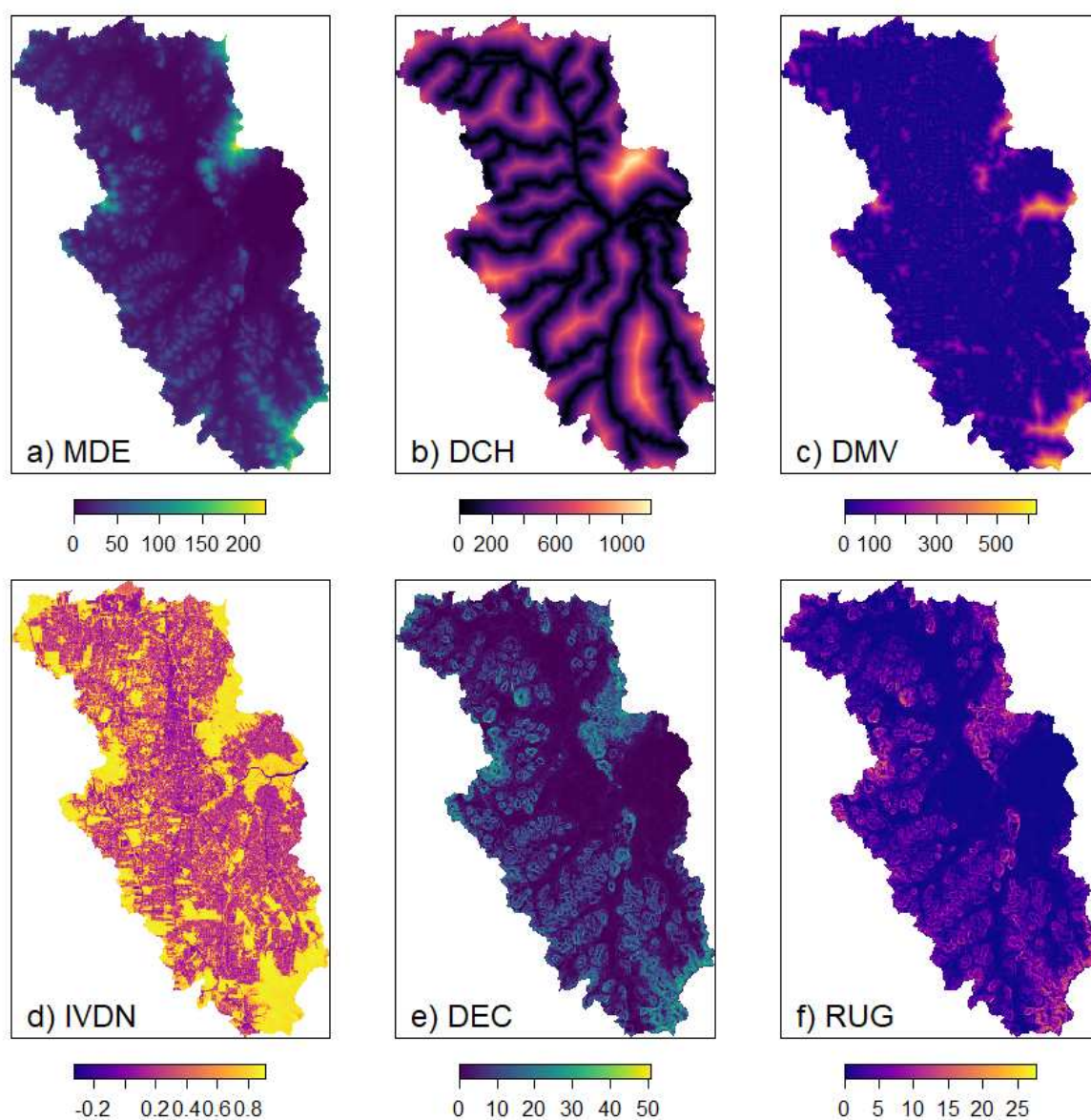
- a) modelo digital de elevação (MDE): base topográfica para análise de gradientes;
- b) distância de corpos hídricos (DCH) e distância da malha viária (DMV): indicadores de proximidade a rios e infraestrutura urbana;
- c) índice de vegetação por diferença normalizada (IVDN): métrica de cobertura vegetal;
- d) declividade do terreno (DEC): derivada do MDE para análise de inclinação;
- e) índice de rugosidade do terreno (RUG): quantificação da heterogeneidade superficial, influenciando a velocidade de escoamento;
- f) índices hidrológicos:
 - índice de umidade topográfico (IUT): estimativa de acumulação de água no solo;
 - índice de potência de escoamento (IPE): representação do potencial erosivo associado ao fluxo hídrico;
- g) orientação do terreno (ASP): também conhecida como aspecto, refere-se à direção cardinal para a qual uma encosta está voltada, influenciando diretamente a exposição à radiação solar, os processos de evapotranspiração, a distribuição da umidade do solo e a dinâmica do escoamento superficial;
- h) precipitações médias anuais (PREC): evento modelado estatisticamente;
- i) curvaturas do terreno: perfil (CUR_PER) e plana (CUR_PLAN), relacionadas à convergência/divergência de fluxos;
- j) atributos pedológicos: tipo de solo (TSO) e uso e ocupação do solo (UOS), obtidos por mapeamento georreferenciado;
- k) influência de marés astronômicas (MARE): modulação de níveis de água em zonas estuarinas. Sendo representado por 0 (sem influência) e 1 (com influência).

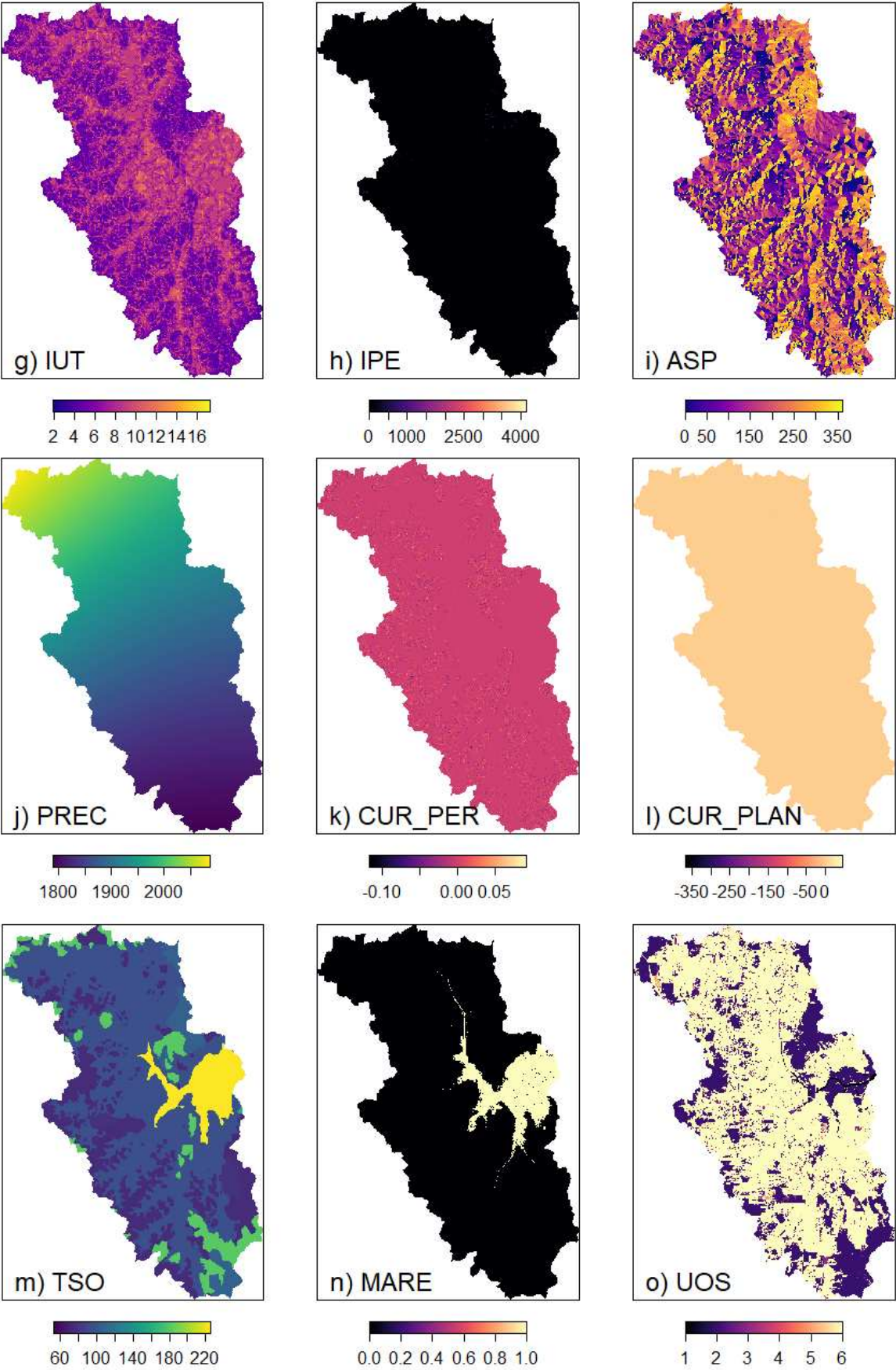
A integração dessas variáveis em um sistema de informação geográfica (SIG) permitiu a espacialização de padrões críticos, garantindo uma análise sistêmica das interações entre fatores naturais e antrópicos nas dinâmicas de inundação. A seleção criteriosa dos parâmetros fundamenta-se em sua relevância comprovada na literatura para modelagem de riscos

hidrológicos, assegurando robustez metodológica ao estudo. Como resultado se obteve os mapas temáticos representados na Figura 32.

Figura 32 - Mapas temáticos das variáveis independentes: a) MDE, b) DCH, c) DMV, d) IVDN, e) DEC, f) RUG, g) IUT, h) IPE, i) ASP, j) PREC, k) CUR_PER, l) CUR_PLAN, m) TSO, n) MARE, o) UOS

(continua)



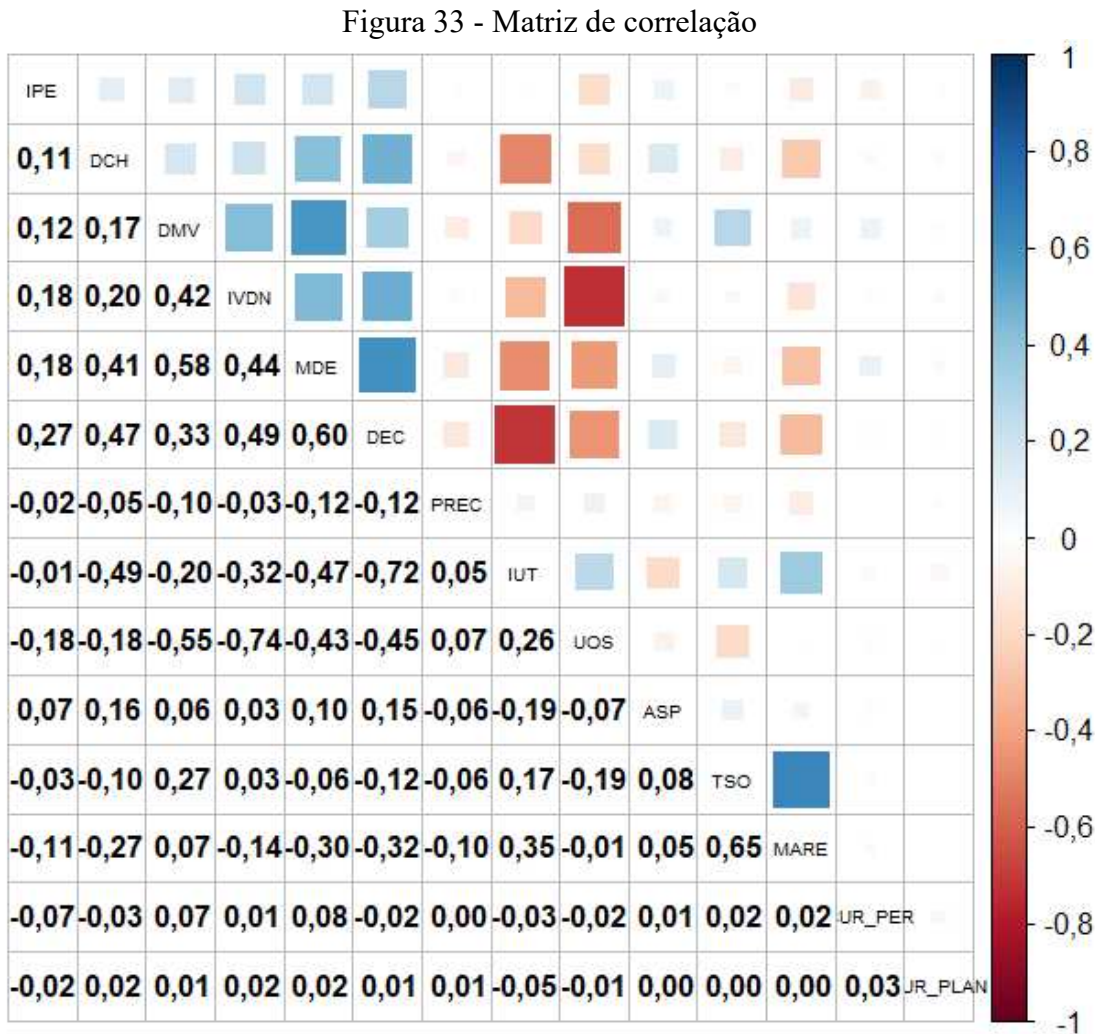


Fonte: Elaborada pelo autor (2025).

A Bacia Hidrográfica do Rio Cachoeira, inserida no contexto urbano de Joinville, apresenta uma dinâmica de uso e ocupação do solo marcada pela predominância de áreas antrópicas, conforme evidenciado pela análise das classes mapeadas (Figura 29-o). A classe predominante é a área urbanizada (classe 6), correspondendo a 69,75% da bacia. Esse valor reflete a intensa pressão antrópica associada à expansão urbana, característica de regiões metropolitanas. A concentração de edificações e infraestrutura tende a amplificar desafios como impermeabilização do solo, aumento do escoamento superficial e redução da capacidade de infiltração, fatores críticos para a ocorrência de inundações. Em segundo lugar, destaca-se a floresta densa (classe 2), que ocupa 24,50% da área.

5.3 RESULTADO DA CORRELAÇÃO

A Figura 33 representa a correlação de Person entre todas as variáveis propostas nesse estudo.



Fonte: Elaborada pelo autor (2025).

No gráfico apresentado, destaca-se a correlação alta (próxima de 0,99) entre declividade (DEC) e rugosidade (RUG), indicando que ambas descrevem, em grande parte, a mesma característica topográfica. Além disso, há uma correlação moderada (cerca de 0,65) entre tipo de solo (TSO) e manchas de maré (MARE), sugerindo que certos tipos de solo podem ocorrer majoritariamente em áreas de influência de maré. Já as demais variáveis, como distância de cursos d'água (DCH), distância da malha viária (DMV) e precipitação (PREC), mostram correlações mais baixas com os fatores topográficos, o que indica que trazem informações adicionais possivelmente relevantes para o modelo. Esses resultados fornecem subsídios para, por exemplo, avaliar a remoção ou combinação de DEC e RUG e, ao mesmo tempo, manter variáveis como PREC e DCH, que parecem contribuir com informações complementares na análise de suscetibilidade.

5.4 RESULTADO DA MULTICOLINEARIDADE

Para análises espaciais, recomenda-se um limiar conservador de $VIF < 10$, embora limiares mais rigorosos ($VIF < 5$) sejam adotados em contextos com alta sensibilidade a variáveis correlacionadas, como em modelos ecológicos ou previsões baseadas em sensoriamento remoto (Dormann *et al.*, 2013; Montgomery *et al.*, 2012).

Com base na primeira análise realizada (Tabela 6), observou-se que as variáveis declividade ($VIF = 49,79$) e índice de rugosidade ($VIF = 47,97$) apresentavam valores extremamente elevados, indicando alta colinearidade. Após a remoção da variável rugosidade, houve redução do VIF da variável declividade para 3,05, além de uma queda geral dos valores das demais variáveis. Essa mudança ocorreu porque a rugosidade compartilhava uma grande quantidade de informações com a variável declividade, elevando a multicolinearidade do modelo como um todo. Ao remover a rugosidade, reduziu-se a redundância, proporcionando uma representação mais clara e independente das variáveis restantes.

Diante dos diagnósticos estatísticos de multicolinearidade ($VIF > 10$) e correlação linear elevada ($r > 0.8$) identificados na matriz de correlação, a variável rugosidade (RUG) foi excluída do modelo preditivo (Dormann *et al.*, 2013; Montgomery *et al.*, 2012). Essa etapa foi fundamental para manter a validade estatística (Belsley *et al.*, 1980), precisão preditiva (James *et al.*, 2013) e interpretação causal inequívoca das variáveis independentes. A remoção de preditores redundantes é particularmente crítica em modelos espaciais aplicados a cenários de avaliação de suscetibilidade à inundação, na qual a sobreposição de fatores morfométricos (ex.: declive, curvatura) pode mascarar relações hidrológicas reais (Meraj *et al.*, 2015). A

rigoriedade nessa fase assegura que o modelo final atenda aos pressupostos de independência entre as variáveis, evitando *overfitting* e viés na inferência paramétrica (Zuur *et al.*, 2010).

Tabela 6 - Análise da multicolinearidade

N	Variável	Fator de inflação da variância:		Fator de inflação da variância:	
		com a variável rugosidade (RUG)		sem a variável rugosidade (RUG)	
1	MDE	2,437966	<10	2,442641	<10
2	DCH	1,165370	<10	1,177543	<10
3	DMV	1,842531	<10	1,937939	<10
4	IVDN	3,048568	<10	3,107651	<10
5	DEC	49,793613	≥ 10	3,053143	<10
6	RUG	47,967396	≥ 10	-	
7	IUT	2,525308	<10	2,436701	<10
8	IPE	1,204835	<10	1,197084	<10
9	ASP	1,045871	<10	1,056187	<10
10	PREC	1,030514	<10	1,031689	<10
11	CUR PER	1,129691	<10	1,099755	<10
12	CUR PLAN	1,176300	<10	1,157477	<10
13	TSO	2,038147	<10	2,114807	<10
14	MARE	2,175749	<10	2,309378	<10
15	UOS	3,238174	<10	3,312950	<10

Fonte: Elaborada pelo autor (2025).

5.5 PROCESSAMENTO DOS MODELOS

Após análises da correlação e da multicolinearidade o processo de seleção resultou em um conjunto final de 14 variáveis preditoras: MDE (Modelo Digital de Elevação); DCH (Distanciamento dos Cursos Hídricos); DMV (Distanciamento da Malha Viária); IVDN (Índice de Vegetação por Diferença Normalizada); DEC (Declividade); IUT (Índice de Umidade Topográfico); IPE (Índice de Potência de Escoamento); ASP (Aspecto); PREC (Precipitação); CUR_PER (Curvatura do Perfil); CUR_PLAN (Curvatura do Plano); TSO (Tipo de Solo); MARE (Manchas de Maré); UOS (Uso e Ocupação do Solo).

Os algoritmos FA e MVS, implementados em linguagem R, foram executados em uma máquina de porte leve com processador Intel i7 (10ª geração). Cada modelo demandou um tempo máximo de 20 segundos para treinamento e validação, com base em parâmetros otimizados. Essa performance demonstra a viabilidade técnica da metodologia para aplicações em escalas regionais, desde que apoiada por infraestrutura computacional compatível com as demandas de memória e armazenamento. O ciclo completo de geração de mapas de suscetibilidade, englobando leitura de variáveis, modelagem estatística e renderização cartográfica, foi concluído em 2 minutos para uma área de 82 km² (820.000 pixels 10x10m),

destacando a colaboração entre algoritmos de aprendizado de máquina e arquiteturas computacionais acessíveis.

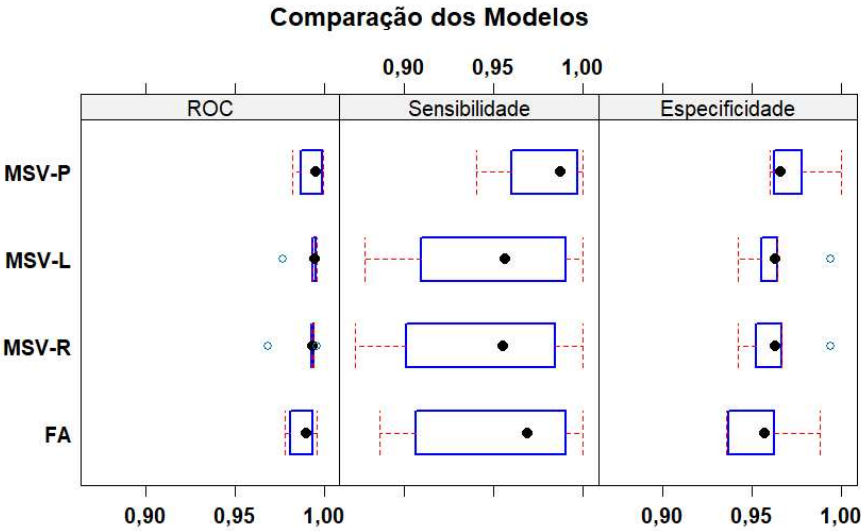
O Apêndice D apresenta o código resumido utilizado na metodologia deste estudo, permitindo a replicação das análises realizadas. Esse código pode ser adaptado para processos automatizados, incorporando etapas como seleção de variáveis com base em critérios estatísticos ($p < 0,01$), diagnóstico de multicolinearidade pelo fator de inflação de variância ($VIF > 10$) e análise de correlação cruzada ($\rho > 0,8$), além de procedimentos de pré-processamento adaptativos para a normalização de dados heterogêneos.

5.6 RESULTADOS DOS MODELOS DE TREINAMENTO

Neste estudo, a construção final dos modelos baseou-se na seleção das 14 variáveis preditoras, definidas mediante análise de relevância estatística e consistência teórica com os processos condicionantes da suscetibilidade à inundação na área de estudo.

Na Figura 34 é apresentado o resultado da função ‘*bwplot*’ (*box-and-whisker plot*) do pacote ‘*caret*’ mostrando a distribuição das métricas AUC, sensibilidade e especificidade, para cada modelo, ao longo de várias reamostragens realizadas no processo de treinamento. Isto facilita a comparação visual entre os modelos, mostrando informações como mediana, intervalo interquartil (a “caixa”) e possíveis valores atípicos. Dessa forma, é possível verificar não apenas a média de desempenho, mas também a variabilidade do modelo, o que contribui para uma análise mais robusta.

Figura 34 - Diagramas de caixa (*box plot*) de sensibilidade e especificidade dos modelos de treinamento



Fonte: Elaborada pelo autor (2024).

Todos os modelos de FA, MSV-L, MSV-R e MSV-P, apresentaram desempenhos elevados nas métricas de ROC, sensibilidade e especificidade. Observa-se que, em termos de ROC, MSV-P exibe valores medianos ligeiramente superiores, o que indica uma maior capacidade de discriminar entre as classes positivas e negativas, enquanto a FA, a MSV-R e a MSV-L também demonstram ROC robusta, embora de forma um pouco inferior. Quanto à sensibilidade (correta identificação de áreas inundáveis), todos os métodos atingiram patamares elevados, com MSV-P e FA registrando medianas marginalmente mais altas, sugerindo que esses algoritmos são menos propensos a falhar na identificação de instâncias positivas. No que diz respeito à especificidade (correta classificação de áreas não inundáveis), os resultados próximos de 1 demonstram uma baixa incidência de falsos positivos, destacando-se a MSV-P e a MSV-L, que apresentam medianas mais altas e menor variabilidade, indicando um bom desempenho na correta identificação de verdadeiros negativos.

De modo geral, todos os algoritmos apresentaram desempenho consistente e estável, com elevada acurácia ($\text{ROC} \geq 0,95$), baixa dispersão e equilíbrio entre sensibilidade e especificidade.

5.7 HIPERPARÂMETROS DOS MODELOS FINAIS

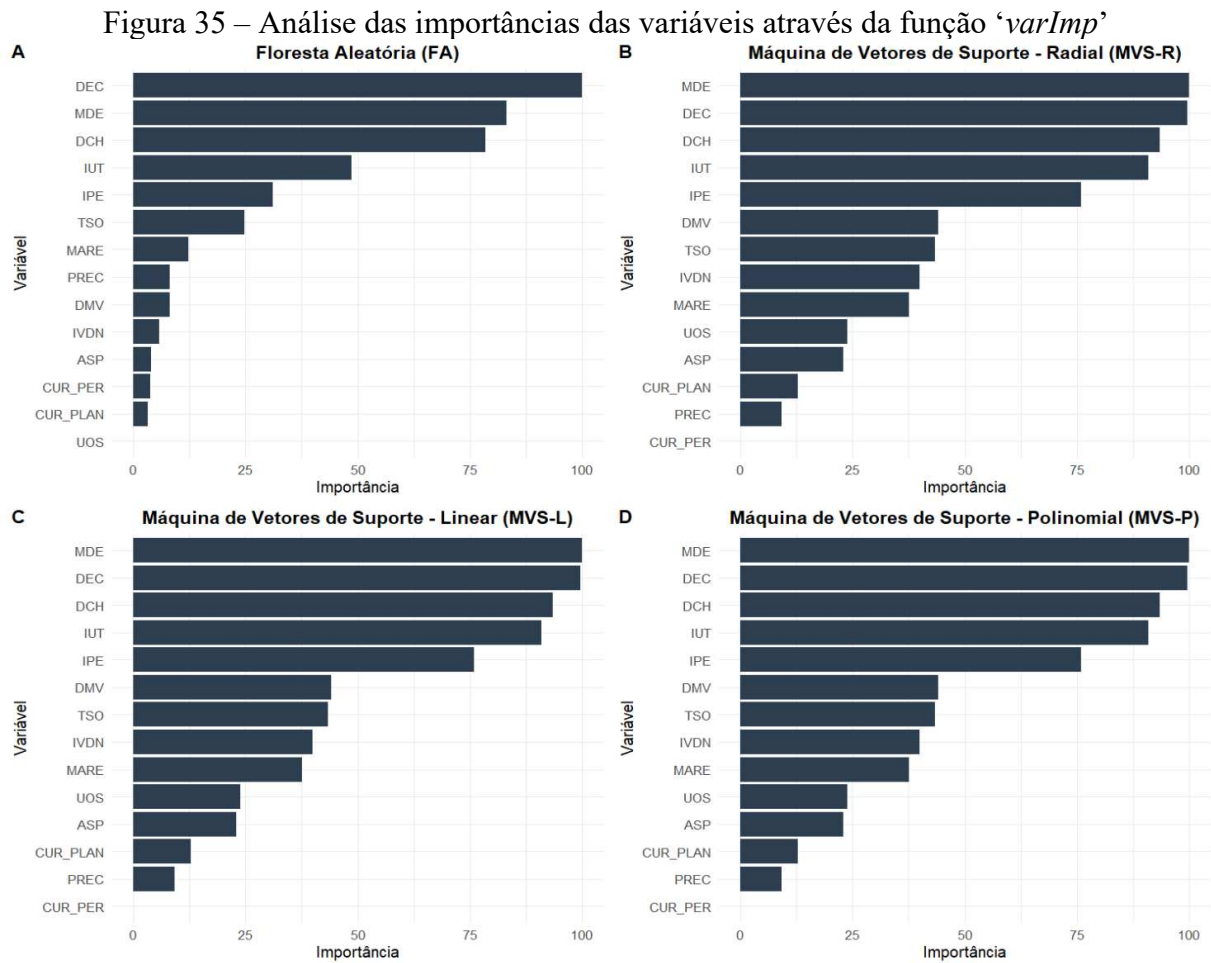
Após a análise das curvas AUC/ROC, os melhores modelos de cada tipo de algoritmo foram selecionados com base em seu desempenho preditivo. A floresta aleatória (FA), configurada com 500 árvores (ntree) e duas variáveis testadas por divisão (mtry = 2), apresentou desempenho consistente segundo a validação cruzada, com taxa de erro fora da amostra (Out-Of-Bag, OOB) extremamente baixa. Esse resultado evidencia a elevada capacidade discriminativa do modelo e sua adequação ao problema proposto.

Para a máquina de vetores de suporte com kernel radial (MVS-R), a otimização dos hiperparâmetros resultou em um custo (C) de 0,25 e sigma igual a 0,07695. O modelo operou com 751 vetores de suporte e obteve uma função objetivo de -135,2979. A validação cruzada indicou um erro de treinamento de 3,20%, reforçando a eficácia do algoritmo MVS-R na classificação.

5.8 ANÁLISE DA IMPORTÂNCIA DAS VARIÁVEIS

Com base nos quatro gráficos de importância das variáveis (Figura 35) para cada modelo, observa-se que há um núcleo de variáveis que se destacam consistentemente na explicação do fenômeno (susceptibilidade à inundação). Em todos os casos, as variáveis DEC (declividade), MDE (modelo digital de elevação) e DCH (distanciamento dos cursos hídricos) aparecem entre as mais importantes, seguidas de IUT (índice de umidade topográfica) e IPE (índice de potência de escoamento).

Os resultados (Figura 35) reforçam que, ao planejar medidas de mitigação ou prever inundações, é fundamental analisar o relevo e suas derivações (declividade, elevação, umidade topográfica), bem como o distanciamento dos cursos hídricos. A variação entre os modelos, por sua vez, mostra que diferentes algoritmos podem valorizar de forma desigual variáveis específicas (como manchas de maré ou distância de rodovias), mas há forte convergência quanto aos principais condicionantes.



Fonte: Elaborada pelo autor (2025).

A análise da importância das variáveis DEC, MDE, DCH, IUT e IPE, revela uma forte convergência com os resultados reportados na literatura. A elevação (MDE) surge como um dos fatores mais frequentemente destacados nas diversas fontes como tendo uma influência significativa na suscetibilidade à inundação. No estudo conduzido por Costache *et al.* (2020), ao aplicar o modelo SVM-IoE, a elevação foi considerada um fator relevante para o modelo (5.71%). Em Vafakhah *et al.* (2020), a altitude foi identificada como o fator mais efetivo para o mapeamento da suscetibilidade à inundação na Província de Gilan, Irã. Khosravi *et al.* (2019) também indica essa importância, mencionando que a altitude foi o fator condicionante mais eficaz para a suscetibilidade à inundação na Bacia de Ningdu, China. Rahman *et al.* (2019) observaram que a elevação teve um impacto considerável na ocorrência de inundações, embora menor do que outros fatores como solo e geologia. Pham *et al.* (2021), também destacam a elevação como um dos fatores mais importantes para a modelagem de inundações em sua área de estudo no Vietnã central. Esses achados em conjunto reforçam a ideia de que a elevação, por influenciar o fluxo de água e a acumulação em áreas mais baixas, é um preditor chave da suscetibilidade à inundação.

A declividade (DEC), também é frequentemente apontada como um fator de grande importância. Tehrany *et al.* (2013) mencionam que áreas com menor ângulo de declive são mais suscetíveis a inundações. Costache (2019) atribuiu a maior importância (0.93) ao ângulo de declive dentro dos fatores condicionantes da inundação no modelo FSVM. Wojtek *et al.* (2021) consideraram a declividade como o fator de maior importância relativa, com declives de 0 a 2 graus sendo os mais suscetíveis à inundação. Pham *et al.* (2021) também identificaram o ângulo de declive como um dos três fatores mais importantes para a modelagem de inundações. Esses resultados sugerem que a declividade influencia a velocidade do escoamento superficial e a capacidade de infiltração, sendo áreas planas ou de baixa declividade mais propensas à acumulação de água.

Para o distanciamento dos cursos hídricos (DCH) Vafakhah *et al.* (2020) identificaram a distância dos rios como o fator mais importante para a suscetibilidade à inundação em sua área de estudo. Costache *et al.* (2020) encontraram que a distância dos rios alcançou a maior importância relativa (17.63%) no modelo SVM-IoE. Wojtek *et al.* (2021) consideraram a distância do rio como o terceiro fator mais importante, com as menores distâncias sendo as mais suscetíveis à inundação. Pham *et al.* (2021) observaram que a ocorrência de inundações se correlaciona fortemente com os menores valores de distância do rio. Esses achados são coerentes do ponto de vista hidrológico, uma vez que áreas adjacentes a corpos d'água

apresentam maior propensão à inundação durante eventos de cheia, devido à sua menor elevação relativa e proximidade às zonas de transbordamento.

Pham *et al.* (2021) observaram que as inundações em sua área de estudo estavam associadas a baixos valores de índice de potência de escoamento (IPE). Contudo a importância do IPE (SPI em inglês) pode variar dependendo da área de estudo (Costache, 2019; Wang *et al.* 2019), pois ele está relacionado à capacidade de erosão e transporte de sedimentos pela água corrente.

O índice de umidade topográfico (IUT), ou *topographic wetness index* (TWI), também é reconhecido como um importante fator condicionante. Pham *et al.* (2021) mostraram que as inundações em sua área de estudo estavam associadas a altos valores de TWI, indicando solos úmidos e lençol freático elevado. Costache *et al.* (2020) encontraram uma importância de 4.37% para o TWI no modelo SVM isolado e 1.96% no modelo SVM-IoE. Costache (2019) atribuiu um peso de 0.14 ao TWI no modelo FSVM. Esses resultados confirmam que o TWI, ao indicar áreas com potencial de acúmulo de água, desempenha um papel crucial na avaliação da suscetibilidade a inundações.

Os dados apresentados na Figura 35 sugerem a baixa importância da curvatura do terreno (CUR) e do aspecto (ASP), em consonância com estudos que também atribuem menor importância a estes parâmetros. Por exemplo, em Liu *et al.* (2021), o valor de Ganho de Informação (IG) para o aspecto da encosta foi de apenas 0.05, um dos mais baixos entre os nove fatores analisados. Pham *et al.* (2021) removeram o aspecto da encosta (e a direção do fluxo) do processo de modelagem por apresentarem valores de mérito médio (AMIGR) de 0, indicando que gerariam ruído e levariam a problemas de sobreajuste.

A curvatura do terreno também foi considerada menos importante em alguns estudos. Khosravi *et al.* (2019) observaram que a curvatura não teve efeito na ocorrência de inundações na Bacia de Ningdu, sendo removida da análise. Costache e Bui (2019) encontraram valores relativamente baixos de importância para a curvatura plana (0.031 no MLP-WOE) e curvatura do perfil (0.083 no MLP-WOE) em seu estudo. A menor importância da curvatura pode depender da escala de análise e das características específicas do terreno, sendo que em algumas áreas, a convergência ou divergência do fluxo influenciada pela curvatura pode ter um papel mais significativo.

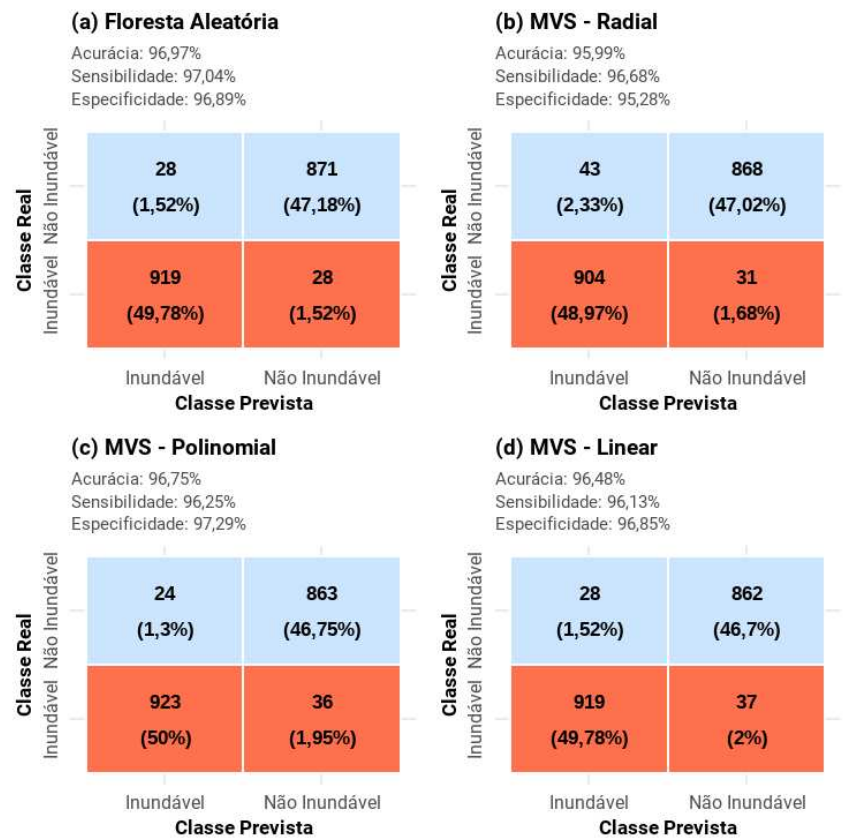
Em suma, a discussão dos resultados, destaca um núcleo de variáveis consistentemente importantes para a suscetibilidade à inundação, e amplamente suportada pelas evidências encontradas na literatura científica. A elevação, a declividade e a distância dos cursos hídricos

emergem como preditores recorrentes em diversos contextos geográficos e utilizando diferentes metodologias de modelagem.

5.9 RESULTADOS DA AMOSTRA DE TESTE

A matriz de confusão (Figura 36) é de grande importância para verificar como os modelos se comportam na amostra de teste, composta por dados que o modelo não havia visto durante o treinamento. Ela permite avaliar a capacidade de generalização do modelo, demonstrando o desempenho por meio de métricas como a acurácia, que mede a proporção de predições corretas entre todas as observações; a sensibilidade, que representa a capacidade do modelo de identificar corretamente os casos positivos; e a especificidade, que reflete a capacidade de identificar corretamente os casos negativos. Após processar as predições com os modelos escolhidos, procedeu-se à análise da matriz de confusão para cada modelo. Essa matriz foi obtida transformando as probabilidades estimadas pelos modelos em classificações binárias (0 ou 1), utilizando como ponto de corte (*threshold*) o valor de 0,5 (Kuhn, 2024).

Figura 36 - Matrizes de confusão da amostra de teste: a) FA, b) MVS-R, c) MVS-P e d) MVS-L



Fonte: Elaborada pelo autor (2025).

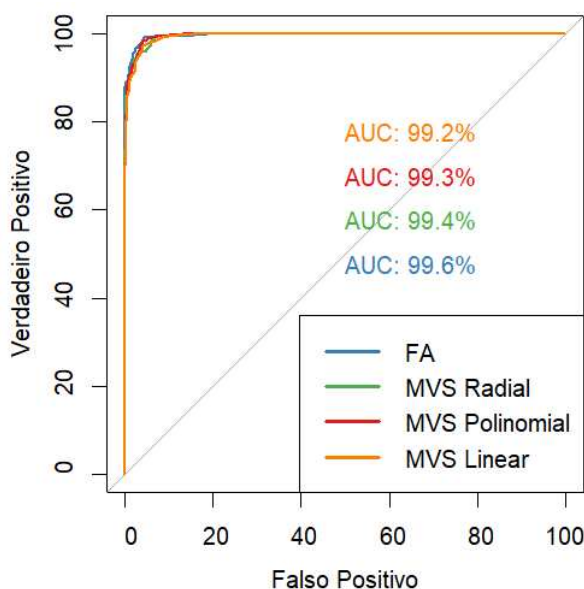
Entre os modelos avaliados, o modelo de floresta aleatória (FA) apresentou o melhor resultado geral, alcançando uma acurácia de aproximadamente 0,9697, sensibilidade de 0,9704 e especificidade de 0,9689, indicando um desempenho ótimo na classificação correta dos eventos. Os modelos de máquina de vetores de suporte (MVS) tiveram resultados próximos, com destaque para a versão polinomial (MVS-P), que atingiu uma acurácia de 0,9675, sensibilidade de 0,9625 e a maior especificidade (0,9729). Já a versão radial (MVS-R) apresentou a menor acurácia (0,9599), mas manteve bons valores de sensibilidade (0,9668) e especificidade (0,9528). Por fim, a versão linear (MVS-L) obteve um bom desempenho nas métricas analisadas, com acurácia de 0,9648, sensibilidade de 0,9613 e especificidade de 0,9685. Esses resultados demonstram que a escolha do modelo mais adequado pode variar conforme as necessidades específicas de aplicação, especialmente considerando-se a importância relativa entre erros do tipo falso positivo e falso negativo.

Os resultados obtidos na matriz de confusão utilizando o ponto de corte (*threshold*) de 0,5 demonstram um alto nível de desempenho preditivo dos modelos avaliados. No entanto, para obter uma avaliação mais completa da qualidade geral desses modelos, é fundamental analisar também a curva ROC (*receiver operating characteristic*) e sua respectiva área sob a curva (AUC). A curva ROC avalia o desempenho do modelo considerando todos os *thresholds* possíveis, sendo construída a partir do gráfico da taxa de verdadeiros positivos (sensibilidade) contra a taxa de falsos positivos (especificidade) para cada valor possível do ponto de corte.

A seleção dos modelos ótimos para a etapa de validação foi realizada com base nos maiores valores de AUC observados durante a etapa de treinamento (validação cruzada). A Figura 37 apresenta a curva ROC obtida na fase de teste para cada um dos modelos finalistas, ilustrando claramente a capacidade desses modelos em diferenciar áreas suscetíveis de áreas não suscetíveis a inundações. Esta avaliação leva em consideração as métricas de sensibilidade (capacidade de identificar corretamente áreas suscetíveis) e especificidade (capacidade de identificar corretamente áreas não suscetíveis), essenciais em problemas de classificação. Destaca-se que todos os modelos apresentaram desempenhos excelentes, com valores de AUC variando entre 99,2% e 99,6%, reforçando a robustez e eficiência na classificação dos modelos testados.

Apesar de diferenças sutis, todos superaram a marca de 99%, o que reforça a capacidade de discriminação dos modelos. O melhor desempenho foi obtido pelo algoritmo floresta aleatória (FA) que atingiu o maior AUC (99,6%), indicando uma melhor capacidade de discriminação entre as classes. Por outro lado, máquina de vetores de suporte linear (MVS-L) obteve o menor AUC (99,2%), mas ainda em um nível ótimo de acurácia.

Figura 37 - Curva ROC dados de validação



Fonte: Elaborada pelo autor (2024).

Comparando esses resultados com outros estudos (Choubin *et al.*, 2019; Lee *et al.*, 2017; Mosavi *et al.*, 2020), constata-se que todos os valores de AUC são muito superiores ao mínimo de 0,90 geralmente adotado como referência para boa capacidade preditiva em problemas de inundação. Assim, o desempenho aqui reportado não apenas é considerado aceitável, mas também altamente satisfatório, atendendo aos critérios metodológicos propostos.

Em síntese, a Figura 37 evidencia que cada um dos quatro algoritmos é capaz de distinguir, de forma robusta, as áreas de suscetibilidade à inundação, sendo o modelo FA o que apresenta ligeira vantagem no conjunto de teste. Tais achados são valiosos para aplicações práticas, pois permitem a escolha de modelos com alta precisão, fundamentais em estratégias de gestão e mitigação de riscos em regiões suscetíveis à inundação.

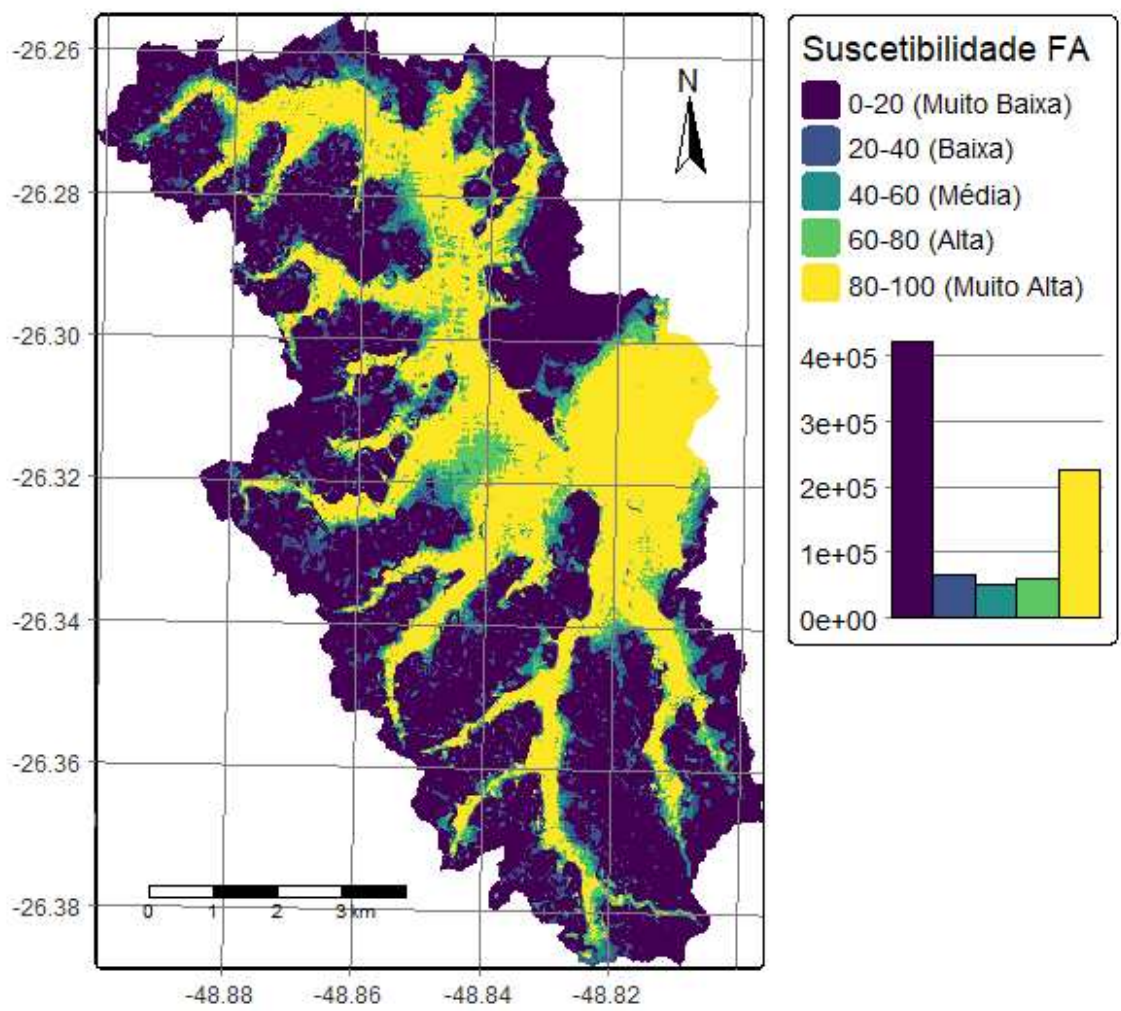
5.10 MAPAS DE SUSCETIBILIDADE À INUNDAÇÃO

Após a validação dos modelos e a obtenção de métricas de desempenho satisfatórias, procedeu-se à aplicação dos algoritmos selecionados para a predição de áreas suscetíveis à inundação em escala espacial para Bacia Hidrográfica do Rio Cachoeira. Para tal, foram selecionados os dois modelos que demonstraram superioridade estatística com base na análise da curva AUC-ROC: Floresta Aleatória (FA), um método de aprendizado *ensemble* baseado em árvores de decisão, e máquina de vetores de suporte com kernel radial (MVS-R). A escolha

desses dois modelos justifica-se por suas diferenças metodológicas intrínsecas, permitindo uma avaliação comparativa de abordagens distintas.

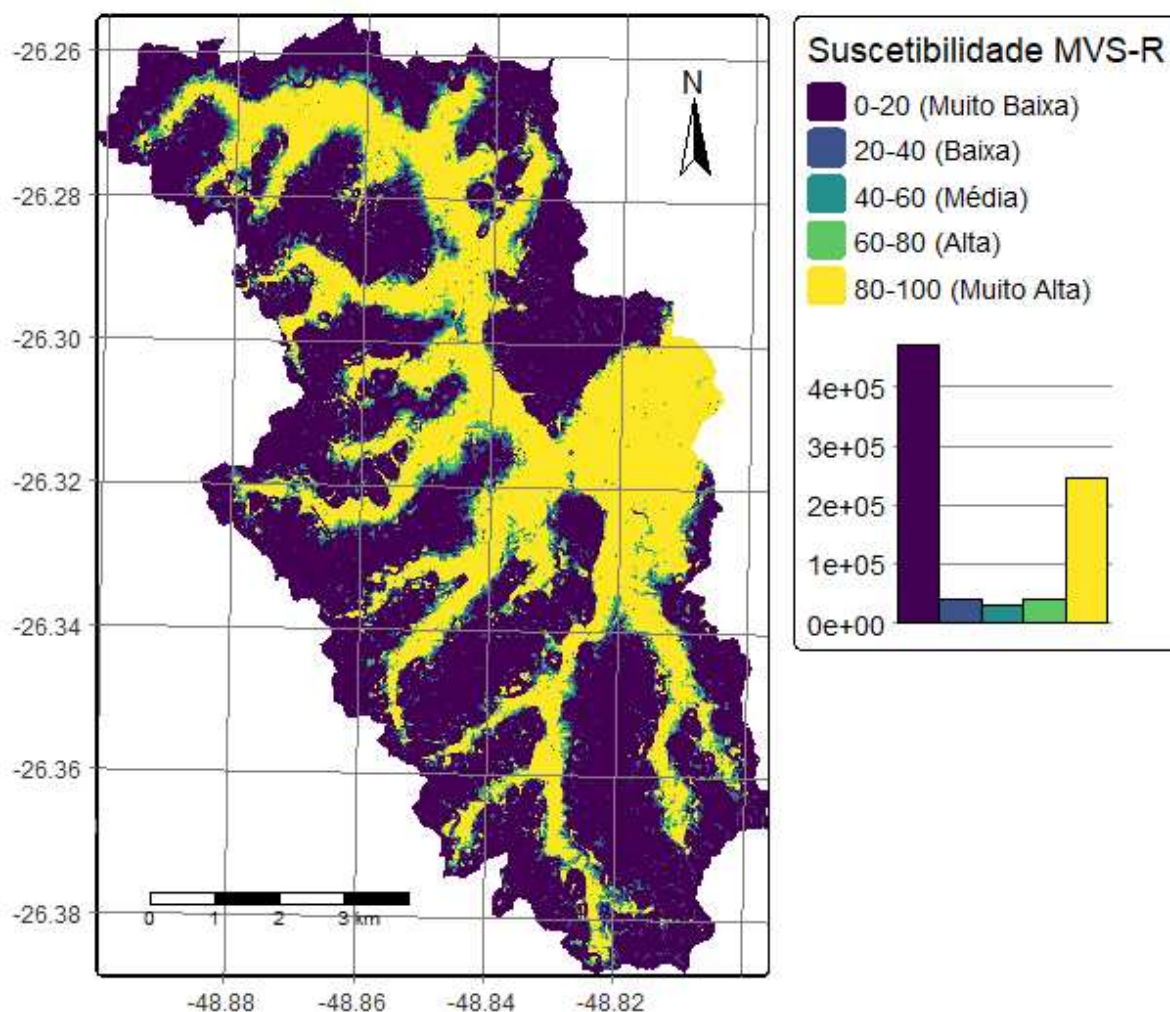
Os resultados espacializados das predições, ilustrados na Figuras 38 (FA) e Figura 39 (MVS-R), representam a suscetibilidade estimada à ocorrência de inundações em escala regional. Esses mapas temáticos foram gerados por classificações probabilísticas, destacando gradientes de suscetibilidade (0 a 1) com resolução espacial compatível com as variáveis preditoras utilizadas. A visualização cartográfica permite identificar padrões geográficos críticos, como regiões de maior suscetibilidade associadas.

Figura 38 - Mapa de suscetibilidade à inundação algoritmo floresta aleatória



Fonte: Elaborada pelo autor (2024).

Figura 39 - Mapa de suscetibilidade à inundação - algoritmo máquina de vetores de suporte



Fonte: Elaborada pelo autor (2024).

As discrepâncias identificadas entre os mapas de suscetibilidade apresentados (Figuras 38 e 39) refletem variações inerentes aos algoritmos empregados na modelagem. Contudo, observa-se uma significativa sobreposição espacial nas áreas classificadas nas faixas extremas de suscetibilidade (0,0–0,4 e 0,6–1,0), conforme demonstrado na Tabela 7. Esta convergência sugere robustez na identificação das regiões com suscetibilidade alta e baixa. Adicionalmente, evidencia-se uma consistente concordância entre os modelos quanto às regiões consideradas mais e menos suscetíveis às inundações, indicando que, apesar das diferenças algorítmicas, ambos os métodos fornecem resultados coerentes e complementares na identificação de áreas prioritárias para monitoramento e intervenções.

Tabela 7 - Comparação de áreas e níveis de suscetibilidade

Classificação de suscetibilidade	FA	MVS Radial
	ÁREA (Km²)	ÁREA (Km²)
Muito baixo [0,0 – 0,2]	42,06	46,93
Baixo [0,2 – 0,4]	6,64	4,13
Média [0,4 – 0,6]	5,01	3,18
Alta [0,6 – 0,8]	6,12	3,95
Muito alta [0,8 – 1,0]	22,70	24,58
Área total	82,53	82,77

Fonte: Elaborada pelo autor (2025).

Analisando a Tabela 7, verifica-se que as áreas classificadas com suscetibilidade "muito baixa" apresentam similaridade entre os modelos, correspondendo a 50,96% (42,06 km²) da área total para floresta aleatória (FA) e 56,69% (46,93 km²) para máquina de vetores de suporte (MVS-R). Observam-se algumas discrepâncias nas classificações "baixa" e "média", com o modelo FA identificando 8,05% (6,64 km²) e 6,07% (5,01 km²), respectivamente, comparado a valores inferiores do MVS Radial, que são de 4,99% (4,13 km²) e 3,84% (3,18 km²). As diferenças mais acentuadas ocorrem na categoria "alta", em que o FA classificou 7,42% (6,12 km²), enquanto o MVS Radial indicou uma área menor, de 4,77% (3,95 km²). Na classe "muito alta", os valores são relativamente próximos, com ligeira superioridade para o MVS Radial (29,69%, 24,58 km²) em comparação ao FA (27,51%, 22,70 km²).

Os resultados apresentados na Tabela 7 revelam que aproximadamente 34,92% (28,82 km²) a 34,46% (28,53 km²) da área total analisada estão classificadas com suscetibilidade "alta" ou "muito alta" à inundação, respectivamente para os modelos de floresta aleatória (FA) e máquina de vetores de suporte (MVS-R). Essa proporção é preocupante, especialmente considerando que as áreas de maior suscetibilidade estão concentradas predominantemente na região do perímetro urbano de Joinville.

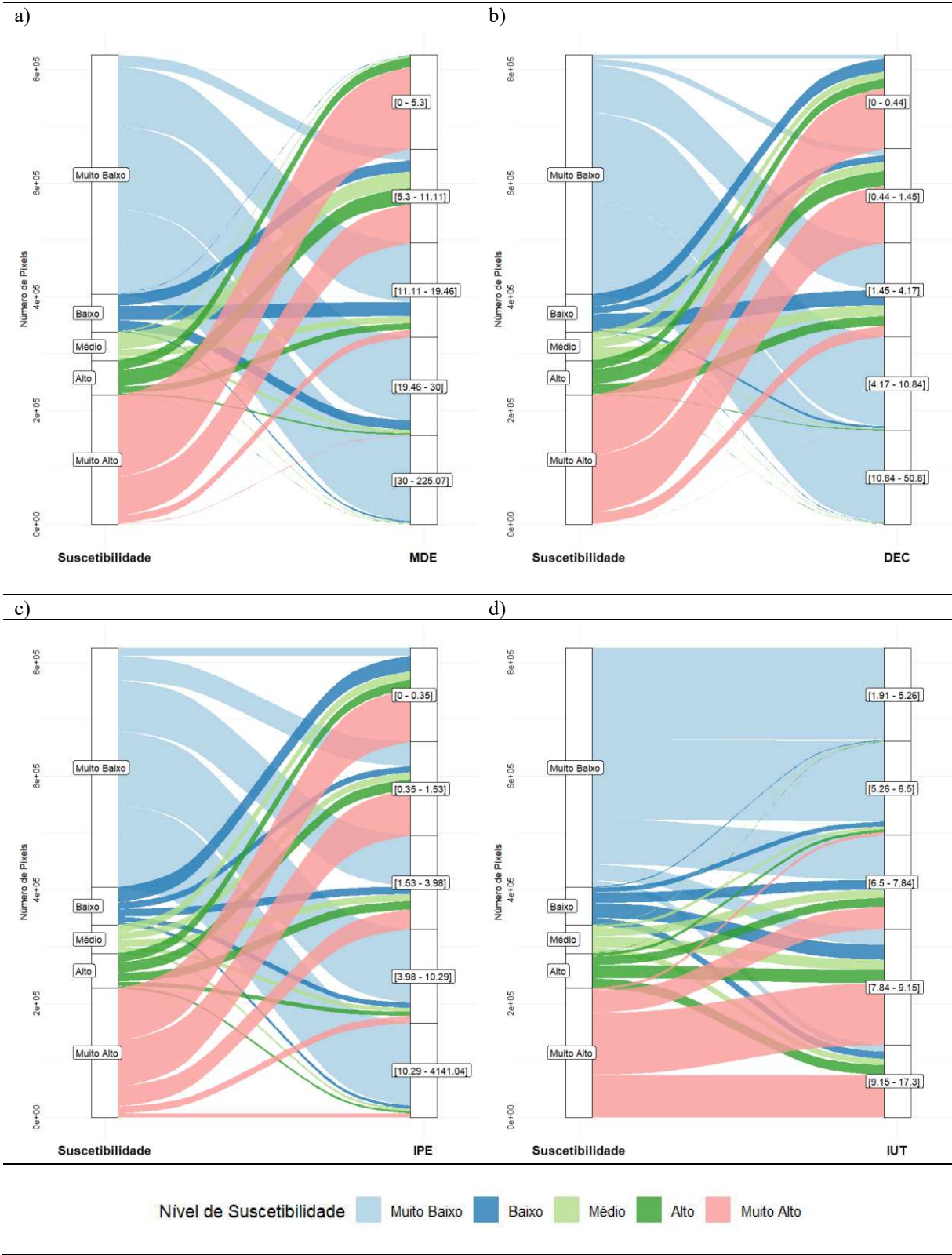
Este padrão espacial de concentração das áreas suscetíveis nas regiões urbanas ressalta uma condição crítica de exposição a eventos de inundação. Tal constatação reforça a necessidade de políticas públicas de planejamento urbano e estratégias integradas de gestão de riscos, visando a mitigação e adaptação a eventos extremos.

5.11 ANÁLISE DA SUSCETIBILIDADE À INUNDAÇÃO EM FUNÇÃO DAS VARIÁVEIS

No Apêndice C, apresenta-se a tabela de contingência (ou tabela cruzada) que relaciona todos os níveis de suscetibilidade do modelo floresta aleatória obtido neste estudo, classificado em cinco categorias de suscetibilidade, com os intervalos ou categorias das variáveis analisadas neste estudo. Através das tabelas de contingência (Apêndice C), torna-se possível a geração de diagramas aluviais (Figura 40 e Figura 41) para representar as relações entre a suscetibilidade à inundação e as variáveis investigadas neste estudo. Esses diagramas permitem representar fluxos e relações entre diferentes categorias, facilitando a análise conjunta de variáveis que influenciam fenômenos complexos, como a suscetibilidade à inundação. Dentro dos modelos, os fatores de elevação, declividade e distância de corpos hídricos emergiram como as variáveis mais fundamentais para a análise da suscetibilidade à inundação, considerando a importância da relação espacial entre esses fatores e a ocorrência de eventos hidrológicos extremos. Os parâmetros de elevação e declividade, derivados por meio de modelos digitais de elevação (MDE), são essenciais por permitir uma quantificação precisa e uma visualização clara da interação entre a topografia local e a vulnerabilidade à inundação. A integração dessas variáveis em diagramas aluviais (Figura 40) possibilita representar graficamente a associação espacial, destacando categorias predominantes nas classes mais suscetíveis às inundações.

A análise da relação entre a suscetibilidade à inundação e o modelo digital de elevação (Figura 40- a), com base na tabela de contingência, evidencia uma clara correlação inversa entre a altitude do terreno e o risco de inundação na BHRC. Observa-se que as áreas classificadas com alta suscetibilidade (classe 5) concentram-se principalmente nas menores cotas altimétricas (intervalo de 0 a 5,3 m), representando 17,5% da área analisada. À medida que a elevação aumenta, ocorre uma redução progressiva da suscetibilidade, sendo que áreas acima de 30m apresentam uma representatividade significativamente menor, chegando a apenas 0,00557%. Esses resultados demonstram que regiões com menores elevações estão claramente mais expostas ao risco de inundação, confirmando a topografia como um fator essencial na avaliação e gestão de riscos hidrológicos.

Figura 40 - Diagramas aluviais das variáveis: a) MDE, b) DEC, c) IPE e d) IUT



A análise da suscetibilidade à inundação em função da declividade (Figura 40-b) revelou um padrão claro de associação entre áreas de menor inclinação e níveis elevados de suscetibilidade. As áreas classificadas como de suscetibilidade muito alta concentram-se nas

faixas de menor declividade, especialmente entre 0 e 1,45% ([0 - 1,45]), representando, juntas, aproximadamente 5,18% da área total analisada. Em contraste, regiões com maior declividade (>10,84%) são predominantemente associadas às classes de suscetibilidade muito baixa ou baixa. Essa distribuição evidencia que terrenos planos ou com suave declividade favorecem a acumulação e o escoamento lento das águas pluviais, potencializando o risco de inundação, enquanto áreas mais inclinadas são menos propensas ao acúmulo de água superficial, reduzindo, portanto, seu grau de suscetibilidade.

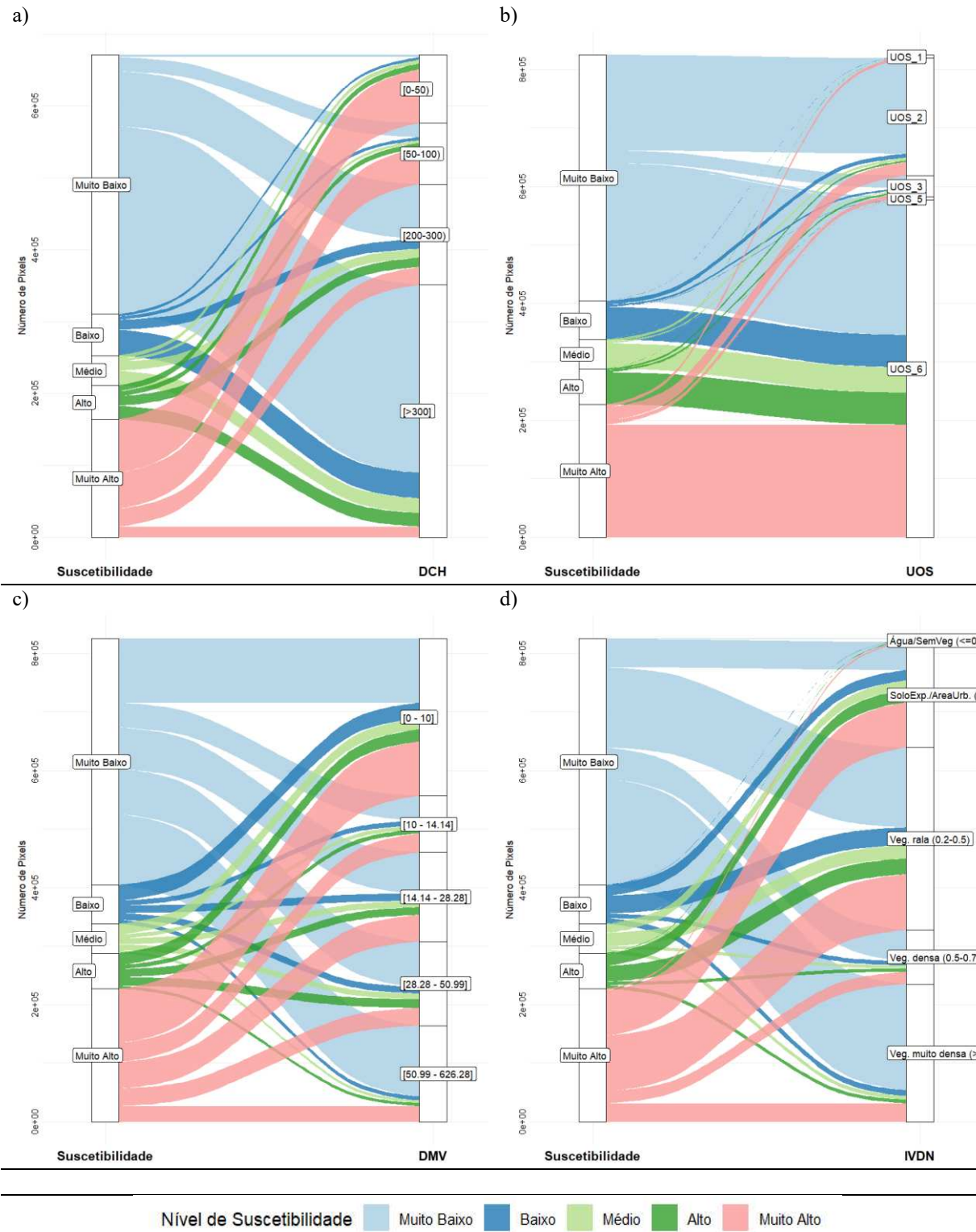
Na Figura 40-c na análise da suscetibilidade à inundação em relação ao índice de potência de escoamento (IPE ou SPI) demonstrou uma associação entre áreas com valores reduzidos desse índice e maiores níveis de suscetibilidade. As regiões de suscetibilidade muito alta concentram-se principalmente nas classes inferiores do IPE, sobretudo no intervalo mais baixo (0 a 0,35), representando 11,2% da área total analisada, e no intervalo intermediário (0,35 a 1,53), com 9,76%. Em contrapartida, as regiões com suscetibilidade muito baixa estão predominantemente associadas a valores elevados do IPE (> 10,29), totalizando 17,6% da área estudada. Esses resultados sugerem que áreas com menor capacidade de escoamento superficial (baixo IPE) são mais vulneráveis à inundação, ao passo que áreas com maior capacidade de escoamento (alto IPE) apresentam menor suscetibilidade, destacando a importância desse índice como fator preditivo para eventos de inundação.

A análise da relação entre a suscetibilidade à inundação e o índice de umidade topográfica (IUT ou TWI) revelou uma tendência clara indicando que áreas com valores elevados deste índice estão associadas a níveis mais altos de suscetibilidade à inundação. Observa-se que as áreas classificadas com suscetibilidade "Muito Alta" concentram-se majoritariamente nos intervalos mais elevados do IUT (entre 0 e 1,53), representando juntos aproximadamente 20,96% da área total estudada. Em contraste, áreas com "Muito Baixa" suscetibilidade predominam nos intervalos superiores do índice (acima de 10,29), abrangendo cerca de 17,6% da área analisada. Esses resultados reforçam a importância do IUT como indicador efetivo para a previsão de áreas suscetíveis à inundação, visto que valores mais altos do índice refletem regiões propensas ao acúmulo de água devido a características topográficas que favorecem processos de saturação e reduzido escoamento superficial.

Para o fator de distância dos canais hídricos (Figura 41 -a) observou-se que as áreas com maior suscetibilidade concentram-se predominantemente nas proximidades imediatas dos cursos d'água, especialmente no intervalo de distância de 0 a 50 metros, correspondendo a cerca de 9,01% da área estudada. A suscetibilidade diminui progressivamente com o aumento da distância dos canais, sendo que apenas 6,08% das áreas de alta suscetibilidade estão entre 50 e

300 metros, e uma pequena porção (0,164%) encontra-se a mais de 300 metros dos canais hídricos.

Figura 41 - Diagramas aluviais das variáveis: a) DCH, b) UOS, c) DMV e d) IVDN



Fonte: Elaborada pelo autor (2024).

Para os fatores antrópicos, a relação euclidiana da distância às vias (DMV) demonstrou uma associação leve entre a proximidade das vias e os níveis elevados de suscetibilidade à inundação. Esse resultado era esperado, dado que a área de estudo (BHRC) está inserida em um perímetro urbano, onde a infraestrutura viária pode influenciar significativamente os padrões de escoamento superficial e a vulnerabilidade do ambiente. Áreas classificadas como de suscetibilidade muito alta (classe 5) estão predominantemente próximas às vias, com aproximadamente 11,2% da área total estudada concentrada a menos de 10 metros destas estruturas. À medida que a distância aumenta, observa-se uma redução gradual das áreas muito suscetíveis, com cerca de 5,49% situadas entre 14 e 28 metros e apenas 3,65% entre 28 e 51 metros. Em contrapartida, regiões classificadas como de suscetibilidade muito baixa (classe 1) estão mais dispersas espacialmente, com maior concentração a distâncias superiores a 14 metros das vias (40,2% do total).

A análise conjunta da suscetibilidade à inundação e do uso e ocupação do solo (Figura 41 - b) indica que, embora a maior parte da área mapeada (50,95%) apresente baixo risco (classe 1), um percentual expressivo (27,49%) encontra-se em zonas de alta suscetibilidade (classe 5). Sobre as áreas construções/edificações (UOS_6), que ocupam cerca de 70% do território (BHRC), com parcela significativa (23,3% de toda a área) disposta em zonas de alta suscetibilidade. Esses resultados ressaltam a necessidade de políticas de ordenamento territorial para minimizar impactos nas áreas mais vulneráveis.

De acordo com o índice de vegetação por diferença normalizada (IVDN/NDVI), as áreas de alta suscetibilidade (classe 5) concentram-se em regiões de menor cobertura vegetal, sendo 9,36% em “Solo exposto/Área urbana” (IVDN de 0,01 a 0,2) e 11,5% em “Vegetação rala” (0,2 a 0,5). Esse padrão sugere que a vegetação densa desempenha um papel fundamental na atenuação do risco de inundação, enquanto solos expostos, áreas urbanizadas e vegetação pouco desenvolvida elevam a suscetibilidade a processos de alagamento, em razão da menor capacidade de interceptação e infiltração de água no solo. Tal resultado era esperado, pois o desenvolvimento de Joinville ocorreu ao longo do Rio Cachoeira (principal rio da bacia), cujas margens são densamente urbanizadas, favorecendo, portanto, valores mais baixos de IVDN e maior risco de inundações.

5.12 COMPARAÇÃO COM OUTROS ESTUDOS

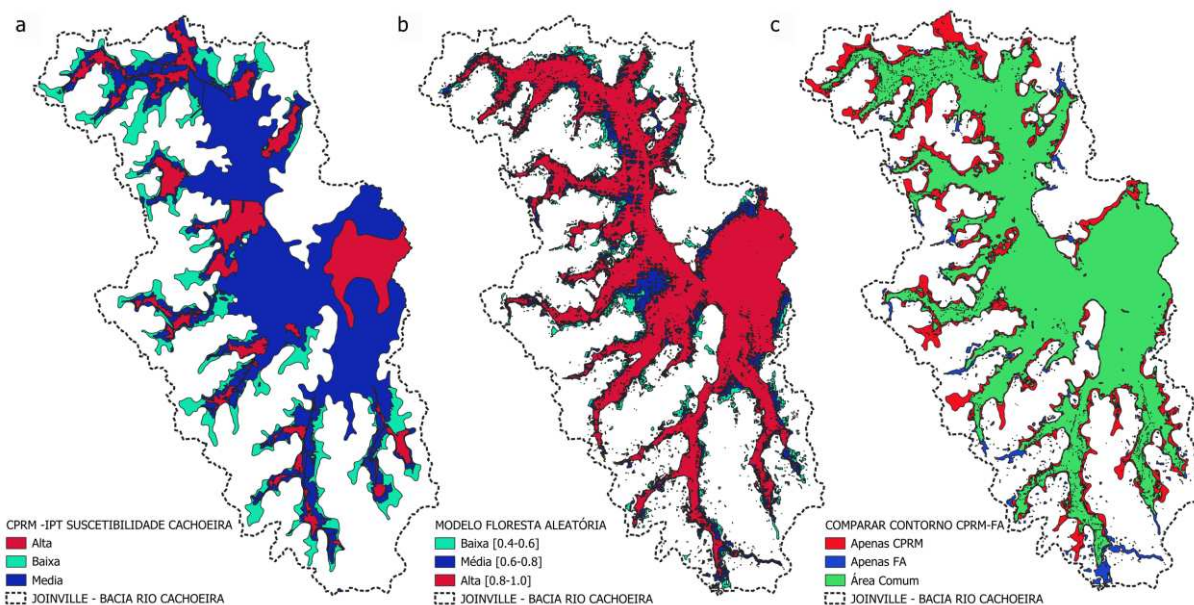
Para esta seção selecionaram-se estudos que elaboraram mapas de suscetibilidade na Bacia Hidrográfica do Rio Cachoeira, com o objetivo de realizar uma análise crítica. A

finalidade é avaliar se o modelo proposto apresenta congruência com os padrões espaciais identificados na literatura, validando sua consistência na identificação de processos geodinâmicos e ambientais em consonância com os resultados estabelecidos.

Em 2015, o Serviço Geológico do Brasil (CPRM) realizou um estudo específico sobre suscetibilidade à inundação no perímetro municipal de Joinville, empregando o método de processo de análise hierárquica (AHP). Nesse estudo, foram definidas três classes de suscetibilidade: baixa, média e alta. Áreas consideradas sem efeito significativo de inundação não tiveram representação explícita, sendo, portanto, omitidas da classificação gerada pelo CPRM (Fonte).

Para permitir uma comparação adequada com o método desenvolvido neste trabalho, foi necessário compatibilizar as classificações. Dessa forma, adotou-se o mapa de suscetibilidade gerado pelo algoritmo floresta aleatória (FA), estabelecendo critérios específicos para classificação: valores de predição entre 0,4 e 0,6 foram definidos como suscetibilidade baixa; valores entre 0,6 e 0,8 foram classificados como média suscetibilidade; e valores superiores a 0,8 foram considerados como alta suscetibilidade. Tal procedimento permitiu padronizar e comparar as classificações resultantes dos dois métodos. A Figura 42 ilustra os contornos das classificações obtidas pelo CPRM (2015) e floresta aleatória, permitindo uma comparação visual e detalhada das áreas identificadas como suscetíveis à inundação.

Figura 42 - Comparação entre os mapas de suscetibilidade à inundação: a) CPRM, b) FA, c) sobreposição CPRM e FA



Fonte: Elaborada pelo autor (2025).

O método AHP utilizado pelo CPRM (Figura 42-a) classificou as áreas em graus distintos de suscetibilidade (baixa, média e alta) fundamentado na análise multicritério ponderada por especialistas, o que resultou em um mapeamento relativamente generalizado, com regiões amplas de média suscetibilidade, especialmente concentradas ao longo da calha principal dos rios da bacia. Em contraste, o modelo desenvolvido por meio do algoritmo floresta aleatória (Figura 41-b) na qual apresentou um detalhamento por *pixel* (10x10m), com padrões espaciais mais fragmentados, demonstrando maior sensibilidade em captar nuances locais da suscetibilidade.

A Figura 42-c apresenta uma análise comparativa de sobreposição espacial entre o mapa do CPRM e algoritmo floresta aleatória (FA), identificando uma área de convergência de 32,00 km², correspondendo a 76,4% da área total classificada. Além disso, identificou-se que a área exclusiva ao método CPRM foi de 8,11 km², enquanto a área exclusiva ao método floresta aleatória foi de 1,83 km². Essas divergências localizadas (23,6% do total) sugerem influência direta dos critérios de classificação e das abordagens técnicas adotadas, como diferenças na parametrização de algoritmos, tratamento de dados e resolução espacial.

O método *Analytic Hierarchy Process* (AHP), embora amplamente utilizado para mapeamentos de suscetibilidade devido à sua simplicidade conceitual e facilidade de aplicação, apresenta certas limitações relacionadas à subjetividade intrínseca na definição dos pesos dos critérios, altamente dependente da experiência e julgamento dos especialistas envolvidos (Rahman *et al.*, 2019; Rana e Mhanta 2023). Essa subjetividade pode reduzir a precisão espacial e dificultar a reprodução consistente dos resultados, limitando sua aplicação em análises detalhadas e dinâmicas. Por outro lado, o algoritmo de aprendizado de máquina pode superar essas restrições ao adotar uma abordagem baseada em dados, capaz de identificar padrões complexos e não lineares sem depender diretamente de julgamento humano (Nachappa *et al.*, 2019).

De forma complementar, um estudo conduzido por Muller (2012) na mesma bacia, utilizando sistemas de informação geográfica (SIG) com médias ponderadas para integrar parâmetros ambientais críticos (declividade, uso e tipologia do solo, geomorfologia), identificou distribuição heterogênea de suscetibilidade: apenas 2% em classe muito baixa/nula, com incremento progressivo nas categorias subsequentes (21% baixa, 22% média, 24% alta e 32% muito alta). Em contrapartida, a abordagem com aprendizado de máquina, baseada no algoritmo floresta aleatória, apresentou padrão distinto, classificando 51% da área como muito baixa/nula, reduzindo as proporções intermediárias (8% baixa, 6% média, 7% alta) e mantendo 28% na classe máxima. A discrepância acentuada, especialmente a ampliação da classe de

menor suscetibilidade (2% para 51%), sugere que o método ponderado por SIG pode superestimar gradientes lineares de risco, enquanto o aprendizado de máquina ao capturar padrões não lineares e hierarquias entre fatores ambientais, tende a redistribuir as prioridades de suscetibilidade, possivelmente identificando limiares críticos não detectados na abordagem tradicional.

Apesar das diferenças metodológicas observadas na análise comparativa entre os estudos do CPRM e de Müller (2012), as áreas de alta suscetibilidade apresentaram congruência espacial significativa, mesmo considerando que a abordagem de Müller (2012) empregou metodologias simplificadas e um número reduzido de fatores preditivos em relação ao modelo proposto neste trabalho. Essa convergência espacial evidencia que os algoritmos de aprendizado de máquina aplicados na geração de mapas de suscetibilidade não apenas alcançam alta precisão preditiva, mas também reproduzem padrões consistentes na identificação de processos geodinâmicos e ambientais, alinhando-se qualitativamente a métodos consolidados na literatura. Tal resultado reforça a robustez do uso de aprendizado de máquina para replicar fenômenos complexos, validando seu potencial como ferramenta para gerar mapas de suscetibilidade à inundação.

5.13 LIMITAÇÕES

Esta seção tem como objetivo realizar uma análise crítica das limitações intrínsecas ao modelo proposto, bem como discutir os desafios metodológicos associados à replicação de métodos de aprendizado de máquina para gerar mapas de suscetibilidades à inundação em diferentes contextos geográficos.

5.13.1 Avaliação da quantidade adequada de fatores

Um aspecto fundamental nesse tipo de estudo reside na escolha dos fatores que influenciam a ocorrência de inundações. Conforme Termeh *et al.* (2018), a seleção de fatores substanciais é um pré-requisito essencial para a avaliação da suscetibilidade. No entanto, não existe um consenso universal entre os pesquisadores sobre quantos fatores devem ser incluídos nos modelos (Vojtek *et al.*, 2021). Enquanto Mahmoud e Gan (2018) sugerem o uso de mais de seis fatores em estudos com MCDA para evitar a superestimação de alguns deles, outros estudos empregam um número menor (Lollo *et al.*, 2018; Falah *et al.*, 2019). Tehrany *et al.* (2019)

argumentam que um maior número de fatores não garante necessariamente maior precisão, enfatizando a influência do processo de modelagem. A própria disponibilidade de dados pode restringir o número de fatores considerados. Além disso, fatores selecionados podem apresentar redundância de informação, sendo necessário realizar análises de multi-colinearidade para identificar e mitigar esse problema, conforme mencionado em Rana e Mahanta (2023) na qual foi utilizando indicadores como VIF e tolerância. A remoção de fatores com valor preditivo nulo também é uma etapa importante para melhorar o desempenho do modelo (Phan *et al.*, 2021). A escolha inadequada ou a exclusão de fatores relevantes podem comprometer a precisão dos mapas de suscetibilidade.

Neste estudo foram inicialmente considerados 15 fatores, dos quais 14 foram mantidos após a análise da multicolinearidade para compor o modelo final. Utilizando-se imagens do satélite Sentinel-2, integradas a outras fontes de dados auxiliares, foi possível obter uma quantidade expressiva de fatores para a modelagem preditiva. Ainda que a quantidade final de 14 fatores supere a média usualmente utilizada na literatura (11 fatores conforme análise bibliométrica desse estudo) é possível que outros fatores não estudados tenham efeito na variável dependente.

5.13.2 Limitações decorrentes das fontes de dados e sua qualidade

A elaboração de mapas de suscetibilidade à inundação depende fundamentalmente da disponibilidade e qualidade dos dados utilizados para representar os fatores condicionantes das inundações. No entanto, diversas limitações inerentes às fontes de dados e à sua qualidade podem impactar significativamente a acurácia e a confiabilidade dos resultados obtidos.

Uma das principais limitações reside na diversidade de fontes de dados e suas características intrínsecas. Estudos de suscetibilidade à inundação frequentemente integram informações provenientes de diversas fontes, como modelos digitais de elevação (MDE) obtidos por diferentes métodos e com variadas resoluções espaciais, dados climáticos (precipitação) com diferentes períodos de registro e granularidade temporal, mapas de uso e cobertura do solo com distintas classificações e níveis de detalhe, dados geológicos e de solos em diferentes escalas, informações sobre a rede hidrográfica extraídas de cadastros com diferentes níveis de precisão, e inventários históricos de inundações com variações na exatidão espacial e temporal.

Essa heterogeneidade nas características dos dados, especialmente em termos de resolução espacial e escala original dos mapas, representa uma limitação (Vojtek *et al.*, 2021).

Idealmente, a modelagem de suscetibilidade à inundação se beneficiaria de dados de entrada com a mesma escala original e resolução espacial. Contudo, alcançar essa homogeneidade é frequentemente difícil na prática (Ali *et al.*, 2020). No presente estudo, o modelo digital de elevação (MDE), utilizado para derivar variáveis topográficas, e as imagens multiespectrais do Sentinel-2, foram obtidos em resolução espacial de 10x10 m, considerada adequada para análises em escala regional. Entretanto, em contextos operacionais diversos, limitações técnicas ou logísticas podem elevar essa resolução para valores superiores, reduzindo a precisão das análises. Já a qualidade dos fatores de distância de rios e vias, gerados a partir de bases cartográficas atualizadas, garantiu parâmetros espacialmente consistentes e robustos para o modelo preditivo.

Contudo, a qualidade temporal dos dados coletados apresentou limitações devido à indisponibilidade de informações atualizadas, além de desafios inerentes à escala espacial de alguns fatores do modelo, como a escassez de dados pluviométricos contínuos e a distribuição insuficiente de estações meteorológicas na região, resultando em um mapa pluviométrico (isoieta) com resolução espacial reduzida. Outro aspecto crítico refere-se às mudanças dinâmicas no uso e cobertura do solo ao longo do período, as quais podem alterar parâmetros hidrológicos fundamentais, como taxas de infiltração e padrões de escoamento superficial, introduzindo incertezas nas relações causa-efeito modeladas.

Entretanto, as limitações identificadas neste estudo podem ser mitigadas mediante a integração de fluxos de dados com maior resolução espacial e atualização temporal contínua, vantagem inerente dos métodos de aprendizado de máquina, que permitem assimilação iterativa de novos dados. A capacidade desses algoritmos de recalibrar parâmetros automaticamente, adaptando-se a informações recentes e refinadas, ampliaria a aderência do modelo às condições espaço-temporais atuais, gerando simulações mais fidedignas. Além disso, a natureza dinâmica do aprendizado de máquina possibilita a implementação de sistemas de alimentação contínua, em que atualizações incrementais refinam previsões e reduzem incertezas sem exigir reconstruções completas do modelo. Essa flexibilidade metodológica não apenas garantiria respostas ágeis as mudanças ambientais e urbanas, mas também transformaria a ferramenta em um recurso iterativo e resiliente.

5.13.3 Limitações da base de dados do inventário de inundações

A construção de uma base de dados robusta para o inventário de inundações representa um pilar fundamental na modelagem da suscetibilidade, conforme enfatizado por diversos

autores (Nachappa *et al.*, 2020; Liu *et al.*, 2021; Yassen *et al.*, 2022). No entanto, essa etapa enfrenta limitações que podem comprometer a precisão e a aplicabilidade dos modelos resultantes. Uma das principais restrições reside na disponibilidade e na qualidade dos registros históricos de inundações. Em muitas regiões, os dados sobre eventos passados podem ser incompletos, imprecisos ou até mesmo inexistentes para determinados períodos ou áreas geográficas (Liu *et al.*, 2021). A coleta dessas informações frequentemente depende de fontes diversas, como relatos de órgãos governamentais, notícias de jornais, interpretação de imagens de satélite, e levantamentos de campo (Costache e Bui, 2019; Wang *et al.*, 2019). Cada uma dessas fontes possui suas próprias limitações em termos de cobertura espacial e temporal, nível de detalhe e confiabilidade. Por exemplo, registros mais antigos podem carecer de coordenadas geográficas precisas, enquanto dados obtidos por sensoriamento remoto podem ser afetados pela cobertura de nuvens ou pela resolução espacial do sensor. Além disso, a definição do que constitui um "evento de inundação" pode variar entre diferentes registros, introduzindo inconsistências na base de dados. A dificuldade em determinar as fronteiras e a altura das inundações passadas, especialmente para eventos mais antigos, é outra limitação prática mencionada na literatura (Liu *et al.*, 2021).

Outra limitação importante das bases de dados de inventário de inundações se refere ao tipo de informação que é tipicamente registrada. Frequentemente, o foco principal recai sobre a localização geográfica dos eventos de inundação (Khosravi *et al.*, 2019; Nachappa *et al.*, 2020; Pham *et al.*, 2021), utilizando-se pontos centrais de áreas afetadas ou polígonos que representam a extensão da inundação. Embora essa informação espacial seja essencial para a modelagem, outros atributos relevantes, como a profundidade da água, a duração da inundação, a velocidade do fluxo e as causas específicas do evento (por exemplo, precipitação intensa, rompimento de barragem), podem não estar consistentemente disponíveis ou detalhados nos registros (Liu *et al.*, 2021). A ausência dessas informações adicionais limita a capacidade dos modelos em descrever a complexidade dos processos de inundação e de diferenciar eventos com características distintas.

A metodologia adotada neste estudo considera a suscetibilidade à inundação como uma variável binária para fins de modelagem e validação (presença ou ausência de eventos em unidades espaciais), embora o resultado final represente uma estimativa contínua de probabilidade da suscetibilidade à inundação. Essa abordagem, ao desconsiderar a dimensão temporal, restringe a análise das relações causais, especialmente em cenários onde mudanças graduais (e.g., alterações no uso do solo ou variabilidade pluviométrica) influenciam a suscetibilidade ao longo do tempo. Para otimizar a modelagem, sugere-se a elaboração de um

inventário hidrológico espaço-temporal para a bacia, combinando séries históricas georreferenciadas com classificações descritivas dos eventos (intensidade, duração, extensão espacial). Essa estrutura permitiria a categorização multivariada dos episódios, viabilizando a calibração de modelos probabilísticos que estimem não apenas a suscetibilidade geral, mas também cenários condicionados a precipitações de intensidade específica. Tal aprimoramento integraria variáveis hidrodinâmicas e temporais, ampliando a capacidade preditiva para estratificar riscos por classe de evento e quantificar vulnerabilidades associadas a projeções climáticas extremas, assegurando maior robustez e aplicabilidade em contextos dinâmicos.

5.13.4 Limitações da base de dados do inventário de não inundações

A criação de uma base de dados de inventário de não inundações, essencial para contrastar com as áreas afetadas e treinar modelos de suscetibilidade, enfrenta desafios que podem impactar a robustez das análises. Uma das principais limitações reside na definição e na seleção das áreas consideradas como "não inundadas". Frequentemente, essas áreas são definidas por exclusão, ou seja, são os locais onde não foram registrados eventos de inundação (Costache; Bui 2019; Vafakhah *et al.*, 2020).

No entanto, a ausência de registros não garante que uma área seja intrinsecamente segura contra futuras inundações. A falta de dados históricos abrangentes ou a limitação das metodologias de coleta podem levar à classificação errônea de áreas que, embora não tenham sofrido inundações no passado recente, possuem características biofísicas ou de infraestrutura que as tornam vulneráveis. A seleção aleatória de pontos não inundados, como mencionado em algumas metodologias (Vafakhah *et al.*, 2020), pode não capturar adequadamente a heterogeneidade das áreas não suscetíveis, potencialmente enviesando o treinamento dos modelos. Além disso, a natureza dinâmica do risco de inundação implica que áreas atualmente não inundadas podem tornar-se suscetíveis devido a mudanças no uso do solo, alterações climáticas ou falhas em sistemas de proteção, tornando a classificação estática de "não inundado" uma simplificação da realidade.

Este estudo combinou critérios ativos e passivos para a construção do inventário de áreas de não inundação, integrando pontos de um mapa de rotas seguras contra inundações (fornecido pela Defesa Civil) com parâmetros geomorfológicos quantitativos, como cotas altimétricas elevadas, declividades acentuadas e distância mínima de corpos hídricos. Essa abordagem híbrida, configura uma estratégia complementar à definição puramente por exclusão,

minimizando a subjetividade inerente à seleção de áreas de referência. Contudo, persiste um viés metodológico: a seleção de áreas "seguras" pressupõe implicitamente a ausência de registros históricos de inundação nessas unidades espaciais, lacuna que pode comprometer a validação do modelo em cenários de eventos extremos ou alterações hidrodinâmicas não mapeadas. Para contornar essa limitação, propõe-se um protocolo de monitoramento contínuo que inclua não apenas a catalogação espacial de eventos de inundação, mas também o georreferenciamento sistemático de zonas estáveis durante episódios críticos. Essa dupla verificação, com registro de inundações e identificação de áreas resilientes, permitiria a calibração iterativa dos critérios geomorfológicos, incorporando dinâmicas temporais e validando empiricamente os pressupostos de segurança. Adicionalmente, a integração de dados de sensoriamento remoto em tempo real e séries históricas de precipitação possibilitaria a correlação entre parâmetros morfométricos e limiares pluviométricos críticos, refinando a modelagem probabilística de suscetibilidade sob diferentes cenários climáticos.

5.13.5 Limitações metodológicas dos modelos de aprendizado de máquina

O aprendizado de máquina tem emergido como uma ferramenta poderosa para a geração de mapas de suscetibilidade à inundação, oferecendo uma alternativa ou complemento aos métodos estatísticos tradicionais e à análise multicritério de decisão (Rahman *et al.*, 2019; Bera *et al.*, 2022). No entanto, muitos modelos possuem o risco de *overfitting* (Vojtek *et al.*, 2021). Modelos complexos de AM, com muitos parâmetros, podem aprender o ruído presente nos dados de treinamento em vez dos padrões subjacentes, resultando em alta precisão nos dados de treinamento, mas baixa capacidade de generalização para dados não vistos (Nachappa *et al.*, 2020). Técnicas como validação cruzada e a utilização de conjuntos de teste independentes são essenciais para avaliar a capacidade de generalização dos modelos. Outro problema está na escolha dos hiperparâmetros dos modelos de AM, pois também influencia seu desempenho e requer um ajuste cuidadoso, muitas vezes através de métodos de otimização (Wang *et al.*, 2019; Pham *et al.*, 2021). A falta de uma teoria universal para a seleção de modelos de AM para mapeamento de suscetibilidade à inundação significa que diferentes algoritmos podem apresentar desempenhos distintos dependendo da área de estudo e dos dados disponíveis (Nachappa, *et al.*, 2020).

A interpretabilidade limitada de alguns modelos de AM, como redes neurais profundas e SVM com kernels não lineares, representa outra limitação metodológica (Liu *et al.*, 2021).

Embora esses modelos possam alcançar alta precisão preditiva, a dificuldade em compreender os mecanismos internos que levam a uma determinada previsão pode limitar a confiança dos usuários e a capacidade de extrair informações sobre a importância relativa dos diferentes fatores condicionantes. Em contraste, modelos baseados em árvores de decisão oferecem maior transparência e permitem avaliar a importância das variáveis (Vojtek *et al.*, 2021). No entanto, mesmo para modelos mais interpretáveis, a complexidade das interações não lineares entre os fatores de inundação pode ser difícil de capturar e explicar completamente.

Também é importante reconhecer que os mapas de suscetibilidade gerados por AM indicam a probabilidade espacial da suscetibilidade à inundação, mas não fornecem informações detalhadas sobre a extensão da inundação, a profundidade da água ou a velocidade do fluxo (Khosravi *et al.*, 2019; Liu *et al.*, 2021). Para análises mais detalhadas em escalas menores, como setores de rios, a integração de modelos de AM com modelos hidrológicos e hidráulicos pode ser necessária para levar em consideração as influências antropogênicas e estimar as características da inundação para diferentes cenários de probabilidade (Khosravi *et al.*, 2019).

A seleção metodológica deste estudo fundamentou-se em uma revisão sistemática de algoritmos de aprendizado de máquina aplicados em modelos de suscetibilidade, priorizando floresta aleatória e máquinas de vetores de suporte devido à sua comprovada eficácia em problemas de classificação espacial. A robustez da modelagem foi assegurada por meio de validação cruzada, seguida de avaliação de desempenho em conjunto de teste independente, cuja área sob a curva ROC (AUC) excedeu 99%, indicando alta capacidade discriminatória e generalização dos modelos para a área de estudo. Contudo, o desempenho ótimo dos algoritmos no contexto analisado não implica transferibilidade automática para outras regiões, uma vez que a calibração dos hiperparâmetros e a representatividade das variáveis preditoras são sensíveis à heterogeneidade geoespacial e à qualidade dos dados de entrada.

5.13.6 Limitações quanto ao tipo de inundação modelada

Uma limitação significativa nos modelos de suscetibilidade à inundação reside na sua aplicabilidade frequentemente direcionada a um tipo específico de evento, principalmente as inundações fluviais (ribeirinhas). Muitos estudos e metodologias apresentados nas fontes concentram-se em fatores condicionantes que são mais diretamente relevantes para a dinâmica de cheias de rios, como a distância a cursos d'água, a elevação em relação à rede hidrográfica,

a declividade da superfície e a acumulação de fluxo (Costache; Bui 2019; Vafakhah *et al.*, 2020; Bera *et al.*, 2022). Consequentemente, a capacidade desses modelos em mapear a suscetibilidade a outros tipos de inundações, como as rápidas (*flash floods*), pluviais (urbanas, por saturação do solo e precipitação intensa), costeiras (por ação de marés e ressacas) ou em lâmina (*sheetfloods*), torna-se limitada (Tehrany, *et al.*, 2013; Vojtek *et al.*, 2021). A própria seleção e ponderação dos fatores condicionantes, frequentemente baseada em dados históricos de inundações fluviais, pode não ser apropriada para caracterizar a probabilidade de ocorrência de eventos com mecanismos causais distintos (Vojtek *et al.*, 2021).

Para modelar adequadamente a suscetibilidade a tipos de inundação não fluviais, torna-se necessário adaptar as metodologias existentes, o que pode envolver a inclusão de diferentes fatores condicionantes e a modificação de seus esquemas de reclassificação (Vojtek *et al.*, 2021). Por exemplo, a modelagem de inundações rápidas pode exigir maior ênfase em características da bacia de drenagem, intensidade e duração da precipitação, e características do solo que influenciam a rápida geração de escoamento superficial (Janizadeh *et al.*, 2019). Da mesma forma, a avaliação da suscetibilidade a inundações costeiras demandaria a consideração de fatores como a elevação em relação ao nível do mar, a proximidade da costa e dados sobre eventos de tempestade e marés (Bera *et al.*, 2022).

A caracterização da bacia hidrográfica do Rio Cachoeira revela uma dinâmica de inundações multifatorial, integrando eventos pluviométricos extremos (com transbordamentos fluviais) e deficiências estruturais de drenagem urbana, agravadas pela expansão urbana desordenada (Silveira *et al.*, 2009). Contudo, a singularidade geomorfológica desta bacia reside em sua localização na zona costeira adjacente à Baía da Babitonga, sujeita a pulsos de inundação por marés astronômicas e meteorológicas, que induzem retroalimentação hidrológica em períodos de aumento da maré. Para discriminar espacialmente a influência maregráfica, incorporou-se ao modelo um layer binário georreferenciado (valor 1 para áreas expostas a inundações por maré; 0 para áreas não afetadas), derivado de simulações hidrodinâmicas validadas e com registros históricos.

Além dos fatores tradicionais (declividade, uso do solo, proximidade a canais fluviais), a inclusão deste parâmetro costeiro permitiu uma análise multivariada da suscetibilidade, segmentando contribuições antrópicas, hidrogeomorfológicas e oceanográficas. Essa abordagem estratificada não apenas ampliou a resolução espacial do modelo, mas também aumentou sua fidedignidade operacional, ao capturar sinergias entre processos distintos.

6. CONCLUSÃO

Este estudo investigou a metodologia de aprendizado de máquina (AM) para a geração de mapas de suscetibilidade à inundação (MSI), integrando uma análise bibliométrica, revisão sistemática de fatores preditivos e avaliação de ferramentas computacionais, culminando em uma aplicação prática de previsão da suscetibilidade à inundação na bacia hidrográfica do Rio Cachoeira (Joinville/SC). A análise bibliométrica (período 2013–2023) revelou um crescimento no uso de algoritmos como floresta aleatória, máquina de vetores de suporte e redes neurais artificiais em estudos de suscetibilidade, destacando-se a preferência por modelos *ensemble* e híbridos para capturar relações não lineares entre variáveis ambientais e hidrológicas (Sachdeva *et al.*, 2017; Zhao *et al.*, 2018; Nachappa *et al.*, 2020, Rana; Mahanta, 2022).

A análise dos softwares empregados nos estudos selecionados revela uma predominância no uso de sistemas de informação geográfica (SIG), com ênfase no ArcGIS devido à sua robustez em manipulação de dados espaciais e geoprocessamento, aliados a ambientes de programação como MATLAB e R, que oferecem pacotes estatísticos e bibliotecas especializadas em aprendizado de máquina. Essa sinergia metodológica permite a integração multidimensional de dados geoespaciais, a implementação algorítmica de modelos preditivos e a visualização dinâmica de resultados, assegurando precisão analítica e reprodutibilidade computacional. A interoperabilidade entre essas ferramentas não apenas otimiza a modelagem de fenômenos complexos, como também sustenta a rigorosidade metodológica, ao combinar a flexibilidade de linguagens de programação com a capacidade de análise espacial avançada dos SIGs. Tal abordagem evidencia um paradigma contemporâneo na ciência de dados ambientais, no qual a convergência entre técnicas estatísticas, inteligência artificial e geoprocessamento amplia a eficiência na resolução de problemas de escala real, garantindo análises escaláveis e adaptáveis a diferentes contextos geográficos e temáticos.

O avanço tecnológico associado ao aumento da capacidade de processamento computacional e surgimento de técnicas sofisticadas de aprendizado de máquina e *deep learning* tem viabilizado a análise de grandes volumes de dados geoespaciais em escalas temporais reduzidas. Este estudo demonstrou tal potencial ao processar uma bacia hidrográfica de 83,12 km² por meio de um *script* desenvolvido em linguagem R, integrado a ferramentas de sistemas de informação geográfica (QGIS), tudo em um ambiente *open-source*. A combinação dessas tecnologias proporcionou análises espaciais avançadas, com alta precisão e reprodutibilidade, sem a dependência de licenças comerciais onerosas. Essa abordagem evidencia a viabilidade de soluções acessíveis para órgãos públicos e governamentais, capazes

de otimizar recursos financeiros sem comprometer a qualidade técnica, alinhando-se às demandas por métodos sustentáveis e economicamente viáveis em geoprocessamento.

Além disso, este trabalho oferece um arcabouço metodológico replicável, detalhado no Apêndice C, que permite a adaptação do *script* a outras regiões ou contextos ambientais. A modularidade e a adaptabilidade do código facilitam a customização conforme variáveis específicas, como características hidrológicas, geomorfológicas ou climáticas, essenciais para a modelagem de fenômenos complexos. Essa flexibilidade não apenas amplia o escopo de aplicação do método, mas também fortalece a democratização de ferramentas analíticas, contribuindo para a padronização de protocolos científicos e a otimização de políticas públicas baseadas em evidências.

A análise crítica da literatura científica evidencia que alguns modelos de aprendizado de máquina superam abordagens tradicionais em precisão, especialmente em contextos complexos (Janizadeh *et al.*, 2018; Khosravi *et al.*, 2019; Vojtek *et al.*, 2021). Contudo, a quantificação ótima de variáveis preditoras necessárias para garantir a robustez dos modelos permanece um desafio metodológico, dada a interdependência não linear entre fatores ambientais. Alguns estudos destacam a relevância de parâmetros essenciais para modelagem da suscetibilidade à inundação, como o modelo digital de elevação (MDE), índice de declividade (DEC), e proximidade a corpos hídricos (DCH), ratificando sua influência estatística na predição de riscos (Janizadeh *et al.*, 2019; Pham *et al.*, 2021; Vafakhah *et al.*, 2021). Adicionalmente, a inclusão de variáveis ambientais complementares, como uso do solo, permeabilidade do solo e padrões pluviométricos, demonstra potencial sinérgico para aprimorar a acurácia preditiva, particularmente na detecção de zonas críticas em microescala (Pham *et al.*, 2021). Essa vantagem emerge da capacidade intrínseca dos algoritmos de aprendizado de máquina em decodificar relações não lineares e dependências espaciais, aspectos intrinsecamente negligenciados por modelos heurísticos baseados em interpretação humana. Assim, a flexibilidade analítica desses sistemas permite a identificação de padrões ocultos em grandes volumes de dados, estabelecendo um paradigma superior para a avaliação de suscetibilidade em cenários de alta heterogeneidade espacial.

Os resultados da modelagem para a BHRC, mediante a aplicação de algoritmos de aprendizado de máquina com otimização hiperparamétrica (*tuning*) e validação cruzada *k-fold*, demonstraram desempenho excepcional nas amostras de teste (30% do *dataset*), com valores de AUC-ROC superiores a 0.99, indicando excelência discriminativa na classificação de suscetibilidade. Esse desempenho elevado pode ser atribuído à robustez metodológica na seleção de amostras, garantindo representatividade espacial e fidedignidade estatística, aliada à

homogeneidade estrutural da bacia estudada, particularmente a distribuição equilibrada de parâmetros geomorfométricos, que favoreceu a redução de viés e a captura de padrões não lineares intrínsecos ao sistema. A convergência entre essas características permitiu que os modelos identificassem nuances espaciais críticas, como microdepressões e zonas de convergência de fluxo, otimizando a precisão preditiva.

Contudo, a generalização dessa acurácia para bacias com maior heterogeneidade geomorfológica ou variabilidade ambiental requer cautela. Estudos futuros devem investigar a aplicabilidade da abordagem em contextos geográficos distintos, incorporando técnicas de *data augmentation* e validação externa para mitigar riscos de *overfitting* e assegurar uma robustez *cross-basin* das predições (Liu *et al.*, 2024, Nai *et al.*, 2024).

Em relação aos achados da pesquisa sobre a suscetibilidade de inundações na bacia hidrográfica do Rio Cachoeira, destaca-se a distribuição percentual estratificada dos níveis de suscetibilidade. A análise quantitativa revelou que, embora 28% da área urbana esteja situado em zonas de baixa suscetibilidade (níveis 1 e 2), um dado crítico emerge: 33% das edificações (equivalente a 23,3% da área total analisada) localizam-se em áreas classificadas como Classe 5 (alta suscetibilidade), conforme metodologia de classificação multinível adotada. Esse cenário evidencia uma disparidade espacial na distribuição de riscos, com expressiva densidade de ocupação do solo em zonas críticas, o que amplifica a vulnerabilidade socioambiental e demanda atenção prioritária em políticas de ordenamento territorial. A concentração de infraestrutura em classe 5 sugere exposição sistêmica a danos materiais e humanos durante episódios hidrológicos extremos, reforçando a necessidade de estratégias de mitigação. Este processo está associado a processos históricos de urbanização não planejada (Silveira *et al.*, 2009). Esses resultados reforçam a urgência de integrar critérios urbanístico rígidos em planos diretores, visando reduzir assimetrias na resiliência urbana frente a cenários climáticos futuros.

Em síntese, este estudo comprovou a viabilidade metodológica da aplicação dos algoritmos de floresta aleatória (FA) e máquina de vetores de suporte (MVS), na modelagem espacial de suscetibilidade à inundação na Bacia Hidrográfica do Rio Cachoeira. Os mapas temáticos gerados, validados por métricas de desempenho e análises multicritério, apresentaram uma estratificação robusta entre áreas de alta e baixa suscetibilidade, evidenciando resolução espacial superior quando comparados a abordagens convencionais. A convergência dos resultados com estudos precedentes na mesma área (Muller, 2012; CPRM, 2015), reforça a eficácia do aprendizado de máquina como paradigma analítico para riscos hidrológicos, destacando sua capacidade de integrar variáveis ambientais em modelos preditivos escaláveis.

A adoção dessas técnicas não apenas otimiza a previsão de eventos extremos, como também subsidia políticas públicas baseadas em evidências, mitigando vulnerabilidades socioambientais e promovendo estratégias de adaptação urbana alinhadas à sustentabilidade.

No entanto, pode-se reconhecer que os mapas gerados por FA e MVS-R incorporam incertezas inerentes à simplificação de sistemas naturais multifatoriais. Tais limitações emergem da heterogeneidade espacial não capturada, da qualidade dos dados de entrada e da estocasticidade ambiental, fatores que podem influenciar a generalização dos modelos. Apesar disso, a precisão estatística obtida ($AUC > 0,99$), valida sua utilidade como ferramentas heurísticas para a gestão de riscos, permitindo a simulação probabilística de cenários críticos e a priorização de intervenções. Futuros estudos devem explorar a integração de dados de alta resolução temporal e espacial, além de técnicas híbridas (incorporação de modelos hidrológicos e hidráulicos), para reduzir vieses e aprimorar a representatividade dos modelos, garantindo assim decisões mais resilientes frente à crescente variabilidade climática.

REFERÊNCIAS

- ABEDI, R.; COSTACHE, R.; SHAFIZADEH-MOGHADAM, H.; PHAM, Q. B. Flash-flood susceptibility mapping based on XGBoost, random forest and boosted regression trees. **Geocarto International**, v. 37, n. 19, p. 5479–5496, 2022. DOI: 10.1080/10106049.2021.1920636.
- ABU-SALIH, B.; WONGTHONGTHAM, P.; COUTINHO, K.; QADDOURA, R.; ALSHAWEEESH, O.; WEDYAN, M. The development of a road network flood risk detection model using optimised ensemble learning. **Engineering Applications of Artificial Intelligence**, v. 122, p. 106081, 2023. DOI: 10.1016/j.engappai.2023.106081.
- ADNAN, Mohammed Sarfaraz Gani; SIAM, Zakaria Shams; KABIR, Irfat; KABIR, Zobaidul; AHMED, M. Razu; HASSAN, Quazi K.; RAHMAN, Rashedur M.; DEWAN, Ashraf. A novel framework for addressing uncertainties in machine learning-based geospatial approaches for flood prediction. **Journal of Environmental Management**, v. 326, p. 116813, 2023. DOI: 10.1016/j.jenvman.2022.116813.
- AGRAWAL, R.; SRIKANT, R. Fast Algorithms for Mining Association Rules in Large Databases. *In*: INTERNATIONAL CONFERENCE ON VERY LARGE DATA BASES, 20., 1994, San Francisco. **Proceedings [...]**. San Francisco: Morgan Kaufmann, 1994. p. 487-499.
- ÁGUAS DE SANTA CATARINA – SISTEMA DE INFORMAÇÕES DE RECURSOS HÍDRICOS. **Base documental: Comitê Rio Cubatão Norte – Publicações**. [s.d.]. Disponível em: <https://www.aguas.sc.gov.br/base-documental/documentos?jsmallfib=1&dir=JSROOT/Comite%20Rio%20Cubatao%20Norte/Publicacoes>. Acesso em: 01 jul. 2025.
- AHMADLOU, Mohammad; AL-FUGARA, A'kif; AL-SHABEEB, Abdel Rahman; ARORA, Aman; AL-ADAMAT, Rida; PHAM, Quoc Bao; AL-ANSARI, Nadhir; LINH, Nguyen Thi Thuy; SAJEDI, Hedieh. Flood susceptibility mapping and assessment using a novel deep learning model combining multilayer perceptron and autoencoder neural networks. **Journal of Flood Risk Management**, v. 14, n. 1, p. e12683, 2021. DOI: 10.1111/jfr3.12683.
- AHMED, I. A.; TALUKDAR, S.; SHAHFAHAD; PARVEZ, A.; RIHAN, M.; BAIG, M. R. I.; RAHMAN, A. Flood susceptibility modeling in the urban watershed of Guwahati using improved metaheuristic-based ensemble machine learning algorithms. **Geocarto International**, v. 37, n. 26, p. 12238–12266, 2022. DOI: 10.1080/10106049.2022.2066200.
- AHMED, N.; HOQUE, M. A.; ARABAMERI, A.; PAL, S. C.; CHAKRABORTTY, R.; JUI, J. Flood susceptibility mapping in Brahmaputra floodplain of Bangladesh using deep boost, deep learning neural network, and artificial neural network. **Geocarto International**, v. 37, n. 25, p. 8770–8791, 2022. DOI: 10.1080/10106049.2021.2005698.
- AL-ABADI, Alaa M. Mapping flood susceptibility in an arid region of southern Iraq using ensemble machine learning classifiers: a comparative study. **Arabian Journal of Geosciences**, v. 11, n. 9, p. 218, 2018. DOI: 10.1007/s12517-018-3584-5.

AL-AIZARI, A. R.; AL-MASNAY, Y. A.; AYDDA, A.; ZHANG, J.; ULLAH, K.; ISLAM, A. R. M. T.; HABIB, T.; KAKU, D. U.; NIZEYIMANA, J. C.; AL-SHAIBAH, B. Assessment Analysis of Flood Susceptibility in Tropical Desert Area: A Case Study of Yemen. **Remote Sensing**, v. 14, n. 16, p. 4050, 2022. DOI: 10.3390/rs14164050.

AL-AREEQ, Ahmed M.; ABBA, S. I.; YASSIN, Mohamed A.; BENAAFI, Mohammed; GHALEB, Mustafa; ALJUNDI, Isam H. Computational Machine Learning Approach for Flood Susceptibility Assessment Integrated with Remote Sensing and GIS Techniques from Jeddah, Saudi Arabia. **Remote Sensing**, v. 14, n. 21, p. 5515, 2022. DOI: 10.3390/rs14215515.

AL-AREEQ, Ahmed M.; SALEH, Radhwan A. A.; GHANIM, Abdulnoor A. J.; GHALEB, Mustafa; AL-AREEQ, Nabil M.; AL-WAJIH, Ebrahim. Flood hazard assessment in Yemen using a novel hybrid approach of Grey Wolf and Levenberg Marquardt optimizers. **Geocarto International**, v. 38, n. 1, p. 2243884, 2023. DOI: 10.1080/10106049.2023.2243884.

ALDIANSYAH, Septianto; WARDANI, Farida. Evaluation of flood susceptibility prediction based on a resampling method using machine learning. **Journal of Water and Climate Change**, v. 14, n. 3, p. 937–961, 2023. DOI: 10.2166/wcc.2023.494.

ALI, Sk Ajim; PARVIN, Farhana; PHAM, Quoc Bao; VOJTEK, Matej; VOJTEKOVÁ, Jana; COSTACHE, Romulus; NGUYEN, Thi Thuy Linh; NGUYEN, Hong Quan; AHMAD, Ateeque; GHORBANI, Mohammad Ali. GIS-based comparative assessment of flood susceptibility mapping using hybrid multi-criteria decision-making approach, naïve Bayes tree, bivariate statistics and logistic regression: A case of Topľa basin, Slovakia. **Ecological Indicators**, v. 117, p. 106620, 2020. DOI: 10.1016/j.ecolind.2020.106620.

ANDARYANI, Soghra; NOURANI, Vahid; HAGHIGHI, Ali Torabi; KEESSTRA, Saskia. Integration of hard and soft supervised machine learning for flood susceptibility mapping. **Journal of Environmental Management**, v. 291, p. 112731, 2021. DOI: 10.1016/j.jenvman.2021.112731.

APPELHANS, T.; DETSCH, F.; REUDENBACH, C.; WOELLAUER, S. **mapview**: Interactive Viewing of Spatial Data in R. R package, versão 2.11.2.9001. [S. l.], 2025. Disponível em: <https://github.com/r-spatial/mapview>. Acesso em: 01 mar. 2025.

ARABAMERI, Alireza; SAHA, Sunil; CHEN, Wei; ROY, Jagabandhu; PRADHAN, Biswajeet; BUI, Dieu Tien. Flash flood susceptibility modelling using functional tree and hybrid ensemble techniques. **Journal of Hydrology**, v. 587, p. 125007, 2020. DOI: 10.1016/j.jhydrol.2020.125007.

ARABAMERI, Alireza; SEYED DANESH, Amir; SANTOSH, M.; CERDA, Artemi; PAL, Subodh Chandra; GHORBANZADEH, Omid; ROY, Paramita; CHOWDHURI, Indrajit. Flood susceptibility mapping using meta-heuristic algorithms. **Geomatics, Natural Hazards and Risk**, Oxford, v. 13, n. 1, p. 949-974, 2022. DOI: 10.1080/19475705.2022.2060138.

ARIA, Massimo; CUCCURULLO, Corrado. *bibliometrix*: An R-tool for comprehensive science mapping analysis. **Journal of Informetrics**, v. 11, n. 4, p. 959–975, 2017. DOI: 10.1016/j.joi.2017.08.007.

ARIA, Massimo; CUCCURULLO, Corrado. **bibliometrix**: Comprehensive Science Mapping

Analysis. R package. 2023. Disponível em: <https://cran.r-project.org/web/packages/bibliometrix/index.html>. Acesso em: 30 abr. 2023.

ARORA, Aman; ARABAMERI, Alireza; PANDEY, Manish; SIDDIQUI, Masood A.; SHUKLA, U. K.; BUI, Dieu Tien; MISHRA, Varun Narayan; BHARDWAJ, Anshuman. Optimization of state-of-the-art fuzzy-metaheuristic ANFIS-based machine learning models for flood susceptibility prediction mapping in the Middle Ganga Plain, India. **Science of The Total Environment**, v. 750, p. 141565, 2021. DOI: 10.1016/j.scitotenv.2020.141565.

ARULKUMARAN, Kai; DEISENROTH, Marc Peter; BRUNDAGE, Miles; BHARATH, Anil Anthony. Deep Reinforcement Learning: A Brief Survey. **IEEE Signal Processing Magazine**, v. 34, n. 6, p. 26–38, 2017. DOI: 10.1109/MSP.2017.2743240.

ASIRI, Mashael M.; ALDEHIM, Ghadah; ALRUWAIS, Nuha; ALLAFI, Randa; ALZAHIRANI, Ibrahim; NOURI, Amal M.; ASSIRI, Mohammed; AHMED, Noura Abdelaziz. Coastal Flood risk assessment using ensemble multi-criteria decision-making with machine learning approaches. **Environmental Research**, v. 245, p. 118042, 2024. DOI: 10.1016/j.envres.2023.118042.

ASKAR, Shavan; ZERAAT PEYMA, Sajjad; YOUSEF, Mohanad Mohsen; PRODANOVA, Natalia Alekseevna; MUDA, Iskandar; ELSAHABI, Mohamed; HATAMIAFKOUEIEH, Javad. Flood Susceptibility Mapping Using Remote Sensing and Integration of Decision Table Classifier and Metaheuristic Algorithms. **Water**, v. 14, n. 19, p. 3062, 2022. DOI: 10.3390/w14193062.

AVAND, Mohammadtaghi; KURIQI, Alban; KHAZAEI, Majid; GHORBANZADEH, Omid. DEM resolution effects on machine learning performance for flood probability mapping. **Journal of Hydro-environment Research**, v. 40, p. 1–16, 2022. DOI: 10.1016/j.jher.2021.10.002.

AYDIN, Halit Enes; IBAN, Muzaffer Can. Predicting and analyzing flood susceptibility using boosting-based ensemble machine learning algorithms with SHapley Additive exPlanations. **Natural Hazards**, v. 116, n. 3, p. 2957–2991, 2023. DOI: 10.1007/s11069-022-05793-y.

BADAREU, Gabriela; CÂRSTINA, Silviu; MILITARU, Felicia; SIMINICĂ, Marian Ilie; CÎRCIUMARU, Daniel. What Is the Current Status of Research on the Impact of Digitalization in Medicine? A Bibliometric Analysis. **Healthcare**, v. 13, n. 2, p. 93, 2025. DOI: 10.3390/healthcare13020093.

BALOGUN, Abdul-Lateef; SHENG, Tan Yong; SALLEHUDDIN, Muhammad Helmy; AINA, Yusuf A.; DANO, Umar Lawal; PRADHAN, Biswajeet; YEKEEN, Shamsudeen; TELLA, Abdulwaheed. Assessment of data mining, multi-criteria decision making and fuzzy-computing techniques for spatial flood susceptibility mapping: a comparative study. **Geocarto International**, v. 37, n. 26, p. 12989–13015, 2022. DOI: 10.1080/10106049.2022.2076910.

BAND, Shahab S.; JANIZADEH, Saeid; CHANDRA PAL, Subodh; SAHA, Asish; CHAKRABORTTY, Rabin; MELESSE, Assefa M.; MOSAVI, Amirhosein. Flash Flood Susceptibility Modeling Using New Approaches of Hybrid and Ensemble Tree-Based Machine Learning Algorithms. **Remote Sensing**, v. 12, n. 21, p. 3568, 2020. DOI: 10.3390/rs12213568.

BASGALUPP, Márcio Porto. **LEGAL-Tree: um algoritmo genético multi-objetivo lexicográfico para indução de árvores de decisão**. 2010. 94 f. Tese (Doutorado em Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2010.

BEDIENT, Philip B.; HUBER, Wayne C.; VIEUX, Baxter E. **Hydrology and Floodplain Analysis**. Pearson Education, 2015.

BELSLEY, David A.; KUH, Edwin; WELSCH, Roy E. **Regression Diagnostics: Identifying Influential Data and Sources of Collinearity**. 1. ed. [s.l.]: Wiley, 1980. DOI: 10.1002/0471725153.

BENGIO, Yoshua; LECUN, Yann; HINTON, Geoffrey. Deep learning for AI. **Commun. ACM**, v. 64, n. 7, p. 58–65, 2021. DOI: 10.1145/3448250.

BENTIVOGLIO, Roberto; ISUFI, Elvin; JONKMAN, Sebastian Nicolaas; TAORMINA, Riccardo. Deep learning methods for flood mapping: a review of existing applications and future research directions. **Hydrology and Earth System Sciences**, v. 26, n. 16, p. 4345–4378, 2022. DOI: 10.5194/hess-26-4345-2022.

BERA, Subhas; DAS, Arup; MAZUMDER, Taraknath. Evaluation of machine learning, information theory and multi-criteria decision analysis methods for flood susceptibility mapping under varying spatial scale of analyses. **Remote Sensing Applications: Society and Environment**, v. 25, p. 100686, 2022. DOI: 10.1016/j.rsase.2021.100686.

BERGSTRA, James; BENGIO, Yoshua. Random search for hyper-parameter optimization. **Journal of Machine Learning Research**, v. 13, p. 281–305, 2012. DOI: 10.5555/2188385.2188395.

BHUIYAN, Shelina A.; BATAILLE, Clement P.; MCGRATH, Heather. Harmonizing and Extending Fragmented 100 Year Flood Hazard Maps in Canada's Capital Region Using Random Forest Classification. **Water**, v. 14, n. 23, p. 3801, 2022. DOI: 10.3390/w14233801.

BISHOP, Christopher M. **Pattern Recognition and Machine Learning**. New York: Springer, 2006.

BIVAND, R.; KEITT, T.; ROWLINGSON, B. **rgdal: Bindings for the 'Geospatial' Data Abstraction Library**. R package version 1.5-23. 2021. Disponível em: <https://CRAN.R-project.org/package=rgdal>. Acesso em: 01 mar. 2025.

BIVAND, R.; RUNDEL, C. **rgeos: Interface to Geometry Engine - Open Source ('GEOS')**. R package version 0.5-9. 2021. Disponível em: <https://CRAN.R-project.org/package=rgeos>. Acesso em: 01 mar. 2025.

BLOCH, Robin; LAMOND, Jessica. **Cities and flooding: a guide to integrated urban flood risk management for the 21st century**. Washington, D.C: World Bank, 2012.

BOX, G. E. P.; WILSON, K. B. On the Experimental Attainment of Optimum Conditions. **Journal of the Royal Statistical Society: Series B (Methodological)**, v. 13, n. 1, p. 1–38, 1951. DOI: 10.1111/j.2517-6161.1951.tb00067.x.

BRABB, Earl E.; PAMPEYAN, Earl H.; BONILLA, Manuel G. **Landslide susceptibility in San Mateo County, California**. Washington, D.C.: U.S. Geological Survey, 1972. 1 mapa, escala 1:62.500. (Miscellaneous Field Studies Map, MF-360). DOI: 10.3133/mf360.

BRADLEY, Andrew P. The use of the area under the ROC curve in the evaluation of machine learning algorithms. **Pattern Recognition**, v. 30, n. 7, p. 1145–1159, 1997. DOI: 10.1016/S0031-3203(96)00142-2.

BREIMAN, Leo. Random Forests. **Machine Learning**, v. 45, n. 1, p. 5–32, 2001. DOI: 10.1023/A:1010933404324.

BREUNIG, Markus M.; KRIEGEL, Hans-Peter; NG, Raymond T.; SANDER, Jörg. LOF: identifying density-based local outliers. **SIGMOD Rec.**, v. 29, n. 2, p. 93–104, 2000. DOI: 10.1145/335191.335388.

BROADUS, R. N. Toward a definition of “bibliometrics”. **Scientometrics**, v. 12, n. 5–6, p. 373–379, 1987. DOI: 10.1007/BF02016680.

BROWN, Tom B. et al. **Language Models are Few-Shot Learners**. arXiv, 2020. DOI: 10.48550/arXiv.2005.14165. Disponível em: <http://arxiv.org/abs/2005.14165>. Acesso em: 02 mar. 2025.

BUCKLAND, Michael; GEY, Fredric. The relationship between Recall and Precision. **Journal of the American Society for Information Science**, v. 45, n. 1, p. 12–19, 1994. DOI: 10.1002/(SICI)1097-4571(199401)45:1<12::AID-ASIS2>3.0.CO;2-L.

BUI, Dieu Tien; NGO, Phuong-Thao Thi; PHAM, Tien Dat; JAAFARI, Abolfazl; MINH, Nguyen Quang; HOA, Pham Viet; SAMUI, Pijush. A novel hybrid approach based on a swarm intelligence optimized extreme learning machine for flash flood susceptibility mapping. **CATENA**, v. 179, p. 184–196, 2019. DOI: 10.1016/j.catena.2019.04.009.

BURROUGH, P. A.; MCDONNELL, Rachael A.; LLOYD, Christopher D. **Principles of Geographical Information Systems**. OUP Oxford, 2015.

BURTON, Ian; KATES, Robert William; WHITE, Gilbert F. **The Environment as Hazard**. Oxford University Press, 1978.

CARRARA, Alberto. Multivariate models for landslide hazard evaluation. **Journal of the International Association for Mathematical Geology**, v. 15, n. 3, p. 403–426, 1983. DOI: 10.1007/BF01031290.

CARUANA, Rich; LOU, Yin; GEHRKE, Johannes; KOCH, Paul; STURM, Marc; ELHADAD, Noemie. Intelligible Models for HealthCare: Predicting Pneumonia Risk and Hospital 30-day Readmission. In: ACM SIGKDD INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING, 21., 2015, Sydney. **Anais [...]**. New York: Association for Computing Machinery, 2015. p. 1721–1730. DOI: 10.1145/2783258.2788613.

ÇELİK, Hüseyin E.; COSKUN, Gonca; CIGIZOGLU, H. Kerem; AĞIRALIOĞLU, Necati; AYDIN, Abdurrahim; ESIN, A. İlker. The analysis of 2004 flood on Kozdere Stream in

Istanbul. **Natural Hazards**, v. 63, n. 2, p. 461–477, 2012. DOI: 10.1007/s11069-012-0165-x.

CENTRE FOR RESEARCH ON THE EPIDEMIOLOGY OF DISASTERS (CRED). **Disasters in numbers 2022**. Brussels, 2023. Disponível em: <https://www.emdat.be/publications>. Acesso em: 23 abr. 2023.

CERBELAUD, Arnaud; BLANCHET, Gwendoline; ROUPIOZ, Laure; BREIL, Pascal; BRIOTTET, Xavier. Mapping Pluvial Flood-Induced Damages with Multi-Sensor Optical Remote Sensing: A Transferable Approach. **Remote Sensing**, v. 15, n. 9, p. 2361, 2023. DOI: 10.3390/rs15092361.

CHAKRABORTTY, Rabin; PAL, Subodh Chandra; REZAIE, Fatemeh; ARABAMERI, Alireza; LEE, Saro; ROY, Paramita; SAHA, Asish; CHOWDHURI, Indrajit; MOAYEDI, Hossein. Flash-flood hazard susceptibility mapping in Kangsabati River Basin, India. **Geocarto International**, v. 37, n. 23, p. 6713–6735, 2022. DOI: 10.1080/10106049.2021.195361

CHAN, Jireh Yi-Le; LEOW, Steven Mun Hong; BEA, Khean Thye; CHENG, Wai Khuen; PHOONG, Seuk Wai; HONG, Zeng-Wei; CHEN, Yen-Lin. Mitigating the Multicollinearity Problem and Its Machine Learning Approach: A Review. **Mathematics**, v. 10, n. 8, p. 1283, 2022. DOI: 10.3390/math10081283.

CHAPI, Kamran; SINGH, Vijay P.; SHIRZADI, Ataollah; SHAHABI, Himan; BUI, Dieu Tien; PHAM, Binh Thai; KHOSRAVI, Khabat. A novel hybrid artificial intelligence approach for flood susceptibility assessment. **Environmental Modelling & Software**, v. 95, p. 229–245, 2017. DOI: 10.1016/j.envsoft.2017.06.012.

CHAWLA, N. V.; BOWYER, K. W.; HALL, L. O.; KEGELMEYER, W. P. SMOTE: Synthetic Minority Over-sampling Technique. **Journal of Artificial Intelligence Research**, v. 16, p. 321–357, 2002. DOI: 10.1613/jair.953.

CHEN, Wei; HONG, Haoyuan; LI, Shaojun; SHAHABI, Himan; WANG, Yi; WANG, Xiaojing; AHMAD, Baharin Bin. Flood susceptibility modelling using novel hybrid approach of reduced-error pruning trees with bagging and random subspace ensembles. **Journal of Hydrology**, v. 575, p. 864–873, 2019. DOI: 10.1016/j.jhydrol.2019.05.089.

CHEN, Yuguo; ZHANG, Xinyi; YANG, Kejun; ZENG, Shiyi; HONG, Anyu. Modeling rules of regional flash flood susceptibility prediction using different machine learning models. **Frontiers in Earth Science**, v. 11, 2023. DOI: 10.3389/feart.2023.1117004.

CHENG, J.; KARAMBELKAR, B.; XIE, Y. **leaflet**: Create Interactive Web Maps with the JavaScript 'Leaflet' Library. R package version 2.0.4.1. 2021. Disponível em: <https://CRAN.R-project.org/package=leaflet>. Acesso em: 01 mar. 2025.

CHOUBIN, Bahram; MORADI, Ehsan; GOLSHAN, Mohammad; ADAMOWSKI, Jan; SAJEDI-HOSSEINI, Farzaneh; MOSAVI, Amir. An ensemble prediction of flood susceptibility using multivariate discriminant analysis, classification and regression trees, and support vector machines. **Science of The Total Environment**, v. 651, p. 2087–2096, 2019. DOI: 10.1016/j.scitotenv.2018.10.064.

CHOWDHURI, Indrajit; PAL, Subodh Chandra; CHAKRABORTTY, Rabin. Flood

susceptibility mapping by ensemble evidential belief function and binomial logistic regression model on river basin of eastern India. **Advances in Space Research**, v. 65, n. 5, p. 1466–1489, 2020. DOI: 10.1016/j.asr.2019.12.003.

CONROY, Gemma; MALLAPATY, Smriti. How China created AI model DeepSeek and shocked the world. **Nature**, v. 638, n. 8050, p. 300–301, 2025. DOI: 10.1038/d41586-025-00259-0.

CORTES, Corinna; VAPNIK, Vladimir. Support-vector networks. **Machine Learning**, v. 20, n. 3, p. 273–297, 1995. DOI: 10.1007/BF00994018.

COSTA, Fabio Henrique Lima. **Estudo da suscetibilidade a inundações na bacia hidrográfica do Rio Alcântara (RJ)**: subsídios ao desenvolvimento de metodologias para avaliação de enchentes urbanas. 2017. Dissertação (Mestrado em Geografia) – Faculdade de Formação de Professores, Universidade do Estado do Rio de Janeiro, São Gonçalo, 2017.

COSTACHE, Romulus et al. Flood hazard potential evaluation using decision tree state-of-the-art models. **Risk Analysis**, v. 44, n. 2, p. 439–458, 2024. DOI: 10.1111/risa.14179.

COSTACHE, Romulus et al. New neural fuzzy-based machine learning ensemble for enhancing the prediction accuracy of flood susceptibility mapping. **Hydrological Sciences Journal**, v. 65, n. 16, p. 2816–2837, 2020. b. DOI: 10.1080/02626667.2020.1842412.

COSTACHE, Romulus et al. Novel hybrid models between bivariate statistics, artificial neural networks and boosting algorithms for flood susceptibility assessment. **Journal of Environmental Management**, v. 265, p. 110485, 2020. a. DOI: 10.1016/j.jenvman.2020.110485.

COSTACHE, Romulus. Flood Susceptibility Assessment by Using Bivariate Statistics and Machine Learning Models - A Useful Tool for Flood Risk Management. **Water Resources Management**, v. 33, n. 9, p. 3239–3256, 2019. DOI: 10.1007/s11269-019-02301-z.

COSTACHE, Romulus; ABDO, Hazem Ghassan; PRATAP MISHRA, Arun; PAL, Subodh Chandra; ISLAM, Abu Reza Md. Towfiqul; PANDE, Chaitanya B.; ALMOHAMAD, Hussein; ABDULLAH AL DUGHAIIRI, Ahmed; ALBANAI, Jasem A. Using fuzzy and machine learning iterative optimized models to generate the flood susceptibility maps: case study of Prahova River basin, Romania. **Geomatics, Natural Hazards and Risk**, v. 14, n. 1, p. 2281241, 2023. DOI: 10.1080/19475705.2023.2281241.

COSTACHE, Romulus; ARABAMERI, Alireza; BLASCHKE, Thomas; PHAM, Quoc Bao; PHAM, Binh Thai; PANDEY, Manish; ARORA, Aman; LINH, Nguyen Thi Thuy; COSTACHE, Iulia. Flash-Flood Potential Mapping Using Deep Learning, Alternating Decision Trees and Data Provided by Remote Sensing Sensors. **Sensors**, v. 21, n. 1, p. 280, 2021. a. DOI: 10.3390/s21010280.

COSTACHE, Romulus; ARABAMERI, Alireza; COSTACHE, Iulia; CRĂCIUN, Anca; PHAM, Binh Thai. New Machine Learning Ensemble for Flood Susceptibility Estimation. **Water Resources Management**, v. 36, n. 12, p. 4765–4783, 2022. DOI: 10.1007/s11269-022-03276-0.

COSTACHE, Romulus; ARABAMERI, Alireza; ELKHRACHY, Ismail; GHORBANZADEH, Omid; PHAM, Quoc Bao. Detection of areas prone to flood risk using state-of-the-art machine learning models. **Geomatics, Natural Hazards and Risk**, v. 12, n. 1, p. 1488–1507, 2021. DOI: 10.1080/19475705.2021.1920480.

COSTACHE, Romulus; NGO, Phuong Thao Thi; BUI, Dieu Tien. Novel Ensembles of Deep Learning Neural Network and Statistical Learning for Flash-Flood Susceptibility Mapping. **Water**, v. 12, n. 6, p. 1549, 2020. DOI: 10.3390/w12061549.

COSTACHE, Romulus; PHAM, Quoc Bao; SHARIFI, Ehsan; LINH, Nguyen Thi Thuy; ABBA, S. I.; VOJTEK, Matej; VOJTEKOVÁ, Jana; NHI, Pham Thi Thao; KHOI, Dao Nguyen. Flash-Flood Susceptibility Assessment Using Multi-Criteria Decision Making and Machine Learning Supported by Remote Sensing and GIS Techniques. **Remote Sensing**, v. 12, n. 1, p. 106, 2020. c. DOI: 10.3390/rs12010106.

COSTACHE, Romulus; POPA, Mihnea Cristian; TIEN BUI, Dieu; DIACONU, Daniel Constantin; CIUBOTARU, Nicu; MINEA, Gabriel; PHAM, Quoc Bao. Spatial predicting of flood potential areas using novel hybridizations of fuzzy decision-making, bivariate statistics, and machine learning. **Journal of Hydrology**, v. 585, p. 124808, 2020. d. DOI: 10.1016/j.jhydrol.2020.124808.

COSTACHE, Romulus; TIEN BUI, Dieu. Identification of areas prone to flash-flood phenomena using multiple-criteria decision-making, bivariate statistics, machine learning and their ensembles. **Science of The Total Environment**, v. 712, p. 136492, 2020. DOI: 10.1016/j.scitotenv.2019.136492.

COSTACHE, Romulus; TIEN BUI, Dieu. Spatial prediction of flood potential using new ensembles of bivariate statistics and artificial intelligence: A case study at the Putna river catchment of Romania. **Science of The Total Environment**, v. 691, p. 1098–1118, 2019. DOI: 10.1016/j.scitotenv.2019.07.197.

COVER, T.; HART, P. Nearest neighbor pattern classification. **IEEE Transactions on Information Theory**, v. 13, n. 1, p. 21–27, 1967. DOI: 10.1109/TIT.1967.1053964.

CRED. **EM-DAT - The international disaster database**. [s.d.]. Disponível em: <https://www.emdat.be/>. Acesso em: 14 abr. 2025.

DARABI, Hamid et al. Development of a novel hybrid multi-boosting neural network model for spatial prediction of urban flood. **Geocarto International**, v. 37, n. 19, p. 5716–5741, 2022. DOI: 10.1080/10106049.2021.1920629.

DARABI, Hamid; TORABI HAGHIGHI, Ali; RAHMATI, Omid; JALALI SHAHROOD, Abolfazl; ROUZBEH, Sajad; PRADHAN, Biswajeet; TIEN BUI, Dieu. A hybridized model based on neural network and swarm intelligence-grey wolf algorithm for spatial prediction of urban flood-inundation. **Journal of Hydrology**, v. 603, p. 126854, 2021. DOI: 10.1016/j.jhydrol.2021.126854.

DAS, Sumit. Flood susceptibility mapping of the Western Ghat coastal belt using multi-source geospatial data and analytical hierarchy process (AHP). **Remote Sensing Applications: Society and Environment**, v. 20, p. 100379, 2020. DOI: 10.1016/j.rsase.2020.100379.

DAWSON, Geoffrey; BUTT, Junaid; JONES, Anne; FRACCARO, Paolo. Flood susceptibility mapping at the country scale using machine learning approaches. **Applied AI Letters**, v. 4, n. 4, p. e88, 2023. DOI: 10.1002/ail2.88.

DEANE-MAYER, Z. A.; KNOWLES, J. E. **caretEnsemble: Ensembles of Caret Models**. R package version 2.0.1. 2019. Disponível em: <https://CRAN.R-project.org/package=caretEnsemble>. Acesso em: 01 jul. 2024.

DEBNATH, Jatan et al. Evaluating Flood Susceptibility in the Brahmaputra River Basin: An Insight into Asia's Eastern Himalayan Floodplains Using Machine Learning and Multi-Criteria Decision-Making. **Earth Systems and Environment**, v. 7, n. 4, p. 733–760, 2023. DOI: 10.1007/s41748-023-00358-w.

DEROLIYA, Prakhar; GHOSH, Mousumi; MOHANTY, Mohit P.; GHOSH, Subimal; RAO, K. H. V. Durga; KARMAKAR, Subhankar. A novel flood risk mapping approach with machine learning considering geomorphic and socio-economic vulnerability dimensions. **Science of The Total Environment**, v. 851, p. 158002, 2022. DOI: 10.1016/j.scitotenv.2022.158002.

DODANGEH, Esmaeel; CHOUBIN, Bahram; EIGDIR, Ahmad Najafi; NABIPOUR, Narjes; PANAHI, Mehdi; SHAMSHIRBAND, Shahaboddin; MOSAVI, Amir. Integrated machine learning methods with resampling algorithms for flood susceptibility prediction. **Science of The Total Environment**, v. 705, p. 135983, 2020. DOI: 10.1016/j.scitotenv.2019.135983.

DOMINGOS, Pedro; PAZZANI, Michael. On the Optimality of the Simple Bayesian Classifier under Zero-One Loss. **Machine Learning**, v. 29, n. 2, p. 103–130, 1997. DOI: 10.1023/A:1007413511361.

DOMÍNGUEZ-CUESTA, María José. Susceptibility. In: BOBROWSKY, Peter T. (org.). **Encyclopedia of Natural Hazards**. Dordrecht: Springer Netherlands, 2013. p. 988–988. DOI: 10.1007/978-1-4020-4399-4_340. Disponível em: https://doi.org/10.1007/978-1-4020-4399-4_340. Acesso em: 09 fev. 2025.

DORDI, Truzaar; HENSTRA, Daniel; THISTLETHWAITE, Jason. Flood risk management and governance: A bibliometric review of the literature. **Journal of Flood Risk Management**, v. 15, n. 2, p. e12797, 2022. DOI: 10.1111/jfr3.12797.

DORMAN, Michael; GRASER, Anita; NOWOSAD, Jakub; LOVELACE, Robin. **Geocomputation with Python**. New York: Chapman and Hall/CRC, 2025. DOI: 10.1201/9781003379911.

DORMANN, Carsten F. et al. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. **Ecography**, v. 36, n. 1, p. 27–46, 2013. DOI: 10.1111/j.1600-0587.2012.07348.x.

DUDA, Richard O.; HART, Peter E.; STORK, David G. **Pattern Classification**. 2. ed. New York: Wiley-Interscience, 2001.

EDAMO, Muluneh Legesse; UKUMO, Tigistu Yisihak; LOHANI, Tarun Kumar; AYANA, Melkamu Teshome; AYELE, Mesfin Amaru; MADA, Zerihun Makayno; ABDI, Dawit Midagsa. A comparative assessment of multi-criteria decision-making analysis and machine

learning methods for flood susceptibility mapping and socio-economic impacts on flood risk in Abela-Abaya floodplain of Ethiopia. **Environmental Challenges**, v. 9, p. 100629, 2022. DOI: 10.1016/j.envc.2022.100629.

EKMEKCIOĞLU, Ömer; KOC, Kerim; ÖZGER, Mehmet; IŞIK, Zeynep. Exploring the additional value of class imbalance distributions on interpretable flash flood susceptibility prediction in the Black Warrior River basin, Alabama, United States. **Journal of Hydrology**, v. 610, p. 127877, 2022. DOI: 10.1016/j.jhydrol.2022.127877.

EKWUEME, Benjamin Nnamdi. Machine Learning Based Prediction of Urban Flood Susceptibility from Selected Rivers in a Tropical Catchment Area. **Civil Engineering Journal**, v. 8, n. 9, p. 1857–1878, 2022. DOI: 10.28991/CEJ-2022-08-09-08.

EL-HADDAD, Bosy A.; YOUSSEF, Ahmed M.; POURGHASEMI, Hamid R.; PRADHAN, Biswajeet; EL-SHATER, Abdel-Hamid; EL-KHASHAB, Mohamed H. Flood susceptibility prediction using four machine learning techniques and comparison of their performance at Wadi Qena Basin, Egypt. **Natural Hazards**, v. 105, n. 1, p. 83–114, 2021. DOI: 10.1007/s11069-020-04296-y.

ELMAHDY, Samy; ALI, Tarig; MOHAMED, Mohamed. Flash Flood Susceptibility Modeling and Magnitude Index Using Machine Learning and Geohydrological Models: A Modified Hybrid Approach. **Remote Sensing**, v. 12, n. 17, p. 2695, 2020. DOI: 10.3390/rs12172695.

ESFANDIARI, Morteza; ABDI, Ghasem; JABARI, Shabnam; MCGRATH, Heather; COLEMAN, David. Flood Hazard Risk Mapping Using a Pseudo Supervised Random Forest. **Remote Sensing**, v. 12, n. 19, p. 3206, 2020. DOI: 10.3390/rs12193206.

ESLAMINEZHAD, Seyed Ahmad; EFTEKHARI, Mobin; AZMA, Aliasghar; KIYANFAR, Ramin; AKBARI, Mohammad. Assessment of flood susceptibility prediction based on optimized tree-based machine learning models. **Journal of Water and Climate Change**, v. 13, n. 6, p. 2353–2385, 2022. DOI: 10.2166/wcc.2022.435.

ESTER, Martin; KRIEGEL, Hans-Peter; SANDER, Jörg; XU, Xiaowei. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In: INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING, 2., 1996, Portland. **Proceedings** [...]. Menlo Park: AAAI Press, 1996. p. 226-231. Disponível em: <https://ui.adsabs.harvard.edu/abs/1996kddm.conf..226E>. Acesso em: 02 mar. 2025.

FALAH, Fatemeh; RAHMATI, Omid; ROSTAMI, Mohammad; AHMADISHARAF, Ebrahim; DALIAKOPOULOS, Ioannis N.; POURGHASEMI, Hamid Reza. Artificial Neural Networks for Flood Susceptibility Mapping in Data-Scarce Urban Areas. In: POURGHASEMI, Hamid Reza; GOKCEOGLU, Candan (Org.). **Spatial Modeling in GIS and R for Earth and Environmental Sciences**. Amsterdam: Elsevier, 2019. cap. 14, p. 323-336. DOI: 10.1016/B978-0-12-815226-3.00014-4.

FANG, Lei; HUANG, Jinliang; CAI, Juntao; NITIVATTANANON, Vilas. Hybrid approach for flood susceptibility assessment in a flood-prone mountainous catchment in China. **Journal of Hydrology**, v. 612, p. 128091, 2022. DOI: 10.1016/j.jhydrol.2022.128091.

FANG, Zhice; WANG, Yi; PENG, Ling; HONG, Haoyuan. Predicting flood susceptibility using LSTM neural networks. **Journal of Hydrology**, v. 594, p. 125734, 2021. DOI: 10.1016/j.jhydrol.2020.125734.

FAWCETT, Tom. An introduction to ROC analysis. **Pattern Recognition Letters**, ROC Analysis in Pattern Recognition. v. 27, n. 8, ROC Analysis in Pattern Recognition, p. 861–874, 2006. DOI: 10.1016/j.patrec.2005.10.010.

FRIEDMAN, Jerome H. Greedy function approximation: A gradient boosting machine. **The Annals of Statistics**, v. 29, n. 5, p. 1189–1232, 2001. DOI: 10.1214/aos/1013203451.

FUJIWARA, Masami; TAKADA, Takenori. Environmental Stochasticity. In: **Encyclopedia of Life Sciences**. Chichester: John Wiley & Sons, 2017. p. 1-8. DOI: 10.1002/9780470015902.a0021220.pub2.

GALLINA, Valentina; TORRESAN, Silvia; CRITTO, Andrea; SPEROTTO, Anna; GLADE, Thomas; MARCOMINI, Antonio. A review of multi-risk methodologies for natural hazards: Consequences and challenges for a climate change impact assessment. **Journal of Environmental Management**, v. 168, p. 123–132, 2016. DOI: 10.1016/j.jenvman.2015.11.011.

GANAI, M. A.; HU, Minghui; MALIK, A. K.; TANVEER, M.; SUGANTHAN, P. N. Ensemble deep learning: A review. **Engineering Applications of Artificial Intelligence**, v. 115, p. 105151, 2022. DOI: 10.1016/j.engappai.2022.105151.

GAURKHEDE, N. T.; ADANE, V. S. FLOOD SUSCEPTIBILITY ANALYSIS USING FREELY AVAILABLE DATA, GIS, AND FREQUENCY RATIO MODEL FOR NAGPUR, INDIA. **Applied Ecology and Environmental Research**, v. 21, n. 3, p. 2341–2361, 2023. DOI: 10.15666/aer/2103_23412361.

GEOKIRIRI. **Formas de relevo da Bacia Hidrográfica do Rio Cachoeira**. GEOKIRIRI, [s.d.]. Disponível em: <https://geokiriri.com/geologia-geomorfologia-pedologia-e-hidrografia-da-bacia-hidrografica-do-rio-cachoeira/ipt/>. Acesso em: 15 maio. 2024.

GHAHRAMAN, Kaveh; NAGY, Balázs; NOOSHIN NOKHANDAN, Fatemeh. Flood-Prone Zones of Meandering Rivers: Machine Learning Approach and Considering the Role of Morphology (Kashkan River, Western Iran). **Geosciences**, v. 13, n. 9, p. 267, 2023. DOI: 10.3390/geosciences13090267.

Global assessment report on disaster risk reduction 2015 | UNDRR. 2015. Disponível em: <https://www.undrr.org/publication/global-assessment-report-disaster-risk-reduction-2015>. Acesso em: 17 mar. 2025.

Glossary of Terms. **Machine Learning**, v. 30, n. 2, p. 271–274, 1998. DOI: 10.1023/A:1017181826899.

GÖKCEOGLU, C.; AKSOY, H. Landslide susceptibility mapping of the slopes in the residual soils of the Mengen region (Turkey) by deterministic stability analyses and image processing techniques. **Engineering Geology**, v. 44, n. 1, p. 147–161, 1996. DOI: 10.1016/S0013-7952(97)81260-4.

GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. **Deep Learning**. MIT Press, 2016.

GREEN, David Marvin; SWETS, John A.; SWETS, John Arthur. **Signal Detection Theory and Psychophysics**. R. E. Krieger Publishing Company, 1974.

GUO, Zifeng; MOOSAVI, Vahid; LEITÃO, João P. Data-driven rapid flood prediction mapping with catchment generalizability. **Journal of Hydrology**, v. 609, p. 127726, 2022. DOI: 10.1016/j.jhydrol.2022.127726.

HA, Hang; LUU, Chinh; BUI, Quynh Duy; PHAM, Duy-Hoa; HOANG, Tung; NGUYEN, Viet-Phuong; VU, Minh Tuan; PHAM, Binh Thai. Flash flood susceptibility prediction mapping for a road network using hybrid machine learning models. **Natural Hazards**, v. 109, n. 1, p. 1247–1270, 2021. DOI: 10.1007/s11069-021-04877-5.

HA, Minh Cuong; VU, Phuong Lan; NGUYEN, Huu Duy; HOANG, Tich Phuc; DANG, Dinh Duc; DINH, Thi Bao Hoa; ŞERBAN, Gheorghe; RUS, Ioan; BREȚCAN, Petre. Machine Learning and Remote Sensing Application for Extreme Climate Evaluation: Example of Flood Susceptibility in the Hue Province, Central Vietnam Region. **Water**, v. 14, n. 10, p. 1617, 2022. DOI: 10.3390/w14101617.

HAARNOJA, Tuomas; ZHOU, Aurick; ABBEEL, Pieter; LEVINE, Sergey. **Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor**. arXiv, 2018. DOI: 10.48550/arXiv.1801.01290. Disponível em: <http://arxiv.org/abs/1801.01290>. Acesso em: 03 mar. 2025.

HABIBI, A.; DELAVAR, M. R.; SADEGHIAN, M. S.; NAZARI, B. Susceptibility mapping and assessment using regularized random forest and naïve Bayes algorithms. **ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences**, v. X-4-W1-2022, p. 241–248, 2023. a. DOI: 10.5194/isprs-annals-X-4-W1-2022-241-2023.

HABIBI, Alireza; DELAVAR, Mahmoud Reza; SADEGHIAN, Mohammad Sadegh; NAZARI, Borzoo; PIRASTEH, Saeid. A hybrid of ensemble machine learning models with RFE and Boruta wrapper-based algorithms for flash flood susceptibility assessment. **International Journal of Applied Earth Observation and Geoinformation**, v. 122, p. 103401, 2023. b. DOI: 10.1016/j.jag.2023.103401.

HADIPOUR, Vahid; VAFAIE, Freydoon; DEILAMI, Kaveh. Coastal Flooding Risk Assessment Using a GIS-Based Spatial Multi-Criteria Decision Analysis Approach. **Water**, v. 12, n. 9, p. 2379, 2020. DOI: 10.3390/w12092379.

HAN, Jiawei; PEI, Jian; YIN, Yiwen. Mining frequent patterns without candidate generation. **SIGMOD Rec.**, v. 29, n. 2, p. 1–12, 2000. DOI: 10.1145/335191.335372.

HANLEY, J. A.; MCNEIL, B. J. The meaning and use of the area under a receiver operating characteristic (ROC) curve. **Radiology**, v. 143, n. 1, p. 29–36, 1982. DOI: 10.1148/radiology.143.1.7063747.

HARTNETT, Michael; NASH, Stephen. High-resolution flood modeling of urban areas using MSN_Flood. **Water Science and Engineering**, v. 10, n. 3, p. 175–183, 2017. DOI:

10.1016/j.wse.2017.10.003.

HASAN, Md. Hasibul; AHMED, Asib; NAFEE, K. M.; HOSSEN, Md. Amzed. Use of machine learning algorithms to assess flood susceptibility in the coastal area of Bangladesh. **Ocean & Coastal Management**, v. 236, p. 106503, 2023. DOI: 10.1016/j.ocecoaman.2023.106503.

HASANUZZAMAN, Md; ISLAM, Aznarul; BERA, Biswajit; SHIT, Pravat Kumar. A comparison of performance measures of three machine learning algorithms for flood susceptibility mapping of river Silabati (tropical river, India). **Physics and Chemistry of the Earth, Parts A/B/C**, v. 127, p. 103198, 2022. DOI: 10.1016/j.pce.2022.103198.

HASANUZZAMAN, Md; SHIT, Pravat Kumar; BERA, Biswajit; ISLAM, Aznarul. Characterizing recurrent flood hazards in the Himalayan foothill region through data-driven modelling. **Advances in Space Research**, v. 71, n. 12, p. 5311–5326, 2023. DOI: 10.1016/j.asr.2023.02.028.

HASTIE, Trevor; TIBSHIRANI, Robert; FRIEDMAN, Jerome. **The Elements of Statistical Learning**. New York, NY: Springer, 2009. DOI: 10.1007/978-0-387-84858-7. Disponível em: <http://link.springer.com/10.1007/978-0-387-84858-7>. Acesso em: 04 mar. 2025.

HE, Haibo; GARCIA, Edwardo A. Learning from Imbalanced Data. **IEEE Transactions on Knowledge and Data Engineering**, v. 21, n. 9, p. 1263–1284, 2009. DOI: 10.1109/TKDE.2008.239.

HIDAYAH, Entin; INDARTO; LEE, Wei-Koon; HALIK, Gusfan; PRADHAN, Biswajeet. Assessing Coastal Flood Susceptibility in East Java, Indonesia: Comparison of Statistical Bivariate and Machine Learning Techniques. **Water**, v. 14, n. 23, p. 3869, 2022. DOI: 10.3390/w14233869.

HIJMANS, R. J. **raster: Geographic Data Analysis and Modeling**. R package version 3.5-15. 2021. Disponível em: <https://CRAN.R-project.org/package=raster>. Acesso em: 01 mar. 2025.

HINTON, G. E.; SALAKHUTDINOV, R. R. Reducing the Dimensionality of Data with Neural Networks. **Science**, v. 313, n. 5786, p. 504–507, 2006. DOI: 10.1126/science.1127647.

HITOURI, Sliman; MERIAME, Mohajane; SK AJIM, Ali; RENATA PACHECO, Quevedo; NGUYEN-HUY, Thong; QUOC BAO, Pham; ELKHRACHY, Ismail; VARASANO, Antonietta. Gully erosion mapping susceptibility in a Mediterranean environment: A hybrid decision-making model. **International Soil and Water Conservation Research**, v. 12, n. 2, p. 279–297, 2024. a. DOI: 10.1016/j.iswcr.2023.09.008.

HITOURI, Sliman; MOHAJANE, Meriame; LAHSAINI, Meriam; ALI, Sk Ajim; SETARGIE, Tadesual Asamin; TRIPATHI, Gaurav; D'ANTONIO, Paola; SINGH, Suraj Kumar; VARASANO, Antonietta. Flood Susceptibility Mapping Using SAR Data and Machine Learning Algorithms in a Small Watershed in Northwestern Morocco. **Remote Sensing**, v. 16, n. 5, p. 858, 2024. b. DOI: 10.3390/rs16050858.

HO, Jonathan; JAIN, Ajay; ABBEEL, Pieter. **Denoising Diffusion Probabilistic Models**.

arXiv, 2020. DOI: 10.48550/arXiv.2006.11239. Disponível em: <http://arxiv.org/abs/2006.11239>. Acesso em: 02 mar. 2025.

HOANG, Duc-Vinh; LIOU, Yuei-An. Assessing the influence of human activities on flash flood susceptibility in mountainous regions of Vietnam. **Ecological Indicators**, v. 158, p. 111417, 2024. DOI: 10.1016/j.ecolind.2023.111417.

HOSMER, David W.; LEMESHOW, Stanley. Application of Logistic Regression with Different Sampling Models. *In: Applied Logistic Regression*. John Wiley & Sons, Ltd, 2000. p. 203–222. DOI: 10.1002/0471722146.ch6.

HOUDIJK, R. Handbook mitigating spatial relevant risks in European regions and towns. Zwolle: ITINERIS, 2012.

HUANG, Yingchun; BÁRDOSY, András; ZHANG, Ke. Sensitivity of hydrological models to temporal and spatial resolutions of rainfall data. **Hydrology and Earth System Sciences**, v. 23, n. 6, p. 2647–2663, 2019. DOI: 10.5194/hess-23-2647-2019.

HUSSAIN, Muhammad; TAYYAB, Muhammad; ULLAH, Kashif; ULLAH, Safi; RAHMAN, Zahid Ur; ZHANG, Jiquan; AL-SHAIBAH, Bazel. Development of a new integrated flood resilience model using machine learning with GIS-based multi-criteria decision analysis. **Urban Climate**, v. 50, p. 101589, 2023. DOI: 10.1016/j.uclim.2023.101589.

HUTTER, Frank; KOTTHOFF, Lars; VANSCHOREN, Joaquin (ORG.). **Automated Machine Learning: Methods, Systems, Challenges**. Cham: Springer International Publishing, 2019. DOI: 10.1007/978-3-030-05318-5. Disponível em: <http://link.springer.com/10.1007/978-3-030-05318-5>. Acesso em: 3 mar. 2025.

IBRAHIM, Karim Sherif Mostafa Hassan; HUANG, Yuk Feng; AHMED, Ali Najah; KOO, Chai Hoon; EL-SHAFIE, Ahmed. A review of the hybrid artificial intelligence and optimization modelling of hydrological streamflow forecasting. **Alexandria Engineering Journal**, v. 61, n. 1, p. 279–303, 2022. DOI: 10.1016/j.aej.2021.04.100.

IGHILE, Eseosa Halima; SHIRAKAWA, Hiroaki; TANIKAWA, Hiroki. **Application of GIS and Machine Learning to Predict Flood Areas in Nigeria**. SUSTAINABILITYST ALBAN-ANLAGE 66, CH-4052 BASEL, SWITZERLANDMDPI, 2022. DOI: 10.3390/su14095039.

ILIA, Ioanna; TSANGARATOS, Paraskevas; TZAMPOGLOU, Ploutarchos; CHEN, Wei; HONG, Haoyuan. Flash flood susceptibility mapping using stacking ensemble machine learning models. **Geocarto International**, v. 37, n. 27, p. 15010–15036, 2022. DOI: 10.1080/10106049.2022.2093990.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). **Censo Demográfico 2022**. Rio de Janeiro: IBGE, 2023. Disponível em: <https://www.ibge.gov.br/cidades-e-estados/sc/joinville.html>. Acesso em: 10 maio 2024.

ISLAM, Rakibul; CHOWDHURY, Padmanabha. Local-scale flash flood susceptibility assessment in northeastern Bangladesh using machine learning algorithms. **Environmental Challenges**, v. 14, p. 100833, 2024. DOI: 10.1016/j.envc.2023.100833.

JAIN, MANOJ K.; KOTHYARI, UMESH C. Estimation of soil erosion and sediment yield using GIS. **Hydrological Sciences Journal**, v. 45, n. 5, p. 771–786, 2000. DOI: 10.1080/02626660009492376.

JAMES, Gareth; WITTEN, Daniela; HASTIE, Trevor; TIBSHIRANI, Robert. **An Introduction to Statistical Learning: with Applications in R**. New York, NY: Springer US, 2021. DOI: 10.1007/978-1-0716-1418-1.

JANIZADEH, Saeid; AVAND, Mohammadtaghi; JAAFARI, Abolfazl; PHONG, Tran Van; BAYAT, Mahmoud; AHMADISHARAF, Ebrahim; PRAKASH, Indra; PHAM, Binh Thai; LEE, Saro. Prediction Success of Machine Learning Methods for Flash Flood Susceptibility Mapping in the Tafresh Watershed, Iran. **Sustainability**, v. 11, n. 19, p. 5426, 2019. DOI: 10.3390/su11195426.

JANIZADEH, Saeid; VAFAKHAH, Mehdi; KAPELAN, Zoran; MOBARGHAE DINAN, Naghmeh. Hybrid XGboost model with various Bayesian hyperparameter optimization algorithms for flood hazard susceptibility modeling. **Geocarto International**, v. 37, n. 25, p. 8273–8292, 2022. DOI: 10.1080/10106049.2021.1996641.

JENSEN, J. R. **Remote Sensing of the Environment: An Earth Resource Perspective**. 2. ed. Boston: Pearson, 2015.

JOHNSON, Stephen C. Hierarchical clustering schemes. **Psychometrika**, v. 32, n. 3, p. 241–254, 1967. DOI: 10.1007/BF02289588.

JOINVILLE (SC). Prefeitura. **Joinville cidade em dados 2020**. Joinville, 26 jun. 2020. Disponível em: <https://www.joinville.sc.gov.br/publicacoes/joinville-cidade-em-dados-2020/>. Acesso em: 20 mar. 2024.

JOINVILLE (SC). Prefeitura. **PDDU: Plano Diretor de Drenagem Urbana da Bacia Hidrográfica do Rio Cachoeira**. [Joinville], 2011. Disponível em: https://www.joinville.sc.gov.br/wp-content/uploads/2017/11/PDDU-BHRC_Cartilha-Drenagem-Urbana-Joinville-enfrenta-o-desafio_2011-00-00.pdf. Acesso em: 20 mar. 2024.

Joinville Cidade em Dados 2020 - Prefeitura de Joinville. 2020. Disponível em: <https://www.joinville.sc.gov.br/publicacoes/joinville-cidade-em-dados-2020/>. Acesso em: 20 mar. 2024.

Joinville Cidade em Dados 2022 - Prefeitura de Joinville. 2022. Disponível em: <https://www.joinville.sc.gov.br/publicacoes/joinville-cidade-em-dados-2022/>. Acesso em: 20 mar. 2024.

JORDAN, Michael I. Graphical Models. **Statistical Science**, v. 19, n. 1, p. 140–155, 2004. DOI: 10.1214/088342304000000026.

KAISER, Maria; GÜNNEMANN, Stephan; DISSE, Markus. Regional-scale prediction of pluvial and flash flood susceptible areas using tree-based classifiers. **Journal of Hydrology**, v. 612, p. 128088, 2022. DOI: 10.1016/j.jhydrol.2022.128088.

KARGER, D. N. **climate Stability: Quantitative Methods for Climate Stability Analyses**. R package version 1.1.1. 2021. Disponível em: <https://CRAN.R-project.org/package=climateStability>. Acesso em: 01 mar. 2025.

KASSAMBARA, A. **ggpubr: 'ggplot2' Based Publication Ready Plots**. R package version 0.4.0. 2020. Disponível em: <https://CRAN.R-project.org/package=ggpubr>. Acesso em: 1 jul. 2025.

KAVEH, Neda; EBRAHIMI, Ataollah; ASADI, Esmail. Comparative analysis of random forest, exploratory regression, and structural equation modeling for screening key environmental variables in evaluating rangeland above-ground biomass. **Ecological Informatics**, v. 77, p. 102251, 2023. DOI: 10.1016/j.ecoinf.2023.102251.

KAZAKIS, Nerantzis; KOUGIAS, Ioannis; PATSIALIS, Thomas. Assessment of flood hazard areas at a regional scale using an index-based approach and Analytical Hierarchy Process: Application in Rhodope–Evros region, Greece. **Science of The Total Environment**, v. 538, p. 555–563, 2015. DOI: 10.1016/j.scitotenv.2015.08.055.

KHOSRAVI, Khabat et al. A comparative assessment of flood susceptibility modeling using Multi-Criteria Decision-Making Analysis and Machine Learning Methods. **Journal of Hydrology**, v. 573, p. 311–323, 2019. DOI: 10.1016/j.jhydrol.2019.03.073.

KHOSRAVI, Khabat; PHAM, Binh Thai; CHAPI, Kamran; SHIRZADI, Ataollah; SHAHABI, Himan; REVHAUG, Inge; PRAKASH, Indra; TIEN BUI, Dieu. A comparative assessment of decision trees algorithms for flash flood susceptibility modeling at Haraz watershed, northern Iran. **Science of The Total Environment**, v. 627, p. 744–755, 2018. DOI: 10.1016/j.scitotenv.2018.01.266.

KHOSRAVI, Khabat; POURGHASEMI, Hamid Reza; CHAPI, Kamran; BAHRI, Masoumeh. Flash flood susceptibility analysis and its mapping using different bivariate models in Iran: a comparison between Shannon's entropy, statistical index, and weighting factor models. **Environmental Monitoring and Assessment**, v. 188, n. 12, p. 656, 2016. DOI: 10.1007/s10661-016-5665-9.

KIA, Masoud Bakhtyari; PIRASTEH, Saied; PRADHAN, Biswajeet; MAHMUD, Ahmad Rodzi; SULAIMAN, Wan Nor Azmin; MORADI, Abbas. An artificial neural network model for flood simulation using GIS: Johor River Basin, Malaysia. **Environmental Earth Sciences**, v. 67, n. 1, p. 251–264, 2012. DOI: 10.1007/s12665-011-1504-z.

KOHANSARBAZ, Ashkan; KOHANSARBAZ, Afshin; SHABANLOU, Saeid; YOSEFVAND, Fariborz; RAJABI, Ahmad. Modelling flood susceptibility in northern Iran: Application of five well-known machine-learning models. **Irrigation and Drainage**, v. 71, n. 5, p. 1332–1350, 2022. DOI: 10.1002/ird.2745.

KRON, Wolfgang. Flood Risk = Hazard • Values • Vulnerability. **Water International**, 2005. DOI: 10.1080/02508060508691837.

KUHN, M. **caret: Classification and Regression Training**. R package version 6.0-86. 2020. Disponível em: <https://CRAN.R-project.org/package=caret>. Acesso em: 01 mar. 2025.

KUHN, M.; HETZEL, G.; VAUGHAN, D. **corr: Correlations in R**. R package version 0.4.3. 2020. Disponível em: <https://CRAN.R-project.org/package=corr>. Acesso em: 01 mar. 2025.

KULITHALAI SHIYAM SUNDAR, Parthasarathy; KUNDAPURA, Subrahmanya. Spatial Mapping of Flood Susceptibility Using Decision Tree–Based Machine Learning Models for the Vembanad Lake System in Kerala, India. **Journal of Water Resources Planning and Management**, v. 149, n. 10, p. 04023052, 2023. DOI: 10.1061/JWRMD5.WRENG-5858.

KURUGAMA, Kumudu Madhawa; KAZAMA, So; HIRAGA, Yusuke; SAMARASURIYA, Chaminda. A Novel Hybrid Approach of Spatial Prediction of Flood Susceptibility in Sri Lanka Using Bivariate Index of Entropy and Support Vector Machine Models. **Journal of JSCE**, v. 11, n. 2, 2023. DOI: 10.2208/journalofjsce.23-27037.

LAI, Tuan Anh; NGUYEN, Ngoc-Thach; BUI, Quang-Thanh. Hyper-parameter optimization of gradient boosters for flood susceptibility analysis. **Transactions in GIS**, v. 27, n. 1, p. 224–238, 2023. DOI: 10.1111/tgis.13023.

LAMOVEC, Peter; VELKANOVSKI, Tatjana; MIKOS, Matjaz; OSIR, Kristof. Detecting flooded areas with machine learning techniques: case study of the Selška Sora river flash flood in September 2007. **Journal of Applied Remote Sensing**, v. 7, n. 1, p. 073564, 2013. DOI: 10.1117/1.JRS.7.073564.

LANGFORD, John. Efficient Exploration in Reinforcement Learning. In: SAMMUT, Claude; WEBB, Geoffrey I. (org.). **Encyclopedia of Machine Learning**. Boston, MA: Springer US, 2010. p. 309–311. DOI: 10.1007/978-0-387-30164-8_244.

LECUN, Yann; BENGIO, Yoshua; HINTON, Geoffrey. Deep learning. **Nature**, v. 521, n. 7553, p. 436–444, 2015. DOI: 10.1038/nature14539.

LEE, Saro; RYU, Joo-Hyung; LEE, Moun-Jin; WON, Joong-Sun. Use of an artificial neural network for analysis of the susceptibility to landslides at Boun, Korea. **Environmental Geology**, v. 44, n. 7, p. 820–833, 2003. DOI: 10.1007/s00254-003-0825-y.

LEE, Sunmin; KIM, Jeong-Cheol; JUNG, Hyung-Sup; LEE, Moun-Jin; LEE, Saro. Spatial prediction of flood susceptibility using random-forest and boosted-tree models in Seoul metropolitan city, Korea. **Geomatics, Natural Hazards and Risk**, v. 8, n. 2, p. 1185–1203, 2017. DOI: 10.1080/19475705.2017.1308971.

LEVINE, Sergey; FINN, Chelsea; DARRELL, Trevor; ABBEEL, Pieter. **End-to-End Training of Deep Visuomotor Policies**. arXiv, 2016. DOI: 10.48550/arXiv.1504.00702. Disponível em: <http://arxiv.org/abs/1504.00702>. Acesso em: 3 mar. 2025.

LI, Jianuo; ZHANG, Hongyan; ZHAO, Jianjun; GUO, Xiaoyi; RIHAN, Wu; DENG, Guorong. Embedded Feature Selection and Machine Learning Methods for Flash Flood Susceptibility-Mapping in the Mainstream Songhua River Basin, China. **Remote Sensing**, v. 14, n. 21, p. 5523, 2022. a. DOI: 10.3390/rs14215523.

LI, Pin-Ching; DEY, Sayan; MERWADE, Venkatesh. Analyzing the effects of data splitting and covariate shift on machine learning based streamflow prediction in ungauged basins. **Journal of Hydrology**, v. 653, p. 132731, 2025. DOI: 10.1016/j.jhydrol.2025.132731.

LI, Wenkai; LIU, Yuanchi; LIU, Ziyue; GAO, Zhen; HUANG, Huabing; HUANG, Weijun. A Positive-Unlabeled Learning Algorithm for Urban Flood Susceptibility Modeling. **Land**, v. 11, n. 11, p. 1971, 2022. b. DOI: 10.3390/land11111971.

LI, Yao; OSEI, Frank Badu; HU, Tangao; STEIN, Alfred. Urban flood susceptibility mapping based on social media data in Chengdu city, China. **Sustainable Cities and Society**, v. 88, p. 104307, 2023. DOI: 10.1016/j.scs.2022.104307.

LIAO, Mingyong; WEN, Haijia; YANG, Ling; WANG, Guilin; XIANG, Xuekun; LIANG, Xiaowen. Improving the model robustness of flood hazard mapping based on hyperparameter optimization of random forest. **Expert Systems with Applications**, v. 241, p. 122682, 2024. DOI: 10.1016/j.eswa.2023.122682.

LILLICRAP, Timothy P.; HUNT, Jonathan J.; PRITZEL, Alexander; HEESS, Nicolas; EREZ, Tom; TASSA, Yuval; SILVER, David; WIERSTRA, Daan. **Continuous control with deep reinforcement learning**. arXiv, 2019. DOI: 10.48550/arXiv.1509.02971. Disponível em: <http://arxiv.org/abs/1509.02971>. Acesso em: 03 mar. 2025.

LIMA, Gabriel da Silva; AMORIM, Raul Reis. Suscetibilidade e vulnerabilidade: um impasse conceitual que dificulta a responsabilização pelo desastre. In: EVENTO EM COMEMORAÇÃO AOS 20 ANOS DO PROGRAMA DE PÓS-GRADUAÇÃO EM GEOGRAFIA (IG-UNICAMP), 2023, Campinas. Anais [...]. Campinas: IG-UNICAMP, 2023. p. 50-65.

LIU, Chia-Fu; HUANG, Lipai; YIN, Kai; BRODY, Sam; MOSTAFAVI, Ali. Flood Damage Cast: Building flood damage nowcasting with machine-learning and data augmentation. **International Journal of Disaster Risk Reduction**, v. 114, p. 104971, 2024. DOI: 10.1016/j.ijdr.2024.104971.

LIU, Fei Tony; TING, Kai Ming; ZHOU, Zhi-Hua. Isolation Forest. In: IEEE INTERNATIONAL CONFERENCE ON DATA MINING, 8., 2008, Pisa. **Anais** [...]. New York: IEEE, 2008. p. 413–422. DOI: 10.1109/ICDM.2008.17.

LIU, Jun; WANG, Jiyan; XIONG, Junnan; CHENG, Weiming; SUN, Huaizhang; YONG, Zhiwei; WANG, Nan. Hybrid Models Incorporating Bivariate Statistics and Machine Learning Methods for Flash Flood Susceptibility Assessment Based on Remote Sensing Datasets. **Remote Sensing**, v. 13, n. 23, p. 4945, 2021. a. DOI: 10.3390/rs13234945.

LIU, Jun; WANG, Jiyan; XIONG, Junnan; CHENG, Weiming; SUN, Huaizhang; YONG, Zhiwei; WANG, Nan. Hybrid Models Incorporating Bivariate Statistics and Machine Learning Methods for Flash Flood Susceptibility Assessment Based on Remote Sensing Datasets. **Remote Sensing**, Basel, v. 13, n. 23, p. 1-24, 2021. DOI: 10.3390/rs13234945.

LIU, Jun; XIONG, Junnan; CHEN, Yangbo; SUN, Huaizhang; ZHAO, Xueqiang; TU, Fengmiao; GU, Yu. A new avenue to improve the performance of integrated modeling for flash flood susceptibility assessment: Applying cluster algorithms. **Ecological Indicators**, v. 146, p. 109785, 2023. DOI: 10.1016/j.ecolind.2022.109785.

LIU, Junfei; LIU, Kai; WANG, Ming. A Residual Neural Network Integrated with a Hydrological Model for Global Flood Susceptibility Mapping Based on Remote Sensing

Datasets. **Remote Sensing**, v. 15, n. 9, p. 2447, 2023. DOI: 10.3390/rs15092447.

LMEIDA, Eliana Ribeiro. **Modelação hidrológica de inundações urbanas baseada em dados geoespaciais de alta resolução**. 2015. Dissertação (Mestrado em Engenharia do Ambiente) – Instituto Superior Técnico, Universidade de Lisboa, Lisboa, 2015. Disponível em: <https://repositorio.ulisboa.pt/handle/10451/18471>. Acesso em: 11 fev. 2025.

LOLLO, José Augusto de; MARTELI, Alice N.; LORANDI, Reinaldo. Flooding Susceptibility Identification Using the HAND Algorithm Tool Supported by Land Use/Land Cover Data. *In: IAEG/AEG annual meeting proceedings, San Francisco, California, 2018: volume 2*. Cham: Springer International Publishing, 2019. p. 107–112. DOI: 10.1007/978-3-319-93127-2_16.

LUO, Ziyuan; TIAN, Jian; ZENG, Jian; PILLA, Francesco. Resilient landscape pattern for reducing coastal flood susceptibility. **Science of The Total Environment**, v. 856, p. 159087, 2023. DOI: 10.1016/j.scitotenv.2022.159087.

LUU, Chinh et al. GIS-based ensemble computational models for flood susceptibility prediction in the Quang Binh Province, Vietnam. **Journal of Hydrology**, v. 599, p. 126500, 2021. a. DOI: 10.1016/j.jhydrol.2021.126500.

LUU, Chinh; BUI, Quynh Duy; COSTACHE, Romulus; NGUYEN, Luan Thanh; NGUYEN, Thu Thuy; VAN PHONG, Tran; VAN LE, Hiep; PHAM, Binh Thai. Flood-prone area mapping using machine learning techniques: a case study of Quang Binh province, Vietnam. **Natural Hazards**, v. 108, n. 3, p. 3229–3251, 2021. b. DOI: 10.1007/s11069-021-04821-7.

LUU, Chinh; NGUYEN, Duc-Dam; VAN PHONG, Tran; PRAKASH, Indra; PHAM, Binh Thai. Using Decision Tree J48 Based Machine Learning Algorithm for Flood Susceptibility Mapping: A Case Study in Quang Binh Province, Vietnam. *In: CIGOS 2021: emerging technologies and applications for green infrastructure*. Singapore: Springer Nature, 2022. p. 1927–1935. DOI: 10.1007/978-981-16-7160-9_195.

LYU, Hai-Min; YIN, Zhen-Yu. Flood susceptibility prediction using tree-based machine learning models in the GBA. **Sustainable Cities and Society**, v. 97, p. 104744, 2023. DOI: 10.1016/j.scs.2023.104744.

MADHURI, R.; SISTLA, S.; SRINIVASA RAJU, K. Application of machine learning algorithms for flood susceptibility assessment and risk management. **Journal of Water and Climate Change**, v. 12, n. 6, p. 2608–2623, 2021. DOI: 10.2166/wcc.2021.051.

MADUSHANI, Jathun Arachchige Thilini; WITHANAGE, Neel Chaminda; MISHRA, Prabuddh Kumar; MERAJ, Gowhar; KIBEBE, Caxton Griffith; KUMAR, Pankaj. Thematic and Bibliometric Review of Remote Sensing and Geographic Information System-Based Flood Disaster Studies in South Asia During 2004–2024. **Sustainability**, v. 17, n. 1, p. 217, 2025. DOI: 10.3390/su17010217.

MAHATO, Susanta; PAL, Swades; TALUKDAR, Swapan; SAHA, Tamal Kanti; MANDAL, Parikshit. Field based index of flood vulnerability (IFV): A new validation technique for flood susceptible models. **Geoscience Frontiers**, v. 12, n. 5, p. 101175, 2021. DOI: 10.1016/j.gsf.2021.101175.

MAHDIZADEH GHARAKHANLOU, Navid; PEREZ, Liliana. Flood susceptible prediction through the use of geospatial variables and machine learning methods. **Journal of Hydrology**, v. 617, p. 129121, 2023. DOI: 10.1016/j.jhydrol.2023.129121.

MAHDIZADEH GHARAKHANLOU, Navid; PEREZ, Liliana. Spatial Prediction of Current and Future Flood Susceptibility: Examining the Implications of Changing Climates on Flood Susceptibility Using Machine Learning Models. **Entropy**, v. 24, n. 11, p. 1630, 2022. DOI: 10.3390/e24111630.

MAHMOUD, Shereif H.; GAN, Thian Yew. Multi-criteria approach to develop flood susceptibility maps in arid regions of Middle East. **Journal of Cleaner Production**, v. 196, p. 216–229, 2018. DOI: 10.1016/j.jclepro.2018.06.047.

MALIK, Sadhan; CHANDRA PAL, Subodh; CHOWDHURI, Indrajit; CHAKRABORTTY, Rabin; ROY, Paramita; DAS, Biswajit. Prediction of highly flood prone areas by GIS based heuristic and statistical model in a monsoon dominated region of Bengal Basin. **Remote Sensing Applications: Society and Environment**, v. 19, p. 100343, 2020. DOI: 10.1016/j.rsase.2020.100343.

MANSOR, Shattri; PRADHAN, Biswajeet; DAUD, Mohamed; JAMALUDIN, Normalina; KHUZAIMAH, Zailani. Landslide susceptibility analysis using an artificial neural network model. *In: REMOTE SENSING FOR ENVIRONMENTAL MONITORING, GIS APPLICATIONS, AND GEOLOGY*, 7., 2007, Florença. **Proceedings [...]**. Bellingham: SPIE, 2007. p. 67490J. DOI: 10.1117/12.738462.

MCGRATH, Heather; GOHL, Piper Nora. Accessing the Impact of Meteorological Variables on Machine Learning Flood Susceptibility Mapping. **Remote Sensing**, Basel, v. 14, n. 7, p. 1–19, 2022. DOI: 10.3390/rs14071656.

MELLO, Yara Rúbia De; SIMM, Mariele; VIEIRA, Celso Voos. Características físicas da Bacia Hidrográfica do Rio Cachoeira, Joinville (SC). **Acta Biológica Catarinense**, v. 4, n. 3, p. 5–17, 2017. DOI: 10.21726/abc.v4i3.390.

MENG, Caisu; JIN, Hailiang. A Comparison of Machine Learning Models for Predicting Flood Susceptibility Based on the Enhanced NHAND Method. **Sustainability**, v. 15, n. 20, p. 14928, 2023. DOI: 10.3390/su152014928.

MERAJ, Gowhar; ROMSHOO, Shakil A.; YOUSUF, A. R.; ALTAF, Sadaff; ALTAF, Farrukh. Assessing the influence of watershed characteristics on the flood vulnerability of Jhelum basin in Kashmir Himalaya. **Natural Hazards**, v. 77, n. 1, p. 153–175, 2015. DOI: 10.1007/s11069-015-1605-1.

MEYER, H. **CAST: 'Caret' Applications for Spatial-Temporal Models**. R package version 0.4.0. 2021. Disponível em: <https://CRAN.R-project.org/package=CAST>. Acesso em: 01 mar. 2025.

MFONDOUM, Alfred Homère Ngandam et al. Stepwise integration of analytical hierarchy process with machine learning algorithms for landslide, gully erosion and flash flood susceptibility mapping over the North-Moungo perimeter, Cameroon. **Geoenvironmental Disasters**, v. 10, n. 1, p. 22, 2023. DOI: 10.1186/s40677-023-00254-5.

MIA, Md Uzzal; CHOWDHURY, Tahmida Naher; CHAKRABORTTY, Rabin; PAL, Subodh Chandra; AL-SADOON, Mohammad Khalid; COSTACHE, Romulus; ISLAM, Abu Reza Md Towfiqul. Flood Susceptibility Modeling Using an Advanced Deep Learning-Based Iterative Classifier Optimizer. **Land**, v. 12, n. 4, p. 810, 2023. DOI: 10.3390/land12040810.

MILETI, Dennis. **Disasters by Design: A Reassessment of Natural Hazards in the United States**. National Academies Press, 1999.

MINELLA, J. P. G.; MERTEN, G. H. Monitoramento de bacias hidrográficas para identificar fontes de sedimentos em suspensão. **Ciência Rural**, Santa Maria, v. 41, n. 3, p. 424-432, mar. 2011. DOI: 10.1590/S0103-84782011000300010.

MISHRA, Bibhu Prasad; GHOSE, Dillip Kumar; SATAPATHY, Deba Prakash. Geospatial modeling using hybrid machine learning approach for flood susceptibility. **Earth Science Informatics**, v. 15, n. 4, p. 2619–2636, 2022. DOI: 10.1007/s12145-022-00872-x.

MISHRA, Surendra Kumar; SINGH, Vijay P. **Soil Conservation Service Curve Number (SCS-CN) Methodology**. Dordrecht: Springer Netherlands, 2003. (Water Science and Technology Library, v. 42). DOI: 10.1007/978-94-017-0147-1.

MONAGHAN, Thomas F.; RAHMAN, Syed N.; AGUDELO, Christina W.; WEIN, Alan J.; LAZAR, Jason M.; EVERAERT, Karel; DMOCHOWSKI, Roger R. Foundational Statistical Principles in Medical Research: Sensitivity, Specificity, Positive Predictive Value, and Negative Predictive Value. **Medicina (Kaunas, Lithuania)**, v. 57, n. 5, p. 503, 2021. DOI: 10.3390/medicina57050503.

MONTGOMERY, Douglas C.; PECK, Elizabeth A.; VINING, G. Geoffrey. **Introduction to Linear Regression Analysis**. John Wiley & Sons, 2012.

MORADKHANI, H.; SOROOSHIAN, S. General review of rainfall-runoff modeling: model calibration, data assimilation, and uncertainty analysis. In: SINGH, V. P.; SOROOSHIAN, S.; SCHUMANN, A. H. (Eds.). **Hydrological Modelling and the Water Cycle**. Berlin: Springer, 2008. p. 1-24.

MOSAVI, Amirhosein; GOLSHAN, Mohammad; JANIZADEH, Saeid; CHOUBIN, Bahram; MELESSE, Assefa M.; DINEVA, Adrienn A. Ensemble models of GLM, FDA, MARS, and RF for flood and erosion susceptibility mapping: a priority assessment of sub-basins. **Geocarto International**, Oxford, v. 37, n. 9, p. 2541–2560, 2022. DOI: 10.1080/10106049.2020.1829101.

MUHADI, Nur Atirah; ABDULLAH, Ahmad Fikri; BEJO, Siti Khairunniza; MAHADI, Muhammad Razif; MIJIC, Ana. The Use of LiDAR-Derived DEM in Flood Applications: A Review. **Remote Sensing**, v. 12, n. 14, p. 2308, 2020. DOI: 10.3390/rs12142308.

MULLER, Cristiane Regina. **Avaliação de suscetibilidade a inundações utilizando geotecnologias para a bacia hidrográfica do rio Cachoeira - Joinville/SC**. 2012. Dissertação - Universidade do Estado de Santa Catarina, 2012. Disponível em: https://www.faed.udesc.br/arquivos/id_submenu/868/cristiane_regina_muller.pdf.

MURPHY, Kevin P. **Machine Learning: A Probabilistic Perspective**. MIT Press, 2012.

NACHAPPA, Thimmaiah Gudiyangada; TAVAKKOLI PIRALILOU, Sepideh; GHOLAMNIA, Khalil; GHORBANZADEH, Omid; RAHMATI, Omid; BLASCHKE, Thomas. Flood susceptibility mapping with machine learning, multi-criteria decision analysis and ensemble using Dempster Shafer Theory. **Journal of Hydrology**, v. 590, p. 125275, 2020. DOI: 10.1016/j.jhydrol.2020.125275.

NAI, Congyi; LIU, Xingcai; TANG, Qihong; LIU, Liu; SUN, Siao; GAFFNEY, Paul P. J. A Novel Strategy for Automatic Selection of Cross-Basin Data to Improve Local Machine Learning-Based Runoff Models. **Water Resources Research**, v. 60, n. 5, p. e2023WR035051, 2024. DOI: 10.1029/2023WR035051.

NAKABI, Taha Abdelhalim; TOIVANEN, Pekka. Deep reinforcement learning for energy management in a microgrid with flexible demand. **Sustainable Energy, Grids and Networks**, v. 25, p. 100413, 2021. DOI: 10.1016/j.segan.2020.100413.

NEYMAN, Jerzy; PEARSON, Egon Sharpe; PEARSON, Karl. IX. On the problem of the most efficient tests of statistical hypotheses. **Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character**, v. 231, n. 694–706, p. 289–337, 1933. DOI: 10.1098/rsta.1933.0009.

NGO, Phuong-Thao Thi; HOANG, Nhat-Duc; PRADHAN, Biswajeet; NGUYEN, Quang Khanh; TRAN, Xuan Truong; NGUYEN, Quang Minh; NGUYEN, Viet Nghia; SAMUI, Pijush; TIEN BUI, Dieu. A Novel Hybrid Swarm Optimized Multilayer Neural Network for Spatial Prediction of Flash Floods in Tropical Areas Using Sentinel-1 SAR Imagery and Geospatial Data. **Sensors**, v. 18, n. 11, p. 3704, 2018. DOI: 10.3390/s18113704.

NGO, Phuong-Thao Thi; PHAM, Tien Dat; NHU, Viet-Ha; LE, Thu Trang; TRAN, Dang An; PHAN, Duong Cao; HOA, Pham Viet; AMARO-MELLADO, José Lázaro; BUI, Dieu Tien. A novel hybrid quantum-PSO and credal decision tree ensemble for tropical cyclone induced flash flood susceptibility mapping with geospatial data. **Journal of Hydrology**, v. 596, p. 125682, 2021. DOI: 10.1016/j.jhydrol.2020.125682.

NGUYEN, Huu Duy et al. A novel hybrid approach to flood susceptibility assessment based on machine learning and land use change. Case study: a river watershed in Vietnam. **Hydrological Sciences Journal**, v. 67, n. 7, p. 1065–1083, 2022. a. DOI: 10.1080/02626667.2022.2060108.

NGUYEN, Huu Duy. Flood susceptibility assessment using hybrid machine learning and remote sensing in Quang Tri province, Vietnam. **Transactions in GIS**, v. 26, n. 7, p. 2776–2801, 2022. a. DOI: 10.1111/tgis.12980.

NGUYEN, Huu Duy. GIS-based hybrid machine learning for flood susceptibility prediction in the Nhat Le–Kien Giang watershed, Vietnam. **Earth Science Informatics**, v. 15, n. 4, p. 2369–2386, 2022. b. DOI: 10.1007/s12145-022-00825-4.

NGUYEN, Huu Duy; VAN, Chien Pham; DO, Anh Duc. Application of hybrid model-based deep learning and swarm-based optimizers for flood susceptibility prediction in Binh Dinh province, Vietnam. **Earth Science Informatics**, v. 16, n. 2, p. 1173–1193, 2023. DOI: 10.1007/s12145-023-00954-4.

NGUYEN, Huu Duy; VU, Phuong Lan; HA, Minh Cuong; DINH, Thi Bao Hoa; NGUYEN, Thuy Hang; HOANG, Tich Phuc; DOAN, Quang Cuong; PHAM, Van Manh; DANG, Dinh Kha. Flood susceptibility mapping using advanced hybrid machine learning and CyGNSS: a case study of Nghe An province, Vietnam. **Acta Geophysica**, v. 70, n. 6, p. 2785–2803, 2022. DOI: 10.1007/s11600-022-00940-2.

NORALLAHI, Mahya; SEYED KABOLI, Hesam. Urban flood hazard mapping using machine learning models: GARP, RF, MaxEnt and NB. **Natural Hazards**, v. 106, n. 1, p. 119–137, 2021. DOI: 10.1007/s11069-020-04453-3.

O'CALLAGHAN, John F.; MARK, David M. The extraction of drainage networks from digital elevation data. **Computer Vision, Graphics, and Image Processing**, v. 28, n. 3, p. 323–344, 1984. DOI: 10.1016/S0734-189X(84)80011-0.

OPELLA, Joe Marlou A.; HERNANDEZ, Alexander A. Developing a Flood Risk Assessment Using Support Vector Machine and Convolutional Neural Network: A Conceptual Framework. *In: IEEE INTERNATIONAL COLLOQUIUM ON SIGNAL PROCESSING & ITS APPLICATIONS*, 15., 2019, Penang. **Anais [...]**. New York: IEEE, 2019. p. 260–265. DOI: 10.1109/CSPA.2019.8695980.

ÖZDEMİR, Hüseyin; BADUNA KOÇYIĞIT, Müsteyde; AKAY, Diyar. Flood susceptibility mapping with ensemble machine learning: a case of Eastern Mediterranean basin, Türkiye. **Stochastic Environmental Research and Risk Assessment**, v. 37, n. 11, p. 4273–4290, 2023. DOI: 10.1007/s00477-023-02507-z.

PAL, Swades; SINGHA, Pankaj. Analyzing sensitivity of flood susceptible model in a flood plain river basin. **Geocarto International**, v. 37, n. 24, p. 7186–7219, 2022. DOI: 10.1080/10106049.2021.1967464.

PANDEY, Manish et al. Flood Susceptibility Modeling in a Subtropical Humid Low-Relief Alluvial Plain Environment: Application of Novel Ensemble Machine Learning Approach. **Frontiers in Earth Science**, v. 9, 2021. DOI: 10.3389/feart.2021.659296.

PAPAIIOANNOU, G.; VASILIADES, L.; LOUKAS, A. Multi-Criteria Analysis Framework for Potential Flood Prone Areas Mapping. **Water Resources Management**, v. 29, n. 2, p. 399–418, 2015. DOI: 10.1007/s11269-014-0817-6.

PASSAS, Ioannis. Bibliometric Analysis: The Main Steps. **Encyclopedia**, v. 4, n. 2, p. 1014–1025, 2024. DOI: 10.3390/encyclopedia4020065.

PEARSON, Karl. LIII. On lines and planes of closest fit to systems of points in space. **The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science**, v. 2, n. 11, p. 559–572, 1901. DOI: 10.1080/14786440109462720.

PEARSON, Karl. **On the Theory of Contingency and Its Relation to Association and Normal Correlation**. Cambridge University Press, 1904.

PEBESMA, Edzer; BIVAND, Roger. **Spatial Data Science: With Applications in R**. New York: Chapman and Hall/CRC, 2023. DOI: 10.1201/9780429459016.

PEDREGOSA, Fabian et al. Scikit-learn: Machine Learning in Python. **Journal of Machine Learning Research**, v. 12, n. 85, p. 2825–2830, 2011.

PERPIÑÁN, O.; HIJMAN, R. J. **rasterVis: Visualization Methods for the Raster Package**. R package version 0.50.3. 2021. Disponível em: <https://CRAN.R-project.org/package=rasterVis>. Acesso em: 01 mar. 2025.

PETERSON, W.; BIRDSALL, T.; FOX, W. The theory of signal detectability. **Transactions of the IRE Professional Group on Information Theory**, v. 4, n. 4, p. 171–212, 1954. DOI: 10.1109/TIT.1954.1057460.

PETERSON, W.; BIRDSALL, T.; FOX, W. The theory of signal detectability. **Transactions of the IRE Professional Group on Information Theory**, New York, v. 4, n. 4, p. 171–212, set. 1953. DOI: 10.1109/TIT.1954.1057460.

PHAM, Binh Thai et al. Can deep learning algorithms outperform benchmark machine learning algorithms in flood susceptibility modeling? **Journal of Hydrology**, v. 592, p. 125615, 2021. a. DOI: 10.1016/j.jhydrol.2020.125615.

PHAM, Binh Thai et al. GIS Based Hybrid Computational Approaches for Flash Flood Susceptibility Assessment. **Water**, v. 12, n. 3, p. 683, 2020. DOI: 10.3390/w12030683.

PHAM, Binh Thai; JAAFARI, Abolfazl; PHONG, Tran Van; YEN, Hoang Phan Hai; TUYEN, Tran Thi; LUONG, Vu Van; NGUYEN, Huu Duy; LE, Hiep Van; FOONG, Loke Kok. Improved flood susceptibility mapping using a best first decision tree integrated with ensemble learning techniques. **Geoscience Frontiers**, v. 12, n. 3, p. 101105, 2021. b. DOI: 10.1016/j.gsf.2020.11.003.

PHAM, Binh Thai; LUU, Chinh; PHONG, Tran Van; NGUYEN, Huu Duy; LE, Hiep Van; TRAN, Thai Quoc; TA, Huong Thu; PRAKASH, Indra. Flood risk assessment using hybrid artificial intelligence models integrated with multi-criteria decision analysis in Quang Nam Province, Vietnam. **Journal of Hydrology**, v. 592, p. 125815, 2021. c. DOI: 10.1016/j.jhydrol.2020.125815.

PHONG, Tran Van; NGUYEN, Duc Dam; PHAM, Binh Thai. Modeling and Mapping of Flood Susceptibility at Que Son District, Quang Nam Province, Vietnam using CatBoost. **IOP Conference Series: Materials Science and Engineering**, v. 1289, n. 1, p. 012019, 2023. DOI: 10.1088/1757-899X/1289/1/012019.

Plano Diretor de Drenagem Urbana (PDDU) da Bacia Hidrográfica do Rio Cachoeira - Prefeitura de Joinville. 2017. Disponível em: <https://www.joinville.sc.gov.br/publicacoes/plano-diretor-de-drenagem-urbana-pddu-da-bacia-hidrografica-do-rio-cachoeira/>. Acesso em: 19 abr. 2025.

PLATARIDIS, Konstantinos; MALLIOS, Zisis. Flood susceptibility mapping using hybrid models optimized with Artificial Bee Colony. **Journal of Hydrology**, v. 624, p. 129961, 2023. DOI: 10.1016/j.jhydrol.2023.129961.

POURGHASEMI, Hamid Reza; POUYAN, Soheila; BORDBAR, Mojgan; GOLKAR, Foroogh; CLAGUE, John J. Flood, landslides, forest fire, and earthquake susceptibility maps

using machine learning techniques and their combination. **Natural Hazards**, v. 116, n. 3, p. 3797–3816, 2023. DOI: 10.1007/s11069-023-05836-y.

POURZANGBAR, Ali; OBERLE, Peter; KRON, Andreas; FRANCA, Mário J. On the Application of Machine Learning into Flood Modeling: Data Consideration and Modeling Algorithm. *In*: GOURBESVILLE, Philippe; CAIGNAERT, Guy (Org.). **Advances in hydroinformatics: SimHydro 2023, volume 1**. Singapore: Springer Nature, 2024. p. 167–182. DOI: 10.1007/978-981-97-4072-7_11.

POWERS, David M. W. **Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation**. 2020. Disponível em: <https://arxiv.org/abs/2010.16061v1>. Acesso em: 13 maio. 2024.

POWERS, David Martin. Evaluation: from Precision, Recall and F-measure to ROC, Informedness, Markedness and Correlation. 2011. Disponível em: <http://hdl.handle.net/2328/27165>. Acesso em: 13 maio. 2024.

POWERS, David. Evaluation: From Precision, Recall and F-Factor to ROC, Informedness, Markedness & Correlation. **Mach. Learn. Technol.**, v. 2, 2008.

PRADHAN, B.; YOUSSEF, A. M. A 100-year maximum flood susceptibility mapping using integrated hydrological and hydrodynamic models: Kelantan River Corridor, Malaysia: A 100-year maximum flood susceptibility mapping. **Journal of Flood Risk Management**, v. 4, n. 3, p. 189–202, 2011. DOI: 10.1111/j.1753-318X.2011.01103.x.

PRADHAN, Biswajeet. Flood susceptible analysis at Kelantan river basin using remote sensing and logistic regression model. v. 37, p. 2496, 2008.

PRADHAN, Biswajeet; LEE, Saro; DIKSHIT, Abhirup; KIM, Hyesu. Spatial flood susceptibility mapping using an explainable artificial intelligence (XAI) model. **Geoscience Frontiers**, v. 14, n. 6, p. 101625, 2023. DOI: 10.1016/j.gsf.2023.101625.

PRASAD, Pankaj; LOVESON, Victor Joseph; DAS, Bappa; KOTHA, Mahender. Novel ensemble machine learning models in flood susceptibility mapping. **Geocarto International**, Oxford, v. 37, n. 16, p. 4571–4593, 2022. DOI: 10.1080/10106049.2021.1892209.

PROBST, Philipp; BOULESTEIX, Anne-Laure; BISCHL, Bernd. Tunability: importance of hyperparameters of machine learning algorithms. **Journal of Machine Learning Research**, Brookline, v. 20, n. 53, p. 1-32, 2019. DOI: 10.48550/arXiv.1802.09596.

QAZI, Atika; HASAN, Najmul; MAO, Rui; ELHAG MOHAMED ABO, Mohamed; KUMAR DEY, Samrat; HARDAKER, Glenn. Machine Learning-Based Opinion Spam Detection: A Systematic Literature Review. **IEEE Access**, v. 12, p. 143485–143499, 2024. DOI: 10.1109/ACCESS.2024.3399264.

QUINN, P.; BEVEN, K.; CHEVALLIER, P.; PLANCHON, O. The prediction of hillslope flow paths for distributed hydrological modelling using digital terrain models. **Hydrological Processes**, v. 5, n. 1, p. 59–79, 1991. DOI: 10.1002/hyp.3360050106.

R CORE TEAM. **R: A Language and Environment for Statistical Computing**. Vienna, Austria: R Foundation for Statistical Computing, 2024. Disponível em: <https://www.R-project.org/>. Acesso em: 1 jul. 2025.

RABAMERI, Alireza; DANESH, Amir Seyed; SANTOSH, M.; CERDA, Artemi; PAL, Subodh Chandra; GHORBANZADEH, Omid; ROY, Paramita; CHOWDHURI, Indrajit. Flood susceptibility mapping using meta-heuristic algorithms. **Geomatics, Natural Hazards and Risk**, v. 13, n. 1, p. 949–974, 2022. DOI: 10.1080/19475705.2022.2060138.

RAHMAN, Mahfuzur et al. Flooding and its relationship with land cover change, population growth, and road density. **Geoscience Frontiers**, v. 12, n. 6, p. 101224, 2021. DOI: 10.1016/j.gsf.2021.101224.

RAHMAN, Mahfuzur; NINGSHENG, Chen; ISLAM, Md Monirul; DEWAN, Ashraf; IQBAL, Javed; WASHAKH, Rana Muhammad Ali; SHUFENG, Tian. Flood Susceptibility Assessment in Bangladesh Using Machine Learning and Multi-criteria Decision Analysis. **Earth Systems and Environment**, v. 3, n. 3, p. 585–601, 2019. DOI: 10.1007/s41748-019-00123-y.

RAHMATI, Omid et al. Development of novel hybridized models for urban flood susceptibility mapping. **Scientific Reports**, v. 10, n. 1, p. 12937, 2020. DOI: 10.1038/s41598-020-69703-7.

RAHMATI, Omid; POURGHASEMI, Hamid Reza. Identification of Critical Flood Prone Areas in Data-Scarce and Ungauged Regions: A Comparison of Three Data Mining Models. **Water Resources Management**, v. 31, n. 5, p. 1473–1487, 2017. DOI: 10.1007/s11269-017-1589-6.

RAHMATI, Omid; POURGHASEMI, Hamid Reza; ZEINIVAND, Hossein. Flood susceptibility mapping using frequency ratio and weights-of-evidence models in the Golastan Province, Iran. **Geocarto International**, v. 31, n. 1, p. 42–70, 2016. DOI: 10.1080/10106049.2015.1041559.

RANA, Manish Singh; MAHANTA, Chandan. Flash-flood susceptibility modelling in a data-scarce region using a novel hybrid approach and trend analysis of precipitation. **Hydrological Sciences Journal**, v. 68, n. 16, p. 2336–2356, 2023. b. DOI: 10.1080/02626667.2023.2259887.

RANA, Manish Singh; MAHANTA, Chandan. Spatial prediction of flash flood susceptible areas using novel ensemble of bivariate statistics and machine learning techniques for ungauged region. **Natural Hazards**, v. 115, n. 1, p. 947–969, 2023. a. DOI: 10.1007/s11069-022-05580-9.

RATHNASIRI, Pavithra; ADENIYI, Onaopepo; THURAIRAJAH, Niraj. Data-driven approaches to built environment flood resilience: A scientometric and critical review. **Advanced Engineering Informatics**, v. 57, p. 102085, 2023. DOI: 10.1016/j.aei.2023.102085.

RAZAVI-TERMEH, Seyed Vahid; SADEGHI-NIARAKI, Abolghasem; CHOI, Soo-Mi. A new approach based on biology-inspired metaheuristic algorithms in combination with random forest to enhance the flood susceptibility mapping. **Journal of Environmental Management**, v. 345, p. 118790, 2023. DOI: 10.1016/j.jenvman.2023.118790.

RAZAVI-TERMEH, Seyed Vahid; SADEGHI-NIARAKI, Abolghasem; SEO, MyoungBae;

CHOI, Soo-Mi. Application of genetic algorithm in optimization parallel ensemble-based machine learning algorithms to flood susceptibility mapping using radar satellite imagery. **Science of The Total Environment**, v. 873, p. 162285, 2023. DOI: 10.1016/j.scitotenv.2023.162285.

RE, Matteo; VALENTINI, Giorgio. Ensemble methods: A review. *Em: Advances in Machine Learning and Data Mining for Astronomy*. [s.l: s.n.]. p. 563–594.

RENTSCHLER, Jun; SALHAB, Melda; JAFINO, Bramka Arga. Flood exposure and poverty in 188 countries. **Nature Communications**, v. 13, n. 1, p. 3527, 2022. DOI: 10.1038/s41467-022-30727-4.

RIAZI, Mostafa; KHOSRAVI, Khabat; SHAHEDI, Kaka; AHMAD, Sajjad; JUN, Changhyun; BATENI, Sayed M.; KAZAKIS, Nerantzis. Enhancing flood susceptibility modeling using multi-temporal SAR images, CHIRPS data, and hybrid machine learning algorithms. **Science of The Total Environment**, v. 871, p. 162066, 2023. DOI: 10.1016/j.scitotenv.2023.162066.

RIBEIRO, J. M. G. O.; NOVAIS, T. M. **Cartilha Geográfica Bacias Hidrográficas dos Rios Cubatão (Norte) e Cachoeira Joinville**. 1. ed. [Joinville]: Mercado de Comunicação, 2014.

RIBEIRO, José Mário Gomes; OLIVEIRA, Therezinha Maria Novais De. **Bacias hidrográficas dos rios Cubatão (Norte) e Cachoeira**. 2014. Disponível em: <https://www.comitebabitonga.org.br/pagina/comite/publicacoes>.

RICHE, Abdelkader; DRIAS, Ammar; GUERMOUI, Mawloud; GHERIB, Tarek; BOULMAIZ, Tayeb; SOUISSI, Boularbah; MELGANI, Farid. A Novel Hybrid Deep-Learning Approach for Flood-Susceptibility Mapping. **Remote Sensing**, v. 16, n. 19, p. 3673, 2024. DOI: 10.3390/rs16193673.

ROBIN, X. et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. **BMC Bioinformatics**, v. 12, n. 77, 2011. Disponível em: <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-12-77>. Acesso em: 1 jul. 2025.

ROCHA, Jorge; ABRANTES, Patrícia (Org.). **Geographic Information Systems and Science**. London: IntechOpen, 2019. DOI: 10.5772/intechopen.78455.

ROKACH, Lior. Ensemble-based classifiers. **Artificial Intelligence Review**, v. 33, n. 1–2, p. 1–39, 2010. DOI: 10.1007/s10462-009-9124-7.

ROTIMI, F. E.; KALATEHJARI, R.; MOSHOOD, T. D.; DOKYI, G. A comprehensive review of bibliometric and methodological approaches in flood mitigation studies: current trends and future directions. **Journal of Safety Science and Resilience**, Amsterdam, v. 6, p. 280-296, 2025. DOI: 10.1016/j.jnlssr.2024.12.004.

ROUBOS, J. A.; VAN STRATEN, G.; VAN BOXTEL, A. J. B. An evolutionary strategy for fed-batch bioreactor optimization; concepts and performance. **Journal of Biotechnology**, v. 67, n. 2, p. 173–187, 1999. DOI: 10.1016/S0168-1656(98)00174-6.

ROUSE, J. W.; HAAS, R. H.; SCHELL, J. A.; DEERING, D. W. Monitoring vegetation systems in the Great Plains with ERTS. *In*: THIRD ERTS-1 SYMPOSIUM, 3., 1973, Washington, D.C. **Anais** [...]. Greenbelt: NASA, 1974. p. 309-317. Disponível em: <https://ntrs.nasa.gov/citations/19740022614>. Acesso em: 01 mar. 2025.

RSTUDIO TEAM. **RStudio: Integrated Development Environment for R**. Boston, MA: RStudio, PBC, 2024. Disponível em: <https://posit.co/>. Acesso em: 01 mar. 2025.

RUMMERY, G. A.; NIRANJAN, Mahesan. **On-line Q-learning Using Connectionist Systems**. University of Cambridge, Department of Engineering, 1994. Disponível em: https://mi.eng.cam.ac.uk/reports/svr-ftp/auto-pdf/rummery_tr166.pdf.

SAATY, Thomas L. Axiomatic Foundation of the Analytic Hierarchy Process. **Management Science**, v. 32, n. 7, p. 841–855, 1986.

SAATY, Thomas L. **The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation**. McGraw-Hill International Book Company, 1980.

SAATY, Thomas L. What is the Analytic Hierarchy Process? *In*: MITRA, Gautam; GREENBERG, Harvey J.; LOOTSMA, Freerk A.; RIJKAERT, Marcel J.; ZIMMERMANN, Hans J. (Org.). **Mathematical models for decision support**. Berlin: Springer, 1988. p. 109–121. DOI: 10.1007/978-3-642-83555-1_5.

SABER, Mohamed; BOULMAIZ, Tayeb; GUERMOUI, Mawloud; ABDRABO, Karim I.; KANTOUSH, Sameh A.; SUMI, Tetsuya; BOUTAGHANE, Hamouda; NOHARA, Daisuke; MABROUK, Emad. Examining LightGBM and CatBoost models for wadi flash flood susceptibility prediction. **Geocarto International**, v. 37, n. 25, p. 7462–7487, 2022. DOI: 10.1080/10106049.2021.1974959.

SACHDEVA, Shruti; BHATIA, Tarunpreet; VERMA, A. K. Flood susceptibility mapping using GIS-based support vector machine and particle swarm optimization: A case study in Uttarakhand (India). *In*: INTERNATIONAL CONFERENCE ON COMPUTING, COMMUNICATION AND NETWORKING TECHNOLOGIES, 8., 2017, Delhi. **Anais** [...]. New York: IEEE, 2017. p. 1–7. DOI: 10.1109/ICCCNT.2017.8204182.

SACHDEVA, Shruti; KUMAR, Bijendra. Flood susceptibility mapping using extremely randomized trees for Assam 2020 floods. **Ecological Informatics**, v. 67, p. 101498, 2022. DOI: 10.1016/j.ecoinf.2021.101498.

SAHA, Asish; PAL, Subodh Chandra; ARABAMERI, Alireza; BLASCHKE, Thomas; PANAHI, Somayeh; CHOWDHURI, Indrajit; CHAKRABORTTY, Rabin; COSTACHE, Romulus; ARORA, Aman. Flood Susceptibility Assessment Using Novel Ensemble of Hyperpipes and Support Vector Regression Algorithms. **Water**, v. 13, n. 2, p. 241, 2021. a. DOI: 10.3390/w13020241.

SAHA, Sunil; GAYEN, Amiya; BAYEN, Bijoy. Deep learning algorithms to develop Flood susceptibility map in Data-Scarce and Ungauged River Basin in India. **Stochastic Environmental Research and Risk Assessment**, v. 36, n. 10, p. 3295–3310, 2022. DOI: 10.1007/s00477-022-02195-1.

SAHA, Tamal Kanti; PAL, Swades; TALUKDAR, Swapan; DEBANSHI, Sandipta; KHATUN, Rumki; SINGHA, Pankaj; MANDAL, Indrajit. How far spatial resolution affects the ensemble machine learning based flood susceptibility prediction in data sparse region. **Journal of Environmental Management**, v. 297, p. 113344, 2021. b. DOI: 10.1016/j.jenvman.2021.113344.

SAHANA, Mehebab; PATEL, Priyank Pravin. A comparison of frequency ratio and fuzzy logic models for flood susceptibility assessment of the lower Kosi River Basin in India. **Environmental Earth Sciences**, v. 78, n. 10, p. 289, 2019. DOI: 10.1007/s12665-019-8285-1.

SAIKH, Nur Islam; MONDAL, Prolay. GIS-based machine learning algorithm for flood susceptibility analysis in the Pagla river basin, Eastern India. **Natural Hazards Research**, v. 3, n. 3, p. 420–436, 2023. DOI: 10.1016/j.nhres.2023.05.004.

SAINI, Satish Kumar; MAHATO, Susanta; PANDEY, Deep Narayan; JOSHI, Pawan Kumar. Modeling flood susceptibility zones using hybrid machine learning models of an agricultural dominant landscape of India. **Environmental Science and Pollution Research International**, v. 30, n. 43, p. 97463–97485, 2023. DOI: 10.1007/s11356-023-29049-9.

SALEH, Azlan; YUZIR, Ali; SABTU, Nuridah; ABUJAYYAB, Sohaib K. M.; BUNMI, Mudashiru Rofiat; PHAM, Quoc Bao. Flash flood susceptibility mapping in urban area using genetic algorithm and ensemble method. **Geocarto International**, Oxford, v. 37, n. 25, p. 10199–10228, 2022. DOI: 10.1080/10106049.2022.2032394.

SALVATI, Aryan; NIA, Alireza Moghaddam; SALAJEGHEH, Ali; GHADERI, Kayvan; ASL, Dawood Talebpour; AL-ANSARI, Nadhir; SOLAIMANI, Feridon; CLAGUE, John J. Flood susceptibility mapping using support vector regression and hyper-parameter optimization. **Journal of Flood Risk Management**, v. 16, n. 4, p. e12920, 2023. DOI: 10.1111/jfr3.12920.

SAMANTA, Ratan Kumar; BHUNIA, Gouri Sankar; SHIT, Pravat Kumar; POURGHASEMI, Hamid Reza. Flood susceptibility mapping using geospatial frequency ratio technique: a case study of Subarnarekha River Basin, India. **Modeling Earth Systems and Environment**, v. 4, n. 1, p. 395–408, 2018. DOI: 10.1007/s40808-018-0427-z.

SAMUEL, A. L. Some Studies in Machine Learning Using the Game of Checkers. **IBM Journal of Research and Development**, v. 3, n. 3, p. 210–229, 1959. DOI: 10.1147/rd.33.0210.

SARAVANAN, Subbarayan; ABIJITH, Devanantham. Flood susceptibility mapping of Northeast coastal districts of Tamil Nadu India using Multi-source Geospatial data and Machine Learning techniques. **Geocarto International**, v. 37, n. 27, p. 15252–15281, 2022. DOI: 10.1080/10106049.2022.2096702.

SARAVANAN, Subbarayan; ABIJITH, Devanantham; REDDY, Nagireddy Masthan; KSS, Parthasarathy; JANARDHANAM, Niraimathi; SATHIYAMURTHI, Subbarayan; SIVAKUMAR, Vivek. Flood susceptibility mapping using machine learning boosting algorithms techniques in Idukki district of Kerala India. **Urban Climate**, v. 49, p. 101503, 2023. DOI: 10.1016/j.uclim.2023.101503.

SARKAR, Showmitra Kumar; ANSAR, Saifullah Bin; EKRAM, Khondaker Mohammed

Mohiuddin; KHAN, Mehedi Hasan; TALUKDAR, Swapan; NAIKOO, Mohd Waseem; ISLAM, Abu Reza Towfiqul; RAHMAN, Atiqur; MOSAVI, Amir. Developing Robust Flood Susceptibility Model with Small Numbers of Parameters in Highly Fertile Regions of Northwest Bangladesh for Sustainable Flood and Agriculture Management. **Sustainability**, v. 14, n. 7, p. 3982, 2022. DOI: 10.3390/su14073982.

SEYDI, Seyd Teymoor; KANANI-SADAT, Yousef; HASANLOU, Mahdi; SAHRAEI, Roya; CHANUSSOT, Jocelyn; AMANI, Meisam. Comparison of Machine Learning Algorithms for Flood Susceptibility Mapping. **Remote Sensing**, v. 15, n. 1, p. 192, 2023. DOI: 10.3390/rs15010192.

SHAFIZADEH-MOGHADAM, Hossein; VALAVI, Roozbeh; SHAHABI, Himan; CHAPI, Kamran; SHIRZADI, Ataollah. Novel forecasting approaches using combination of machine learning and statistical models for flood susceptibility mapping. **Journal of Environmental Management**, v. 217, p. 1–11, 2018. DOI: 10.1016/j.jenvman.2018.03.089.

SHAHABI, Himan et al. Flood Detection and Susceptibility Mapping Using Sentinel-1 Remote Sensing Data and a Machine Learning Approach: Hybrid Intelligence of Bagging Ensemble Based on K-Nearest Neighbor Classifier. **Remote Sensing**, v. 12, n. 2, p. 266, 2020. DOI: 10.3390/rs12020266.

SHAHABI, Himan; SHIRZADI, Ataollah; RONOUD, Somayeh; ASADI, Shahrokh; PHAM, Binh Thai; MANSOURIPOUR, Fatemeh; GEERTSEMA, Marten; CLAGUE, John J.; BUI, Dieu Tien. Flash flood susceptibility mapping using a novel deep learning model based on deep belief network, back propagation and genetic algorithm. **Geoscience Frontiers**, v. 12, n. 3, p. 101100, 2021. DOI: 10.1016/j.gsf.2020.10.007.

SHAHIRI TABARESTANI, Ehsan; HADIAN, Sanaz; PHAM, Quoc Bao; ALI, Sk Ajim; PHUNG, Dung Tri. Flood potential mapping by integrating the bivariate statistics, multi-criteria decision-making, and machine learning techniques. **Stochastic Environmental Research and Risk Assessment**, v. 37, n. 4, p. 1415–1430, 2023. DOI: 10.1007/s00477-022-02342-8.

SHRESTHA, Sujan; DAHAL, Dewasis; POUDEL, Bishal; BANJARA, Mandip; KALRA, Ajay. Flood Susceptibility Analysis with Integrated Geographic Information System and Analytical Hierarchy Process: A Multi-Criteria Framework for Risk Assessment and Mitigation. **Water**, v. 17, n. 7, p. 937, 2025. DOI: 10.3390/w17070937.

SIAM, Zakaria Shams; HASAN, Rubyat Tasnuva; RAHMAN, Rashedur M. Effects of label noise on regression performances and model complexities for hybridized machine learning based spatial flood susceptibility modelling. **Cybernetics and Systems**, Oxford, v. 53, n. 4, p. 362–379, 2022. DOI: 10.1080/01969722.2021.1988446.

SILVEIRA, Wivian Nereida; KOBİYAMA, Masato; GOERL, Roberto Fabris; BRANDENBURG, Brigitte. **História das inundações em Joinville, 1851-2008**. Organic Trading, 2009.

SINGH, Ravindra Kumar; SONI, Ashish; KUMAR, Satish; PASUPULETI, Srinivas; GOVIND, Vasanta. Zonation of flood prone areas by an integrated framework of a hydrodynamic model and ANN. **Water Supply**, v. 21, n. 1, p. 80–97, 2020. DOI:

10.2166/ws.2020.252.

SINGH, Vijay P. (ORG.). **Hydrology of Disasters**. Dordrecht: Springer Netherlands, 1996. v. 24 DOI: 10.1007/978-94-015-8680-1. Disponível em: <http://link.springer.com/10.1007/978-94-015-8680-1>. Acesso em: 23 abr. 2023.

SINGHA, Chiranjit; SWAIN, Kishore Chandra; MELIHO, Modeste; ABDO, Hazem Ghassan; ALMOHAMAD, Hussein; AL-MUTIRY, Motirh. Spatial Analysis of Flood Hazard Zoning Map Using Novel Hybrid Machine Learning Technique in Assam, India. **Remote Sensing**, v. 14, n. 24, p. 6229, 2022. DOI: 10.3390/rs14246229.

SMAGALLA, Jessica; BECKER, Isabela Orsi; ZIMMERMANN, Heloisa Cristina; MONTEIRO, Leonardo Romero; VANELLI, Franciele Maria. História recente das inundações em Joinville: 2009-2019. In: ENCONTRO NACIONAL DE DESASTRES, 2., 2020, *Online*. **Anais** [...]. Porto Alegre: ABRHidro, 2020. Disponível em: <https://anais.abrhidro.org.br/job.php?Job=7297>. Acesso em: 07 fev. 2025

SMITH, Keith; WARD, Roy. **Floods: Physical Processes and Human Impacts**. Chichester: John Wiley & Sons, 1998.

SNOEK, Jasper; LAROCHELLE, Hugo; ADAMS, Ryan P. **Practical Bayesian Optimization of Machine Learning Algorithms**. arXiv, 2012. DOI: 10.48550/arXiv.1206.2944. Disponível em: <http://arxiv.org/abs/1206.2944>. Acesso em: 3 mar. 2025.

SOKOLOVA, Marina; LAPALME, Guy. A systematic analysis of performance measures for classification tasks. **Information Processing & Management**, v. 45, n. 4, p. 427–437, 2009. DOI: 10.1016/j.ipm.2009.03.002.

SOLTANI, Keyvan; EBTEHAJ, Isa; AMIRI, Afshin; AZARI, Arash; GHARABAGHI, Bahram; BONAKDARI, Hossein. Mapping the spatial and temporal variability of flood susceptibility using remotely sensed normalized difference vegetation index and the forecasted changes in the future. **Science of The Total Environment**, v. 770, p. 145288, 2021. DOI: 10.1016/j.scitotenv.2021.145288.

SØRENSEN, R.; ZINKO, U.; SEIBERT, J. On the calculation of the topographic wetness index: evaluation of different methods based on field observations. **Hydrology and Earth System Sciences**, v. 10, n. 1, p. 101–112, 2006. DOI: 10.5194/hess-10-101-2006.

STONE, M. Cross-Validatory Choice and Assessment of Statistical Predictions. **Journal of the Royal Statistical Society: Series B (Methodological)**, v. 36, n. 2, p. 111–133, 1974. DOI: 10.1111/j.2517-6161.1974.tb00994.x.

SUPPAWIMUT, Worawit. GIS-Based Flood Susceptibility Mapping Using Statistical Index and Weighting Factor Models: 10.32526/enrj/19/2021003. **Environment and Natural Resources Journal**, v. 19, n. 6, p. 481–493, 2021.

SWAIN, Kishore Chandra; SINGHA, Chiranjit; NAYAK, Laxmikanta. Flood Susceptibility Mapping through the GIS-AHP Technique Using the Cloud. **ISPRS International Journal of Geo-Information**, v. 9, n. 12, p. 720, 2020. DOI: 10.3390/ijgi9120720.

SWETS, John A. Measuring the Accuracy of Diagnostic Systems. **Science**, v. 240, n. 4857, p. 1285–1293, 1988. DOI: 10.1126/science.3287615.

TALUKDAR, Swapan; GHOSE, Bonosri; SHAHFAHAD; SALAM, Roquia; MAHATO, Susanta; PHAM, Quoc Bao; LINH, Nguyen Thi Thuy; COSTACHE, Romulus; AVAND, Mohammadtaghi. Flood susceptibility modeling in Teesta River basin, Bangladesh using novel ensembles of bagging algorithms. **Stochastic Environmental Research and Risk Assessment**, v. 34, n. 12, p. 2277–2300, 2020. DOI: 10.1007/s00477-020-01862-5.

TAMIRU, Habtamu; WAGARI, Meseret. Machine-learning and HEC-RAS integrated models for flood inundation mapping in Baro River Basin, Ethiopia. **Modeling Earth Systems and Environment**, v. 8, n. 2, p. 2291–2303, 2022. DOI: 10.1007/s40808-021-01175-8.

TANDON, Ayesha. **Analysis: The lack of diversity in climate-science research**. **Carbon Brief**, 2021. Disponível em: <https://www.carbonbrief.org/analysis-the-lack-of-diversity-in-climate-science-research/>. Acesso em: 6 mai. 2025.

TANG, Xianzhe; LI, Jiufeng; LIU, Minnan; LIU, Wei; HONG, Haoyuan. Flood susceptibility assessment based on a novel random Naïve Bayes method: A comparison between different factor discretization methods. **CATENA**, v. 190, p. 104536, 2020. DOI: 10.1016/j.catena.2020.104536.

TAPPERT, Charles C. Who Is the Father of Deep Learning? *In*: INTERNATIONAL CONFERENCE ON COMPUTATIONAL SCIENCE AND COMPUTATIONAL INTELLIGENCE, 2019, Las Vegas. **Anais [...]**. New York: IEEE, 2019. p. 343–348. DOI: 10.1109/CSCI49370.2019.00067.

TAY, Yi; DEHGHANI, Mostafa; BAHRI, Dara; METZLER, Donald. **Efficient Transformers: A Survey**. 2022. DOI: 10.48550/arXiv.2009.06732. Disponível em: <http://arxiv.org/abs/2009.06732>. Acesso em: 01 mar. 2025.

TEHRANY, Mahyat Shafapour; JONES, Simon; SHABANI, Farzin. Identifying the essential flood conditioning factors for flood prone area mapping using machine learning techniques. **CATENA**, v. 175, p. 174–192, 2019. DOI: 10.1016/j.catena.2018.12.011.

TEHRANY, Mahyat Shafapour; KUMAR, Lalit. The application of a Dempster–Shafer-based evidential belief function in flood susceptibility mapping and comparison with frequency ratio and logistic regression methods. **Environmental Earth Sciences**, v. 77, n. 13, p. 490, 2018. DOI: 10.1007/s12665-018-7667-0.

TEHRANY, Mahyat Shafapour; KUMAR, Lalit; SHABANI, Farzin. A novel GIS-based ensemble technique for flood susceptibility mapping using evidential belief function and support vector machine: Brisbane, Australia. **PeerJ**, v. 7, p. e7653, 2019. DOI: 10.7717/peerj.7653.

TEHRANY, Mahyat Shafapour; LEE, Moun-Jin; PRADHAN, Biswajeet; JEBUR, Mustafa Neamah; LEE, Saro. Flood susceptibility mapping using integrated bivariate and multivariate statistical models. **Environmental Earth Sciences**, v. 72, n. 10, p. 4001–4015, 2014. DOI: 10.1007/s12665-014-3289-3.

TEHRANY, Mahyat Shafapour; PRADHAN, Biswajeet; JEBUR, Mustafa Neamah. Spatial prediction of flood susceptible areas using rule based decision tree (DT) and a novel ensemble bivariate and multivariate statistical models in GIS. **Journal of Hydrology**, v. 504, p. 69–79, 2013. DOI: 10.1016/j.jhydrol.2013.09.034.

TEHRANY, Mahyat Shafapour; PRADHAN, Biswajeet; JEBUR, Mustafa Neamah. Flood susceptibility mapping using a novel ensemble weights-of-evidence and support vector machine models in GIS. **Journal of Hydrology**, v. 512, p. 332–343, 2014. DOI: 10.1016/j.jhydrol.2014.03.008.

TEHRANY, Mahyat Shafapour; PRADHAN, Biswajeet; JEBUR, Mustafa Neamah. Flood susceptibility analysis and its verification using a novel ensemble support vector machine and frequency ratio method. **Stochastic Environmental Research and Risk Assessment**, v. 29, n. 4, p. 1149–1165, 2015. DOI: 10.1007/s00477-015-1021-9.

TEHRANY, Mahyat Shafapour; PRADHAN, Biswajeet; MANSOR, Shattri; AHMAD, Noordin. Flood susceptibility assessment using GIS-based support vector machine model with different kernel types. **CATENA**, v. 125, p. 91–101, 2015. DOI: 10.1016/j.catena.2014.10.017.

TENNEKES, M. **tmap: Thematic Maps**. R package version 2.3-2. 2018. Disponível em: <https://CRAN.R-project.org/package=tmap>. Acesso em: 01 mar. 2025.

TERMEH, Seyed Vahid Razavi; KORNEJADY, Aiding; POURGHASEMI, Hamid Reza; KEESSTRA, Saskia. Flood susceptibility mapping using novel ensembles of adaptive neuro fuzzy inference system and metaheuristic algorithms. **Science of The Total Environment**, v. 615, p. 438–451, 2018. DOI: 10.1016/j.scitotenv.2017.09.262.

THI THUY LINH, Nguyen; PANDEY, Manish; JANIZADEH, Saeid; SANKAR BHUNIA, Gouri; NOROUZI, Akbar; ALI, Shoaib; BAO PHAM, Quoc; TRAN ANH, Duong; AHMADI, Kourosh. Flood susceptibility modeling based on new hybrid intelligence model: Optimization of XGboost model using GA metaheuristic algorithm. **Advances in Space Research**, v. 69, n. 9, p. 3301–3318, 2022. DOI: 10.1016/j.asr.2022.02.027.

THI, Huong et al. Prediction of Flash Flood Susceptibility of Hilly Terrain Using Deep Neural Network: A Case Study of Vietnam. **Computer Modeling in Engineering & Sciences**, v. 135, n. 3, p. 2219–2241, 2022. DOI: 10.32604/cmes.2023.022566.

THRUN, Sebastian B. **Efficient exploration in reinforcement learning**. Pittsburgh: Carnegie-Mellon University, 1992. (Relatório técnico, n. CMU-CS-92-102). Disponível em: https://www.ri.cmu.edu/pub_files/pub1/thrun_sebastian_1992_1/thrun_sebastian_1992_1.pdf. Acesso em: 06 mai. 2025.

TIEN BUI, Dieu et al. New Hybrids of ANFIS with Several Optimization Algorithms for Flood Susceptibility Modeling. **Water**, v. 10, n. 9, p. 1210, 2018. DOI: 10.3390/w10091210.

TIEN BUI, Dieu; HOANG, Nhat-Duc; MARTÍNEZ-ÁLVAREZ, Francisco; NGO, Phuong-Thao Thi; HOA, Pham Viet; PHAM, Tien Dat; SAMUI, Pijush; COSTACHE, Romulus. A novel deep learning neural network approach for predicting flash flood susceptibility: A case study at a high frequency tropical storm area. **Science of The Total Environment**, v. 701, p. 134413, 2020. DOI: 10.1016/j.scitotenv.2019.134413.

TIEN BUI, Dieu; HOANG, Nhat-Duc; PHAM, Tien-Dat; NGO, Phuong-Thao Thi; HOA, Pham Viet; MINH, Nguyen Quang; TRAN, Xuan-Truong; SAMUI, Pijush. A new intelligence approach based on GIS-based Multivariate Adaptive Regression Splines and metaheuristic optimization for predicting flash flood susceptible areas at high-frequency tropical typhoon area. **Journal of Hydrology**, v. 575, p. 314–326, 2019. DOI: 10.1016/j.jhydrol.2019.05.046.

TOWFIQUL ISLAM, Abu Reza Md; TALUKDAR, Swapan; MAHATO, Susanta; KUNDU, Sonali; EIBEK, Kutub Uddin; PHAM, Quoc Bao; KURIQI, Alban; LINH, Nguyen Thi Thuy. Flood susceptibility modelling using advanced ensemble machine learning models. **Geoscience Frontiers**, v. 12, n. 3, p. 101075, 2021. DOI: 10.1016/j.gsf.2020.09.006.

TUROĞLU, Hüseyin; DÖLEK, İskender. Floods and their likely impacts on ecological environment in the Bolaman River Basin (Ordu, Turkey). **Research Journal of Agricultural Science**, p. 7, 2011. DOI: 0.5555/20123209221.

UNITED STATES. Geological Survey. **Flood definitions**. Reston, [s.d.]. Disponível em: <https://www.usgs.gov/media/files/flood-definitions>. Acesso em: 14 maio 2024

UNITED STATES. Department of Agriculture. Soil Conservation Service (SCS). **Urban hydrology for small watersheds**. Washington, D.C.: U.S. Department of Agriculture, 1986.

VAFAKHAH, Mehdi; MOHAMMAD HASANI LOOR, Sajad; POURGHASEMI, Hamidreza; KATEBIKORD, Azadeh. Comparing performance of random forest and adaptive neuro-fuzzy inference system data mining models for flood susceptibility mapping. **Arabian Journal of Geosciences**, v. 13, n. 11, p. 417, 2020. DOI: 10.1007/s12517-020-05363-1.

VAJEETHAVEESIN, Thanat; PANBOONYUEN, Teerapong; LAWAWIRONJWONG, Siam; SRESTASATHIERN, Panu; JAIYEN, Saichon; JITKAJORNWANICH, Kulsawasd. A Performance Comparison between GIS-based and Neuron Network Methods for Flood Susceptibility Assessment in Ayutthaya Province. **Trends in Sciences**, v. 19, n. 2, p. 2038–2038, 2022. DOI: 10.48048/tis.2022.2038.

VAN WESTEN, C. J. **Application of geographic information systems to landslide hazard zonation**. 1993. 251 p. Tese (Doutorado) – Delft University of Technology, Delft, 1993. Disponível em: <https://research.utwente.nl/en/publications/application-of-geographic-information-systems-to-landslide-hazard>. Acesso em: 25 abr. 2023.

VARNES, David Joseph; SLOPES, International Association of Engineering Geology Commission on Landslides and Other Mass Movements On. **Landslide Hazard Zonation: A Review of Principles and Practice**. Unesco, 1984.

VASWANI, Ashish; SHAZEER, Noam; PARMAR, Niki; USZKOREIT, Jakob; JONES, Llion; GOMEZ, Aidan N.; KAISER, Lukasz; POLOSUKHIN, Illia. **Attention Is All You Need**. arXiv, 2023. DOI: 10.48550/arXiv.1706.03762. Disponível em: <http://arxiv.org/abs/1706.03762>. Acesso em: 02 mar. 2025.

VASWANI, Ashish; SHAZEER, Noam; PARMAR, Niki; USZKOREIT, Jakob; JONES, Llion; GOMEZ, Aidan N.; KAISER, Łukasz; POLOSUKHIN, Illia. Attention is all you need. In: **ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS**, 31., 2017, Long Beach. **Anais [...]**. Red Hook: Curran Associates, Inc., 2017. p. 5998–6008.

VIANA, Cláudia M.; ABRANTES, Patrícia; ROCHA, Jorge; VIANA, Cláudia M.; ABRANTES, Patrícia; ROCHA, Jorge. Introductory Chapter. *Em: Geographic Information Systems and Science*. IntechOpen, 2019. DOI: 10.5772/intechopen.86121.

VINCENT, Amala Mary; K.S.S., Parthasarathy; JIDESH, P. Flood susceptibility mapping using AutoML and a deep learning framework with evolutionary algorithms for hyperparameter optimization. **Applied Soft Computing**, v. 148, p. 110846, 2023. DOI: 10.1016/j.asoc.2023.110846.

VIRGÜEZ, Edgar; LEON, Lina; FREESE, Lyssa M. The climate sciences need representation from the Global South. **One Earth**, v. 7, n. 3, p. 370–373, 2024. DOI: 10.1016/j.oneear.2024.01.016.

VOJTEK, Matej; VOJTEKOVÁ, Jana; COSTACHE, Romulus; PHAM, Quoc Bao; LEE, Sunmin; ARSHAD, Arfan; SAHOO, Satiprasad; LINH, Nguyen Thi Thuy; ANH, Duong Tran. Comparison of multi-criteria-analytical hierarchy process and machine learning-boosted tree models for regional flood susceptibility mapping: a case study from Slovakia. **Geomatics, Natural Hazards and Risk**, v. 12, n. 1, p. 1153–1180, 2021. DOI: 10.1080/19475705.2021.1912835.

VU, Van Tich; NGUYEN, Huu Duy; VU, Phuong Lan; HA, Minh Cuong; BUI, Van Dong; NGUYEN, Thi Oanh; HOANG, Van Hiep; NGUYEN, Thanh Kim Hue. Predicting land use effects on flood susceptibility using machine learning and remote sensing in coastal Vietnam. **Water Practice and Technology**, v. 18, n. 6, p. 1543–1555, 2023. DOI: 10.2166/wpt.2023.088.

WACHAL, David J.; HUDAK, Paul F. Mapping landslide susceptibility in Travis County, Texas, USA. **GeoJournal**, v. 51, n. 3, p. 245–253, 2000. DOI: 10.1023/A:1017524604463.

WAHBA, Mohamed; EL-RAWY, Mustafa; AL-ARIFI, Nassir; MANSOUR, Mahmoud M. A Novel Estimation of the Composite Hazard of Landslides and Flash Floods Utilizing an Artificial Intelligence Approach. **Water**, v. 15, n. 23, p. 4138, 2023. DOI: 10.3390/w15234138.

WANG, Changpeng; LIN, Yangchun; TAO, Zhiwen; ZHAN, Jiayin; LI, Wenkai; HUANG, Huabing. An Inverse-Occurrence Sampling Approach for Urban Flood Susceptibility Mapping. **Remote Sensing**, v. 15, n. 22, p. 5384, 2023. DOI: 10.3390/rs15225384.

WANG, Yi et al. Flood susceptibility mapping in Dingnan County (China) using adaptive neuro-fuzzy inference system with biogeography based optimization and imperialistic competitive algorithm. **Journal of Environmental Management**, v. 247, p. 712–729, 2019. DOI: 10.1016/j.jenvman.2019.06.102.

WANG, Yi; FANG, Zhice; HONG, Haoyuan; COSTACHE, Romulus; TANG, Xianzhe. Flood susceptibility mapping by integrating frequency ratio and index of entropy with multilayer perceptron and classification and regression tree. **Journal of Environmental Management**, v. 289, p. 112449, 2021. DOI: 10.1016/j.jenvman.2021.112449.

WANG, Yi; FANG, Zhice; HONG, Haoyuan; PENG, Ling. Flood susceptibility mapping using convolutional neural network frameworks. **Journal of Hydrology**, v. 582, p. 124482, 2020. DOI: 10.1016/j.jhydrol.2019.124482.

WANG, Ze; LYU, Heng; ZHANG, Chi. Pluvial flood susceptibility mapping for data-scarce urban areas using graph attention network and basic flood conditioning factors. **Geocarto International**, v. 38, n. 1, p. 2275692, 2023. DOI: 10.1080/10106049.2023.2275692.

WARREN, Craig. MATLAB for Engineers: Development of an Online, Interactive, Self-study Course. **Engineering Education**, v. 9, n. 1, p. 86–93, 2014. DOI: 10.11120/ened.2014.00026.

WATKINS, Christopher J. C. H.; DAYAN, Peter. Q-learning. **Machine Learning**, v. 8, n. 3, p. 279–292, 1992. DOI: 10.1007/BF00992698.

WEI, T.; SIMKO, V. **corrplot: Visualization of a Correlation Matrix**. R package version 0.90. 2021. Disponível em: <https://CRAN.R-project.org/package=corrplot>. Acesso em: 01 jul. 2025.

WEISS, A. D. **Topographic position and landforms analysis**. In: ESRI INTERNATIONAL USER CONFERENCE, 21., 2001, San Diego. **Proceedings...** San Diego, CA: ESRI, 2001.

WICKHAM, H. et al. Welcome to the tidyverse. **Journal of Open Source Software**, v. 4, n. 43, p.1686, 2019. Disponível em: <https://joss.theoj.org/papers/10.21105/joss.01686>. Acesso em: 01 jul. 2025.

WICKHAM, H. **ggplot2: Elegant Graphics for Data Analysis**. New York: Springer-Verlag, 2016.

WICKHAM, H.; BRYAN, J. **readxl: Read Excel Files**. R package version 1.3.1. 2019. Disponível em: <https://CRAN.R-project.org/package=readxl>. Acesso em: 01 jul. 2025.

WILKE, C. O. **cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'**. R package version 1.1.1. 2020. Disponível em: <https://CRAN.R-project.org/package=cowplot>. Acesso em: 01 jul. 2025.

WILLIAMS, Ronald J. Simple statistical gradient-following algorithms for connectionist reinforcement learning. **Machine Learning**, v. 8, n. 3, p. 229–256, 1992. DOI: 10.1007/BF00992696.

WILSON, John P.; GALLANT, John C. **Terrain Analysis: Principles and Applications**. [s.l.]: John Wiley & Sons, 2000.

WU, Yingying; SHE, Dunxian; XIA, Jun; SONG, Jiyun; XIAO, Tong; ZHOU, Yan. The quantitative assessment of impact of pumping capacity and LID on urban flood susceptibility based on machine learning. **Journal of Hydrology**, v. 617, p. 129116, 2023. DOI: 10.1016/j.jhydrol.2023.129116.

YARIYAN, Peyman; AVAND, Mohammadtaghi; ABBASPOUR, Rahim Ali; TORABI HAGHIGHI, Ali; COSTACHE, Romulus; GHORBANZADEH, Omid; JANIZADEH, Saeid; BLASCHKE, Thomas. Flood susceptibility mapping using an improved analytic network process with statistical models. **Geomatics, Natural Hazards and Risk**, v. 11, n. 1, p. 2282–2314, 2020. DOI: 10.1080/19475705.2020.1836036.

YASEEN, Andaleeb; LU, Jianzhong; CHEN, Xiaoling. Flood susceptibility mapping in an arid

region of Pakistan through ensemble machine learning model. **Stochastic Environmental Research and Risk Assessment**, v. 36, n. 10, p. 3041–3061, 2022. DOI: 10.1007/s00477-022-02179-1.

YIN, Yongqiang; ZHANG, Xiaoxiang; GUAN, Zheng; CHEN, Yuehong; LIU, Changjun; YANG, Tao. Flash flood susceptibility mapping based on catchments using an improved Blending machine learning approach. **Hydrology Research**, v. 54, n. 4, p. 557–579, 2023. DOI: 10.2166/nh.2023.139.

YOUSEFI, Saleh; AVAND, Mohammadtaghi; YARIYAN, Peyman; POURGHASEMI, Hamid Reza; KEESSTRA, Saskia; TAVANGAR, Shahla; TABIBIAN, Sahar. A novel GIS-based ensemble technique for rangeland downward trend mapping as an ecological indicator change. **Ecological Indicators**, v. 117, p. 106591, 2020. a. DOI: 10.1016/j.ecolind.2020.106591.

YOUSEFI, Saleh; POURGHASEMI, Hamid Reza; EMAMI, Sayed Naeim; RAHMATI, Omid; TAVANGAR, Shahla; POUYAN, Soheila; TIEFENBACHER, John P.; SHAMSODDINI, Shahbaz; NEKOEIMEHR, Mohammad. Assessing the susceptibility of schools to flood events in Iran. **Scientific Reports**, v. 10, n. 1, p. 18114, 2020. b. DOI: 10.1038/s41598-020-75291-3.

YOUSSEF, Ahmed M.; POURGHASEMI, Hamid Reza; EL-HADDAD, Bosy A. Advanced machine learning algorithms for flood susceptibility modeling — performance comparison: Red Sea, Egypt. **Environmental Science and Pollution Research**, v. 29, n. 44, p. 66768–66792, 2022. DOI: 10.1007/s11356-022-20213-1.

YOUSSEF, Ahmed M.; PRADHAN, Biswajeet; DIKSHIT, Abhirup; MAHDI, Ali M. Comparative study of convolutional neural network (CNN) and support vector machine (SVM) for flood susceptibility mapping: a case study at Ras Gharib, Red Sea, Egypt. **Geocarto International**, v. 37, n. 26, p. 11088–11115, 2022. DOI: 10.1080/10106049.2022.2046866.

YU, Han; LUO, Zengliang; WANG, Lunche; DING, Xiangyi; WANG, Shaoqiang. Improving the Accuracy of Flood Susceptibility Prediction by Combining Machine Learning Models and the Expanded Flood Inventory Data. **Remote Sensing**, v. 15, n. 14, p. 3601, 2023. DOI: 10.3390/rs15143601.

ZHANG, Yuzhen; LIU, Jingjing; SHEN, Wenjuan. A Review of Ensemble Learning Algorithms Used in Remote Sensing Applications. **Applied Sciences**, v. 12, n. 17, p. 8654, 2022. DOI: 10.3390/app12178654.

ZHAO, Gang; PANG, Bo; XU, Zongxue; CUI, Lizhuang; WANG, Jingjing; ZUO, Depeng; PENG, Dingzhi. Improving urban flood susceptibility mapping using transfer learning. **Journal of Hydrology**, v. 602, p. 126777, 2021. DOI: 10.1016/j.jhydrol.2021.126777.

ZHAO, Gang; PANG, Bo; XU, Zongxue; PENG, Dingzhi; XU, Liyang. Assessment of urban flood susceptibility using semi-supervised machine learning model. **Science of The Total Environment**, v. 659, p. 940–949, 2019. DOI: 10.1016/j.scitotenv.2018.12.217.

ZHAO, Gang; PANG, Bo; XU, Zongxue; PENG, Dingzhi; ZUO, Depeng. Urban flood susceptibility assessment based on convolutional neural networks. **Journal of Hydrology**, v. 590, p. 125235, 2020. DOI: 10.1016/j.jhydrol.2020.125235.

ZHAO, Gang; PANG, Bo; XU, Zongxue; YUE, Jiajia; TU, Tongbi. Mapping flood susceptibility in mountainous areas on a national scale in China. **Science of The Total Environment**, v. 615, p. 1133–1142, 2018. DOI: 10.1016/j.scitotenv.2017.10.037.

ZHOU, Zhi-Hua. **Ensemble Methods: Foundations and Algorithms**. New York: Chapman and Hall/CRC, 2012. DOI: 10.1201/b12207.

ZUPIC, Ivan; ČATER, Tomaž. Bibliometric Methods in Management and Organization. **Organizational Research Methods**, v. 18, n. 3, p. 429–472, 2015. DOI: 10.1177/1094428114562629.

ZUUR, Alain F.; IENO, Elena N.; ELPHICK, Chris S. A protocol for data exploration to avoid common statistical problems. **Methods in Ecology and Evolution**, v. 1, n. 1, p. 3–14, 2010. DOI: 10.1111/j.2041-210X.2009.00001.x.

APÊNDICE A – LISTA DE ARTIGOS ANALISADOS NA BIBLIOMETRIA

Este apêndice apresenta a relação completa dos artigos científicos incluídos na análise bibliométrica, os quais se dedicam à aplicação de técnicas de aprendizado de máquina no mapeamento da suscetibilidade à inundação. A compilação abrange um período de dez anos (2013-2023), oferecendo um panorama abrangente e atualizado da pesquisa na área. Os artigos estão listados em ordem cronológica crescente de publicação.

Quadro 5 - Lista de artigos analisados na bibliometria

(continua)

Id	Citação	Título
1	Lamovec <i>et al.</i> (2013)	<i>Detecting flooded areas with machine learning techniques: case study of the Selska Sora river flash flood in September 2007</i>
2	Tehrany <i>et al.</i> (2013)	<i>Spatial prediction of flood susceptible areas using rule based decision tree (DT) and a novel ensemble bivariate and multivariate statistical models in GIS</i>
3	Tehrany <i>et al.</i> (2014)	<i>Flood susceptibility mapping using integrated bivariate and multivariate statistical models</i>
4	Tehrany <i>et al.</i> (2014)	<i>Flood susceptibility mapping using a novel ensemble weights-of-evidence and support vector machine models in GIS</i>
5	Tehrany <i>et al.</i> (2015)	<i>Flood susceptibility assessment using GIS-based support vector machine model with different kernel types</i>
6	Tehrany <i>et al.</i> (2015)	<i>Flood susceptibility analysis and its verification using a novel ensemble support vector machine and frequency ratio method</i>
7	Sachdeva <i>et al.</i> (2017)	<i>Flood Susceptibility Mapping using GIS-based Support Vector Machine and Particle Swarm Optimization: A case study in Uttarakhand (India)</i>
8	Chapi <i>et al.</i> (2017)	<i>A novel hybrid artificial intelligence approach for flood susceptibility assessment</i>
9	Lee <i>et al.</i> (2017)	<i>Spatial prediction of flood susceptibility using random-forest and boosted-tree models in Seoul metropolitan city, Korea</i>
10	Sachdeva <i>et al.</i> (2017)	<i>Flood susceptibility mapping using GIS-based support vector machine and particle swarm optimization: A case study in Uttarakhand (India)</i>

Fonte: Elaborada pelo autor (2024).

Quadro 5 - Lista de artigos analisados na bibliometria

(continuação)

Id	Citação	Título
11	Rahmati e Pourghasemi (2017)	<i>Identification of Critical Flood Prone Areas in Data-Scarce and Ungauged Regions: A Comparison of Three Data Mining Models</i>
12	Khosravi <i>et al.</i> (2018)	<i>A comparative assessment of decision trees algorithms for flash flood susceptibility modeling at Haraz watershed, northern Iran</i>
13	Al-Abadi (2018)	<i>Mapping flood susceptibility in an arid region of southern Iraq using ensemble machine learning classifiers: a comparative study</i>
14	Shafizadeh-Moghadam <i>et al.</i> (2018)	<i>Novel forecasting approaches using combination of machine learning and statistical models for flood susceptibility mapping</i>
15	Bui <i>et al.</i> (2018)	<i>New Hybrids of ANFIS with Several Optimization Algorithms for Flood Susceptibility Modeling</i>
16	Samanta <i>et al.</i> (2018)	<i>Flood susceptibility mapping using geospatial frequency ratio technique: a case study of Subarnarekha River Basin, India</i>
17	Zhao <i>et al.</i> (2018)	<i>Mapping flood susceptibility in mountainous areas on a national scale in China</i>
18	Ngo <i>et al.</i> (2018)	<i>A Novel Hybrid Swarm Optimized Multilayer Neural Network for Spatial Prediction of Flash Floods in Tropical Areas Using Sentinel-1 SAR Imagery and Geospatial Data</i>
19	Bui <i>et al.</i> (2019)	<i>A novel hybrid approach based on a swarm intelligence optimized extreme learning machine for flash flood susceptibility mapping</i>
20	Costache (2019)	<i>Flood Susceptibility Assessment by Using Bivariate Statistics and Machine Learning Models - A Useful Tool for Flood Risk Management</i>
21	Chen <i>et al.</i> (2019)	<i>Flood susceptibility modelling using novel hybrid approach of reduced-error pruning trees with bagging and random subspace ensembles</i>
22	Rahman <i>et al.</i> (2019)	<i>Flood Susceptibility Assessment in Bangladesh Using Machine Learning and Multi-criteria Decision Analysis</i>

Fonte: Elaborada pelo autor (2024).

Quadro 5 - Lista de artigos analisados na bibliometria

(continuação)

Id	Citação	Título
23	Janizadeh <i>et al.</i> (2019)	<i>Prediction Success of Machine Learning Methods for Flash Flood Susceptibility Mapping in the Tafresh Watershed, Iran</i>
24	Zhao <i>et al.</i> (2019)	<i>Assessment of urban flood susceptibility using semi-supervised machine learning model</i>
25	Wang <i>et al.</i> (2019)	<i>Flood susceptibility mapping in Dingnan County (China) using adaptive neuro-fuzzy inference system with biogeography based optimization and imperialistic competitive algorithm</i>
26	Sahana e Patel (2019)	<i>A comparison of frequency ratio and fuzzy logic models for flood susceptibility assessment of the lower Kosi River Basin in India</i>
27	Choubin <i>et al.</i> (2019)	<i>An ensemble prediction of flood susceptibility using multivariate discriminant analysis, classification and regression trees, and support vector machines</i>
28	Tehrany <i>et al.</i> (2019)	<i>A novel GIS-based ensemble technique for flood susceptibility mapping using evidential belief function and support vector machine: Brisbane, Australia</i>
29	Khosravi <i>et al.</i> (2019)	<i>A comparative assessment of flood susceptibility modeling using Multi-Criteria Decision-Making Analysis and Machine Learning Methods</i>
30	Bui <i>et al.</i> (2019)	<i>A new intelligence approach based on GIS-based Multivariate Adaptive Regression Splines and metaheuristic optimization for predicting flash flood susceptible areas at high-frequency tropical typhoon area</i>
31	Costache e Bui (2019)	<i>Spatial prediction of flood potential using new ensembles of bivariate statistics and artificial intelligence: A case study at the Putna river catchment of Romania</i>
32	Tehrany <i>et al.</i> (2019)	<i>Identifying the essential flood conditioning factors for flood prone area mapping using machine learning techniques</i>
33	Yousefi <i>et al.</i> (2020)	<i>Assessing the susceptibility of schools to flood events in Iran</i>
34	Ali <i>et al.</i> (2020)	<i>GIS-based comparative assessment of flood susceptibility mapping using hybrid multi-criteria decision-making approach, naive Bayes tree, bivariate statistics and logistic regression: A case of Topla basin, Slovakia</i>
35	Pham <i>et al.</i> (2020)	<i>GIS Based Hybrid Computational Approaches for Flash Flood Susceptibility Assessment</i>

Fonte: Elaborada pelo autor (2024).

Quadro 5 - Lista de artigos analisados na bibliometria

(continuação)

Id	Citação	Título
36	Costache <i>et al.</i> (2020)	<i>Spatial predicting of flood potential areas using novel hybridizations of fuzzy decision-making, bivariate statistics, and machine learning</i>
37	Costache e Bui (2020)	<i>Identification of areas prone to flash-flood phenomena using multiple-criteria decision-making, bivariate statistics, machine learning and their ensembles</i>
38	Costache <i>et al.</i> (2020)	<i>New neural fuzzy-based machine learning ensemble for enhancing the prediction accuracy of flood susceptibility mapping</i>
39	Talukdar <i>et al.</i> (2020)	<i>Flood susceptibility modeling in Teesta River basin, Bangladesh using novel ensembles of bagging algorithms</i>
40	Pham <i>et al.</i> (2020)	<i>A Comparative Study of Kernel Logistic Regression, Radial Basis Function Classifier, Multinomial Naive Bayes, and Logistic Model Tree for Flash Flood Susceptibility Mapping</i>
41	Nachappa <i>et al.</i> (2020)	<i>Flood susceptibility mapping with machine learning, multi-criteria decision analysis and ensemble using Dempster Shafer Theory</i>
42	Shahabi <i>et al.</i> (2020)	<i>Flood Detection and Susceptibility Mapping Using Sentinel-1 Remote Sensing Data and a Machine Learning Approach: Hybrid Intelligence of Bagging Ensemble Based on K-Nearest Neighbor Classifier</i>
43	Costache <i>et al.</i> (2020)	<i>Flash-Flood Susceptibility Assessment Using Multi-Criteria Decision Making and Machine Learning Supported by Remote Sensing and GIS Techniques</i>
44	Elmahdy <i>et al.</i> (2020)	<i>Flash Flood Susceptibility Modeling and Magnitude Index Using Machine Learning and Geohydrological Models: A Modified Hybrid Approach</i>
45	Band <i>et al.</i> (2020)	<i>Flash Flood Susceptibility Modeling Using New Approaches of Hybrid and Ensemble Tree-Based Machine Learning Algorithms</i>
46	Dodangeh <i>et al.</i> (2020)	<i>Integrated machine learning methods with resampling algorithms for flood susceptibility prediction</i>
47	Zhao <i>et al.</i> (2020)	<i>Urban flood susceptibility assessment based on convolutional neural networks</i>
48	Yousefi <i>et al.</i> (2020)	<i>A novel GIS-based ensemble technique for rangeland downward trend mapping as an ecological indicator change</i>

Fonte: Elaborada pelo autor (2024).

Quadro 5 - Lista de artigos analisados na bibliometria

(continuação)

Id	Citação	Título
49	Costache <i>et al.</i> (2020)	<i>Novel hybrid models between bivariate statistics, artificial neural networks and boosting algorithms for flood susceptibility assessment</i>
50	Rahmati <i>et al.</i> (2020)	<i>Development of novel hybridized models for urban flood susceptibility mapping</i>
51	Wang <i>et al.</i> (2020)	<i>Flood susceptibility mapping using convolutional neural network frameworks</i>
52	Vafakhah <i>et al.</i> (2020)	<i>Comparing performance of random forest and adaptive neuro-fuzzy inference system data mining models for flood susceptibility mapping (vol 13, 417, 2020)</i>
53	Vafakhah <i>et al.</i> (2020)	<i>Comparing performance of random forest and adaptive neuro-fuzzy inference system data mining models for flood susceptibility mapping</i>
54	Chowdhuri <i>et al.</i> (2020)	<i>Flood susceptibility mapping by ensemble evidential belief function and binomial logistic regression model on river basin of eastern India</i>
55	Esfandiari <i>et al.</i> (2020)	<i>Flood Hazard Risk Mapping Using a Pseudo Supervised Random Forest</i>
56	Tang <i>et al.</i> (2020)	<i>Flood susceptibility assessment based on a novel random Naive Bayes method: A comparison between different factor discretization methods</i>
57	Arabameri <i>et al.</i> (2020)	<i>Flash flood susceptibility modelling using functional tree and hybrid ensemble techniques</i>
58	Yariyan <i>et al.</i> (2020)	<i>Flood susceptibility mapping using an improved analytic network process with statistical models</i>
59	Bui} <i>et al.</i> (2020)	<i>A novel deep learning neural network approach for predicting flash flood susceptibility: A case study at a high frequency tropical storm area</i>
60	Costache <i>et al.</i> (2020)	<i>Novel ensembles of deep learning neural network and statistical learning for flash-flood susceptibility mapping</i>
61	Costache <i>et al.</i> (2020)	<i>Flash-flood susceptibility assessment using multi-criteria decision making and machine learning supported by remote sensing and GIS techniques</i>
62	Hadipour <i>et al.</i> (2020)	<i>Coastal flooding risk assessment using a GIS-based spatial multi-criteria decision analysis approach</i>

Fonte: Elaborada pelo autor (2024).

Quadro 5 - Lista de artigos analisados na bibliometria

(continuação)

Id	Citação	Título
63	Tang <i>et al.</i> (2021)	<i>A novel index to evaluate discretization methods: A case study of flood susceptibility assessment based on random forest</i>
64	Pham <i>et al.</i> (2021)	<i>Improved flood susceptibility mapping using a best first decision tree integrated with ensemble learning techniques</i>
65	Rahman <i>et al.</i> (2021)	<i>Paper Flooding and its relationship with land cover change, population growth, and road density</i>
66	Wang <i>et al.</i> (2021)	<i>Flood susceptibility mapping by integrating frequency ratio and index of entropy with multilayer perceptron and classification and regression tree</i>
67	Mahato <i>et al.</i> (2021)	<i>Field based index of flood vulnerability (IFV): A new validation technique for flood susceptible models</i>
68	Arora <i>et al.</i> (2021)	<i>Optimization of state-of-the-art fuzzy-metaheuristic ANFIS-based machine learning models for flood susceptibility prediction mapping in the Middle Ganga Plain, India</i>
69	Pandey <i>et al.</i> (2021)	<i>Flood Susceptibility Modeling in a Subtropical Humid Low-Relief Alluvial Plain Environment: Application of Novel Ensemble Machine Learning Approach</i>
70	Luu <i>et al.</i> (2021)	<i>Flood-prone area mapping using machine learning techniques: a case study of Quang Binh province, Vietnam</i>
71	El-Haddad <i>et al.</i> (2021)	<i>Flood susceptibility prediction using four machine learning techniques and comparison of their performance at Wadi Qena Basin, Egypt</i>
72	Islam <i>et al.</i> (2021)	<i>Flood susceptibility modelling using advanced ensemble machine learning models</i>
73	Andaryani <i>et al.</i> (2021)	<i>Integration of hard and soft supervised machine learning for flood susceptibility mapping</i>
74	Vojtek <i>et al.</i> (2021)	<i>Comparison of multi-criteria-analytical hierarchy process and machine learning-boosted tree models for regional flood susceptibility mapping: a case study from Slovakia</i>
75	Ha <i>et al.</i> (2021)	<i>Flash flood susceptibility prediction mapping for a road network using hybrid machine learning models</i>
76	Pham <i>et al.</i> (2021)	<i>Can deep learning algorithms outperform benchmark machine learning algorithms in flood susceptibility modeling?</i>

Fonte: Elaborada pelo autor (2024).

Quadro 5 - Lista de artigos analisados na bibliometria

(continuação)

Id	Citação	Título
77	Norallahi e Kaboli (2021)	<i>Urban flood hazard mapping using machine learning models: GARP, RF, MaxEnt and NB</i>
78	Liu <i>et al.</i> (2021)	<i>Hybrid Models Incorporating Bivariate Statistics and Machine Learning Methods for Flash Flood Susceptibility Assessment Based on Remote Sensing Datasets</i>
79	Saha <i>et al.</i> (2021)	<i>How far spatial resolution affects the ensemble machine learning based flood susceptibility prediction in data sparse region</i>
80	Costache <i>et al.</i> (2021)	<i>Detection of areas prone to flood risk using state-of-the-art machine learning models</i>
81	Madhuri <i>et al.</i> (2021)	<i>Application of machine learning algorithms for flood susceptibility assessment and risk management</i>
82	Ahmadlou <i>et al.</i> (2021)	<i>Flood susceptibility mapping and assessment using a novel deep learning model combining multilayer perceptron and autoencoder neural networks</i>
83	Fang <i>et al.</i> (2021)	<i>Predicting flood susceptibility using LSTM neural networks</i>
84	Costache <i>et al.</i> (2021)	<i>Flash-Flood Potential Mapping Using Deep Learning, Alternating Decision Trees and Data Provided by Remote Sensing Sensors</i>
85	Darabi <i>et al.</i> (2021)	<i>A hybridized model based on neural network and swarm intelligence-grey wolf algorithm for spatial prediction of urban flood-inundation</i>
86	Soltani <i>et al.</i> (2021)	<i>Mapping the spatial and temporal variability of flood susceptibility using remotely sensed normalized difference vegetation index and the forecasted changes in the future</i>
87	Singh <i>et al.</i> (2021)	<i>Zonation of flood prone areas by an integrated framework of a hydrodynamic model and ANN</i>
88	Saha <i>et al.</i> (2021)	<i>Flood Susceptibility Assessment Using Novel Ensemble of Hyperpipes and Support Vector Regression Algorithms</i>
89	Shahabi <i>et al.</i> (2021)	<i>Flash flood susceptibility mapping using a novel deep learning model based on deep belief network, back propagation and genetic algorithm</i>

Fonte: Elaborada pelo autor (2024).

Quadro 5 - Lista de artigos analisados na bibliometria

(continuação)

Id	Citação	Título
90	Ngo <i>et al.</i> (2021)	<i>A novel hybrid quantum-PSO and credal decision tree ensemble for tropical cyclone induced flash flood susceptibility mapping with geospatial data</i>
91	Pham <i>et al.</i> (2021)	<i>Flood risk assessment using hybrid artificial intelligence models integrated with multi-criteria decision analysis in Quang Nam Province, Vietnam</i>
92	Zhao <i>et al.</i> (2021)	<i>Improving urban flood susceptibility mapping using transfer learning</i>
93	Luu <i>et al.</i> (2021)	<i>GIS-based ensemble computational models for flood susceptibility prediction in the Quang Binh Province, Vietnam</i>
94	Abedi <i>et al.</i> (2022)	<i>Flash-flood susceptibility mapping based on XGBoost, random forest and boosted regression trees</i>
95	Fang <i>et al.</i> (2022)	<i>Hybrid approach for flood susceptibility assessment in a flood-prone mountainous catchment in China</i>
96	Kaiser <i>et al.</i> (2022)	<i>Regional-scale prediction of pluvial and flash flood susceptible areas using tree-based classifiers</i>
97	Bhuiyan <i>et al.</i> (2022)	<i>Harmonizing and Extending Fragmented 100 Year Flood Hazard Maps in Canada's Capital Region Using Random Forest Classification</i>
98	Arabameri <i>et al.</i> (2022)	<i>Flood susceptibility mapping using meta-heuristic algorithms</i>
99	Mosavi <i>et al.</i> (2022)	<i>Ensemble models of GLM, FDA, MARS, and RF for flood and erosion susceptibility mapping: a priority assessment of sub-basins</i>
100	Darabi <i>et al.</i> (2022)	<i>Development of a novel hybrid multi-boosting neural network model for spatial prediction of urban flood</i>
101	Ekmekcioglu <i>et al.</i> (2022)	<i>Exploring the additional value of class imbalance distributions on interpretable flash flood susceptibility prediction in the Black Warrior River basin, Alabama, United States</i>
102	Al-Aizari <i>et al.</i> (2022)	<i>Assessment Analysis of Flood Susceptibility in Tropical Desert Area: A Case Study of Yemen</i>
103	Sarkar <i>et al.</i> (2022)	<i>Developing Robust Flood Susceptibility Model with Small Numbers of Parameters in Highly Fertile Regions of Northwest Bangladesh for Sustainable Flood and Agriculture Management</i>

Fonte: Elaborada pelo autor (2024).

Quadro 5 - Lista de artigos analisados na bibliometria

(continuação)

Id	Citação	Título
104	Ha <i>et al.</i> (2022)	<i>Machine Learning and Remote Sensing Application for Extreme Climate Evaluation: Example of Flood Susceptibility in the Hue Province, Central Vietnam Region</i>
105	Chakraborty <i>et al.</i> (2022)	<i>Flash-flood hazard susceptibility mapping in Kangsabati River Basin, India</i>
106	Sachdeva e Kumar (2022)	<i>Flood susceptibility mapping using extremely randomized trees for Assam 2020 floods</i>
107	Nguyen <i>et al.</i> (2022)	<i>A novel hybrid approach to flood susceptibility assessment based on machine learning and land use change. Case study: a river watershed in Vietnam</i>
108	Bera <i>et al.</i> (2022)	<i>Evaluation of machine learning, information theory and multi-criteria decision analysis methods for flood susceptibility mapping under varying spatial scale of analyses</i>
109	Pal e Singha (2022)	<i>Analyzing sensitivity of flood susceptible model in a flood plain river basin</i>
110	Yaseen <i>et al.</i> (2022)	<i>Flood susceptibility mapping in an arid region of Pakistan through ensemble machine learning model</i>
111	Nguyen (2022)	<i>Flood susceptibility assessment using hybrid machine learning and remote sensing in Quang Tri province, Vietnam</i>
112	Li <i>et al.</i> (2022)	<i>Embedded Feature Selection and Machine Learning Methods for Flash Flood Susceptibility-Mapping in the Mainstream Songhua River Basin, China</i>
113	Saha <i>et al.</i> (2022)	<i>Deep learning algorithms to develop Flood susceptibility map in Data-Scarce and Ungauged River Basin in India</i>
114	Hidayah <i>et al.</i> (2022)	<i>Assessing Coastal Flood Susceptibility in East Java, Indonesia: Comparison of Statistical Bivariate and Machine Learning Techniques</i>
115	Li <i>et al.</i> (2022)	<i>A Positive-Unlabeled Learning Algorithm for Urban Flood Susceptibility Modeling</i>
116	Al-Areeq <i>et al.</i> (2022)	<i>Computational Machine Learning Approach for Flood Susceptibility Assessment Integrated with Remote Sensing and GIS Techniques from Jeddah, Saudi Arabia</i>
117	Avand <i>et al.</i> (2022)	<i>DEM resolution effects on machine learning performance for flood probability mapping</i>

Fonte: Elaborada pelo autor (2024).

Quadro 5 - Lista de artigos analisados na bibliometria

(continuação)

Id	Citação	Título
118	Prasad <i>et al.</i> (2022)	<i>Novel ensemble machine learning models in flood susceptibility mapping</i>
119	Kohansarbaz <i>et al.</i> (2022)	<i>Modelling flood susceptibility in northern Iran: Application of five well-known machine-learning models</i>
120	Ekwueme (2022)	<i>Machine Learning Based Prediction of Urban Flood Susceptibility from Selected Rivers in a Tropical Catchment Area</i>
121	Costache <i>et al.</i> (2022)	<i>New Machine Learning Ensemble for Flood Susceptibility Estimation</i>
122	Hasanuzzaman <i>et al.</i> (2022)	<i>A comparison of performance measures of three machine learning algorithms for flood susceptibility mapping of river Silabati (tropical river, India)</i>
123	Luu <i>et al.</i> (2022)	<i>Using Decision Tree J48 Based Machine Learning Algorithm for Flood Susceptibility Mapping: A Case Study in Quang Binh Province, Vietnam</i>
124	Singha <i>et al.</i> (2022)	<i>Spatial Analysis of Flood Hazard Zoning Map Using Novel Hybrid Machine Learning Technique in Assam, India</i>
125	Nguyen <i>et al.</i> (2022)	<i>Flood susceptibility mapping using advanced hybrid machine learning and CyGNSS: a case study of Nghe An province, Vietnam</i>
126	Youssef <i>et al.</i> (2022)	<i>Advanced machine learning algorithms for flood susceptibility modeling - performance comparison: Red Sea, Egypt</i>
127	Ighile <i>et al.</i> (2022)	<i>Application of GIS and Machine Learning to Predict Flood Areas in Nigeria</i>
128	Nguyen (2022)	<i>GIS-based hybrid machine learning for flood susceptibility prediction in the Nhat Le-Kien Giang watershed, Vietnam</i>
129	Mishra <i>et al.</i> (2022)	<i>Geospatial modeling using hybrid machine learning approach for flood susceptibility</i>
130	Eslaminezhad <i>et al.</i> (2022)	<i>Assessment of flood susceptibility prediction based on optimized tree-based machine learning models</i>
131	Balogun <i>et al.</i> (2022)	<i>Assessment of data mining, multi-criteria decision making and fuzzy-computing techniques for spatial flood susceptibility mapping: a comparative study</i>

Fonte: Elaborada pelo autor (2024).

Quadro 5 - Lista de artigos analisados na bibliometria

(continuação)

Id	Citação	Título
132	Pandey <i>et al.</i> (2022)	<i>Flood susceptibility modeling based on new hybrid intelligence model: Optimization of XGboost model using GA metaheuristic algorithm</i>
133	Guo <i>et al.</i> (2022)	<i>Data-driven rapid flood prediction mapping with catchment generalizability</i>
134	Yao <i>et al.</i> (2022)	<i>Applications of Stacking/Blending ensemble learning approaches for evaluating flash flood susceptibility</i>
135	Askar <i>et al.</i> (2022)	<i>Flood Susceptibility Mapping Using Remote Sensing and Integration of Decision Table Classifier and Metaheuristic Algorithms</i>
136	Ahmed <i>et al.</i> (2022)	<i>Flood susceptibility modeling in the urban watershed of Guwahati using improved metaheuristic-based ensemble machine learning algorithms</i>
137	Ilia <i>et al.</i> (2022)	<i>Flash flood susceptibility mapping using stacking ensemble machine learning models</i>
138	Janizadeh <i>et al.</i> (2022)	<i>Hybrid XGboost model with various Bayesian hyperparameter optimization algorithms for flood hazard susceptibility modeling</i>
139	Saber <i>et al.</i> (2022)	<i>Examining LightGBM and CatBoost models for wadi flash flood susceptibility prediction</i>
140	Saleh <i>et al.</i> (2022)	<i>Flash flood susceptibility mapping in urban area using genetic algorithm and ensemble method</i>
141	Saravanan e Abijith (2022)	<i>Flood susceptibility mapping of Northeast coastal districts of Tamil Nadu India using Multi-source Geospatial data and Machine Learning techniques</i>
142	Youssef <i>et al.</i> (2022)	<i>Comparative study of convolutional neural network (CNN) and support vector machine (SVM) for flood susceptibility mapping: a case study at Ras Gharib, Red Sea, Egypt</i>
143	Siam <i>et al.</i> (2022)	<i>Effects of Label Noise on Regression Performances and Model Complexities for Hybridized Machine Learning Based Spatial Flood Susceptibility Modelling</i>
144	McGrath e Gohl (2022)	<i>Assessing the Impact of Meteorological Variables on Machine Learning Flood Susceptibility Mapping</i>
145	Gharakhanlou e Perez (2022)	<i>Spatial Prediction of Current and Future Flood Susceptibility: Examining the Implications of Changing Climates on Flood Susceptibility Using Machine Learning Models</i>

Fonte: Elaborada pelo autor (2024).

Quadro 5 - Lista de artigos analisados na bibliometria

(continuação)

Id	Citação	Título
146	Deroliya <i>et al.</i> (2022)	<i>A novel flood risk mapping approach with machine learning considering geomorphic and socio-economic vulnerability dimensions</i>
147	Edamo <i>et al.</i> (2022)	<i>A comparative assessment of multi-criteria decision-making analysis and machine learning methods for flood susceptibility mapping and socio-economic impacts on flood risk in Abela-Abaya floodplain of Ethiopia</i>
148	Vajeethaveesin <i>et al.</i> (2022)	<i>A Performance Comparison between GIS-based and Neuron Network Methods for Flood Susceptibility Assessment in Ayutthaya Province</i>
149	Ahmed <i>et al.</i> (2022)	<i>Flood susceptibility mapping in Brahmaputra floodplain of Bangladesh using deep boost, deep learning neural network, and artificial neural network</i>
150	Ngo <i>et al.</i> (2023)	<i>Prediction of Flash Flood Susceptibility of Hilly Terrain Using Deep Neural Network: A Case Study of Vietnam</i>
151	Liu <i>et al.</i> (2023)	<i>A new avenue to improve the performance of integrated modeling for flash flood susceptibility assessment: Applying cluster algorithms</i>
152	Rana e Mahanta (2023)	<i>Spatial prediction of flash flood susceptible areas using novel ensemble of bivariate statistics and machine learning techniques for ungauged region</i>
153	Hasan <i>et al.</i> (2023)	<i>Use of machine learning algorithms to assess flood susceptibility in the coastal area of Bangladesh</i>
154	Wu <i>et al.</i> (2023)	<i>The quantitative assessment of impact of pumping capacity and LID on urban flood susceptibility based on machine learning</i>
155	Razavi-Termeh <i>et al.</i> (2023)	<i>Application of genetic algorithm in optimization parallel ensemble-based machine learning algorithms to flood susceptibility mapping using radar satellite imagery</i>
156	Gharakhanlou e Perez (2023)	<i>Flood susceptible prediction through the use of geospatial variables and machine learning methods</i>
157	Adnan <i>et al.</i> (2023)	<i>A novel framework for addressing uncertainties in machine learning-based geospatial approaches for flood prediction</i>
158	Chen <i>et al.</i> (2023)	<i>Modeling rules of regional flash flood susceptibility prediction using different machine learning models</i>

Fonte: Elaborada pelo autor (2024).

Quadro 5 - Lista de artigos analisados na bibliometria

(conclusão)

Id	Citação	Título
159	Seydi <i>et al.</i> (2023)	<i>Comparison of Machine Learning Algorithms for Flood Susceptibility Mapping</i>
160	Abu-Salih <i>et al.</i> (2023)	<i>The development of a road network flood risk detection model using optimised ensemble learning</i>
161	Luo <i>et al.</i> (2023)	<i>Resilient landscape pattern for reducing coastal flood susceptibility</i>
162	Aldiansyah e Wardani (2023)	<i>Evaluation of flood susceptibility prediction based on a resampling method using machine learning</i>
163	Nguyen <i>et al.</i> (2023)	<i>Application of hybrid model-based deep learning and swarm-based optimizers for flood susceptibility prediction in Binh Dinh province, Vietnam</i>
164	Tabarestani} <i>et al.</i> (2023)	<i>Flood potential mapping by integrating the bivariate statistics, multi-criteria decision-making, and machine learning techniques</i>
165	Aydin e Iban (2023)	<i>Predicting and analyzing flood susceptibility using boosting-based ensemble machine learning algorithms with SHapley Additive exPlanations</i>
166	Hasanuzzaman <i>et al.</i> (2023)	<i>Characterizing recurrent flood hazards in the Himalayan foothill region through data-driven modelling</i>
167	Li <i>et al.</i> (2023)	<i>Urban flood susceptibility mapping based on social media data in Chengdu city, China</i>
168	Saravanan <i>et al.</i> (2023)	<i>Flood susceptibility mapping using machine learning boosting algorithms techniques in Idukki district of Kerala India</i>
169	Riazi <i>et al.</i> (2023)	<i>Enhancing flood susceptibility modeling using multi-temporal SAR images, CHIRPS data, and hybrid machine learning algorithms</i>

Fonte: Elaborada pelo autor (2024).

APÊNDICE B – REGISTROS E LOCALIZAÇÃO DO INVENTÁRIO DE PONTOS DE INUNDAÇÃO E NÃO INUNDAÇÃO

A seguir são apresentados os dados levantados de todos os pontos da variável dependente (Tabela 8). As coordenadas X e Y encontram-se no sistema EPSG:31982 (SIRGAS 2000 / UTM zona 22S). O valor “1” indica pontos com registro de inundação e “0” indica pontos sem registro de inundação. A coluna “ID” corresponde à sequência de registros, “TIPO” identifica a origem do mapeamento e “CI” é a classificação de inundação (1 para área inundada e 0 para área não inundada).

Os dados completos, assim como o arquivo original do projeto QGIS, podem ser obtidos em:

- <https://github.com/nionmaron/FloodSusceptibilityML-Dissertation-Nion>

Siglas utilizadas:

- DC-DAA: Dados da Defesa Civil referentes ao relatório de declaração de áreas afetadas por desastre natural, utilizado para liberação de FGTS em evento de inundação no ano de 2018.
- AC-RMA: Área consolidada de corpos d’água (rios e massas de água) oriunda do mapeamento hidrográfico realizado pela Prefeitura Municipal de Joinville para a Bacia Hidrográfica do Rio Cachoeira.
- AC-MM: Área consolidada de mancha de maré (cota 2,5 m), conforme o Plano Diretor de Drenagem Urbana (PDDU) da Bacia Hidrográfica do Rio Cachoeira.
- RS-CG: Rotas seguras em caso de inundações (Defesa Civil) e características geomorfológicas de áreas não inundáveis (como altitude segura e inclinação do terreno).

Tabela 8 - Inventário de pontos de inundação e não inundação

(continua)

ID	X	Y	TIPO	CI					
11	717803,9	7086574,0	DC-DAA	1					
12	717715,9	7085445,0	DC-DAA	1					
13	717695,9	7085790,0	DC-DAA	1					
14	717676,9	7086704,0	DC-DAA	1					
15	717619,9	7085406,0	DC-DAA	1					
16	717551,9	7086340,0	DC-DAA	1					
17	717489,9	7086529,0	DC-DAA	1					
18	717465,9	7086259,0	DC-DAA	1					
19	717448,9	7086592,0	DC-DAA	1					
20	717310,9	7085441,0	DC-DAA	1					
21	717306,9	7085275,0	DC-DAA	1					

22	717305,9	7085719,0	DC-DAA	1	75	714852,9	7087869,0	DC-DAA	1
23	717255,9	7085093,0	DC-DAA	1	76	714834,9	7092730,0	DC-DAA	1
24	717179,9	7088240,0	DC-DAA	1	77	714828,9	7084983,0	DC-DAA	1
25	717143,9	7087320,0	DC-DAA	1	78	714818,9	7092644,0	DC-DAA	1
26	717106,9	7079743,0	DC-DAA	1	79	714808,9	7088868,0	DC-DAA	1
27	716832,9	7084652,0	DC-DAA	1	80	714776,9	7092371,0	DC-DAA	1
28	716801,9	7084548,0	DC-DAA	1	81	714772,9	7092503,0	DC-DAA	1
29	716797,9	7082891,0	DC-DAA	1	82	714756,9	7091193,0	DC-DAA	1
30	716745,9	7081498,0	DC-DAA	1	83	714721,9	7087886,0	DC-DAA	1
31	716743,9	7084154,0	DC-DAA	1	84	714534,9	7084669,0	DC-DAA	1
32	716548,9	7082895,0	DC-DAA	1	85	714532,9	7084949,0	DC-DAA	1
33	716546,9	7083938,0	DC-DAA	1	86	714504,9	7090024,0	DC-DAA	1
34	716534,9	7081507,0	DC-DAA	1	87	714504,9	7090163,0	DC-DAA	1
35	716483,9	7083040,0	DC-DAA	1	88	714490,9	7092321,0	DC-DAA	1
36	716453,9	7081535,0	DC-DAA	1	89	714465,9	7092448,0	DC-DAA	1
37	716407,9	7083260,0	DC-DAA	1	90	714444,9	7087555,0	DC-DAA	1
38	716354,9	7082714,0	DC-DAA	1	91	714435,9	7087433,0	DC-DAA	1
39	716220,9	7086588,0	DC-DAA	1	92	714420,9	7090126,0	DC-DAA	1
40	715839,9	7086713,0	DC-DAA	1	93	714361,9	7092392,0	DC-DAA	1
41	715736,9	7091041,0	DC-DAA	1	94	714338,9	7087420,0	DC-DAA	1
42	715728,9	7086666,0	DC-DAA	1	95	714258,9	7092608,0	DC-DAA	1
43	715664,9	7086192,0	DC-DAA	1	96	714119,9	7087902,0	DC-DAA	1
44	715647,9	7088667,0	DC-DAA	1	97	714104,9	7086970,0	DC-DAA	1
45	715611,9	7086091,0	DC-DAA	1	98	714025,9	7092399,0	DC-DAA	1
46	715524,9	7088320,0	DC-DAA	1	99	713962,9	7092704,0	DC-DAA	1
47	715509,9	7090832,0	DC-DAA	1	100	713880,9	7092481,0	DC-DAA	1
48	715480,9	7085700,0	DC-DAA	1	101	713797,9	7087743,0	DC-DAA	1
49	715474,9	7086141,0	DC-DAA	1	102	713774,9	7092725,0	DC-DAA	1
50	715404,9	7085702,0	DC-DAA	1	103	713526,9	7092746,0	DC-DAA	1
51	715398,9	7090689,0	DC-DAA	1	104	713473,9	7092956,0	DC-DAA	1
52	715393,9	7088889,0	DC-DAA	1	105	713441,9	7086751,0	DC-DAA	1
53	715354,9	7089100,0	DC-DAA	1	106	713427,9	7093235,0	DC-DAA	1
54	715338,9	7090278,0	DC-DAA	1	107	713343,9	7086606,0	DC-DAA	1
55	715310,9	7089702,0	DC-DAA	1	108	713267,9	7092821,0	DC-DAA	1
56	715292,9	7089914,0	DC-DAA	1	109	713202,9	7086597,0	DC-DAA	1
57	715285,9	7090787,0	DC-DAA	1	110	713200,9	7086491,0	DC-DAA	1
58	715251,9	7089702,0	DC-DAA	1	111	713148,9	7093096,0	DC-DAA	1
59	715199,9	7088279,0	DC-DAA	1	112	712869,9	7086807,0	DC-DAA	1
60	715171,9	7081936,0	DC-DAA	1	113	712837,9	7093080,0	DC-DAA	1
61	715145,9	7090970,0	DC-DAA	1	114	712733,9	7092430,0	DC-DAA	1
62	715143,9	7090692,0	DC-DAA	1	115	712732,9	7092766,0	DC-DAA	1
63	715130,9	7091896,0	DC-DAA	1	116	712719,9	7092332,0	DC-DAA	1
64	715122,9	7090233,0	DC-DAA	1	117	712433,9	7091995,0	DC-DAA	1
65	715119,9	7085449,0	DC-DAA	1	118	712339,9	7093041,0	DC-DAA	1
66	715110,9	7081805,0	DC-DAA	1	119	712302,9	7092964,0	DC-DAA	1
67	715085,9	7087999,0	DC-DAA	1	120	712186,9	7092786,0	DC-DAA	1
68	715046,9	7089008,0	DC-DAA	1	121	712148,9	7092952,0	DC-DAA	1
69	715003,9	7090962,0	DC-DAA	1	122	712118,9	7091350,0	DC-DAA	1
70	714980,9	7087956,0	DC-DAA	1	123	711893,9	7092784,0	DC-DAA	1
71	714964,9	7091180,0	DC-DAA	1	124	711859,9	7092636,0	DC-DAA	1
72	714952,9	7085155,0	DC-DAA	1	125	711701,9	7092428,0	DC-DAA	1
73	714877,9	7091279,0	DC-DAA	1	126	710915,9	7093039,0	DC-DAA	1
74	714865,9	7090977,0	DC-DAA	1	127	710799,9	7092953,0	DC-DAA	1

128	710624,9	7092700,0	DC-DAA	1	250	717305,9	7085675,0	AC-RMA	1
129	714975,0	7092980,5	DC-DAA	1	251	718005,9	7087865,0	AC-RMA	1
130	713929,0	7092137,7	AC-RMA	1	252	714165,9	7083365,0	AC-RMA	1
200	712185,9	7092835,0	AC-RMA	1	253	716575,9	7083895,0	AC-RMA	1
201	713985,9	7086755,0	AC-RMA	1	254	712725,9	7090625,0	AC-RMA	1
202	715545,9	7088905,0	AC-RMA	1	255	711255,9	7093215,0	AC-RMA	1
203	716195,9	7088025,0	AC-RMA	1	256	714495,9	7090035,0	AC-RMA	1
204	717385,9	7088245,0	AC-RMA	1	257	714425,9	7087845,0	AC-RMA	1
205	717655,9	7086295,0	AC-RMA	1	300	716235,9	7087425,0	AC-MM	1
206	716505,9	7082885,0	AC-RMA	1	301	719105,9	7088565,0	AC-MM	1
207	713245,9	7087715,0	AC-RMA	1	302	717105,9	7088515,0	AC-MM	1
208	717665,9	7080645,0	AC-RMA	1	303	718125,9	7088025,0	AC-MM	1
209	718705,9	7087815,0	AC-RMA	1	304	717615,9	7088015,0	AC-MM	1
210	716865,9	7084755,0	AC-RMA	1	305	717855,9	7087125,0	AC-MM	1
211	712045,9	7090285,0	AC-RMA	1	306	717695,9	7089185,0	AC-MM	1
212	717415,9	7087445,0	AC-RMA	1	307	715075,9	7089235,0	AC-MM	1
213	713185,9	7088455,0	AC-RMA	1	308	717635,9	7086265,0	AC-MM	1
214	718335,9	7082855,0	AC-RMA	1	309	716475,9	7087835,0	AC-MM	1
215	718285,9	7084785,0	AC-RMA	1	310	718665,9	7086975,0	AC-MM	1
216	718575,9	7083785,0	AC-RMA	1	311	715365,9	7089565,0	AC-MM	1
217	710305,9	7092415,0	AC-RMA	1	312	715495,9	7088395,0	AC-MM	1
218	714755,9	7082955,0	AC-RMA	1	313	716175,9	7087015,0	AC-MM	1
219	715555,9	7085975,0	AC-RMA	1	314	715895,9	7088265,0	AC-MM	1
220	718565,9	7081715,0	AC-RMA	1	315	718885,9	7088215,0	AC-MM	1
221	718475,9	7088535,0	AC-RMA	1	316	717115,9	7087535,0	AC-MM	1
222	716065,9	7090575,0	AC-RMA	1	317	718365,9	7088365,0	AC-MM	1
223	716325,9	7092295,0	AC-RMA	1	318	717745,9	7088545,0	AC-MM	1
224	715125,9	7081575,0	AC-RMA	1	319	716805,9	7087145,0	AC-MM	1
225	716195,9	7086975,0	AC-RMA	1	320	717885,9	7087575,0	AC-MM	1
226	713905,9	7088945,0	AC-RMA	1	321	715715,9	7088805,0	AC-MM	1
227	715175,9	7091955,0	AC-RMA	1	322	717415,9	7086965,0	AC-MM	1
228	713395,9	7090295,0	AC-RMA	1	323	717315,9	7085625,0	AC-MM	1
229	712545,9	7089305,0	AC-RMA	1	324	714985,9	7091485,0	AC-MM	1
230	715105,9	7085425,0	AC-RMA	1	325	719005,9	7089025,0	AC-MM	1
231	714315,9	7084455,0	AC-RMA	1	326	718155,9	7088805,0	AC-MM	1
232	718535,9	7086935,0	AC-RMA	1	327	718695,9	7089305,0	AC-MM	1
233	716505,9	7081945,0	AC-RMA	1	328	718355,9	7087475,0	AC-MM	1
234	715185,9	7088155,0	AC-RMA	1	329	717825,9	7085295,0	AC-MM	1
235	713715,9	7092775,0	AC-RMA	1	330	718245,9	7089385,0	AC-MM	1
236	712415,9	7092045,0	AC-RMA	1	331	716065,9	7087875,0	AC-MM	1
237	715175,9	7091145,0	AC-RMA	1	332	716685,9	7084305,0	AC-MM	1
238	716285,9	7085655,0	AC-RMA	1	333	715575,9	7086345,0	AC-MM	1
239	713745,9	7091925,0	AC-RMA	1	334	717025,9	7088025,0	AC-MM	1
240	711445,9	7091935,0	AC-RMA	1	335	715775,9	7086715,0	AC-MM	1
241	715215,9	7090185,0	AC-RMA	1	336	718595,9	7088745,0	AC-MM	1
242	713215,9	7086705,0	AC-RMA	1	337	718155,9	7086625,0	AC-MM	1
243	713795,9	7085245,0	AC-RMA	1	338	715425,9	7090475,0	AC-MM	1
244	713045,9	7093145,0	AC-RMA	1	339	715305,9	7090965,0	AC-MM	1
245	719165,9	7088385,0	AC-RMA	1	340	717725,9	7086685,0	AC-MM	1
246	718675,9	7080405,0	AC-RMA	1	341	718625,9	7087885,0	AC-MM	1
247	711855,9	7086645,0	AC-RMA	1	342	718525,9	7086295,0	AC-MM	1
248	714805,9	7089055,0	AC-RMA	1	343	717225,9	7085115,0	AC-MM	1
249	719025,9	7083065,0	AC-RMA	1	344	714295,9	7092495,0	AC-MM	1

345	715365,9	7090005,0	AC-MM	1	550	715425,9	7085145,0	RS-CG	0
346	715215,9	7088095,0	AC-MM	1	551	714125,9	7090655,0	RS-CG	0
347	717325,9	7088915,0	AC-MM	1	552	710465,9	7093385,0	RS-CG	0
500	717245,9	7083995,0	RS-CG	0	553	718045,9	7083185,0	RS-CG	0
501	716415,9	7090295,0	RS-CG	0	554	716955,9	7081755,0	RS-CG	0
502	713235,9	7086835,0	RS-CG	0	555	717135,9	7083535,0	RS-CG	0
503	714425,9	7082835,0	RS-CG	0	556	713195,9	7091265,0	RS-CG	0
504	710035,9	7093085,0	RS-CG	0	557	712905,9	7088655,0	RS-CG	0
505	718985,9	7086515,0	RS-CG	0	558	718585,9	7080355,0	RS-CG	0
506	712125,9	7087765,0	RS-CG	0	559	716575,9	7091275,0	RS-CG	0
507	713445,9	7091825,0	RS-CG	0	560	716995,9	7081085,0	RS-CG	0
508	714555,9	7093265,0	RS-CG	0	561	716145,9	7083535,0	RS-CG	0
509	716145,9	7080715,0	RS-CG	0	562	717145,9	7085945,0	RS-CG	0
510	715845,9	7085825,0	RS-CG	0	563	715475,9	7082635,0	RS-CG	0
511	716125,9	7082225,0	RS-CG	0	564	714915,9	7084155,0	RS-CG	0
512	718385,9	7081165,0	RS-CG	0	565	711425,9	7092495,0	RS-CG	0
513	713105,9	7093575,0	RS-CG	0	566	711835,9	7086425,0	RS-CG	0
514	712615,9	7086715,0	RS-CG	0	567	718635,9	7079775,0	RS-CG	0
515	712935,9	7089455,0	RS-CG	0	568	717185,9	7079885,0	RS-CG	0
516	714805,9	7082395,0	RS-CG	0	569	715895,9	7089765,0	RS-CG	0
517	714015,9	7088705,0	RS-CG	0	570	715925,9	7084125,0	RS-CG	0
518	714005,9	7088245,0	RS-CG	0	571	712825,9	7086015,0	RS-CG	0
519	717205,9	7084415,0	RS-CG	0	572	716725,9	7083595,0	RS-CG	0
520	713565,9	7090515,0	RS-CG	0	573	713235,9	7085155,0	RS-CG	0
521	718405,9	7082735,0	RS-CG	0	574	717715,9	7089865,0	RS-CG	0
522	712655,9	7090195,0	RS-CG	0	575	715805,9	7091535,0	RS-CG	0
523	715595,9	7089245,0	RS-CG	0	576	714395,9	7089525,0	RS-CG	0
524	717355,9	7082935,0	RS-CG	0	577	714165,9	7087705,0	RS-CG	0
525	712545,9	7089755,0	RS-CG	0	578	714245,9	7086075,0	RS-CG	0
526	716685,9	7090735,0	RS-CG	0	579	716145,9	7088265,0	RS-CG	0
527	715545,9	7083045,0	RS-CG	0	580	713755,9	7085765,0	RS-CG	0
528	716015,9	7089285,0	RS-CG	0	581	716345,9	7084835,0	RS-CG	0
529	711515,9	7091475,0	RS-CG	0	582	719305,9	7084015,0	RS-CG	0
530	716845,9	7088985,0	RS-CG	0	583	714535,9	7084435,0	RS-CG	0
531	719745,9	7082215,0	RS-CG	0	584	717975,9	7083835,0	RS-CG	0
532	713785,9	7089335,0	RS-CG	0	585	715355,9	7083465,0	RS-CG	0
533	718085,9	7081655,0	RS-CG	0	586	710195,9	7092595,0	RS-CG	0
534	711875,9	7093645,0	RS-CG	0	587	713005,9	7085545,0	RS-CG	0
535	717485,9	7080315,0	RS-CG	0	588	715045,9	7086685,0	RS-CG	0
536	716445,9	7089455,0	RS-CG	0	589	718685,9	7081615,0	RS-CG	0
537	717725,9	7081855,0	RS-CG	0	590	713935,9	7085065,0	RS-CG	0
538	712925,9	7094025,0	RS-CG	0	591	716595,9	7092325,0	RS-CG	0
539	717325,9	7082295,0	RS-CG	0	592	716915,9	7086865,0	RS-CG	0
540	717835,9	7084255,0	RS-CG	0	593	716755,9	7086045,0	RS-CG	0
541	714915,9	7086115,0	RS-CG	0	594	715605,9	7091985,0	RS-CG	0
542	712595,9	7088085,0	RS-CG	0	595	711935,9	7090095,0	RS-CG	0
543	713245,9	7092375,0	RS-CG	0	596	718915,9	7084585,0	RS-CG	0
544	712485,9	7091455,0	RS-CG	0	597	715455,9	7093685,0	RS-CG	0
545	713775,9	7087395,0	RS-CG	0	598	715555,9	7081375,0	RS-CG	0
546	710775,9	7092715,0	RS-CG	0	599	711725,9	7091905,0	RS-CG	0
547	716815,9	7080715,0	RS-CG	0	600	712665,9	7087665,0	RS-CG	0
548	716075,9	7093205,0	RS-CG	0	601	715875,9	7092565,0	RS-CG	0
549	716245,9	7088925,0	RS-CG	0	602	715085,9	7093485,0	RS-CG	0

603	710505,9	7093795,0	RS-CG	0
604	713275,9	7087385,0	RS-CG	0
605	717905,9	7082325,0	RS-CG	0
606	713515,9	7093755,0	RS-CG	0
607	717265,9	7089085,0	RS-CG	0
608	713815,9	7086315,0	RS-CG	0
609	716495,9	7093315,0	RS-CG	0
610	716155,9	7083025,0	RS-CG	0
611	719605,9	7081815,0	RS-CG	0
612	716975,9	7085125,0	RS-CG	0
613	711155,9	7087215,0	RS-CG	0
614	715865,9	7084605,0	RS-CG	0
615	716495,9	7092875,0	RS-CG	0
616	711965,9	7093245,0	RS-CG	0
617	718055,9	7080765,0	RS-CG	0
618	716545,9	7084475,0	RS-CG	0
619	712585,9	7088915,0	RS-CG	0
620	716525,9	7085215,0	RS-CG	0
621	711235,9	7093115,0	RS-CG	0
622	716900,4	7090131,1	RS-CG	0
623	712605,9	7093315,0	RS-CG	0
624	719445,9	7082565,0	RS-CG	0
625	719025,9	7082265,0	RS-CG	0
626	715505,9	7082035,0	RS-CG	0
627	710325,9	7092085,0	RS-CG	0
628	716865,9	7088575,0	RS-CG	0
629	711115,9	7092075,0	RS-CG	0
630	717615,9	7081325,0	RS-CG	0
631	719555,9	7084405,0	RS-CG	0
632	717405,9	7089505,0	RS-CG	0
633	713995,9	7093525,0	RS-CG	0
634	712715,9	7091985,0	RS-CG	0
635	713385,9	7094335,0	RS-CG	0
636	716575,9	7088045,0	RS-CG	0
637	714725,9	7090595,0	RS-CG	0
638	719035,9	7082735,0	RS-CG	0
639	716145,9	7081475,0	RS-CG	0
640	718735,9	7085215,0	RS-CG	0
641	718485,9	7083235,0	RS-CG	0
642	713465,9	7088995,0	RS-CG	0
643	712785,9	7091075,0	RS-CG	0
644	716795,9	7085525,0	RS-CG	0
645	713955,9	7084635,0	RS-CG	0
646	713225,9	7085935,0	RS-CG	0
647	714855,9	7088725,0	RS-CG	0
648	714535,9	7093705,0	RS-CG	0
649	716955,9	7082755,0	RS-CG	0
650	714955,9	7083515,0	RS-CG	0
651	713235,9	7090815,0	RS-CG	0
652	714825,9	7088195,0	RS-CG	0
653	717185,9	7090035,0	RS-CG	0
654	714545,9	7083515,0	RS-CG	0
655	713075,9	7084755,0	RS-CG	0

656	718415,9	7085515,0	RS-CG	0
657	715005,9	7084895,0	RS-CG	0
658	716285,9	7081095,0	RS-CG	0
659	715985,9	7092155,0	RS-CG	0
660	718765,9	7080875,0	RS-CG	0
661	713825,9	7091655,0	RS-CG	0
662	715735,9	7090275,0	RS-CG	0
663	713875,9	7086985,0	RS-CG	0
664	718345,9	7084265,0	RS-CG	0
665	712425,9	7090815,0	RS-CG	0
666	713515,9	7084125,0	RS-CG	0
667	711875,9	7090535,0	RS-CG	0
668	713405,9	7083725,0	RS-CG	0
669	715265,9	7092755,0	RS-CG	0
670	714345,9	7092825,0	RS-CG	0
671	715295,9	7084575,0	RS-CG	0
672	709755,9	7092475,0	RS-CG	0
673	717171,4	7086412,4	RS-CG	0
674	716365,9	7081885,0	RS-CG	0
675	718355,9	7085085,0	RS-CG	0
676	716685,9	7091805,0	RS-CG	0
677	718125,9	7079695,0	RS-CG	0
678	713915,9	7089825,0	RS-CG	0
679	711905,9	7090975,0	RS-CG	0
680	712065,9	7092415,0	RS-CG	0
681	719105,9	7081665,0	RS-CG	0
682	711315,9	7086745,0	RS-CG	0
683	716515,9	7080415,0	RS-CG	0
684	714705,9	7081905,0	RS-CG	0
685	714835,9	7082795,0	RS-CG	0
686	719025,9	7080275,0	RS-CG	0
687	713675,9	7087995,0	RS-CG	0
688	712615,9	7085655,0	RS-CG	0
689	715955,9	7088645,0	RS-CG	0
690	715025,9	7081265,0	RS-CG	0
691	715025,9	7085665,0	RS-CG	0
692	718725,9	7086025,0	RS-CG	0
693	719525,9	7083025,0	RS-CG	0
694	716665,9	7081355,0	RS-CG	0
695	717655,9	7084665,0	RS-CG	0
696	715435,9	7084035,0	RS-CG	0
697	715275,9	7092325,0	RS-CG	0
698	712275,9	7093555,0	RS-CG	0
699	716875,9	7089425,0	RS-CG	0
700	710395,9	7091675,0	RS-CG	0
701	714385,9	7086515,0	RS-CG	0
702	714375,9	7090995,0	RS-CG	0
703	712115,9	7087365,0	RS-CG	0
704	718185,9	7080235,0	RS-CG	0
705	710735,9	7091945,0	RS-CG	0
706	717635,9	7083575,0	RS-CG	0
707	718655,9	7083635,0	RS-CG	0
708	714335,9	7084895,0	RS-CG	0

709	718175,9	7089795,0	RS-CG	0
710	711135,9	7093915,0	RS-CG	0
711	715895,9	7093655,0	RS-CG	0
712	715765,9	7080525,0	RS-CG	0
713	714375,9	7083965,0	RS-CG	0
714	716975,9	7091055,0	RS-CG	0
715	716915,9	7080295,0	RS-CG	0
716	715875,9	7082735,0	RS-CG	0
717	718965,9	7081265,0	RS-CG	0
718	715295,9	7091575,0	RS-CG	0
719	713785,9	7094065,0	RS-CG	0
720	716835,9	7082355,0	RS-CG	0
721	718645,9	7082045,0	RS-CG	0
722	713595,9	7085385,0	RS-CG	0
723	715345,9	7086355,0	RS-CG	0
724	714185,9	7089135,0	RS-CG	0
725	713805,9	7083475,0	RS-CG	0
726	711375,9	7093575,0	RS-CG	0
727	715785,9	7081015,0	RS-CG	0
728	711535,9	7087425,0	RS-CG	0
729	711075,9	7091265,0	RS-CG	0
730	713035,9	7089065,0	RS-CG	0
731	714665,9	7087125,0	RS-CG	0
732	715865,9	7081805,0	RS-CG	0
733	719395,9	7083575,0	RS-CG	0
734	714244,9	7088487,2	RS-CG	0
735	715465,8	7090218,6	RS-CG	0

Fonte: Elaborada pelo autor (2025).

APÊNDICE C – TABELAS CRUZADAS DE SUSCETIBILIDADE E FATORES

Este apêndice apresenta uma tabela consolidada (Tabela 9) que relaciona os níveis de suscetibilidade do **modelo floresta aleatória**, classificados em cinco categorias (muito baixo, baixo, médio, alto e muito alto), com os diferentes intervalos ou categorias das variáveis analisadas, como MDE, UOS, TSO, entre outras. Por meio desta tabela de contingência, é possível visualizar a distribuição dos percentuais (total de *pixels*), em cada combinação de classe de suscetibilidade e faixa de valor das variáveis.

A reclassificação dos dados foi realizada com critérios específicos para cada variável, garantindo que as características dos diferentes ambientes sejam representadas de forma consistente. Assim, o apêndice fornece uma ferramenta fundamental para a interpretação dos fatores de risco e a identificação das áreas mais suscetíveis aos fenômenos estudados, contribuindo para a fundamentação das análises.

Tabela 9 - Tabela de contingência

(continua)

Variável	Percentual	Intervalo	Nível de Suscetibilidade
MDE	0,03	[0 - 5,3]	Muito Baixo
MDE	2,54	[5,3 - 11,11]	Muito Baixo
MDE	12,25	[11,11 - 19,46]	Muito Baixo
MDE	17,41	[19,46 - 30]	Muito Baixo
MDE	18,25	[30 - 225,07]	Muito Baixo
MDE	0,12	[0 - 5,3]	Baixo
MDE	2,19	[5,3 - 11,11]	Baixo
MDE	2,88	[11,11 - 19,46]	Baixo
MDE	2,08	[19,46 - 30]	Baixo
MDE	0,37	[30 - 225,07]	Baixo
MDE	0,47	[0 - 5,3]	Médio
MDE	3,28	[5,3 - 11,11]	Médio
MDE	1,63	[11,11 - 19,46]	Médio
MDE	0,91	[19,46 - 30]	Médio
MDE	0,19	[30 - 225,07]	Médio
MDE	2,04	[0 - 5,3]	Alto
MDE	3,26	[5,3 - 11,11]	Alto
MDE	1,44	[11,11 - 19,46]	Alto
MDE	0,42	[19,46 - 30]	Alto
MDE	0,04	[30 - 225,07]	Alto

Fonte: Elaborada pelo autor (2025).

Tabela 9 - Tabela de contingência

(continuação)

Variável	Percentual	Intervalo	Nível de Suscetibilidade
MDE	17,4	[0 - 5,3]	Muito Alto
MDE	8,79	[5,3 - 11,11]	Muito Alto
MDE	1,86	[11,11 - 19,46]	Muito Alto
MDE	0,15	[19,46 - 30]	Muito Alto
MDE	0	[30 - 225,07]	Muito Alto
DCH	0,55	[0-50)	Muito Baixo
DCH	2,57	[50-100)	Muito Baixo
DCH	9,08	[200-300)	Muito Baixo
DCH	30,99	[>300]	Muito Baixo
DCH	0,44	[0-50)	Baixo
DCH	0,34	[50-100)	Baixo
DCH	1,31	[200-300)	Baixo
DCH	4,71	[>300]	Baixo
DCH	0,69	[0-50)	Médio
DCH	0,37	[50-100)	Médio
DCH	1,62	[200-300)	Médio
DCH	2,87	[>300]	Médio
DCH	0,79	[0-50)	Alto
DCH	0,8	[50-100)	Alto
DCH	1,67	[200-300)	Alto
DCH	2,3	[>300]	Alto
DCH	9,01	[0-50)	Muito Alto
DCH	6,27	[50-100)	Muito Alto
DCH	3,28	[200-300)	Muito Alto
DCH	1,67	[>300]	Muito Alto
DMV	13,17	[0 - 10]	Muito Baixo
DMV	5,04	[10 - 14,14]	Muito Baixo
DMV	8,43	[14,14 - 28,28]	Muito Baixo
DMV	9,3	[28,28 - 50,99]	Muito Baixo
DMV	14,54	[50,99 - 626,28]	Muito Baixo
DMV	3,1	[0 - 10]	Baixo
DMV	0,99	[10 - 14,14]	Baixo
DMV	1,45	[14,14 - 28,28]	Baixo
DMV	1,29	[28,28 - 50,99]	Baixo
DMV	0,81	[50,99 - 626,28]	Baixo
DMV	2,38	[0 - 10]	Médio
DMV	0,85	[10 - 14,14]	Médio
DMV	1,34	[14,14 - 28,28]	Médio
DMV	1,26	[28,28 - 50,99]	Médio

Fonte: Elaborada pelo autor (2025).

Tabela 9 - Tabela de contingência

(continuação)

Variável	Percentual	Intervalo	Nível de Suscetibilidade
DMV	0,65	[50,99 - 626,28]	Médio
DMV	2,46	[0 - 10]	Alto
DMV	0,86	[10 - 14,14]	Alto
DMV	1,53	[14,14 - 28,28]	Alto
DMV	1,74	[28,28 - 50,99]	Alto
DMV	0,62	[50,99 - 626,28]	Alto
DMV	11,41	[0 - 10]	Muito Alto
DMV	4,01	[10 - 14,14]	Muito Alto
DMV	5,7	[14,14 - 28,28]	Muito Alto
DMV	3,83	[28,28 - 50,99]	Muito Alto
DMV	3,26	[50,99 - 626,28]	Muito Alto
IVDN	0,11	Água/SemVeg (≤ 0)	Muito Baixo
IVDN	5,74	SoloExp./AreaUrb, (0-0,2)	Muito Baixo
IVDN	16,33	Veg, rala (0,2-0,5)	Muito Baixo
IVDN	6,41	Veg, densa (0,5-0,7)	Muito Baixo
IVDN	21,88	Veg, muito densa ($>0,7$)	Muito Baixo
IVDN	0,05	Água/SemVeg (≤ 0)	Baixo
IVDN	2,12	SoloExp./AreaUrb, (0-0,2)	Baixo
IVDN	3,5	Veg, rala (0,2-0,5)	Baixo
IVDN	0,85	Veg, densa (0,5-0,7)	Baixo
IVDN	1,1	Veg, muito densa ($>0,7$)	Baixo
IVDN	0,07	Água/SemVeg (≤ 0)	Médio
IVDN	1,97	SoloExp./AreaUrb, (0-0,2)	Médio
IVDN	2,88	Veg, rala (0,2-0,5)	Médio
IVDN	0,66	Veg, densa (0,5-0,7)	Médio
IVDN	0,91	Veg, muito densa ($>0,7$)	Médio
IVDN	0,09	Água/SemVeg (≤ 0)	Alto
IVDN	2,56	SoloExp./AreaUrb, (0-0,2)	Alto
IVDN	3,17	Veg, rala (0,2-0,5)	Alto
IVDN	0,66	Veg, densa (0,5-0,7)	Alto
IVDN	0,73	Veg, muito densa ($>0,7$)	Alto
IVDN	0,34	Água/SemVeg (≤ 0)	Muito Alto
IVDN	9,47	SoloExp./AreaUrb, (0-0,2)	Muito Alto
IVDN	11,89	Veg, rala (0,2-0,5)	Muito Alto
IVDN	2,62	Veg, densa (0,5-0,7)	Muito Alto
IVDN	3,87	Veg, muito densa ($>0,7$)	Muito Alto
DEC	0,49	[0 - 0,44]	Muito Baixo
DEC	1,15	[0,44 - 1,45]	Muito Baixo
DEC	9,79	[1,45 - 4,17]	Muito Baixo

Fonte: Elaborada pelo autor (2025).

Tabela 9 - Tabela de contingência

(continuação)

Variável	Percentual	Intervalo	Nível de Suscetibilidade
DEC	19,33	[4,17 - 10,84]	Muito Baixo
DEC	19,72	[10,84 - 50,8]	Muito Baixo
DEC	2,47	[0 - 0,44]	Baixo
DEC	1,44	[0,44 - 1,45]	Baixo
DEC	3,3	[1,45 - 4,17]	Baixo
DEC	0,32	[4,17 - 10,84]	Baixo
DEC	0,11	[10,84 - 50,8]	Baixo
DEC	1,88	[0 - 0,44]	Médio
DEC	1,8	[0,44 - 1,45]	Médio
DEC	2,44	[1,45 - 4,17]	Médio
DEC	0,24	[4,17 - 10,84]	Médio
DEC	0,12	[10,84 - 50,8]	Médio
DEC	2,06	[0 - 0,44]	Alto
DEC	3,11	[0,44 - 1,45]	Alto
DEC	1,95	[1,45 - 4,17]	Alto
DEC	0,06	[4,17 - 10,84]	Alto
DEC	0,02	[10,84 - 50,8]	Alto
DEC	13,13	[0 - 0,44]	Muito Alto
DEC	12,54	[0,44 - 1,45]	Muito Alto
DEC	2,53	[1,45 - 4,17]	Muito Alto
DEC	0	[4,17 - 10,84]	Muito Alto
DEC	0	[10,84 - 50,8]	Muito Alto
RUG	0,37	[0 - 0,21]	Muito Baixo
RUG	0,86	[0,21 - 0,69]	Muito Baixo
RUG	10,12	[0,69 - 1,93]	Muito Baixo
RUG	19,36	[1,93 - 4,94]	Muito Baixo
RUG	19,77	[4,94 - 27,68]	Muito Baixo
RUG	2,26	[0 - 0,21]	Baixo
RUG	1,52	[0,21 - 0,69]	Baixo
RUG	3,43	[0,69 - 1,93]	Baixo
RUG	0,31	[1,93 - 4,94]	Baixo
RUG	0,11	[4,94 - 27,68]	Baixo
RUG	1,78	[0 - 0,21]	Médio
RUG	1,92	[0,21 - 0,69]	Médio
RUG	2,42	[0,69 - 1,93]	Médio
RUG	0,24	[1,93 - 4,94]	Médio
RUG	0,12	[4,94 - 27,68]	Médio
RUG	2,04	[0 - 0,21]	Alto
RUG	3,19	[0,21 - 0,69]	Alto
RUG	1,89	[0,69 - 1,93]	Alto

Fonte: Elaborada pelo autor (2025).

Tabela 9 - Tabela de contingência

(continuação)

Variável	Percentual	Intervalo	Nível de Suscetibilidade
RUG	0,07	[1,93 - 4,94]	Alto
RUG	0,02	[4,94 - 27,68]	Alto
RUG	13,5	[0 - 0,21]	Muito Alto
RUG	12,54	[0,21 - 0,69]	Muito Alto
RUG	2,16	[0,69 - 1,93]	Muito Alto
RUG	0	[1,93 - 4,94]	Muito Alto
RUG	0	[4,94 - 27,68]	Muito Alto
IUT	19,55	[1,91 - 5,26]	Muito Baixo
IUT	17,06	[5,26 - 6,5]	Muito Baixo
IUT	9,43	[6,5 - 7,84]	Muito Baixo
IUT	3,17	[7,84 - 9,15]	Muito Baixo
IUT	1,27	[9,15 - 17,3]	Muito Baixo
IUT	0,18	[1,91 - 5,26]	Baixo
IUT	1,05	[5,26 - 6,5]	Baixo
IUT	2,07	[6,5 - 7,84]	Baixo
IUT	2,84	[7,84 - 9,15]	Baixo
IUT	1,5	[9,15 - 17,3]	Baixo
IUT	0,14	[1,91 - 5,26]	Médio
IUT	0,67	[5,26 - 6,5]	Médio
IUT	1,79	[6,5 - 7,84]	Médio
IUT	2,52	[7,84 - 9,15]	Médio
IUT	1,36	[9,15 - 17,3]	Médio
IUT	0,04	[1,91 - 5,26]	Alto
IUT	0,56	[5,26 - 6,5]	Alto
IUT	1,82	[6,5 - 7,84]	Alto
IUT	2,78	[7,84 - 9,15]	Alto
IUT	2,01	[9,15 - 17,3]	Alto
IUT	0,01	[1,91 - 5,26]	Muito Alto
IUT	0,64	[5,26 - 6,5]	Muito Alto
IUT	4,92	[6,5 - 7,84]	Muito Alto
IUT	13,44	[7,84 - 9,15]	Muito Alto
IUT	9,2	[9,15 - 17,3]	Muito Alto
IPE	1,46	[0 - 0,35]	Muito Baixo
IPE	4,96	[0,35 - 1,53]	Muito Baixo
IPE	10,77	[1,53 - 3,98]	Muito Baixo
IPE	15,65	[3,98 - 10,29]	Muito Baixo
IPE	17,66	[10,29 - 4141,04]	Muito Baixo
IPE	2,97	[0 - 0,35]	Baixo
IPE	1,44	[0,35 - 1,53]	Baixo
IPE	1,55	[1,53 - 3,98]	Baixo

Fonte: Elaborada pelo autor (2025).

Tabela 9 - Tabela de contingência

(continuação)

Variável	Percentual	Intervalo	Nível de Suscetibilidade
IPE	1,04	[3,98 - 10,29]	Baixo
IPE	0,63	[10,29 - 4141,04]	Baixo
IPE	2,05	[0 - 0,35]	Médio
IPE	1,51	[0,35 - 1,53]	Médio
IPE	1,54	[1,53 - 3,98]	Médio
IPE	0,86	[3,98 - 10,29]	Médio
IPE	0,52	[10,29 - 4141,04]	Médio
IPE	2,12	[0 - 0,35]	Alto
IPE	2,16	[0,35 - 1,53]	Alto
IPE	1,69	[1,53 - 3,98]	Alto
IPE	0,83	[3,98 - 10,29]	Alto
IPE	0,39	[10,29 - 4141,04]	Alto
IPE	11,42	[0 - 0,35]	Muito Alto
IPE	9,93	[0,35 - 1,53]	Muito Alto
IPE	4,43	[1,53 - 3,98]	Muito Alto
IPE	1,62	[3,98 - 10,29]	Muito Alto
IPE	0,81	[10,29 - 4141,04]	Muito Alto
ASP	5,89	[0 - 42,13]	Muito Baixo
ASP	10,81	[42,13 - 114,62]	Muito Baixo
ASP	11,27	[114,62 - 202,4]	Muito Baixo
ASP	12,1	[202,4 - 286,05]	Muito Baixo
ASP	10,42	[286,05 - 360]	Muito Baixo
ASP	2,5	[0 - 42,13]	Baixo
ASP	1,29	[42,13 - 114,62]	Baixo
ASP	1,14	[114,62 - 202,4]	Baixo
ASP	1,33	[202,4 - 286,05]	Baixo
ASP	1,38	[286,05 - 360]	Baixo
ASP	1,75	[0 - 42,13]	Médio
ASP	1,22	[42,13 - 114,62]	Médio
ASP	1,12	[114,62 - 202,4]	Médio
ASP	1,09	[202,4 - 286,05]	Médio
ASP	1,3	[286,05 - 360]	Médio
ASP	1,64	[0 - 42,13]	Alto
ASP	1,49	[42,13 - 114,62]	Alto
ASP	1,42	[114,62 - 202,4]	Alto
ASP	1,14	[202,4 - 286,05]	Alto
ASP	1,52	[286,05 - 360]	Alto
ASP	8,21	[0 - 42,13]	Muito Alto
ASP	5,18	[42,13 - 114,62]	Muito Alto

Fonte: Elaborada pelo autor (2025).

Tabela 9 - Tabela de contingência

(continuação)

Variável	Percentual	Intervalo	Nível de Suscetibilidade
ASP	5,04	[114,62 - 202,4]	Muito Alto
ASP	4,35	[202,4 - 286,05]	Muito Alto
ASP	5,41	[286,05 - 360]	Muito Alto
PREC	13,58	[1789,73 - 1834,62]	Muito Baixo
PREC	9,51	[1834,62 - 1878,14]	Muito Baixo
PREC	5,86	[1878,14 - 1921,7]	Muito Baixo
PREC	10,88	[1921,7 - 1982,31]	Muito Baixo
PREC	10,66	[1982,31 - 2088,12]	Muito Baixo
PREC	1,67	[1789,73 - 1834,62]	Baixo
PREC	1,47	[1834,62 - 1878,14]	Baixo
PREC	1,18	[1878,14 - 1921,7]	Baixo
PREC	1,23	[1921,7 - 1982,31]	Baixo
PREC	2,08	[1982,31 - 2088,12]	Baixo
PREC	1,19	[1789,73 - 1834,62]	Médio
PREC	1,35	[1834,62 - 1878,14]	Médio
PREC	1,27	[1878,14 - 1921,7]	Médio
PREC	1,29	[1921,7 - 1982,31]	Médio
PREC	1,38	[1982,31 - 2088,12]	Médio
PREC	1,17	[1789,73 - 1834,62]	Alto
PREC	1,34	[1834,62 - 1878,14]	Alto
PREC	1,8	[1878,14 - 1921,7]	Alto
PREC	1,61	[1921,7 - 1982,31]	Alto
PREC	1,29	[1982,31 - 2088,12]	Alto
PREC	2,41	[1789,73 - 1834,62]	Muito Alto
PREC	6,39	[1834,62 - 1878,14]	Muito Alto
PREC	9,97	[1878,14 - 1921,7]	Muito Alto
PREC	4,98	[1921,7 - 1982,31]	Muito Alto
PREC	4,45	[1982,31 - 2088,12]	Muito Alto
CUR_PER	16,9	[-0,12 - 0]	Muito Baixo
CUR_PER	8,47	[0 - 0]	Muito Baixo
CUR_PER	1,06	[0 - 0]	Muito Baixo
CUR_PER	6,93	[0 - 0]	Muito Baixo
CUR_PER	17,12	[0 - 0,09]	Muito Baixo
CUR_PER	1,03	[-0,12 - 0]	Baixo
CUR_PER	1,48	[0 - 0]	Baixo
CUR_PER	2,72	[0 - 0]	Baixo
CUR_PER	1,45	[0 - 0]	Baixo
CUR_PER	0,96	[0 - 0,09]	Baixo
CUR_PER	0,7	[-0,12 - 0]	Médio
CUR_PER	1,47	[0 - 0]	Médio

Fonte: Elaborada pelo autor (2025).

Tabela 9 - Tabela de contingência

(continuação)

Variável	Percentual	Intervalo	Nível de Suscetibilidade
CUR_PER	2,09	[0 - 0]	Médio
CUR_PER	1,62	[0 - 0]	Médio
CUR_PER	0,6	[0 - 0,09]	Médio
CUR_PER	0,56	[-0,12 - 0]	Alto
CUR_PER	1,91	[0 - 0]	Alto
CUR_PER	2,11	[0 - 0]	Alto
CUR_PER	2,15	[0 - 0]	Alto
CUR_PER	0,48	[0 - 0,09]	Alto
CUR_PER	0,8	[-0,12 - 0]	Muito Alto
CUR_PER	6,7	[0 - 0]	Muito Alto
CUR_PER	12,06	[0 - 0]	Muito Alto
CUR_PER	7,87	[0 - 0]	Muito Alto
CUR_PER	0,77	[0 - 0,09]	Muito Alto
CUR_PLAN	13,22	[-367,42 - -0,01]	Muito Baixo
CUR_PLAN	10,62	[-0,01 - 0]	Muito Baixo
CUR_PLAN	0,22	[0 - 0]	Muito Baixo
CUR_PLAN	11,36	[0 - 0,02]	Muito Baixo
CUR_PLAN	15,07	[0,02 - 43,61]	Muito Baixo
CUR_PLAN	1,25	[-367,42 - -0,01]	Baixo
CUR_PLAN	4,28	[-0,01 - 0]	Baixo
CUR_PLAN	0,13	[0 - 0]	Baixo
CUR_PLAN	1,07	[0 - 0,02]	Baixo
CUR_PLAN	0,91	[0,02 - 43,61]	Baixo
CUR_PLAN	1,06	[-367,42 - -0,01]	Médio
CUR_PLAN	3,31	[-0,01 - 0]	Médio
CUR_PLAN	0,21	[0 - 0]	Médio
CUR_PLAN	1,13	[0 - 0,02]	Médio
CUR_PLAN	0,77	[0,02 - 43,61]	Médio
CUR_PLAN	1,2	[-367,42 - -0,01]	Alto
CUR_PLAN	3,36	[-0,01 - 0]	Alto
CUR_PLAN	0,32	[0 - 0]	Alto
CUR_PLAN	1,44	[0 - 0,02]	Alto
CUR_PLAN	0,89	[0,02 - 43,61]	Alto
CUR_PLAN	3,31	[-367,42 - -0,01]	Muito Alto
CUR_PLAN	15,51	[-0,01 - 0]	Muito Alto
CUR_PLAN	2,08	[0 - 0]	Muito Alto
CUR_PLAN	5,01	[0 - 0,02]	Muito Alto
CUR_PLAN	2,29	[0,02 - 43,61]	Muito Alto

Fonte: Elaborada pelo autor (2025).

Tabela 9 - Tabela de contingência

(continuação)

Variável	Percentual	Intervalo	Nível de Suscetibilidade
TSO	19,86	CAMBISSOLO HAPLICO	Muito Baixo
TSO	0,22	GLEISSOLO HAPLICO	Muito Baixo
TSO	15,06	CAMBISSOLO FLUVICO	Muito Baixo
TSO	4	NEOSSOLO LITOLICO	Muito Baixo
TSO	7,66	ARGISSOLO AMARELO	Muito Baixo
TSO	0	SOLOS MANGUE	Muito Baixo
TSO	1,54	CAMBISSOLO HAPLICO	Baixo
TSO	0,13	GLEISSOLO HAPLICO	Baixo
TSO	5,35	CAMBISSOLO FLUVICO	Baixo
TSO	0,05	NEOSSOLO LITOLICO	Baixo
TSO	0,26	ARGISSOLO AMARELO	Baixo
TSO	0,01	SOLOS MANGUE	Baixo
TSO	0,51	CAMBISSOLO HAPLICO	Médio
TSO	0,22	GLEISSOLO HAPLICO	Médio
TSO	5,38	CAMBISSOLO FLUVICO	Médio
TSO	0,06	NEOSSOLO LITOLICO	Médio
TSO	0,13	ARGISSOLO AMARELO	Médio
TSO	0,02	SOLOS MANGUE	Médio
TSO	0,2	CAMBISSOLO HAPLICO	Alto
TSO	0,28	GLEISSOLO HAPLICO	Alto
TSO	6,49	CAMBISSOLO FLUVICO	Alto
TSO	0,03	NEOSSOLO LITOLICO	Alto
TSO	0,06	ARGISSOLO AMARELO	Alto
TSO	0,06	SOLOS MANGUE	Alto
TSO	0,04	CAMBISSOLO HAPLICO	Muito Alto
TSO	1	GLEISSOLO HAPLICO	Muito Alto
TSO	19,82	CAMBISSOLO FLUVICO	Muito Alto
TSO	0,01	NEOSSOLO LITOLICO	Muito Alto
TSO	0,01	ARGISSOLO AMARELO	Muito Alto
TSO	6,95	SOLOS MANGUE	Muito Alto
MARE	50,48	Sem influência - Maré (0)	Muito Baixo
MARE	7,62	Sem influência - Maré (0)	Baixo
MARE	0,01	Com influência - Maré (1)	Baixo
MARE	6,46	Sem influência - Maré (0)	Médio
MARE	0,02	Com influência - Maré (1)	Médio
MARE	7,11	Sem influência - Maré (0)	Alto
MARE	0,1	Com influência - Maré (1)	Alto
MARE	19,61	Sem influência - Maré (0)	Muito Alto
MARE	8,59	Com influência - Maré (1)	Muito Alto

Fonte: Elaborada pelo autor (2025).

Tabela 9 - Tabela de contingência

(conclusão)

Variável	Percentual	Intervalo	Nível de Suscetibilidade
UOS	0,02	UOS_1	Muito Baixo
UOS	19,71	UOS_2	Muito Baixo
UOS	2,56	UOS_3	Muito Baixo
UOS	0,5	UOS_5	Muito Baixo
UOS	27,7	UOS_6	Muito Baixo
UOS	0	UOS_1	Baixo
UOS	0,83	UOS_2	Baixo
UOS	0,38	UOS_3	Baixo
UOS	0,08	UOS_5	Baixo
UOS	6,34	UOS_6	Baixo
UOS	0,02	UOS_1	Médio
UOS	0,63	UOS_2	Médio
UOS	0,28	UOS_3	Médio
UOS	0,03	UOS_5	Médio
UOS	5,53	UOS_6	Médio
UOS	0,03	UOS_1	Alto
UOS	0,37	UOS_2	Alto
UOS	0,33	UOS_3	Alto
UOS	0,02	UOS_5	Alto
UOS	6,45	UOS_6	Alto
UOS	0,63	UOS_1	Muito Alto
UOS	2,78	UOS_2	Muito Alto
UOS	0,85	UOS_3	Muito Alto
UOS	0,02	UOS_5	Muito Alto
UOS	23,93	UOS_6	Muito Alto

Fonte: Elaborada pelo autor (2025).

APÊNDICE D – CÓDIGO DE PROCESSAMENTO DO MODELO

Este apêndice apresenta o código computacional utilizado para implementar o modelo de aprendizado de máquina para gerar os mapas de suscetibilidade à inundação desenvolvido neste estudo, conforme estabelecido na metodologia da pesquisa. O código contempla as etapas de treinamento e predição dos modelos, garantindo a reprodutibilidade dos resultados obtidos. Para maior transparência, o *script* completo com os parâmetros de configuração, está disponível publicamente no repositório GitHub:

- <https://github.com/nionmaron/FloodSusceptibilityML-Dissertation>.

***Script* de aprendizado de máquina para gerar mapas de suscetibilidade à inundação em linguagem de programação R**

```
001: #####
002: # script: mapas de suscetibilidade por aprendizado de máquina
003: # descrição: análise da suscetibilidade à inundação
004: #
005: # autor: Nion Maron Dransfeld
006: # data: 2024-10-12
007: # versão: 1.0
008: #####
009:
010: # -----
011: # 1. preparação do ambiente
012: # -----
013: # limpa todos os objetos do ambiente de trabalho
014: rm(list = ls(all.names = TRUE))
015:
016: # (opcional) define o diretório de trabalho
017: # setwd("nion/diretorio")
018:
019: # exibe o diretório atual
020: getwd()
021:
022: # -----
023: # 2. carregamento dos pacotes necessários
024: # -----
025: {
026:   # 1. importação e manipulação de dados
027:   library(readxl) # leitura de arquivos excel
028:   library(tidyverse) # conjunto de funções diversas
029:   library(tidyr) # manipulação de dados (organização e reshape)
030:   library(knitr) # geração de relatórios dinâmicos
031:   library(devtools) # ferramentas de desenvolvimento
032:   library(units) # trabalhar com unidades de medida
033: }
```

```

034: # 2. dados espaciais e mapeamento
035: library(raster) # leitura e manipulação de imagens raster
036: library(sf) # manipulação de dados espaciais (shapefiles, vetores)
037: library(tmap) # criação de mapas temáticos interativos e estáticos
038: library(leaflet) # criação de mapas interativos
039: library(mapview) # visualização interativa de mapas
040: library(ggspatial) # adição de elementos espaciais
041: library(rasterVis) # visualização de dados raster
042:
043: # 3. visualização e plotagem de gráficos
044: library(ggplot2) # criação de gráficos estatísticos
045: library(cowplot) # arranjo de múltiplos gráficos em grid
046: library(ggpubr) # combinação e aprimoramento de gráficos ggplot
047: library(grid) # estruturação de layouts gráficos
048: library(gridExtra) # arranjo adicional de gráficos
049:
050: # 4. paletas de cores
051: library(viridis) # paletas de cores com alta perceptibilidade
052: library(RColorBrewer) # conjunto de paletas de cores
053:
054: # 5. análise de correlação
055: library(corrplot) # visualização de matrizes de correlação
056: library(corr) # análise de rede de correlação
057:
058: # 6. machine learning e modelagem
059: library(caret) # algoritmos e funções para modelagem preditiva
060: library(CAST) # ferramentas específicas para modelagem
061: library(caretEnsemble) # combinação de múltiplos modelos de ml
062: library(pROC) # geração de curvas roc para avaliação de modelos
063:
064: # 7. diversos
065: library(climateStability) # contém função rescale0to1
066: }
067:
068: # -----
069: # 3. leitura e preparação dos dados espaciais
070: # -----
071: # leitura de limites administrativos (joinville/sc - brasil)
072: limite_municipal <- st_read("database/shapefiles/Limite_Municipal.shp")
073:
074: # leitura dos limites de bacias hidrográficas
075: limite_bacias <- st_read("database/shapefiles/Dados_Bacias.shp")
076:
077: # filtra a bacia hidrográfica do rio cachoeira
078: limite_bacia_cachoeira <- limite_bacias[limite_bacias$NOME == "BACIA
CACHOEIRA", ]
079:
080: # leitura da hidrografia principal
081: rios_principais <- st_read("database/shapefiles/Rios_Principais.shp")
082:
083: # -----
084: # 4. leitura das imagens raster (variáveis independentes e dependentes)
085: # -----
086: # todas as imagens devem estar em epsg:31982 (atenção à compatibilidade)
087:
088: # 4.1. modelo digital de elevação
089: FATOR_01 <- raster("database/independent_variables/01_BHRC_MDE.tif")
090: names(FATOR_01) <- c("MDE")
091:

```

```

092: # 4.2. cadastro da amostra final em raster (variável dependente: 0 e 1)
093: INVENTARIO <- raster("database/dependent_variable/BHRC-flood-inventory.tif")
094: names(INVENTARIO) <- c("AMT")
095: INVENTARIO <- resample(INVENTARIO, FATOR_01, method='bilinear')
096:
097: # 4.3. raster com registros para controlar os pontos
098: REG <- raster("database/dependent_variable/BHRC-reg-inventory.tif")
099: names(REG) <- c("REG")
100: REG <- resample(REG, FATOR_01, method='bilinear')
101:
102: # 4.4. distanciamento dos rios
103: FATOR_02 <- raster("database/independent_variables/02_BHRC_DCH.tif")
104: names(FATOR_02) <- c("DCH")
105:
106: # 4.5. distanciamento da malha viária
107: FATOR_03 <- raster("database/independent_variables/03_BHRC_DMV.tif")
108: names(FATOR_03) <- c("DMV")
109:
110: # 4.6. índice de vegetação por diferença normalizada
111: FATOR_04 <- raster("database/independent_variables/04_BHRC_IVDN.tif")
112: names(FATOR_04) <- c("IVDN")
113: FATOR_04 <- resample(FATOR_04, FATOR_01, method='bilinear')
114:
115: # 4.7. inclinação (declividade) do terreno
116: FATOR_05 <- raster("database/independent_variables/05__BHRC_DEC.tif")
117: names(FATOR_05) <- c("DEC")
118:
119: # 4.8. rugosidade do terreno
120: FATOR_06 <- raster("database/independent_variables/06_BHRC_RUG.tif")
121: names(FATOR_06) <- c("RUG")
122:
123: # 4.9. índice de umidade topográfico (iut)
124: FATOR_07 <- raster("database/independent_variables/07_BHRC_IUT.tif")
125: names(FATOR_07) <- c("IUT")
126:
127: # 4.10. índice de potência de escoamento (ipe)
128: FATOR_08 <- raster("database/independent_variables/08_BHRC_IPE.tif")
129: names(FATOR_08) <- c("IPE")
130:
131: # 4.11. aspecto do terreno
132: FATOR_09 <- raster("database/independent_variables/09_BHRC_ASP.tif")
133: names(FATOR_09) <- c("ASP")
134:
135: # 4.12. precipitação
136: FATOR_10 <- raster("database/independent_variables/10_BHRC_PREC.tif")
137: names(FATOR_10) <- c("PREC")
138: FATOR_10 <- resample(FATOR_10, FATOR_01, method='bilinear')
139:
140: # 4.13. curvatura do perfil
141: FATOR_11 <- raster("database/independent_variables/11_BHRC_CUR_PER_.tif")
142: names(FATOR_11) <- c("CUR_PER")
143:
144: # 4.14. curvatura do plano
145: FATOR_12 <- raster("database/independent_variables/12_BHRC_CUR_PLAN.tif")
146: names(FATOR_12) <- c("CUR_PLAN")
147:
148: # 4.15. tipo de solo
149: FATOR_13 <- raster("database/independent_variables/13_BHRC_TSO.tif")
150: names(FATOR_13) <- c("TSO")

```

```

151:
152: # 4.16. manchas de maré de 2,5 metros (plano diretor de drenagem urbana -
153: # 2011)
154: FATOR_14 <- raster("database/independent_variables/14_BHRC_MARE.tif")
155: names(FATOR_14) <- c("MARE")
156:
157: # 4.17. uso e ocupação do solo (uos)
158: FATOR_15 <- raster("database/independent_variables/15_BHRC_UOS.tif")
159: names(FATOR_15) <- c("UOS")
160:
161: # -----
162: # 5. empilhamento das variáveis em objetos raster
163: # -----
164: FATORES <- stack(INVENTARIO, REG,
165:                 FATOR_01, FATOR_02, FATOR_03, FATOR_04, FATOR_05,
166:                 FATOR_06, FATOR_07, FATOR_08, FATOR_09, FATOR_10,
167:                 FATOR_11, FATOR_12, FATOR_13, FATOR_14, FATOR_15)
168:
169: FATORES_P <- stack(FATOR_01, FATOR_02, FATOR_03, FATOR_04, FATOR_05,
170:                  FATOR_06, FATOR_07, FATOR_08, FATOR_09, FATOR_10,
171:                  FATOR_11, FATOR_12, FATOR_13, FATOR_14, FATOR_15)
172:
173: # exibição rápida das variáveis (fatores)
174: par(mfrow = c(3,5))
175: plot(FATORES_P)
176:
177: # -----
178: # 6. preparação dos dados de treino e teste para o modelo de classificação
179: # -----
180: # converte o stack raster em data frame, incluindo coordenadas xy
181: FATORES_DF <- as.data.frame(FATORES, xy = TRUE, optional = FALSE)
182:
183: # a variável "AMT" será um fator binário para o modelo (inundável ou não)
184: FATORES_DF$AMT <- as.factor(FATORES_DF$AMT)
185: FATORES_DF$REG <- as.factor(FATORES_DF$REG)
186:
187: # renomeia os níveis do fator para melhor compreensão
188: levels(FATORES_DF$AMT) <- c("NAO_INUNDAVEL", "INUNDAVEL")
189:
190: # remove valores NA (pixels sem informação)
191: DADOS_LIMPO <- na.omit(FATORES_DF)
192:
193: # arredonda valores numéricos para reduzir tamanho e otimizar processamento
194: DADOS_LIMPO <- data.frame(lapply(DADOS_LIMPO, function(y) {
195:   if (is.numeric(y)) round(y, 2) else y
196: })))
197:
198: # -----
199: # 7. divisão dos dados entre não inundáveis e inundáveis
200: # -----
201: # semente para reprodutibilidade
202: set.seed(124)
203:
204: # lista de registros do inventário
205: LISTA_PONTOS <- as.numeric(levels(DADOS_LIMPO$REG))
206:
207: # separa registros de não inundação (ni)

```

```

207: PONTOS_NI      <-
sample(as.numeric(levels(factor(DADOS_LIMPO$REG[DADOS_LIMPO$AMT ==
"NAO_INUNDAVEL"]))))))
208: DIV_AMOSTRA00  <- round(length(PONTOS_NI)*0.7, 0)
209: PONTOS_NI_TREINO <- (PONTOS_NI[1:DIV_AMOSTRA00])
210: PONTOS_NI_TESTE  <- PONTOS_NI[(DIV_AMOSTRA00+1):length(PONTOS_NI)]
211:
212: # separa registros de inundação (i)
213: PONTOS_I        <-
sample(as.numeric(levels(factor(DADOS_LIMPO$REG[DADOS_LIMPO$AMT ==
"INUNDAVEL"]))))))
214: DIV_AMOSTRA01   <- round(length(PONTOS_I)*0.7, 0)
215: PONTOS_I_TREINO  <- (PONTOS_I[1:DIV_AMOSTRA01])
216: PONTOS_I_TESTE   <- (PONTOS_I[(DIV_AMOSTRA01+1):length(PONTOS_I)])
217:
218: # -----
219: # 8. criação das amostras de treino e teste
220: # -----
221: treino <- c(PONTOS_NI_TREINO, PONTOS_I_TREINO)
222: for (i in seq_along(treino)) {
223:   treino00 <- DADOS_LIMPO[DADOS_LIMPO$REG == treino[i], ]
224:   if (i == 1) {
225:     AMOSTRA_TREINO <- treino00
226:   } else {
227:     AMOSTRA_TREINO <- rbind(AMOSTRA_TREINO, treino00)
228:   }
229: }
230:
231: teste <- c(PONTOS_NI_TESTE, PONTOS_I_TESTE)
232: for (i in seq_along(teste)) {
233:   teste00 <- DADOS_LIMPO[DADOS_LIMPO$REG == teste[i], ]
234:   if (i == 1) {
235:     AMOSTRA_TESTE <- teste00
236:   } else {
237:     AMOSTRA_TESTE <- rbind(AMOSTRA_TESTE, teste00)
238:   }
239: }
240:
241: # define os preditores (fatores independentes que entrarão no modelo)
242: PREDITORES <- c("MDE", "DCH", "DMV", "IVDN", "DEC",
243:                 "IUT", "IPE", "ASP", "PREC", "CUR_PER",
244:                 "CUR_PLAN", "TSO", "MARE", "UOS")
245:
246: print(paste("o modelo terá as seguintes variáveis: ",
247:             paste(PREDITORES, collapse = ", ")))
248:
249: # -----
250: # 9. definição dos algoritmos de modelagem escolhidos (machine learning)
251: # -----
252: LISTA_MODELOS <- c("rf",
253:                   "svmRadial",
254:                   "svmLinear",
255:                   "svmPoly")
256:
257: # -----
258: # 10. parâmetros iniciais para o treinamento dos modelos
259: # -----
260: # criação de folds no espaço (divisão por regiões - 'REG')
261: indices <- CreateSpacetimeFolds(AMOSTRA_TREINO,

```

```

262:                                     spacevar = "REG",
263:                                     k = 6)
264:
265: # parâmetros de treinamento (cross-validation repetida)
266: trainControl <- trainControl(
267:   method      = "repeatedcv",
268:   number      = 20,
269:   repeats     = 5,
270:   classProbs  = TRUE,
271:   index       = indices$index,
272:   summaryFunction= twoClassSummary,
273:   savePredictions= TRUE # necessário para montagem de curva roc
274: )
275:
276: # -----
277: # 11. treinamento dos modelos (loop)
278: # -----
279: # atenção: este processo pode demandar bastante tempo de processamento
280: for (MODELO in LISTA_MODELOS) {
281:   T01 <- Sys.time()
282:   set.seed(1234)
283:
284:   # cria nome para salvar o modelo
285:   objmodel <- paste(Sys.Date(), "MODEL", MODELO, sep = "_")
286:   print(paste("início do processamento:", objmodel))
287:
288:   # treinamento do modelo
289:   MODELO_00 <- train(
290:     AMOSTRA_TREINO[, PREDITORES],
291:     AMOSTRA_TREINO$AMT,
292:     metric      = "ROC",
293:     preProcess  = c("center", "scale"),
294:     method      = MODELO,
295:     trControl   = trainControl
296:   )
297:
298:   # salva o modelo no diretório indicado
299:   saveRDS(MODELO_00, file = paste0("database/model_training/", objmodel,
300:   ".rds"), version = 3)
301:
302:   # limpeza e cálculo do tempo
303:   rm(MODELO_00)
304:   TEMPO_PROCESSAMENTO <- (Sys.time() - T01)
305:   print(paste("fim:", objmodel,
306:   "tempo de processamento:",
307:   round(TEMPO_PROCESSAMENTO, 2),
308:   "segundos"))
309: }
310: # -----
311: # 12. leitura e comparação dos modelos gerados
312: # -----
313: # para ler os modelos e compará-los (roc, accurácia, etc.)
314: LEITURA_LIST <- LISTA_MODELOS
315: AMD <- "2025-03-12" # data do modelo gerado (ajustar conforme necessário)
316: n <- 0
317:
318: for (MODELO in LEITURA_LIST) {
319:   n <- n + 1

```

```

320: objmodel <- paste(AMD, "MODEL", MODELO, sep = "_")
321: nome      <- paste("MODELO", n, sep = "_")
322:
323: assign(nome, readRDS(paste0("database/model_training/",
324:                             paste(objmodel, "rds", sep = "."))))
325: }
326:
327: # exemplo de conferência do arquivo carregado
328: paste0("database/model_training/",
329:        paste(objmodel, "rds", sep = "."))
330:
331: # -----
332: # 13. previsão dos mapas de suscetibilidade (exemplo FA)
333: # -----
334: # no exemplo, utilizamos o modelo_1 (rf) para gerar o mapa
335: # index=2 retorna a probabilidade de 'inundável' (classe positiva)
336:
337: PRE_MOD_01 <- raster::predict(FATORES_P, MODELO_1, type = "prob", index = 2)
338:
339: # salva o mapa previsto (opcional)
340: # writeRaster(PRE_MOD_01, "database/processing_results/MSI-AM-FA.tif",
341: # overwrite = TRUE)
342:
342: # plota o raster resultante
343: plot(PRE_MOD_01)
344:
345: # renomeia para facilitar a visualização posterior
346: MSI_RF <- PRE_MOD_01
347:
348: # -----
349: # 14. criação de mapa temático de suscetibilidade
350: # -----
351: tm_shape(MSI_RF) +
352:   tm_raster(n = 5,
353:             title = "probabilidade de inundação",
354:             palette = colorRampPalette(c("aquamarine", "darkorange",
355:             "blue"))(5),
356:             legend.hist = TRUE) +
357:   tm_legend(outside = TRUE, hist.width = 3) +
358:   tm_grid(labels.inside.frame = FALSE, n.x = 4, n.y = 5,
359:           projection = "+proj=longlat") +
360:   tm_scale_bar(position = c("left", "bottom")) +
361:   tm_compass(type = "8star")
362:
363: # -----
364: # 15. cálculo simplificado da área por classe de suscetibilidade
365: # -----
366: AREA_PIXEL <- res(MSI_RF)[1] * res(MSI_RF)[2] / 1000000 # área de cada pixel
367: em km²
368:
369: # total de células válidas
370: TC <- ncell(MSI_RF[MSI_RF >= 0]) * AREA_PIXEL
371:
372: # contagem para diferentes faixas de probabilidade
373: MBP <- ncell(MSI_RF[MSI_RF >= 0.0]) * AREA_PIXEL # muito baixa
374: BP <- ncell(MSI_RF[MSI_RF >= 0.2]) * AREA_PIXEL # baixa
375: MP <- ncell(MSI_RF[MSI_RF >= 0.4]) * AREA_PIXEL # média
376: AP <- ncell(MSI_RF[MSI_RF >= 0.6]) * AREA_PIXEL # alta
377: MAP <- ncell(MSI_RF[MSI_RF >= 0.8]) * AREA_PIXEL # muito alta

```

```

376:
377: # cria data frame com a classificação e área correspondente
378: CLASS_DF <- data.frame(
379:   classificacao = c("muito baixa", "baixa", "média", "alta", "muito alta"),
380:   area_km2      = c((MBP - BP), (BP - MP), (MP - AP), (AP - MAP), MAP)
381: )
382: CLASS_DF$area_km2 <- round(CLASS_DF$area_km2, 2)
383:
384: CLASS_DF
385:
386: # -----
387: # 16. mapa interativo MSI
388: # -----
389: mapview(MSI_RF,
390:         alpha.regions = 0.6,
391:         map.types = "OpenStreetMap",
392:         legend = FALSE)
393:
394: # fim do script
395: # -----
396:
397:

```