



UNIVERSIDADE ESTADUAL PAULISTA
“JÚLIO DE MESQUITA FILHO”
Campus de São José dos Campos
Instituto de Ciência e Tecnologia



Cemaden
Centro Nacional de Monitoramento
e Alertas de Desastres Naturais

UNIDADE DE PESQUISA DO

MINISTÉRIO DA
CIÊNCIA, TECNOLOGIA,
INOVAÇÕES E COMUNICAÇÕES



PÁTRIA AMADA
BRASIL
GOVERNO FEDERAL

TATIANY CORREIA DA SILVA SANTOS

INTEGRAÇÃO DE ANÁLISES GEOESPACIAIS E O
MÉTODO FLORESTA ALEATÓRIA PARA A ANÁLISE DE
SUSCETIBILIDADE A ESCORREGAMENTOS: estudo de caso
na Bacia do Córrego Piracuama, município de Campos do Jordão
(SP)

2021

TATIANY CORREIA DA SILVA SANTOS

**INTEGRAÇÃO DE ANÁLISES GEOESPACIAIS E O MÉTODO
FLORESTA ALEATÓRIA PARA A ANÁLISE DA SUSCETIBILIDADE A
ESCORREGAMENTOS: estudo de caso na Bacia do Córrego Piracuama, município
de Campos do Jordão (SP)**

Dissertação apresentada ao Instituto de
Ciência e Tecnologia, Universidade Estadual
Paulista (Unesp), Campus de São José dos Campos;
Centro Nacional de Monitoramento e Alertas de
Desastres Naturais (Cemaden), como parte dos
requisitos para obtenção do título de MESTRE pelo
Programa de Pós-Graduação em DESASTRES
NATURAIS.

Área: Desastres Naturais. Linha de pesquisa:
Desastres associados a eventos extremos, inundações
e movimentos de massa.

Orientador: Prof. Dr. Silvio Jorge Coelho Simões
Coorientadora: Prof.^a Dra. Tatiana Sussel Gonçalves Mendes

São José dos Campos

2021

BANCA EXAMINADORA

Prof. Dr. Silvio Jorge Coelho Simões (Orientador)

Universidade Estadual Paulista (Unesp)

Instituto de Ciência e Tecnologia

Campus de São José dos Campos

Dra. Alessandra Cristina Corsi

Instituto de Pesquisas Tecnológicas - IPT

Seção de Investigações, Riscos e Gerenciamento Ambiental – SIRGA

Prof. Dr. Rogério Galante Negri

Universidade Estadual Paulista (Unesp)

Instituto de Ciência e Tecnologia

Campus de São José dos Campos

São José dos Campos, 22 de setembro de 2021.

*Aos meus pais, Sergio e Joyce,
meus maiores e melhores orientadores na vida.*

AGRADECIMENTOS

Uma recordação marcante que tenho do mestrado é sobre uma fala de um professor, que nos disse, no nosso primeiro dia de aula, que nosso objetivo ali era fazer ciência. E, no cenário que estamos vivendo, diante de tantas adversidades políticas, faltas de incentivos e desvalorização dessa base extremamente importante, agradeço às pessoas e às instituições que continuam a lutar, bravamente, pela ciência.

Aos meus amados pais, meus maiores exemplos, que sempre foram meu combustível e força para lutar por meus objetivos. Obrigada por todo apoio incondicional durante toda a vida! Ao meu irmão, Rafa, que mesmo tão distante e diante de tantas incertezas, sempre me instigou a ir além e a pensar fora da caixa.

Ao Mateus, que elevou a minha definição da palavra companheirismo. Que em todos os momentos foi meu porto seguro, meu incentivador, minha calma e a melhor pessoa que eu poderia escolher para ter ao meu lado.

Aos meus queridos orientadores, Silvio e Tati, grandes exemplos de professores e seres humanos, que toparam a temática desafiadora desse trabalho junto comigo. Agradeço por todas as discussões, reflexões, ensinamentos, pela confiança que depositaram em mim e por estarem presentes em todas as etapas dessa jornada.

A Jennifer e Felícia, grandes parceiras de projeto, pelos desafios compartilhados, pelas inúmeras discussões, vivências e por todo o aprendizado. Foi um imenso prazer trabalhar com mulheres tão inteligentes e empoderadas como vocês.

Ao Professor Rogério, por todos os ensinamentos e por mostrar, de forma tão leve, que é possível desbravar o universo da programação.

A Imagem, que suportou e apoiou a minha decisão de fazer o mestrado. A todos os profissionais que dividiram essa etapa comigo, em especial ao Cadu, pelas ideias, suporte e por compartilhar o seu grande conhecimento.

Aos amigos, por toda força, por terem celebrado comigo todas as etapas, pelo apoio nos momentos de dificuldade, pela leveza nas horas de desconcentração e por entenderem a minha ausência em tantos momentos.

Com muita satisfação encerro essa trajetória tão importante na minha vida que, ao mesmo tempo que foi tão desafiadora, brilha os olhos e traz uma sensação de missão cumprida quando penso em cada etapa vivenciada até aqui. A todos que contribuíram para a minha formação pessoal e profissional, muito obrigada!

“A principal causa do sofrimento é o desejo egoísta do próprio conforto e felicidade.”

Dalai-Lama

SUMÁRIO

RESUMO	13
ABSTRACT	14
1 INTRODUÇÃO	15
1.1 Estrutura da dissertação.....	17
2 FUNDAMENTAÇÃO TEÓRICA	19
2.1 Desastres.....	19
2.1.1 Desastres no mundo	19
2.1.2 Desastres no Brasil	21
2.2 Movimentos de massa	25
2.2.1 Escorregamentos	26
2.2.1.1 Classificação dos escorregamentos	27
2.2.1.2 Fatores condicionantes dos escorregamentos	28
2.2.1.3 Modelos de suscetibilidade a escorregamentos.....	31
2.3 Aprendizado de máquina.....	32
2.3.1 Floresta Aleatória.....	34
3 PROPOSTA DA PESQUISA	38
4 MATERIAIS E MÉTODOS.....	39
4.1 Etapa 1.....	40
4.1.1 Seleção da área de estudo	40
4.1.2 Caracterização da área de estudo	41
4.1.2.1 A Bacia do Córrego Piracuama	41
4.1.2.2 Histórico de movimentos de massa	47
4.1.2.3 Histórico de ocupação	49
4.1.2.4 Clima	49
4.1.2.5 Geologia, pedologia e geomorfologia	50
4.1.2.6 Instrumentações de gestão de risco.....	56
4.1.3 Trabalho de campo.....	58
4.2 Etapa 2.....	62
4.2.1 Levantamento de cicatrizes	63
4.2.2 Levantamento dos fatores condicionantes	64
4.2.2.1 Parâmetros morfométricos.....	65

4.2.2.2 Parâmetros Temáticos	73
4.2.3 Razão de Frequência	76
4.3 Etapa 3.....	78
4.3.1 Definição das amostras de treino e teste.....	79
4.3.2 Identificação da correlação entre as variáveis.....	82
4.3.3 Análise da importância das variáveis condicionantes.....	82
4.3.4 Aplicação do método Floresta Aleatória	84
4.3.5 Métricas de performance.....	85
4.3.6 Geração dos mapas de suscetibilidade	87
5 RESULTADOS.....	88
5.1 Razão da Frequência.....	88
5.2 Identificação da correlação entre as variáveis.....	93
5.3 Importância dos atributos	94
5.4 Construção e aplicação do modelo de predição.....	97
5.4.1. Busca em grade.....	97
5.4.2 Avaliação de desempenho.....	98
5.5 Geração dos mapas de suscetibilidade	100
6 DISCUSSÕES	107
7 CONCLUSÕES	109
REFERÊNCIAS	111

LISTA DE FIGURAS

Figura 1. Desastres naturais no Brasil (1991-2012).....	22
Figura 2. Perdas humanas e população afetada por tipo de desastres no Brasil (1990-2012). 23	
Figura 3. Mortos e afetados por região brasileira (1991-2012).....	25
Figura 4. Esquema das principais rupturas em escorregamentos.	27
Figura 5. Esquema de um escorregamento planar.....	28
Figura 6. Estrutura do método Floresta Aleatória	35
Figura 7. Criação das árvores de decisão a partir da técnica <i>bootstrapping</i>	35
Figura 8. Utilização dos dados “ <i>out-of-bag</i> ” para validação do modelo	36
Figura 9. Fluxograma da dissertação.....	39
Figura 10. Localização da área de estudo	43
Figura 11 Setores de risco localizados na área de estudo	46
Figura 12. Mapa geológico de Campos de Jordão	53
Figura 13. Mapa geológico de Campos de Jordão	54
Figura 14. Carta de suscetibilidade a movimentos gravitacionais de massa.....	57
Figura 15. Pontos coletados em campo	59
Figura 16. Obras estruturais observadas em campo. a) Canaleta de drenagem. b) Escada hidráulica. c) Muro de contenção com canaletas de drenagem, que auxiliam no escoamento das águas infiltradas no terreno	60
Figura 17. Paleossoma do migmatito observado em campo. Observa-se alteração de coloração avermelhada e presença de foliações, derivadas da natureza gnaissíca	60
Figura 18. Neossoma do migmatito observado em campo. Nota-se o produto de alteração de coloração clara, associado à origem granítica.....	61
Figura 19. A) Superfície de contato entre neossoma e paleossoma. B) Estrutura sigmoidal, relacionada às zonas de cisalhamento e que indica a cinemática do movimento	61
Figura 20. a) Cicatriz de deslizamento observada em campo. Nota-se um anfiteatro na localidade, bem como vegetações pós-evento. b) Moradias de baixo padrão construtivo observadas em campo, apresentando fundação irregular	62
Figura 21. Localização das cicatrizes na área de estudo	64
Figura 22. Parâmetros morfométricos utilizados no modelo de acordo com (a) Altimetria; (b) Aspecto; (c) Curvatura Combinada; (d) Curvatura em Perfil; (e) Curvatura em Planta; (f) Declividade; (g) SPI; (h) STI; (i) TPI; (j) TRI;(k) TWI.....	68

Figura 23. Parâmetros temáticos utilizados no modelo de acordo com: (a) Densidade de Lineamentos; (b) Uso e Cobertura da Terra; (c) Vegetação	74
Figura 24. Separação do conjunto de dados para treinamento, validação e teste	80
Figura 25. Amostras de “escorregamentos” e “não escorregamentos” utilizadas no modelo de predição	81
Figura 26. Matriz de Confusão.....	86
Figura 27. Distribuição da RF normalizada dos parâmetros morfométricos de acordo com (a) Altimetria; (b) Aspecto; (c) Curvatura Combinada; (d) Curvatura em Perfil; (e) Curvatura em Planta; (f) Declividade; (g) SPI; (h) STI; (i) TPI; (j) TRI; (k) TWI.....	89
Figura 28. Distribuição da RF normalizada para os parâmetros temáticos: (a) Densidade de Lineamentos; (b) Uso e Cobertura; (c) Vegetação	92
Figura 29. Matriz de distância com os valores de correlação para os fatores condicionantes .	94
Figura 30. Importância dos fatores condicionantes de acordo com o Ganho de Informação ..	95
Figura 31. Diagrama de caixa para os atributos mais relevantes para o modelo: (a) TRI e (b) Declividade e menos relevantes: (c) Altimetria e (d) Curvatura em Perfil	96
Figura 32. Valores médios de acurácia para cada combinação da busca em grade	99
Figura 33. Curva de (a) taxa de sucesso e (b) predição.....	100
Figura 34. Bacia do Córrego Piracuama classificada em áreas de escorregamento e áreas de não escorregamento.....	101
Figura 35. Mapa de suscetibilidade a escorregamentos para a área de estudo.....	102
Figura 36. Distribuição das classes de suscetibilidade na Bacia do Córrego Piracuama.....	102
Figura 37. Comparação entre o mapa de suscetibilidade e os fatores condicionantes que apresentaram maior influência no modelo de predição.....	103
Figura 38. Comparação entre o (a) modelo final e o (b) mapa de suscetibilidade elaborado por CPRM, 2014.....	105
Figura 39. Comparação entre o (a) modelo final e o (b) mapa de risco elaborado por IG, 2014	106

LISTA DE QUADROS

Quadro 1. Classificação geral de desastres naturais adotada pelo EM-DAT	20
Quadro 2. Fatores condicionantes e suas respectivas relevâncias para a análise de suscetibilidade a escorregamentos	29
Quadro 3. Indicadores de resposta de Campos do Jordão	40
Quadro 4. Indicadores de estado ou situação de Campos do Jordão.....	41
Quadro 5. Descrição dos setores de risco do bairro Vila Albertina	44
Quadro 6. Descrição dos setores de risco do bairro Brancas Nuvens	45
Quadro 7. Registros de eventos ocorridos na Vila Albertina e Brancas Nuvens	47
Quadro 8. Unidades litoestratigráficas que abrangem o município Campos do Jordão	52
Quadro 9. Levantamento das cicatrizes na área de estudo	63
Quadro 10. Parâmetros de entrada utilizados na construção do algoritmo Floresta Aleatória	85
Quadro 11. Combinações realizadas durante a busca em grade e seus respectivos valores médios de acurácia e precisão	98

LISTA DE TABELAS

Tabela 1. FR calculado para cada variável utilizada no modelo de predição	76
Tabela 2. Métricas de desempenho para os dados de treinamento/validação e teste	100
Tabela 3. Relação das classes de suscetibilidade com o inventário de cicatrizes	104
Tabela 4. Relação das classes de suscetibilidade com o inventário de cicatrizes	104

LISTA DE ABREVIATURAS E SIGLAS

Cemaden	Centro Nacional de Monitoramento e Alertas de Desastres Naturais
CPRM	Serviço Geológico do Brasil
CRED	<i>Centre for Research on the Epidemiology of Disasters</i>
EM-DAT	<i>The international disasters database</i>
GIS	<i>Geographic Information System</i>
IBGE	Instituto Brasileiro de Geografia e Estatística
IDF	Intensidade - Duração - Frequência
INPE	Instituto Nacional de Pesquisas Espaciais
IPT	Instituto de Pesquisas Tecnológicas
IG	Instituto Geológico
ONU	Organização das Nações Unidas
SIG	Sistema de Informação Geográfica
SPI	<i>Stream Power Index</i>
SRTM	<i>Shuttle Radar Topography Mission</i>
STI	<i>Sediment Transport Index</i>
TPI	<i>Topographic Position Index</i>
TRI	<i>Topographic Ruggedness Index</i>
UNISDR	<i>United Nations Office for Disaster Risk Reduction</i>

SANTOS, T. C. S. INTEGRAÇÃO DE ANÁLISES GEOESPACIAIS E O MÉTODO FLORESTA ALEATÓRIA PARA A ANÁLISE DE SUSCETIBILIDADE A ESCORREGAMENTOS: estudo de caso na Bacia do Córrego Piracuama, município de Campos do Jordão (SP). 2021. Dissertação (Mestrado em Desastres Naturais) - Universidade Estadual Paulista (Unesp), Instituto de Ciência e Tecnologia; Centro Nacional de Monitoramento e Alertas de Desastres Naturais (Cemaden), São José dos Campos, 2021.

RESUMO

O desafio dos entes federativos brasileiros em gerir e conviver com o risco de desastres é uma pauta cada vez mais constante e necessária na administração pública. A região Sudeste do Brasil está significativamente relacionada à ocorrência de movimentos gravitacionais de massa, que representam o segundo maior tipo de desastre que causa óbitos nessa região. Dessa forma, o desenvolvimento e utilização de métodos que auxiliem na previsão desses processos podem ser grandes aliados no planejamento urbano dos municípios. Dentre os diversos métodos utilizados para mapear a suscetibilidade a escorregamentos, as técnicas baseadas em aprendizado de máquina possuem uma grande vantagem, uma vez que permitem lidar com grande volume e diversidade de dados, além de trabalhar amplamente com fatores que condicionam a deflagração de escorregamentos. A combinação dessas técnicas com ambientes SIG permite otimizar e superar limitações de processamento e espacialização dos dados. Diante de diversos algoritmos de aprendizado de máquina, a Floresta Aleatória vem sendo cada vez mais utilizada nessa temática e apresentando bons resultados para a predição de suscetibilidade. Dessa forma, o objetivo desse trabalho consiste em analisar a suscetibilidade a escorregamentos da Bacia do Córrego Piracuama, localizada no município Campos do Jordão, SP, usando o algoritmo Floresta Aleatória. A partir do inventário de cicatrizes, foram definidas 200 amostras de forma aleatória (100 de ocorrência e 100 de não ocorrência), destinadas ao treino/validação e teste do modelo de predição, em uma proporção de 70:30. Quatorze fatores condicionantes de escorregamentos, divididos em morfométricos e temáticos, foram selecionados como variáveis de entrada para a construção do modelo de predição por Floresta Aleatória e, dentre essas variáveis, o Índice de Rugosidade do Terreno (TRI) e a declividade foram as que apresentaram, respectivamente, as melhores importâncias no modelo. Com a finalidade de alcançar a melhor parametrização do algoritmo, foi aplicado a técnica busca em grade, que identificou que os melhores parâmetros para a Floresta Aleatória foram 300 árvores e 4 variáveis testadas em cada nó. Após a construção do modelo, diversas métricas de avaliação foram utilizadas para analisar a sua performance que, de acordo com os resultados, apresentou uma boa capacidade no mapeamento da suscetibilidade a escorregamentos. Por fim, o mapa de suscetibilidade obtido pelo método proposto foi comparado com mapas já existentes para a área, porém produzidos através de diferentes técnicas de mapeamento. Esse trabalho permitiu identificar que modelagens baseadas em aprendizado de máquina devem ser cada vez mais exploradas na gestão de riscos de desastres, visto que configuram ferramentas importantes na elaboração de estratégias, políticas públicas e conhecimento territorial.

Palavras-chave: Suscetibilidade a escorregamentos. Aprendizado de máquina. Floresta Aleatória. Sistema de Informação Geográfica. Desastres naturais.

SANTOS T. C. S. INTEGRATION OF GEOSPATIAL ANALYSIS AND THE RANDOM FOREST METHOD FOR THE ANALYSIS OF LANDSLIDE SUSCEPTIBILITY: case study in the Piracuama Basin, municipality of Campos do Jordão (SP). 2021. Dissertation (Master's degree in Natural Disaster) - São Paulo State University (Unesp), Institute of Science and Technology, National Center for Monitoring and Early Warning of Natural Disasters (Cemaden), São José dos Campos, 2021.

ABSTRACT

The challenge of Brazilian federative entities in managing and living with the risk of disasters is an increasingly constant and necessary agenda in public administration. The Southeast region of Brazil is related to landslides, which represents the second largest type of disaster that causes death in this region. Thus, the development and use of methods that assist in predicting these processes can be great allies in urban planning in municipalities. Among the various methods used to map the landslide susceptibility, techniques based on machine learning have a great advantage, since they allow dealing with a large volume and diversity of data, in addition to working extensively with factors that affect the occurrence of landslides. The combination of these techniques with GIS environments allows optimizing and overcoming limitations in data processing and spatialization. Faced with several machine learning algorithms, the Random Forest has been increasingly used in this theme and this has good results for the prediction of susceptibility. Thus, the objective of this work is to analyze the susceptibility to landslides in the Córrego Piracuama Basin, located in Campos do Jordão, SP, using the Random Forest algorithm. From the landslide inventory, 200 were randomly defined (100 occurrences and 100 non-occurrences), intended for training and testing of the prediction model, in a ratio of 70:30. Fourteen landslide conditioning factors, divided into morphometric and thematic, were selected as input variables for the construction of the Random Forest prediction model and, among these variables, Terrain Ruggedness Index (TRI) and slope were the ones that presented, respectively, the best importance in the model. In order to achieve the best parameterization of the algorithm, the grid search technique was applied, which identified that the best parameters for the Random Forest were 300 trees and 4 variables tested in each node. After building the model, several evaluation metrics were used to analyze its performance, which, according to the results, showed a good ability to map the susceptibility to landslides. Finally, the susceptibility map obtained by the proposed method was compared with existing maps for the area but produced through different mapping techniques. This work allowed us to identify that machine learning-based models should be increasingly explored in disaster risk management, as they constitute important tools in the development of strategies, public policies and territorial knowledge.

Keywords: Landslide susceptibility. Machine learning. Random forest. Geographic Information System. Natural Disasters.

1 INTRODUÇÃO

Os desastres representam uma preocupação mundial cada vez mais crescente. A alta concentração populacional, o intenso processo de urbanização e os impactos ligados às mudanças climáticas são fatores que intensificam o número de pessoas vivendo em áreas de risco e expostas a eventos extremos. De acordo com Centro de Pesquisas de Epidemiologia em Desastres (CRED) e a Estratégia Internacional de Nações Unidas para a Redução de Desastres (UNISDR) (2018), entre 1998 e 2017 foram registradas 7.255 ocorrências de desastres que causaram 1,3 milhões de óbitos e afetaram mais de 4 bilhões de pessoas entre feridos, desabrigados ou que precisaram de ajuda emergencial.

No Brasil, entre 1990 e 2012, os desastres relacionados às inundações e aos movimentos de massa registraram o maior número de vítimas, sendo que a maioria desses óbitos se situaram na região sudeste (CEPED UFSC, 2013). Dentre os fatores que contribuem para a maior concentração nessa região, destaca-se a alta densidade demográfica instalada nesses espaços, cuja ocupação desordenada em áreas de risco é frequente. Quando a construção de moradias é feita em relevos acidentados, os desastres mais comuns estão associados a escorregamentos em encostas. Conforme Santos (2017), são ocupados terrenos que, devido a sua alta instabilidade geológica natural, nunca deveriam ser urbanizados, como é o caso das expansões crescentes na Serra da Mantiqueira.

Devido às características como a alta densidade de drenagem, vertentes muito inclinadas, vales entalhados e formas de relevo muito dissecadas, a Serra da Mantiqueira é identificada como uma unidade com um nível de fragilidade potencial muito alto, estando sujeita a processos erosivos intensos e a ocorrências de movimentos gravitacionais de massa. Além disso, considerando as chuvas como um dos principais agentes deflagradores de escorregamentos, Carvalho et al. (2002) descrevem que a Serra da Mantiqueira está entre as regiões do estado de São Paulo com maiores registros de precipitação anual e de maior intensidade.

Inserido nesse cenário, o município de Campos do Jordão apresenta um histórico amplo de acidentes em áreas de risco, principalmente decorrentes de escorregamentos em encostas. Ahrendt (2005) aborda que a concentração urbana em áreas mais vulneráveis se intensificou após a segunda metade da década de 1970, causando o adensamento das moradias em áreas menos íngremes e a ocupação generalizadas das encostas. A demanda por habitações populares, principalmente pela população de baixa renda, é agravada pela indústria do turismo

no município, bem como a especulação imobiliária e demanda por mão de obra para suprir as carências econômicas locais (VIEIRA; KAZMIERCZAK; MALTA, 2005).

Com o objetivo de reduzir o risco de desastres, as políticas públicas e diretrizes governamentais devem considerar uma abordagem sistêmica do problema, envolvendo todas as fases que abrangem a gestão integral dos riscos de desastres – prevenção, mitigação, preparação, resposta e recuperação (BRASIL, 2012). Conhecer previamente os processos atuantes nos municípios, a partir de caracterizações e diferentes tipos de mapeamentos que variam conforme a escala e detalhamento das informações, é requisito básico no gerenciamento de riscos, pois é a partir dessa análise territorial que será possível identificar os locais necessários para investimentos em medidas estruturais e não-estruturais, elaborar as leis de zoneamento, estruturar o planejamento urbano e refinar o código de obras de cada cidade. Dentre essas ferramentas, o mapeamento da suscetibilidade tem um papel relevante no processo de desenvolvimento das medidas de prevenção, já que tem como objetivo indicar a potencialidade de ocorrência de processos naturais e induzidos (CARVALHO; GALVÃO, 2016).

Diante da natureza complexa dos escorregamentos que envolve uma série de fatores como condição do solo, topografia, geologia, hidrologia e atividades humanas, muitos métodos vêm sendo propostos para identificar e avaliar a suscetibilidade desses processos (BUI et al., 2016). Dentre as diversas metodologias utilizadas na previsão de regiões instáveis, modelos baseados em aprendizado de máquina se tornaram mais populares no âmbito acadêmico a partir de 2010 (MERGHADI et al., 2020). O aprendizado de máquina é um ramo da inteligência artificial que usa algoritmos para analisar e prever informações por meio de informações extraídas dos dados (JORDAN; MITCHELL, 2015).

Os métodos de predição partem do pressuposto de que os fatores que condicionaram movimentos gravitacionais de massa em uma determinada área no passado, serão os mesmos que desencadearão novos processos no futuro. Portanto, uma etapa essencial na utilização desses métodos consiste na elaboração de mapas de inventário de cicatrizes e levantamento de variáveis condicionantes. Vários métodos de aprendizado de máquina vêm sendo aplicados para analisar a suscetibilidade a escorregamentos como as Redes Neurais Artificiais (DOU et al., 2020; TIEN BUI et al., 2016; CHEN et al., 2017), Máquinas de Vetores de Suporte (HONG et al., 2015; TIEN BUI et al., 2016; CHEN et al., 2017) e Floresta Aleatória (CATANI et al., 2013; PARK; KIM, 2019; DANG et al., 2020; SAHIN et al., 2020 e ZHOU et al., 2021). Dentre eles, o algoritmo Floresta Aleatória é conhecido por ter a capacidade de lidar com um grande

volume de dados, que podem ser categóricos e numéricos, e apresentar altos valores de acurácia na modelagem de suscetibilidade a escorregamentos (SHIRVANI, 2020). Em comparação com outros métodos de aprendizado de máquina, como Rede Neural Artificial e Máquina de Vetores de Suporte, a Floresta Aleatória não requer muitos ajustes refinados de parâmetros para a construção do modelo de predição, fazendo com que a configuração padrão desses parâmetros já atinja altos desempenhos, além de ser considerado computacionalmente leve (TAALAB; CHENG; ZHANG, 2018).

Diante desse contexto, métodos de aprendizado de máquina permitem lidar com uma grande diversidade de dados e reconhecer padrões. Quando essas ferramentas são integradas a Sistemas de Informações Geográficas (SIG) os resultados são ainda mais significativos, devido à possibilidade de espacialização das ocorrências de escorregamentos e das variáveis condicionantes utilizadas no modelo, além da capacidade elevada de processamentos dos dados e geração do mapa de suscetibilidade.

Nesse sentido, visto a necessidade das cidades de conviver e gerenciar o risco de desastres, esse trabalho tem como finalidade identificar as áreas suscetíveis a escorregamentos incorporando uma abordagem de análise de dados baseada em aprendizado de máquina, usando o algoritmo Floresta Aleatória. A área selecionada para este estudo compreende a Bacia do Córrego Piracuama, localizada no município de Campos do Jordão e caracterizada pela alta frequência desses movimentos.

1.1 Estrutura da dissertação

O presente trabalho está dividido em sete capítulos, sendo que o primeiro abrange uma visão introdutória acerca da temática de estudo envolvendo, de forma geral, a problemática de desastres e escorregamentos, bem como o mapeamento da suscetibilidade a esses processos através de técnicas de aprendizado de máquina. O segundo capítulo apresenta o embasamento teórico necessário para o desenvolvimento do trabalho, com uma breve abordagem sobre as temáticas de desastre (global e nacional), contextualização sobre os escorregamentos (definições, classificações, fatores condicionantes e modelos de suscetibilidade a escorregamentos) e uma breve abordagem sobre o aprendizado de máquina, com ênfase no algoritmo Floresta Aleatória. O terceiro capítulo apresenta a proposta de pesquisa do trabalho, envolvendo o objetivo principal e os objetivos específicos. O quarto capítulo discorre sobre a metodologia utilizada para o mapeamento de suscetibilidade, dividida em três etapas principais:

escolha e caracterização da área de estudo; levantamento das cicatrizes e fatores condicionantes; e construção dos modelos de suscetibilidade. O quinto capítulo trata os resultados, avaliando e analisando o desempenho do modelo de predição e apresentando o mapa de suscetibilidade. O sexto capítulo envolve as discussões, que trazem reflexões sobre a utilização do modelo, desafios enfrentados e propostas futuras. O sétimo capítulo traz as conclusões e considerações finais acerca dos resultados e discussões apresentadas.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 Desastres

Desastres são definidos pela UNISDR (2009) como uma interrupção do funcionamento de um sistema social, envolvendo impactos ou perdas humanas, materiais, econômicas ou ambientais de grande extensão, que excedem a capacidade de uma comunidade de arcar com as consequências através dos seus próprios recursos. As definições de desastres são vastas e a falta de um consenso mínimo pode prejudicar esse campo de estudo. Os desastres podem ocorrer em qualquer região do mundo e os seus impactos irão variar de acordo com a magnitude e a frequência dos fenômenos, bem como com a vulnerabilidade do sistema social instalado.

2.1.1 Desastres no mundo

A concentração da população em centros urbanos, aliada à falta de planejamento e estruturas nas cidades, urbanização tardia, desigualdades econômicas encontradas em determinados grupos sociais, tendência na ampliação das chamadas megacidades, são alguns dos fatores que potencializam o risco a desastres (SAUSEN; LACRUZ, 2015). Segundo os dados do Departamento das Nações Unidas para Assuntos Econômicos e Sociais (UNDESA, 2018) a população instalada em área urbana em 1950 representava 29% do cenário mundial e, em 2018, este valor subiu para 55%. As projeções ainda permanecem crescentes e, de acordo com as Nações Unidas, em 2050 cerca de 68% da população mundial estará inserida em meios urbanos. Entender as tendências populacionais globais e antecipar as adaptações demográficas de acordo com essas mudanças é crucial para alcançar o desenvolvimento sustentável.

Para que um evento seja registrado pelo banco de dados internacional de desastres (*The international disaster database* – EM-DAT) é preciso que ao menos um dos seguintes critérios sejam atendidos: número mínimo de dez óbitos, cem ou mais pessoas afetadas, decretação de estado de emergência e solicitação de auxílio internacional. Devido a esses padrões exigidos, muitos eventos de menor magnitude não são registrados, não refletindo a realidade de muitos países. Essa tendência foi reforçada pela Organização das Nações Unidas (ONU) (UNISDR – AM, 2013), que divulgou que mais de 50% das mortes e 90% das perdas materiais por acidentes registrados na América estão relacionados a ocorrências de menor magnitude e recorrentes e,

portanto, não contabilizadas pelo EM-DAT. Outro contexto, também relacionado ao registro das informações dos desastres, ocorre com os dados já computados pelo EM-DAT. De acordo com o CRED/UNISDR (2018), 63% dos desastres reportados não contém dados econômicos. Dessa forma, estima-se que o custo real para economia global é em torno de 520 bilhões de dólares por ano, podendo levar anualmente 26 milhões de pessoas para a pobreza. As perdas econômicas causadas por desastres estão consideravelmente relacionadas ao PIB de cada país, sendo que, quanto menor o índice apresentado, maior serão os impactos vivenciados pela população.

Conforme a classificação proposta pelo CRED, os desastres são divididos em naturais (provocados por fenômenos e desequilíbrios da natureza) e tecnológicos (associados a acidentes com produtos químicos perigosos, industriais e de trânsito). Os desastres naturais se subdividem em geofísicos, climatológicos, hidrológicos, meteorológicos, biológicos e extraterrestres, descritos no Quadro 1.

Quadro 1. Classificação geral de desastres naturais adotada pelo EM-DAT

Subgrupo	Definição	Principais tipos de desastre
Geofísicos	Originados da geosfera	Terremoto, erupção vulcânica, movimento de massa (seco)
Meteorológicos	Causados por processos atmosféricos de curta duração (minutos a dias)	Tempestade (tornados, granizos, vendavais), temperaturas extremas
Hidrológicos	Causados por alterações no ciclo normal da água e/ou por transbordamento de sistemas hídricos	Inundação, enxurradas, movimentos de massa (úmido)
Climatológicos	Causados por processos de longa duração (variabilidade climática de intrasazonal a multidecadal)	Secas, incêndios
Biológicos	Causados pela exposição de organismos vivos a doenças e substâncias tóxicas	Epidemias, infestação de insetos
Extraterrestres	Causado por asteróides, meteoróides e cometas quando atingem a Terra	Impactos

Fonte :CRED/EM-DAT [s.d]

As estratégias para lidar com desastres são tardias e, somente em 1990, após a Década Internacional para a Redução de Desastres Naturais, a temática começou a se consolidar e a considerar o risco como parte da construção social, que pode ser quantificado e objetivamente localizado. Em seguida, outras ações foram estabelecidas, fortalecendo os programas de prevenção tais como o Plano de Ação de Yokohama (1994-2004), Estratégia Internacional para Redução de Desastres (EIRD ou *ISDR*) em 2000, Marco de Ação de Hyogo (2005-2015) e Marco de Sendai (2015-2030). O último, que reflete as medidas a serem adotadas no cenário atual, configura um plano aplicado em escala global que visa a redução significativa do risco de desastres e perdas em geral, através da implementação e inclusão de medidas que diminuam a exposição e a vulnerabilidade aos desastres, bem como o aumento da resiliência da população instalada. O Marco de Sendai reforça principalmente a importância de implementação nos países que apresentam menores índices econômicos.

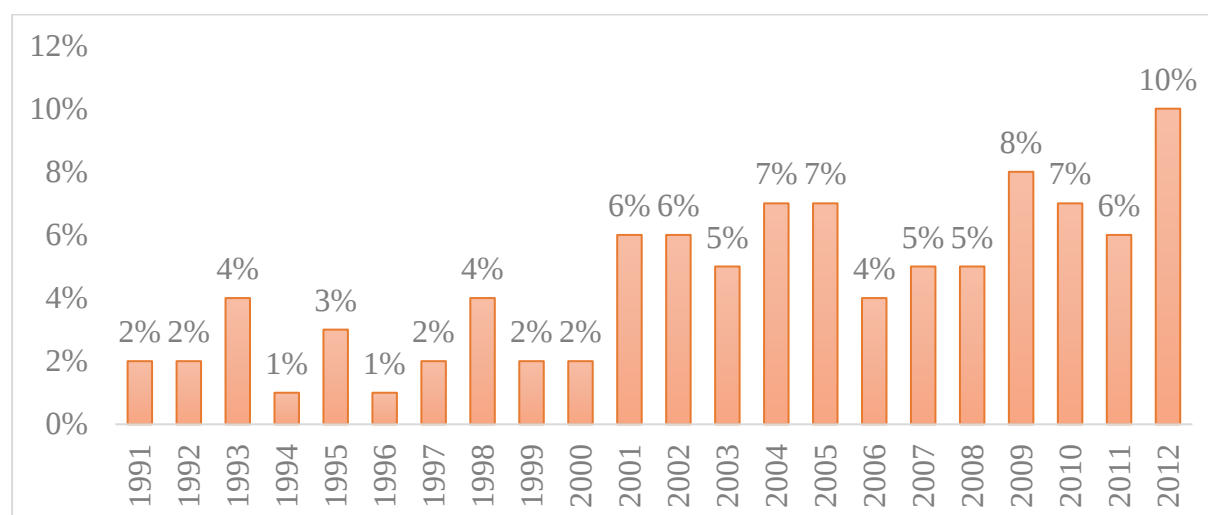
2.1.2 Desastres no Brasil

As ocorrências de desastres no Brasil são derivadas de fenômenos relacionados à dinâmica externa da terra, tais como inundações, movimentos de massa e secas. De acordo com Sausen e Lacruz (2015), historicamente, o contexto brasileiro de desastres esteve associado às ações de resposta e emergência, cujo papel da Defesa Civil se mostrou fundamental no desenvolvimento de materiais técnicos e termos legais.

A temática de desastres, ainda mais recente no Brasil, é prejudicada pelo baixo número de pesquisas, dificuldade de acesso a registros, ausências de padrões de declaração em documentos, ineficiência de métodos de coleta e armazenamento de dados, falta de base histórica, dificuldade no compartilhamento dos dados e divergências de interpretação pelos responsáveis a respeito do tipo do fenômeno ocorrido (CEPED UFSC, 2013). Um grande avanço nesse campo de estudo foi a publicação do *Atlas Brasileiro de Desastres Naturais*, produzido por meio de uma parceria entre a Secretaria Nacional de Defesa Civil (SEDEC) e a Universidade de Santa Catarina, que sistematizou todas as ocorrências de desastres referentes ao período de 1991 e 2012. Outro progresso, em 2013, referiu-se à implementação do Sistema Integrado de Informações sobre Desastres (S2ID) que consiste em um banco de dados digital unificado, com informações que provêm insumos para redução dos riscos de desastres, consumidas principalmente pelos gestores de políticas públicas e o âmbito acadêmico (CEPED UFSC, 2013).

Ao todo, foram registrados pelo documento 38.996 ocorrências de desastres associados à fenômenos e desequilíbrios da natureza que, nos últimos 10 anos analisados, apresentaram uma tendência de aumento a partir de 2001 (Figura 1). Segundo Prieto (2018), é importante ressaltar que o aumento significativo dos registros nos últimos dez anos pode resultar de possíveis fragilidades e deficiências de apontamentos pelos órgãos de defesa civil nos primeiros anos pesquisados (CEPED UFSC, 2013). No entanto, é provável que o número total de apontamentos não reflita a realidade do país, visto que não foram consideradas ocorrências com ausência de decretos de emergência ou estado de calamidade pública (SAUSEN; LACRUZ, 2015).

Figura 1. Desastres naturais no Brasil (1991-2012)



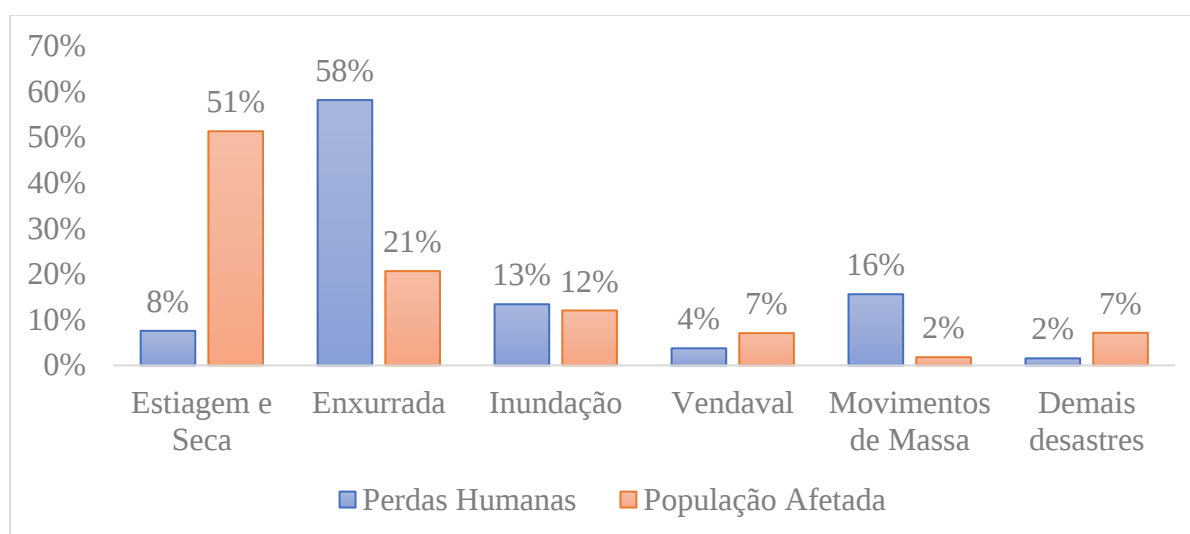
Fonte: CEPED UFSC (2013).

Quando comparadas as ocorrências registradas entre as décadas de 1990 e 2000, os movimentos de massa foram o tipo de desastre que apresentaram o maior acréscimo nos registros – em torno de 21,7 vezes. Entretanto, de acordo com Sausen e Lacruz (2015), cabe destacar que, mesmo com o acréscimo levantado, os registros de movimento de massa podem ser subdimensionados devido às divergências conceituais desses processos que muitas vezes podem ser associados a outros fenômenos (por exemplo, corridas de massa entendidas como inundações bruscas). Em contrapartida, é possível ocorrer o superdimensionamento desses registros, uma vez que as defesas civis podem cadastrar as informações em sistemas distintos e, caso o evento afete mais de um município, os apontamentos podem ser multiplicados.

Em relação à população afetada, ainda com base no período analisado (1990-2012), os

eventos de estiagem e secas, que representaram mais de 50% dos registros, foram os que causaram maiores prejuízos (Figura 2.) Na sequência, estão as enxurradas (21%), inundações (12%), vendavais (7%) e, menos expressivamente, os movimentos de massa (2%). Os demais desastres, que englobam os granizos, erosões, incêndios, tornados, alagamentos e geadas, impactaram juntos 7% da população em questão (CEPED UFSC, 2013). O cenário é diferente quando se trata da análise de óbitos por tipo de desastre (Figura 2). As enxurradas são as principais responsáveis pelas mortes registradas no período, representando 58% do total. Secundariamente, os movimentos de massa, que constituíram apenas 2% das pessoas afetadas, correspondem a 16% dos óbitos nesse período. Em seguida, encontram-se as estiagens e secas (8%), vendavais (4%) e os demais desastres, que somaram 2% das perdas humanas (CEPED UFSC, 2013).

Figura 2. Perdas humanas e população afetada por tipo de desastres no Brasil (1990-2012)

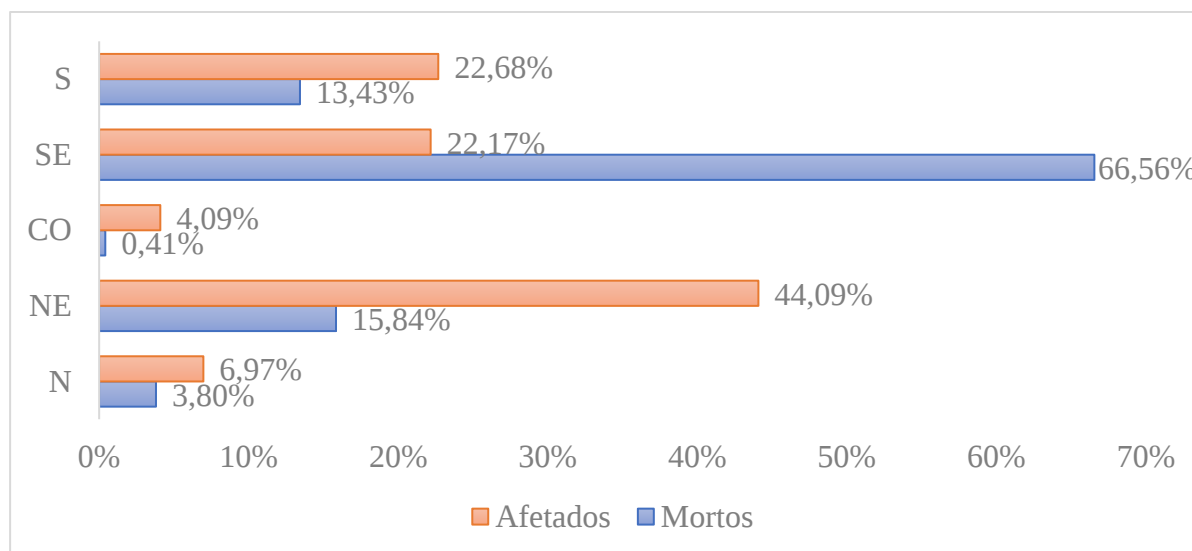


Fonte: Prieto (2018); CEPED UFSC (2013).

Outra análise essencial para a compreensão do contexto brasileiro de desastres é a visibilidade de como as regiões foram impactadas nesse período, segundo os critérios de pessoas afetadas e óbitos. A Figura 3 destaca duas regiões, Nordeste e Sudeste, sendo que a primeira representa o maior número da população afetada, 44,06% do total, e a segunda, com 66,54% dos registros, a maior concentração de mortes (CEPED UFSC, 2013). A região Nordeste está ligada principalmente a eventos de estiagem e seca que podem desencadear uma

série de impactos, tais como redução da agropecuária, desempregos, migrações, agravamentos de doenças e incêndios (SAUSEN; LACRUZ, 2015).

Figura 3. Mortos e afetados por região brasileira (1991-2012).



Fonte: CEPED UFSC (2013).

O alto número de óbitos descritos na região Sudeste possui expressiva relação com a alta densidade demográfica instalada no local. De acordo com o Instituto Brasileiro de Geografia e Estatística (IBGE, 2017), os espaços urbanos concentram 76% da população brasileira sendo que, na região Sudeste, tal proporção é ainda maior, visto que 87% da população vive nas cidades. O acelerado processo de urbanização, intensificado principalmente nas últimas décadas, desencadeou o crescimento desordenado dos municípios e a construção de moradias, muitas vezes precárias, em áreas impróprias à ocupação, aumentando as situações de perigo e risco a desastres (TOMINAGA; SANTORO; AMARAL, 2015).

Em grandes e médias cidades, a expansão urbana vem avançando para relevos acidentados que, naturalmente possuem características geológicas e geomorfológicas desfavoráveis à instalação de moradias, como acontece frequentemente em regiões associadas a Serra do Mar e Serra da Mantiqueira. Entretanto, mesmo em áreas que são passíveis de receber ocupação urbana, podem ser geradas situações de alto risco geotécnico, devido a inadequações técnicas e precariedade construtiva (SANTOS, 2015). Cabe ressaltar que a região sudeste registra o maior número de ocorrências de enxurradas e deslizamentos, os dois tipos de desastres responsáveis pelo maior número de óbitos no país (SAUSEN; LACRUZ, 2015).

Dentre os principais desastres documentados em território brasileiro, destaca-se o megadesastre de 2011, que resultou em movimentos de massa generalizados na região serrana do estado do Rio de Janeiro. O evento catastrófico, que ocasionou 947 perdas humanas e 300 desaparecidos, também foi um gatilho para a reestruturação da defesa civil brasileira, que anteriormente focava principalmente em ações pós-desastre (resposta e recuperação). Diante desse contexto, as estratégias para gerir o risco de desastres foram fortalecidas, levando em conta ações de prevenção, mitigação e preparação pré-desastre (SAUSEN; LACRUZ, 2015).

Em 2011, no âmbito federal, foi criado o Centro Nacional de Monitoramento e Alertas de Desastres Naturais (Cemaden), responsável por realizar o monitoramento das ameaças naturais em áreas de riscos em municípios brasileiros suscetíveis à ocorrência de desastres, bem como a emissão de alertas para subsidiar o gerenciamento de riscos e minimizar possíveis impactos. No ano seguinte, a Lei n.12.608 instituiu a Política Nacional de Proteção e Defesa Civil (PNPDEC), desencadeando diversas mudanças nas estruturas organizacionais brasileiras, tais como a distribuição de competências entre os níveis federais, estaduais e municipais, o cadastro nacional de municípios com áreas suscetíveis à ocorrência de deslizamentos de grandes impactos e inundações bruscas e a participação da sociedade civil na gestão de riscos.

2.2 Movimentos de massa

A evolução geomorfológica das regiões serranas brasileiras está significativamente associada aos movimentos gravitacionais de massa, processos naturais que atuam na dinâmica das encostas e são modeladores do relevo. Esses processos são definidos como movimentos de solo, rocha e/ou vegetação ao longo de uma vertente sob a ação direta da gravidade (TOMINAGA; SANTORO; AMARAL, 2009).

Os movimentos de massa ocorrem em todo o território brasileiro, entretanto são as regiões Sudeste e Sul que registram os maiores números de ocorrências, concentrando 79,8% e 13,6% dos registros respectivamente (CEPED UFSC, 2013). De acordo com Santos (2004), esses processos estão associados à soma de um conjunto de fatores como a alta declividade das encostas, elevados índices pluviométricos, características geológicas e ações antrópicas. Os movimentos de massa que ocorrem na região sudeste são nitidamente sazonais, sendo constantemente associados a épocas de chuvas intensas e concentradas, que geralmente acontecem entre os meses de dezembro a março.

Highland e Bobrowsky (2008) classificam esses fenômenos de acordo com a mecânica

interna de deslocamento da massa do movimento, dividindo-os em quedas, tombamentos, escorregamentos, subsidências e escoamento de fluxo. Os problemas associados a esses processos estão presentes em muitos lugares do mundo, entretanto, em países cujo regime pluvial apresenta características de ambiente tropical e cuja situação socioeconômica seja considerada como de subdesenvolvimento ou em desenvolvimento, os problemas tendem a se intensificar devido à falta de estrutura para evitar ou controlar tal fenômeno (GUERRA, 1994).

2.2.1 Escorregamentos

Dentre os movimentos de massa, os escorregamentos são os processos mais frequentes na região sudeste e são um dos mais fatais e danosos fenômenos naturais que existem no território brasileiro. Tendo como sinônimos de uso popular os termos deslizamento, desmoronamento, queda de barreiras e desbarrancamento, os escorregamentos ocorrem em áreas com elevada declividade, envolvendo materiais que recobrem as encostas como solos, saprólitos, blocos de rocha e vegetação. Esses processos, que são desestabilizados pela ação de chuvas intensas e/ou prolongadas, envolvem uma diversidade de variáveis que abrangem aspectos associados à geologia, geomorfologia, vegetação, clima e uso e ocupação da terra (TOMINAGA; SANTORO; AMARAL, 2009).

As condições essenciais para a deflagração de escorregamentos são a falta de estabilidade da frente das encostas e a existência de superfícies de deslizamento. Tais condições ocasionam movimentos rápidos, com velocidades de metros por hora a metros por segundo, de curta duração e com planos de ruptura bem definidos entre o material deslizado e o não movimentado (GERSCOVICH, 2016). Esses processos ocorrem, basicamente, quando a relação entre a resistência ao cisalhamento do material e a tensão de cisalhamento na superfície com potencial de movimentação decresce até atingir relação igual a uma unidade. No momento de ruptura, a massa em movimento avança com velocidade acelerada e, na medida que o processo acontece, diminuem as forças que determinaram o movimento e a massa tende a se estabilizar. Dessa forma, o movimento cessa ou assume caráter de rastejo (GUIDICINI; NIEBLE, 1984; TOMINAGA, 2007).

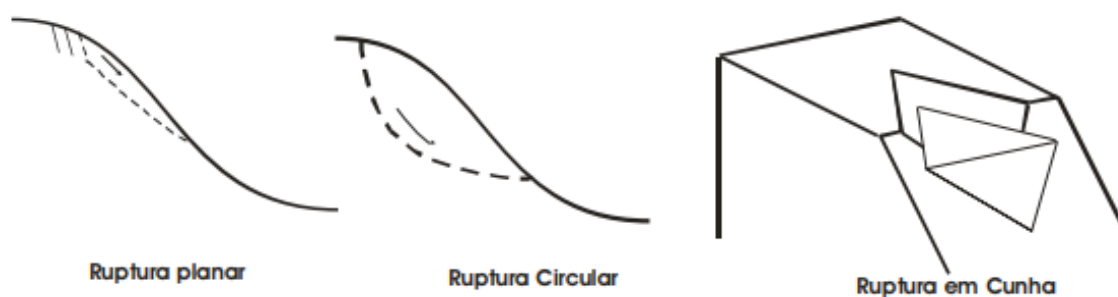
De acordo com Tominaga (2007), as cidades convivem com uma acentuada incidência de escorregamentos induzidos, que estão associados a ações antrópicas como cortes inadequados para implantação de moradias e vias de acesso, lançamento de águas servidas e acúmulos de depósitos de lixo. Mendes et al. (2018) aborda esse cenário para Campos do

Jordão, destacando como fatores antrópicos indutores de escorregamentos as cargas habitacionais, os cortes artificiais e vazamentos de água nos terrenos. A ocorrência de escorregamentos pode estar associada a movimentos isolados ou generalizados, sendo que o último ocorre com menos frequência, afeta grandes áreas e está relacionado a eventos extremos de precipitação.

2.2.1.1 Classificação dos escorregamentos

A classificação dos escorregamentos é baseada na geometria do plano de ruptura (Figura 4), dividindo esses processos geodinâmicos em translacionais (planares), rotacionais (circulares) e em cunha. Os escorregamentos translacionais ou planares (Figura 5) são os movimentos mais comuns na dinâmica das encostas serranas brasileiras, ocorrendo principalmente em encostas com altas declividades e solos pouco desenvolvidos. De acordo com Highland e Bobrowsky (2008), a massa de um escorregamento translacional move-se para fora ou para baixo do talude, ao longo de uma superfície relativamente plana. Esse tipo de movimento pode progredir por distâncias consideráveis, se a superfície da ruptura estiver suficientemente inclinada, ao contrário dos escorregamentos rotacionais, que tendem a restaurar o equilíbrio do deslizamento. Escorregamentos translacionais comumente ocorrem ao longo de descontinuidades geológicas tais como falhas, fraturas, foliações, estratificações, xistosidades ou no ponto de contato entre rocha e solo.

Figura 4. Esquema das principais rupturas em escorregamentos.

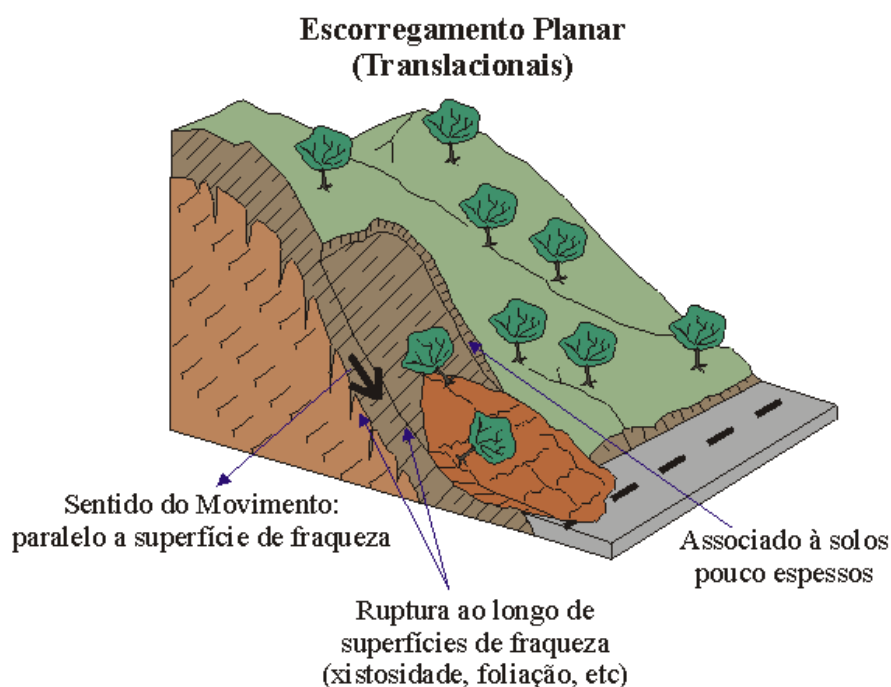


Fonte: Modificado de IPT (1991).

Conforme Mendes (2018), os escorregamentos planares são caracterizados pelo envolvimento de uma fina camada de solo, variando entre 1 a 2 metros de espessura, que desliza

sobre uma superfície geralmente paralela à encosta. Esses processos ocorrem geralmente em encostas de grande amplitude e com inclinações maiores que 30° , abrangendo os horizontes superficiais e, em algumas circunstâncias peculiares, os horizontes mais profundos constituídos por saprólito. Dias e Herrmann (2002) ressaltam que os escorregamentos translacionais costumam ser deflagrados durante eventos de chuvas intensas, que ocasionam o aumento da poropressão do solo. Tal fator ocorre quando a taxa de infiltração se torna superior à taxa de retirada de água do interior da encosta.

Figura 5. Esquema de um escorregamento planar



Fonte: Infanti Jr. e Fornasari Filho (1998); modificada por Reis (2001).

2.2.1.2 Fatores condicionantes dos escorregamentos

Compreender os fatores que condicionam a deflagração dos escorregamentos é fundamental para identificar quais as áreas mais suscetíveis a esses processos, permitindo a adoção de medidas preventivas ou corretivas que, de modo geral, auxiliam na gestão de risco de regiões sujeitas a esses movimentos. Corominas et al. (2014), analisa a relevância de diversos parâmetros – divididos em topográficos, geológicos, hidrológicos, geomorfológicos, antrópicos e climáticos – na ocorrência de escorregamentos e que podem ser considerados no mapeamento

da suscetibilidade (Quadro 2).

Quadro 2. Fatores condicionantes e suas respectivas relevâncias para a análise de suscetibilidade a escorregamentos

(continua)

Grupo	Parâmetros	Relevância para a análise de suscetibilidade
Topográfico	Elevação	As diferenças de elevação resultam em uma maior energia potencial para os escorregamentos
	Declividade	Fator determinante na ocorrência de escorregamentos. A capacidade de transporte de solos e rochas, bem como a velocidade de deslocamento de material nas encostas são diretamente proporcionais à declividade
	Orientação da encosta (aspecto)	Pode implicar em variações de umidade e variações da vegetação
	Comprimento da encosta, forma, curvatura, rugosidade	Importantes para a modelagem da trajetória do fluxo e alcance da massa deslizada
	Direção e acúmulo de fluxo	Aplicados na modelagem hidrológica, por exemplo, no cálculo do Índice Topográfico de Umidade (<i>Topographic Wetness Index</i> - TWI)
Geologia	Litologia	Determina as características geotécnicas das rochas, relacionados à natureza mineralógica da matriz (resistência, alterabilidade, erodibilidade, presença de minerais expansivos)
	Intemperismo	Tipos e graus de intemperismo, que afetam à resistência da rocha
	Descontinuidades	Feições estruturais como falhas, xistosidade, lineamentos, fraturas. Tipos e características das descontinuidades possuem relação com a resistência, orientações e inclinação das vertentes. Além disso, as descontinuidades muitas vezes podem estar preservadas no solo saprolítico, favorecendo a ocorrência de desses processos
Solos	Tipos de solo	A origem do solo determina suas propriedades e geometria
	Profundidade de solos	A profundidade determina a quantidade potencial de material disponível para a mobilização
	Propriedades geotécnicas	Tamanho do grão, coesão, ângulo de atrito, densidade aparente etc.
	Propriedades hidrológicas	Quantidade, tipo e volume dos poros, condutividade hidráulica etc.

Quadro 2. Fatores condicionantes e suas respectivas relevâncias para a análise de suscetibilidade a escorregamentos

(conclusão)

Hidrologia	Profundidade do lençol freático	Variações espaciais e temporais na profundidade do lençol freático, formação de lençóis suspensos, frentes de saturação e poropressão são variáveis importantes
	Umidade do solo	Variações espaciais e temporais no teor de umidade do solo
	Componentes hidrológicos	Interceptação da precipitação, evapotranspiração, fluxos superficiais, infiltração, percolação etc.
	Densidade de drenagem	Em análises em pequena escala, a densidade de drenagem pode ser usada como um indicador de características do terreno
Geomorfologia	Ambientes geomorfológicos	Tropical, glacial, costeiro etc.
	Escorregamentos antigos	Características do terreno e dos materiais podem ter sido alteradas, favorecendo a reativação de escorregamentos antigos. Informações históricas sobre escorregamentos antigos são cruciais para determinar os perigos e riscos a esses processos
Uso e cobertura e fatores antrópicos	Uso atual do solo	Tipo de uso da terra/cobertura da terra, tipo de vegetação, peso da vegetação, profundidade das raízes etc.
	Mudança no uso e ocupação	Variações do uso da terra ao longo do tempo podem influenciar a ocorrência de escorregamentos
	Vias de transporte	Cortes de estradas em áreas declivosas
	Construções	Cortes em encostas para a construção civil
	Redes de irrigação e drenagem	Vazamento nessas redes podem ser uma causa importante na ocorrência de escorregamentos
	Pedreiras e Mineração	Altera a geometria e distribuições de tensões das vertentes. As vibrações decorrentes das detonações também podem causar escorregamentos
Clima	Precipitação	Limiares de chuva, padrões climáticos, chuvas antecedentes, curvas Intensidade-Duração-Frequência (IDF), relações de magnitude, duração e frequência
	Temperatura	Influência importante na hidrologia e no estado da vegetação.

Fonte: traduzido e adaptado de Corominas et al. (2014).

2.2.1.3 Modelos de suscetibilidade a escorregamentos

Diversas metodologias que visam a previsão dos escorregamentos vêm sendo desenvolvidas, dando subsídios a atividades que apoiam a administração pública como planos de evacuação, planejamento urbano, avaliação de traçados de rodovias, entre outros. Dentre elas, os mapas de suscetibilidade trazem uma distribuição espacial das áreas potenciais para ocorrência de escorregamentos, considerando fatores condicionantes relacionados à característica do meio físico e classificando a região de estudo quanto a propensão ao desenvolvimento desses processos.

O mapa de suscetibilidade, conforme IPT (2007), possui as seguintes características: i) é baseado no mapa de inventário; ii) constrói um mapa de fatores condicionantes que influenciam a ocorrência dos eventos e iii) correlaciona os fatores e eventos para, por fim, classificar as unidades de paisagem segundo graus de suscetibilidade. Mesmo com o avanço das modelagens de suscetibilidade, as predições dos escorregamentos ainda continuam sendo um desafio, devido às inúmeras variáveis que influenciam esse tipo de evento, a aquisição de dados de qualidade e a complexidade técnica que esse tipo de modelagem necessita.

De acordo com Park e Kim (2019), muitos métodos vêm sendo propostos para avaliar a suscetibilidade a escorregamentos. Os métodos heurísticos consideram uma combinação de diferentes mapas temáticos (geológico, geomorfológico, geotécnicos, pedológicos, uso e ocupação etc.), aos quais são atribuídos pesos para cada atributo de acordo com a importância relativa na deflagração dos movimentos de massa. Embora seja um método que exija pouco tempo e custo, apresenta como principal desvantagem a grande subjetividade contida na opinião do pesquisador que define, com base nas suas experiências, o peso de cada variável. Já as técnicas matemáticas procuram reproduzir os sistemas ambientais com base em equações físicas, buscando descrever o comportamento dos processos que influenciam os escorregamentos. Esses métodos analisam os processos físicos que desencadeiam a ruptura de uma encosta e, podem ser aplicados em escalas menores, como a de bacia. Tais modelos são divididos em estocásticos ou determinísticos.

As técnicas baseadas em estatística são construídas com o objetivo de descrever a relação funcional de correlação entre fatores de instabilidade, descritos por conjunto de variáveis condicionantes (independentes) e a distribuição da ocorrência de escorregamentos, que configuram as variáveis dependentes. Essa relação funcional é então usada para verificar a propensão do terreno à escorregamentos e prever a suscetibilidade. Segundo Reichenbach et al.

(2018), os métodos estatísticos são indicados para avaliação de suscetibilidade em médias escalas.

Outra possibilidade no mapeamento da suscetibilidade a escorregamentos é a utilização dos métodos de aprendizado de máquina. Esses algoritmos se tornaram populares em diversos campos de pesquisa e, devido ao bom desempenho dos resultados e a possibilidade de lidar com um grande volume de variáveis condicionantes, as pesquisas relacionadas a essa temática tornaram-se cada vez mais usuais nos últimos anos. Reichenbach et al. (2018) realizou uma revisão bibliográfica de modelos de suscetibilidade a escorregamentos no período entre 1983 e 2016 e destacou um grande salto na utilização de técnicas de aprendizado de máquina a partir do ano 2010.

A associação desses métodos aos SIG tem sido cada vez mais comuns e, na avaliação das áreas instáveis, podem ser grandes aliados quando o acesso à área de estudo é difícil ou quando não há disponibilidade de equipes suficientes para o mapeamento. Utilizar um ambiente SIG na análise da suscetibilidade a escorregamentos permite identificar e delinear áreas em que houve a deflagração desses processos, permitindo a criação de um banco de dados geoespacial de ocorrências de escorregamentos, atrelando todas as informações associadas ao evento. Essa ferramenta também pode ser usada para modelar os fatores condicionantes dos escorregamentos e analisar as suas características ao longo do território (MERGHADI et al., 2020).

2.3 Aprendizado de máquina

De acordo com Mohri et al. (2018), o aprendizado de máquina, subcampo da inteligência artificial, consiste em efetuar previsões de situações desconhecidas tendo como base informação previamente conhecida. Samuel (1959), primeiro a utilizar o termo e considerado autor do primeiro programa com capacidade de autoaprendizado, definiu aprendizado de máquina como sendo o "campo de estudos que fornece a computadores a habilidade de aprenderem sem serem explicitamente programados". Em termos práticos, os algoritmos de aprendizado de máquina possuem como finalidade descobrir o relacionamento entre as variáveis de um sistema (entrada/saída) a partir de dados amostrados. Sendo assim, a aplicação desse método é indicada quando não se compreende os relacionamentos entre todas as variáveis envolvidas no problema (entrada/saída), situação frequente quando analisamos problemas reais vivenciados em nosso dia a dia (VON ZUBEN, 2013)

Basicamente, os métodos mais comuns de aprendizado de máquina são subdivididos quanto ao tipo de aprendizado, podendo ser classificados em supervisionados, não-supervisionados e por reforço (VON ZUBEN, 2013). O primeiro busca identificar a classe de uma nova amostra de dados (tendências futuras), a partir do conhecimento adquirido de um conjunto de amostras com classes conhecidas. Dessa forma, esse tipo de classificação possui uma intervenção externa, em que é fornecido ao algoritmo um conjunto de exemplos com rótulos de classes já estabelecidos (MONARD; BARANAUSKAS, 2003).

O aprendizado supervisionado se desmembra em duas subcategorias, as de regressão e classificação. O modelo de regressão é utilizado quando os dados de entrada e saída se relacionam via uma função contínua e, geralmente, os resultados são numéricos (SOUSA, 2019). Já os métodos de aprendizado supervisionado de classificação predizem valores discretos (classes). O processo utiliza um banco de dados de treinamento contendo atributos numéricos e categóricos (atributos preditores) e atributos de classe, que indicam a que classe o registro pertence. A finalidade da tarefa de classificação é extrair do banco de dados de treinamento um modelo que descreva cada classe a partir dos atributos preditores. O modelo gerado pode ser utilizado para prever a classe de registros com classes ainda desconhecidas. Caso necessário, o processo de classificação pode ser repetido e ajustado, por exemplo, adicionando novos atributos ou calibrando certos parâmetros (MONARD; BARANAUSKAS, 2003)

A classificação não supervisionada é voltada para conjuntos de dados que não possuem uma classe determinada, sendo assim, o algoritmo realiza, em linhas gerais, uma busca para identificar padrões de comportamento comuns nestes dados. De acordo com Gameiro (2020), esse tipo de aprendizado, ao contrário do supervisionado, não apresenta uma intervenção externa. Nesse método, é gerado um agrupamento ou *clusterização* dos dados de entrada, através de informações em comum ou que sejam semelhantes entre si.

O aprendizado por reforço ocorre por meio de recompensas ao indutor (GAMEIRO, 2020). Nesse caso, os dados de entrada e saída não precisam ser pré-definidos e o aprendizado se concentra em receber recompensas após a realização de determinadas ações. Esse sinal de recompensa indicará se a decisão tomada foi de fato adequada ou não. Os próximos estágios do ambiente e recompensas são determinados por ações tomadas pelo agente (HARMENDANI; ZANETTI, 2019).

Diversos métodos estão sendo utilizados na análise da suscetibilidade a escorregamentos, principalmente aplicados em escalas regionais, como as Redes Neurais

(*Neural Networks*), Lógica Difusa (*Fuzzy*), Árvores de Decisão (*Decision Trees*), Máquina de Vetores de Suporte (*Support Vector Machine*), entre outros. Com o objetivo de aprimorar as previsões e robustez nas avaliações de suscetibilidade, métodos do tipo aprendizado em conjunto (*ensemble learning*) foram desenvolvidos. Esses métodos são baseados em uma combinação de classificadores e vêm apresentando melhoria no desempenho dos resultados de predição (PARK; KIM, 2019). A Floresta Aleatória (*Random Forest*) é um exemplo dessa técnica, a qual será utilizada no presente trabalho e é descrita brevemente a seguir.

2.3.1 Floresta Aleatória

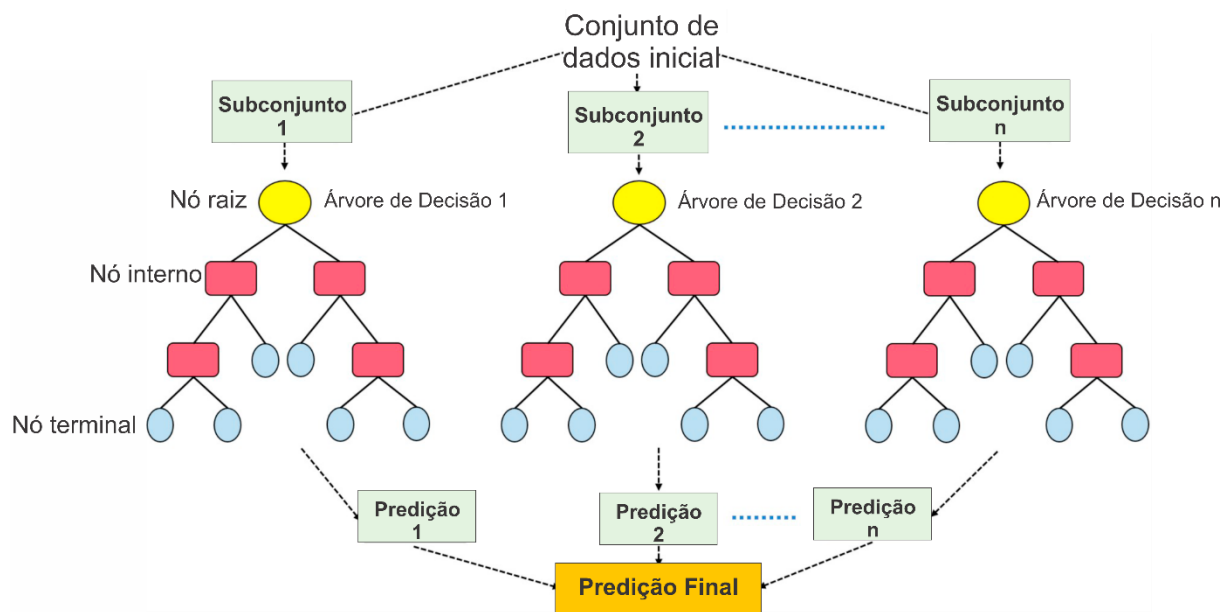
Considerado como um tipo de aprendizado supervisionado, o método Floresta Aleatória, desenvolvido por Breiman (2001), utiliza um conjunto de métodos de aprendizagem buscando obter uma performance maior do que seria alcançada utilizando somente um método. Dessa forma, a Floresta Aleatória combina múltiplas árvores de decisão para melhorar a performance de uma única árvore (YOUSSEF et al., 2016a).

De acordo com Rokach e Maimon (2005), a árvore de decisão possui como objetivo a construção uma estrutura que represente explicitamente o conjunto de dados e possibilite a criação de regras lógicas de decisão a partir de informações disponíveis. Segundo Sato et al. (2013), o seu funcionamento se baseia numa divisão subsequente do conjunto de dados, de forma a separá-lo em um subconjunto homogêneo, com casos de única classe. As informações do banco de dados são apresentadas ao nó raiz da árvore de decisão, que representa um dos atributos (variável decisória) e, a depender do resultado lógico, a árvore ramifica-se, gerando decisões que originam nós internos (intermediários ou decisórios) (SOBRAL, 2003). A repetição desse procedimento ocorre em vários níveis até alcançar um nó terminal que apresenta a variável de decisão.

No método Floresta Aleatória (Figura 6), cada árvore de decisão é treinada com um subconjunto de dados e o algoritmo explora as árvores, geralmente binárias, através de uma técnica chamada *bootstrapping*. Esse método consiste numa seleção aleatória de amostras de treinamento retiradas do conjunto de dados original, que são utilizadas para a construção do modelo. Esses dados reamostrados são chamados de “*in-bag*” e, geralmente, correspondem a 2/3 do conjunto de dados inicial (Figura 7). Os dados não incluídos são referidos como “*out-of-bag*” (OOB) (Figura 8) e são usados para estimar o desempenho do modelo (BREIMAN, 2001). A partir desse conjunto de dados, é possível identificar a taxa de erro do modelo de predição,

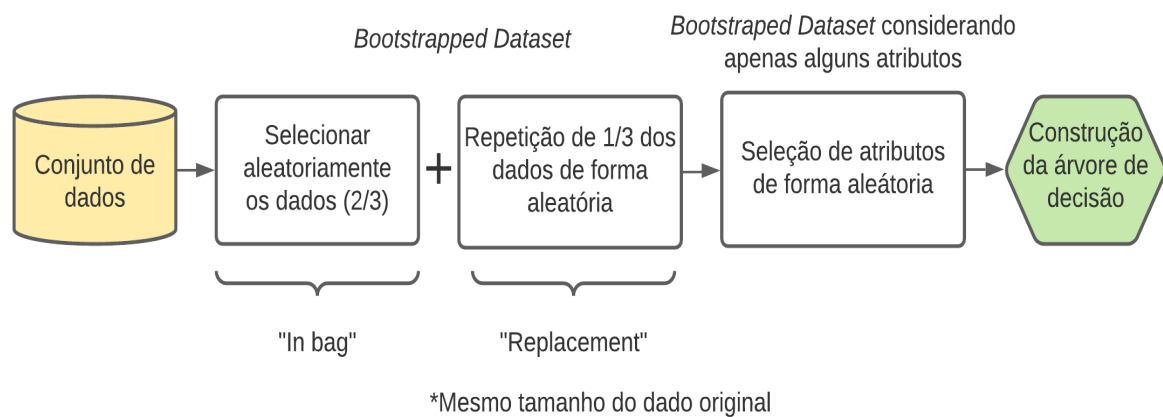
comparando as classes que foram previstas com as classes previamente conhecidas (YOUSSEF et al., 2016b). A proporção de classificações erradas em relação a todos os dados “*out-of-bag*” é chamada de “*out-of-bag-error*” (OOBE). As decisões finais de filiação de classe e a construção do modelo são determinados pela maioria dos votos entre as árvores. (PARK e KIM, 2019).

Figura 6. Estrutura do método Floresta Aleatória



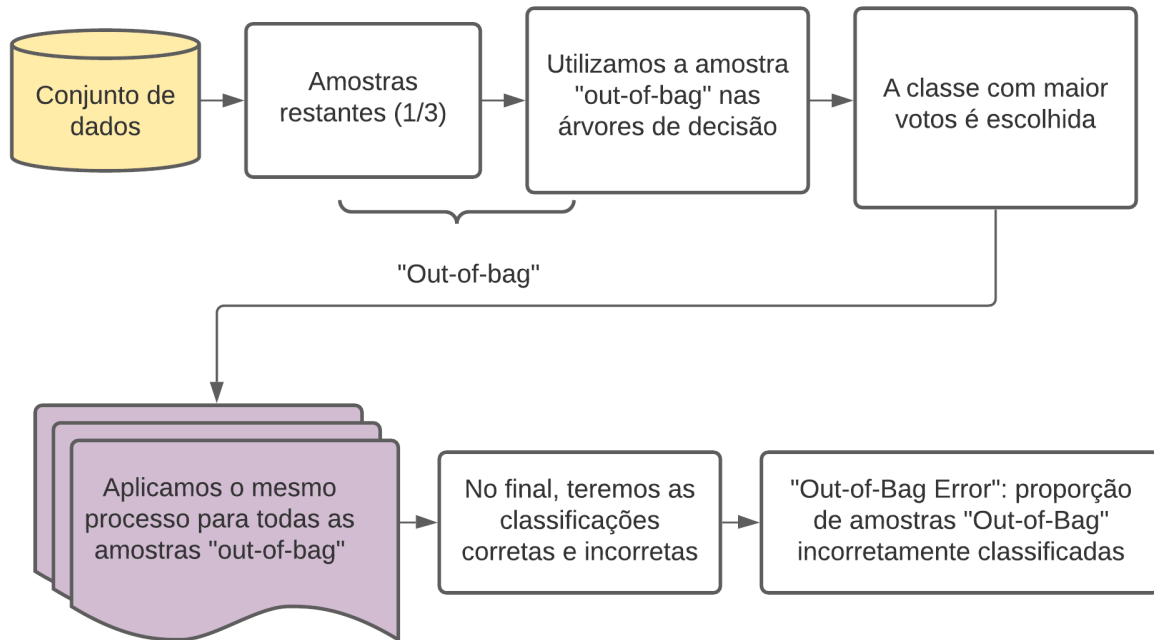
Fonte: adaptado de Merghadi et al. (2020).

Figura 7. Criação das árvores de decisão a partir da técnica *bootstrapping*



Fonte: produção do autor.

Figura 8. Utilização dos dados “out-of-bag” para validação do modelo



Fonte: produção do autor.

De acordo com Park e Kim (2019), dois parâmetros devem ser otimizados prioritariamente para alcançar um bom desempenho no modelo e minimizar o erro OOB – o número de árvores da floresta (*ntree*) e o número de variáveis testadas em cada nó (*mtry*). Taalab et. al (2018) ressaltam que não existem regras específicas sobre o número de árvores exigida em uma Floresta Aleatória e que aumentar o número de árvores não aumenta automaticamente a precisão do modelo, entretanto, exigirá um acréscimo na carga computacional. Os modelos também podem ser sensíveis ao parâmetro *mtry*, visto que usar um maior número de variáveis a cada divisão poderá aumentar a correlação entre as árvores nas florestas. Quando isso ocorre, normalmente, há um aumento na taxa de erros nas previsões.

A vantagem significativa da Floresta Aleatória é que a previsão de uma classe é feita em função da maior concordância e votação entre múltiplas árvores de decisão, melhorando assim as previsões e mitigando erros de *outliers*. Além disso, o método permite a utilização mista de variáveis categóricas e numéricas e é capaz de explicar interações e não linearidades entre variáveis. Na análise de suscetibilidade a escorregamentos, o método permite trabalhar com variáveis de terrenos cujos dados muitas vezes são ausentes ou onde há uma incerteza na atribuição para a classe correta, também nas áreas pesquisadas. Por exemplo, a definição correta

do tipo de solo para uma determinada extensão de área para os quais apenas alguns locais pontuais foram diretamente pesquisados (CATANI et al., 2013).

Catani et al. (2013) utilizaram a Floresta Aleatória para compreender a influência da sensibilidade e questões de escala dos dados de entrada no mapeamento da suscetibilidade. Como resultados, eles indicaram que a escala e o processo de treinamento influenciam fortemente na acurácia do modelo. Segoni et al. (2020) aplicaram o mesmo método para identificar a importância dos parâmetros geológicos como dados de entrada nos modelos de predição. Eles identificaram que a combinação de diferentes parâmetros geológicos (litologia, geocronologia, estrutural etc.) podem apresentar um melhor desempenho do que apenas uma variável. Shirvani (2020) realizou o mapeamento da suscetibilidade a escorregamentos considerando cenários de florestas protegidas e não protegidas. Através dos resultados obtidos, foi possível perceber que a influência dos parâmetros topográficos foi maior na análise das florestas protegidas e que os fatores antrópicos apresentaram maior peso nas áreas de florestas não protegidas. Park e Kim (2019) compararam dois métodos do tipo *ensemble* – Floresta Aleatória e Regressão da Árvore de Decisão Aumentada – considerando variáveis condicionantes topográficas, hidrológicas e florestais. Dentre essas variáveis, as topográficas apresentaram maior importância no modelo. Além disso, o método Floresta Aleatória obteve melhores resultados. Chen et al. (2017b) utilizaram três técnicas de mineração de dados: Floresta Aleatória, Árvore de Classificação e Regressão e Árvore de Modelo Logístico. No geral, todos os três modelos exibiram desempenhos razoáveis, mas o método Floresta Aleatória exibiu a maior capacidade preditiva. Zhang et al. (2017) analisaram modelos de Floresta Aleatória e Árvore de Decisão. Os autores concluíram que a Floresta Aleatória trouxe os resultados mais precisos. Sahin et al. (2020), diante dos desafios encontrados na utilização de diversos softwares para a modelagem e a exigência, muitas vezes, de conhecimento de programação, desenvolveram uma ferramenta específica para aplicação do algoritmo Floresta Aleatória em ambiente ArcGIS.

3 PROPOSTA DA PESQUISA

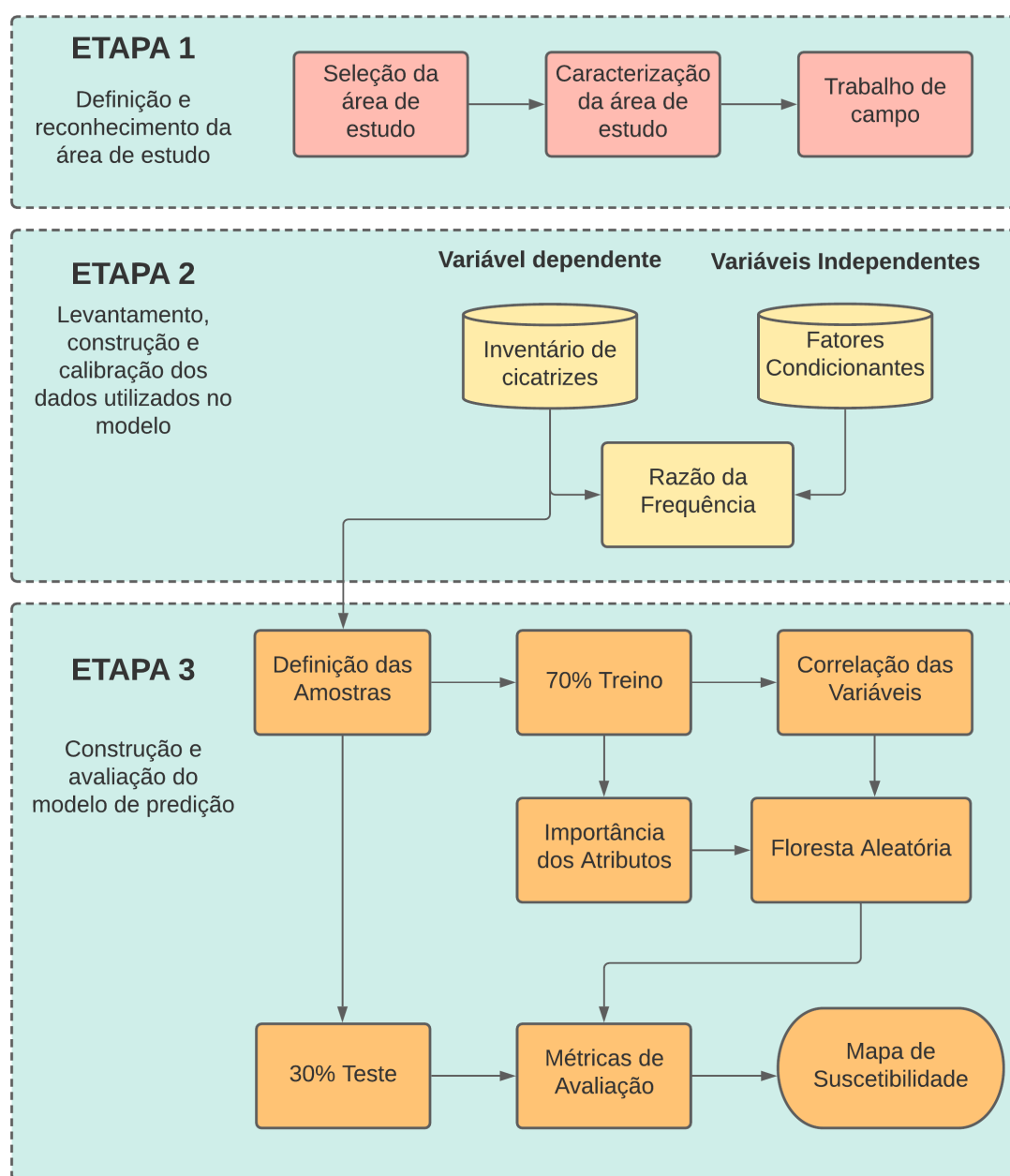
Esse trabalho baseia-se na premissa de que as técnicas de aprendizado de máquina permitem novas possibilidades para obtenção de mapas de suscetibilidade e podem auxiliar na gestão do risco de desastres, principalmente na fase de prevenção. Dessa forma, o trabalho tem como objetivo principal aplicar um método supervisionado de aprendizado de máquina, a Floresta Aleatória, para o mapeamento da suscetibilidade a escorregamentos na Bacia do Córrego Piracuama. Os objetivos específicos são:

- a) Identificar locais em que ocorreram escorregamentos na Bacia do Córrego Piracuama, no período de 2000 até o presente, através de bancos de dados públicos, imagens de satélite e trabalho de campo, resultando em um inventário de cicatrizes;
- b) Identificar quais fatores condicionantes possuem maior relevância no modelo de suscetibilidade a escorregamentos;
- c) Analisar, para cada fator condicionante, quais os intervalos de classe que concentram a maior probabilidade na deflagração dos escorregamentos;
- d) Verificar a acurácia do modelo construído a partir de distintas métricas de avaliação, identificando o desempenho do algoritmo aplicado à área de estudo;
- e) Entender, através da comparação com outros mapas já existentes para a área, quais as diferenças e semelhanças entre os resultados obtidos, levando em consideração a aplicação de metodologias distintas de mapeamento.

4 MATERIAIS E MÉTODOS

A Figura 9 descreve o fluxograma que representa a metodologia do presente trabalho. As três etapas que nortearam esse estudo são divididas em: i) definição e reconhecimento da área de estudo; ii) levantamento, construção e calibração dos dados utilizados no modelo e iii) construção e avaliação dos modelos de predição.

Figura 9. Fluxograma da dissertação



Fonte: produção do autor.

4.1 Etapa 1

4.1.1 Seleção da área de estudo

Um estudo realizado por Brollo e Ferreira (2016) levantou o cenário de desastres do Estado de São Paulo entre 2000 a 2015 e propôs um sistema de indicadores de situação e resposta para tornar mais eficaz a gestão do risco, baseado em cinco variáveis: número de acidentes, número de óbitos, número de pessoas afetadas, número de edificações afetadas e número de municípios com instrumentações de risco (ferramentas que apoiam os municípios a enfrentarem os cenários de riscos de desastres).

Com índices significativamente relevantes, Campos do Jordão se destacou em relação aos demais municípios no total de instrumentações de risco (Quadro 3), visto que se encontra entre os três municípios mapeados com os maiores valores desses recursos e número de acidentes do tipo geológico (dentre os 1.430 acidentes registrados, 24 ocorreram no local), estando entre os dez municípios mais afetados por esse tipo de evento no Estado (Quadro 4). Diante desse contexto, optou-se por escolher o município de Campos do Jordão como área de estudo.

Quadro 3. Indicadores de resposta de Campos do Jordão

Indicadores de Resposta		
Instrumentos de Risco	Quantidade	Ano de Adesão
Mapeamento de Risco	2	2002, 2014
Setorização	0	N/A
Plano Municipal de Redução de Risco (PMRR)	1	2006
Carta de Suscetibilidade	1	2014
Plano Preventivo de Defesa Civil (PPDC)	1	2000
Campanha Cidades Resilientes	1	2014

Fonte: Brollo e Ferreira (2016).

Quadro 4. Indicadores de estado ou situação de Campos do Jordão

Indicador	Variável	Muito alto	Alto	Médio	Baixo	Divisão das Classes com base nas ocorrências/Números de Municípios
Número de Acidentes	Geológicos					Muito alto: 20 a 84 (16); Alta: 6 a 20 (58); Média: 1 a 6 (160); Baixa: 0 a 1 (411)
	Hidrológicos					Muito alto: 1.180 a 1.190 (1); Alta: 40 a 1.180 (33); Média: 10 a 40 (96); Baixa: 0 a 10 (515)
	Meteorológicos					Muito alto: 356 a 356 (1); Alta: 15 a 356 (36); Média: 3 a 15 (143); Baixa: 0 a 3 (465)
	Climatológicos					Muito alto: 47 a 72 (3); Alta: 5 a 47 (43); Média: 1 a 5 (232); Baixa: 0 a 1 (367)
	Total					Muito alto> 1.690 (1); Alta: 60 a 1.690 (38); Média: 10 a 60 (156); Baixa: <10
Número de Óbitos	Óbitos					Muito alto: 16 a 65 (4); Alta: 4 a 16 (35); Média: 1 a 4 (119); Baixa: 0 a 1 (487)
Número de pessoas afetadas	Pessoas afetadas					Muito alto: >118.000 (2); Alta: 17.000 a 118.000 (16); Média: 1.000 a 17.000 (49); Baixa: < 1000(578)
Número de edificações afetadas	Edificações afetadas					Muito alto: 55.200 a 55.300 (1); Alta: 2.100 a 55.200 (7); Média: 100 a 2.100 (62); Baixa: 0 a 100 (575)

Fonte: Brollo e Ferreira (2016).

4.1.2 Caracterização da área de estudo

Conhecido por muitos como a “Suíça Brasileira”, Campos do Jordão está localizado na Unidade de Gerenciamento de Recursos Hídricos (UGRHI) 1 – Serra da Mantiqueira, na porção leste do Estado de São Paulo a uma altitude de 1.628 metros. O município possui uma área de 290,5 km², com uma população estimada de 52.402 habitantes (IBGE, 2020) e é considerado uma das principais estâncias turísticas do Estado de São Paulo, tendo como principais atividades econômicas o turismo, a indústria de confecção de malhas e de chocolate, o artesanato e a exploração da água mineral (ANDRADE et al., 2014).

4.1.2.1 A Bacia do Córrego Piracuama

A escolha por uma área específica e que apresentasse um contexto de risco configurou mais uma etapa dessa pesquisa. Dessa forma, devido ao registro histórico de escorregamentos na área, bem como a presença de fatores que intensificam a suscetibilidade desses processos

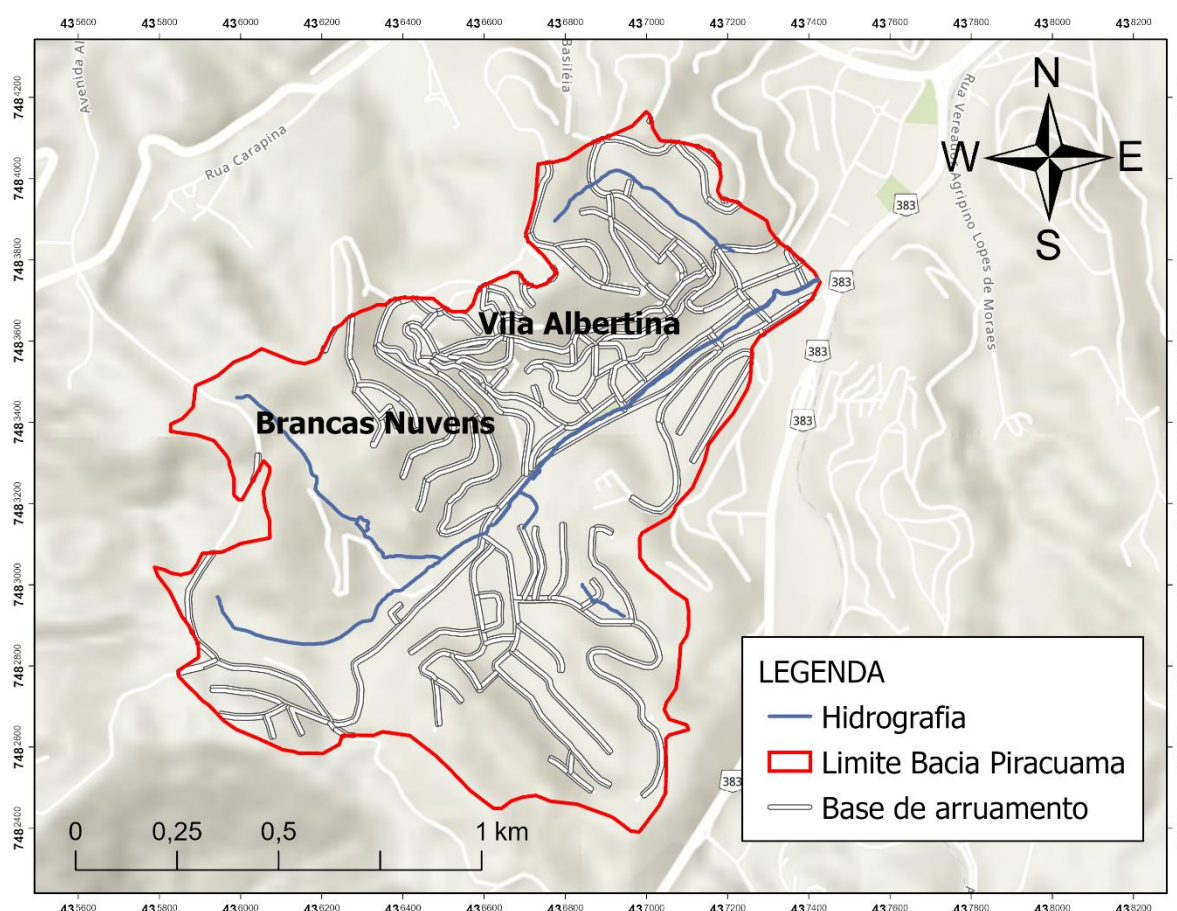
pontuados por König et al. (2019) - declividades acentuadas, cortes verticais nas encostas, encanamentos expostos e com vazamentos, acúmulo de lixo e entulhos etc. – optou-se por concentrar os estudos na Bacia do Córrego Piracuama (Figura 10). A bacia está localizada na porção sudoeste da mancha urbana de Campos do Jordão e sua área de abrangência foi delimitada por Prieto (2018) através de dados topográficos disponibilizados pelas instituições IPT e CPRM.

A Bacia do Córrego Piracuama apresenta uma área de 1,57 km² e compreende os bairros Vila Albertina e Brancas Nuvens. De acordo com Vieira et al. (2015), a ocupação da Vila Albertina é caracterizada pelo predomínio de moradias de alvenaria, madeiras e mistas de baixo a médio padrão construtivo. O bairro encontra-se em consolidação, possuindo 30 a 80% da área ocupada, cujas meias encostas estão associadas a porções não consolidadas e as áreas de base, laterais e topo às porções consolidadas. A ocupação do fundo do vale, próxima ao Córrego Piracuama, apresenta-se consolidada e densa com moradias majoritariamente de alvenaria e padrão construtivo médio. Já o bairro Brancas Nuvens encontra-se em geral em estágio de ocupação consolidado (mais de 80% de ocupação), com porções em consolidação (30 a 80% de ocupação), sendo que os terrenos em geral são grandes e as moradias possuem padrão construtivo variando de muito baixo a bom. O tipo construtivo predominante é de alvenaria, embora também existam moradias em madeira (IG-SMA, 2014).

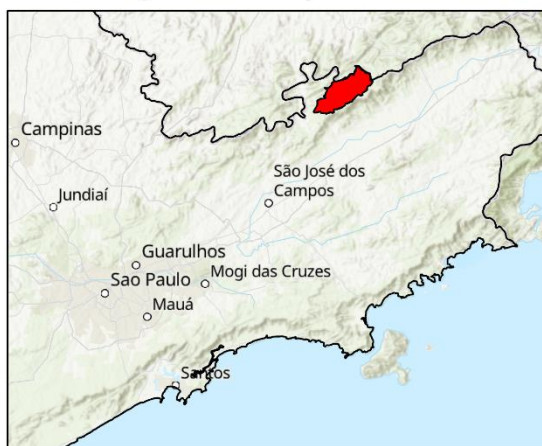
De acordo com o mapeamento realizado pelo Instituto Geológico em 2014, a Vila Albertina compreende a mais extensa área de risco do município de Campos do Jordão, envolvendo aproximadamente 180.000 m². Ao todo, o bairro apresenta 13 setores de risco – nove relacionados a escorregamentos e quatro à inundação – abrangendo 750 moradias em risco. Dos nove setores de risco de escorregamentos, dois são classificados como baixo, cinco como médio, um como alto e um é classificado como muito alto. Os setores de risco de escorregamentos (Quadro 5) abrangem uma área de 165.000 m² e compreendem 706 moradias, sendo que o setor classificado como risco muito alto corresponde a 56% dessa área.

Em relação ao bairro Brancas Nuvens, a área apresenta cinco setores de risco (quatro setores de escorregamento e um de inundação), aos quais se associam 178 moradias em risco. Dos quatro setores de risco de escorregamentos, três deles apresentam risco médio e um apresenta risco alto, aos quais se associam 177 moradias (Quadro 6). Os setores de risco consolidados são apresentados na Figura 11.

Figura 10. Localização da área de estudo

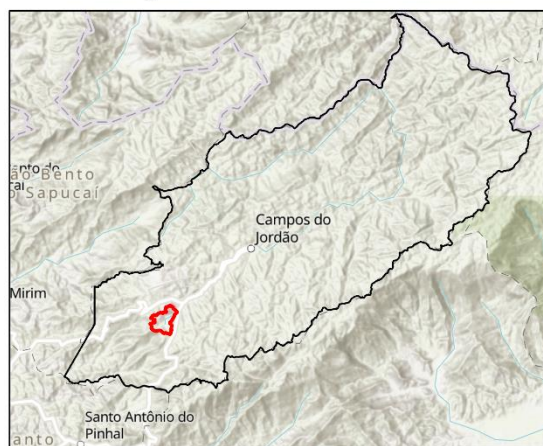


Localização - Campos do Jordão



Sistema de Projeção: UTM 23S
Datum: SIRGAS 2000

Localização - Bacia Piracuama



Fonte dos dados:
Hidrografia: IPT; CPRM, 2014
Limite Bacia: Prieto, 2018
Base de arruamentos: IBGE, 2019

Fonte: produção do autor.

Quadro 5. Descrição dos setores de risco do bairro Vila Albertina

Grau de Risco	Nº de setores/ moradias/moradores	Processos de instabilização esperados	Características do setor
Baixo (A18S1R1ESC e A18S8R1esc)	2; 181; 724	Escorregamento planar raso em taludes de corte e de aterro	Declividade entre 18° a 23°; taludes de corte com alturas entre 3 a 4 m; bom padrão construtivo; presença de sistema de drenagem urbana
Médio (A18S3R2esc, A18S4R2esc, A18S5R2esc, A18S7R2esc e A18S9R2esc)	5; 148; 592	Escorregamento planar raso em taludes de corte Rastejo	Declividade entre 18° a 30°; taludes de corte com alturas entre 3 a 4m; são observadas moradias próximas às bases dos taludes de corte e na encosta; presença de cicatriz de escorregamento; possível área de atingimento caso ocorra deslizamento no setor de risco muito alto
Alto (A18S6R3esc)	1; 27; 108	Escorregamento planar raso em taludes de corte e de aterro	Morfologia da encosta na lateral do anfiteatro; alta declividade (32°); altura excessiva dos taludes de corte (até 7 m)
Muito alto (A18S2R4esc)	1; 350; 1400	Rastejo Escorregamento planar raso Queda e rolamento de blocos em taludes naturais, de corte e de aterro.	Histórico de eventos muito graves em 1972 e 2000; recorrência anual de forma localizada; declividade muito alta (30 a 40°); concentração de linhas de drenagem, dada por anfiteatros e talwegues em situação de alta declividade; ocupação precária com destaque para moradias de alta vulnerabilidade instaladas nos talwegues e ausência de infraestrutura.

Fonte: modificado de IG-SMA (2014) e Andrade et al. (2015).

Dentre as recomendações propostas pelo IG para esses setores, que variam de acordo com o grau de risco, estão: o monitoramento das áreas, serviços de limpeza e recuperação, manutenção dos sistemas de captação e condução das águas pluviais, implantação de obras estruturais, congelamento e fiscalização da área impedindo a ampliação da ocupação, remoções preventivas e imediatas, revegetação da encosta etc.

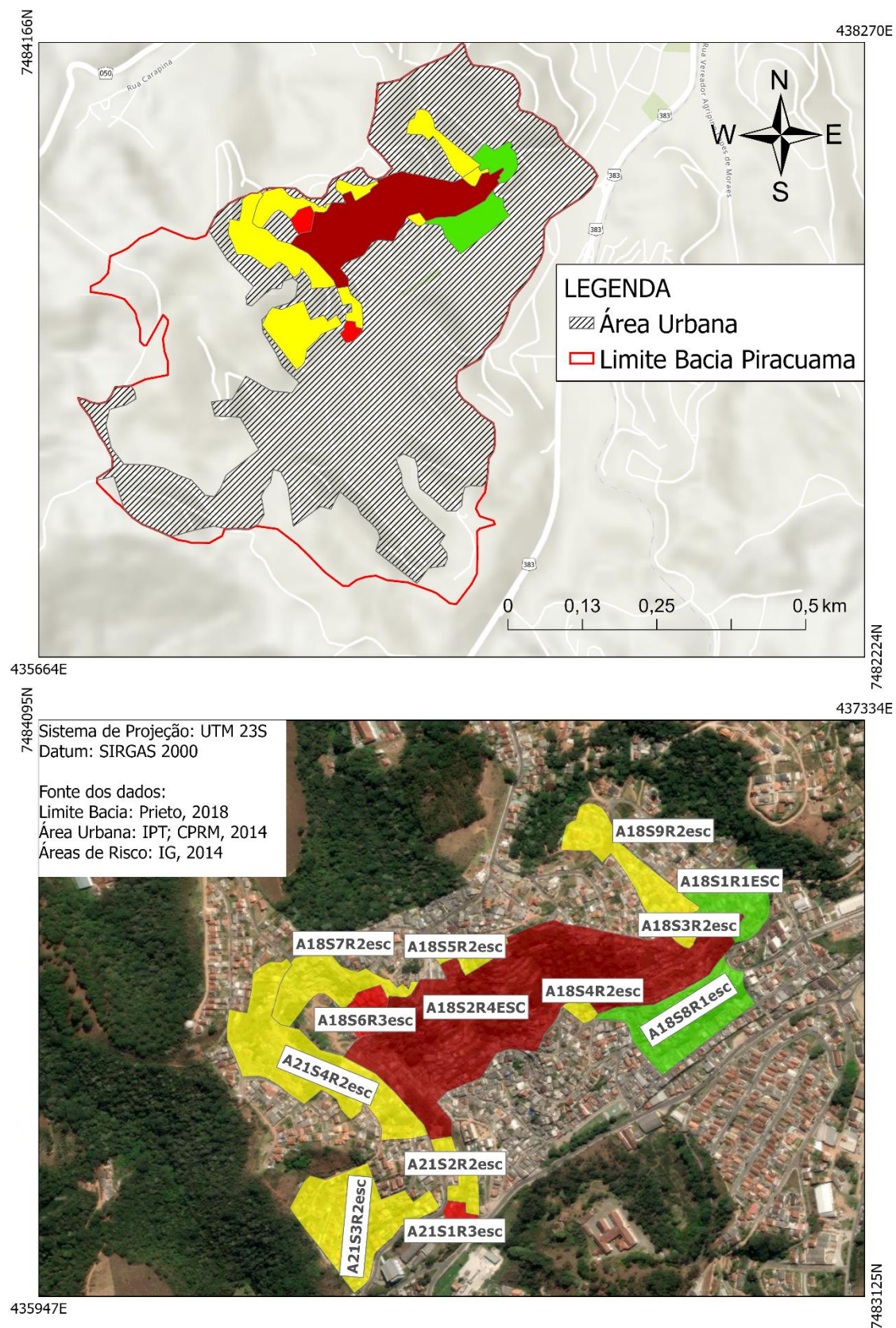
Quadro 6. Descrição dos setores de risco do bairro Brancas Nuvens

Grau de Risco	Nº de setores/ moradias/moradores	Processos de instabilização esperados	Características do setor
Médio (A21S2R2esc, A21S3R2esc e A21S4R2esc)	3; 166; 664	Escorregamento planar raso e erosão em sulcos em talude de corte; rastejo	Declividade alta (entre 24° a 30°); proximidade das moradias ao topo do talude de corte; precariedade do sistema de drenagem; morfologia natural da encosta, com linha de drenagem cortando o setor; taludes de corte com alturas entre 3 a 6m;
Alto (A21S1R3esc)	1; 11; 44	Escorregamento planar raso e erosão em sulcos em talude de corte e de aterro	Alta declividade (28°); alta vulnerabilidade das moradias; proximidade das moradias ao topo do talude de aterro (0 m); concentração de águas em superfície; precariedade do sistema de drenagem.

Fonte: modificado de IG-SMA (2014) e Andrade et al. (2015).

De acordo com Prieto (2018), a ocupação urbana na bacia do Córrego Piracuama é observada significativamente em terrenos com declividades superiores a 15° e em certas localidades em áreas com declividades superiores a 25°, contradizendo a legislação do Parcelamento do Solo Urbano, que não recomenda a ocupação urbana em terrenos com declividades iguais ou superiores a 17° (30% de declividade).

Figura 11 Setores de risco localizados na área de estudo



Fonte: produção do autor.

4.1.2.2 Histórico de movimentos de massa

Campos de Jordão possui registros de ocorrência de movimentos gravitacionais de massa desde a década de 1970. O histórico do município, construído através de artigos, documentos, mídias digitais e do banco de dados S2ID, subsidiaram no conhecimento dos processos atuantes no local, na compreensão das consequências e danos relacionados, na identificação das áreas de risco e no levantamento de cicatrizes que indicam o local de ocorrência desses processos. Os registros de movimentos de massa, descritos no Quadro 7 foram levantados a partir das ocorrências localizadas nos bairros Vila Albertina e Brancas Nuvens.

De acordo com Mendes et al. (2018), os escorregamentos que ocorrem em Campos do Jordão geralmente são classificados como translacionais rasos, sendo caracterizados por superfícies de rupturas menores que 2 metros. Durante a deflagração desses processos, a superfície de ruptura pode ocorrer em solo residual de encostas naturais, em solo residual de um corte em declive ou em uma base de um depósito de aterro.

Quadro 7. Registros de eventos ocorridos na Vila Albertina e Brancas Nuvens

(continua)

Data do Evento	Localização	Danos	Informações Adicionais	Fonte
18/08/1972	Vila Albertina	17 óbitos; 60 casas soterradas; 1000 desabrigados	Deslizamento de cerca de 70.000 m ³ de lama altamente aquosa e rica em matéria orgânica. O material se liquefez devido à carga de um pequeno aterro, talvez auxiliada pela vibração provocada pelo trabalho do trator que transportava a terra (Amaral & Fuck, 1973)	Mendes et.al (2018); Amaral & Fuck (1973)
07/02/1999	Vários bairros, incluindo Vila Albertina	60 famílias desalojadas; prejuízos avaliados em R\$ 2 milhões; mais de 300 pessoas abandonaram suas casas	Em 7/02/1999 foi decretado Estado de Calamidade Pública	IG-SMA (2014)

Quadro 6. Registros de eventos ocorridos na Vila Albertina e Brancas Nuvens

(conclusão)

31/12/1999 a 04/01/2000	Vários bairros, incluindo Vila Albertina	10 óbitos; 1840 desabrigados; 103 feridos	Acumulado de 5 dias de chuvas, atingindo o índice de 453,2 mm.	Mendes et.al (2018) IG-SMA (2014)
16/12/2001	Vila Albertina	1 deslizamento em local desabitado. A Defesa Civil colocou o bairro em Alerta	Volume acumulado em 4 dias: 71.6mm	IG-SMA (2014)
19/01/2005	Brancas Nuvens	1 deslizamento de terra	Não há informações adicionais	IG-SMA (2014)
04/01/2007	Vários bairros, incluindo Vila Albertina	1 muro cedeu; remoção preventiva de 4 casas.	Medida preventiva de remoção contra risco de deslizamento em 50 casas durante o período de chuvas nos bairros.	IG-SMA (2014)
01/01/2010	Vários bairros, incluindo Vila Albertina	04 casas atingidas por deslizamentos; 12 famílias desalojadas no município	No Vale do Paraíba, a chuva matou 5 e desalojou 2500 pessoas	IG-SMA (2014)
11/01/2013	Vários bairros, incluindo Vila Albertina e Brancas Nuvens	05 casas foram atingidas; 67 pessoas desalojadas	Após concluir vistorias iniciais, foram constatados 10 locais de escorregamentos. Acumulado de chuva em 3 dias: 105.5 mm.	S2ID, relatório CONDEC
08/01/2018	Vários bairros, incluindo Vila Albertina e Brancas Nuvens	O caso mais grave é de uma casa que precisou ser interditada na Vila Albertina	Em 3 dias, a cidade registrou acumulado de 133,4 mm na região do Capivari e 138 mm na área do bairro Brancas Nuvens. Foram atendidas 21 ocorrências entre deslizamentos e quedas de árvores.	Jornal RC Vale

Nota: quando a localização engloba mais de um bairro afetado, os danos descritos se referem a todas as áreas.

Fonte: produção do autor.

4.1.2.3 Histórico de ocupação

A ocupação do município teve sua origem no século 18, em decorrência da vinda dos portugueses para a região. Posteriormente, no início do século 20, as ocupações se intensificaram devido ao tratamento oferecido pela cidade para doenças pulmonares, principalmente a tuberculose (MENDES et al., 2018a). Entretanto, a expansão da mancha urbana teve uma significativa aceleração a partir de 1970 em razão da construção da rodovia SP-123, que facilitou a ligação entre a capital e o município. Nessa fase, Campos do Jordão vivenciou diversos processos ligados à urbanização como fluxos migratórios de pessoas, diminuição das atividades rurais e predominância das atividades terciárias e de construção civil (ROSA FILHO, 2006).

A ausência de políticas públicas, planejamento territorial e planos habitacionais populares consistentes intensificaram as ocupações desordenadas em áreas irregulares, principalmente pela população de baixo poder aquisitivo. Desse modo, como forma de contornar o alto custo de solo das regiões centrais e planas, a construção de moradias em áreas de encosta tornou-se comum. Essas habitações são caracterizadas pelo baixo padrão construtivo e precário, cujas condições geotécnicas são desfavoráveis à ocupação.

Ahrendt (2005) ressalta que as encostas se encontram muitas vezes ocupadas de maneira inadequada, visto que os aterros e cortes, frequentemente, são realizados pelos próprios moradores, resultando em cortes praticamente sub-verticais e aterros pouco compactados, compostos por materiais heterogêneos. Além disso, o destino do lixo doméstico, muitas vezes dispostos ao ar livre e em terrenos baldios por vezes íngremes, também aumentam a instabilidade dessas encostas. Outro fator agravante citado por IPT (2002) diz respeito à baixa infraestrutura urbana desses locais, caracterizada pela ausência dos sistemas de drenagem pluvial, pavimentações e sarjetas que, devido à concentração de água, intensificam e aceleram processos erosivos e aumentam a suscetibilidade a movimentos gravitacionais de massa.

4.1.2.4 Clima

De acordo com Modenesi-Gauttieri e Hiruma (2004), a região apresenta um clima tropical de altitude, caracterizado por temperaturas médias anuais de 14,3° C e precipitação anual entre 1.205 e 2.800mm. As chuvas que ocorrem no período de outubro a março

representam 80% da concentração pluviométrica total e estão relacionadas principalmente às frentes frias, que incidem sobretudo na face sul do Planalto de Campos do Jordão.

Dados de 40 anos de observação apontam a frequência das geadas nos meses de inverno, variando de 9 a 70 dias por ano (MODENESI-GAUTTIERI; HIRUMA, 2004). O clima é caracterizado por verões brandos e úmidos e invernos secos.

4.1.2.5 Geologia, pedologia e geomorfologia

De acordo com Nunes et al. (2020), o município de Campos do Jordão está inserido em uma região transicional entre as Províncias Mantiqueira (dividida nos cinturões Araçuaí, Ribeira e Tijucas) e Tocantins (dividida nos cinturões Brasília, Araguaia e Paraguai), complexas entidades geotectônicas resultantes de diferentes cenários evolutivos (HASUI, 2010). A formação dessas entidades ocorreu durante o Ciclo Brasileiro (900-490 Ma), período que compreende a fragmentação do Supercontinente Rodínia e a colagem do Supercontinente Gondwana, resultando em diversos eventos de convergência, fechamento de oceanos que separavam os blocos continentais e retrabalhamento de rochas mais antigas. A consolidação final desse período se deu com o colapso dos orógenos, acompanhada de intrusões e vulcanismos pós-colisionais, ativa erosão e formação de riftes (HASUI, 2010). As províncias guardam o registro de uma longa e complexa evolução, preservando rochas de idade neoproterozóica e núcleos de rochas remanescentes de idade Arqueana, Paleoproterozoica e Mesoproterozoica (HEILBRON et al., 2004).

Após o Ciclo Brasileiro, a região ainda foi submetida a outros episódios tectônicos relacionados a diferentes períodos na escala geológica, como o desenvolvimento de sistemas estruturais extensionais durante o NeoJurássico, devido a ruptura do Gondwana Ocidental e a abertura do Oceano Atlântico Sul, bem como a reativação de antigas zonas de cisalhamento durante o Paleógeno, que originou o *Rift Continental do Sudeste do Brasil* (RSCB). Na porção leste do Estado de São Paulo, a evolução dessa feição possibilitou a configuração atual do relevo, representado pelas Serras do Mar e Mantiqueira, como feições soerguidas, e a Bacia de Taubaté, como feição rebaixada (HIRUMA; RICCOMINI; MODENESI-GAUTTIERI, 2001).

Em relação aos tipos de rochas, a região é caracterizada por afloramentos de terrenos cristalinos, representados predominantemente por gnaisses e migmatitos, intercalados com granitos, xistos, quartzitos, calcários, calciossilicáticas e anfibolitos (IPT, 2002). O mapeamento geológico mais recente que abrange o município de Campos do Jordão foi

realizado pelo Serviço Geológico do Brasil (CPRM) em parceria com a Universidade Federal do Rio de Janeiro (UFRJ) em 2020. Essas duas instituições fazem parte do Programa Nacional de Geologia (PRONAGEO) que visa promover avanços na cartografia geológica básica na escala 1:100.000 em todo território nacional, através de levantamentos de dados geológicos de campo, laboratoriais e da literatura, análises via sensoriamento remoto e interpretação de dados aerogeofísicos de alta resolução. O mapa geológico do município é resultante da junção de duas folhas topográficas: Campos do Jordão (SF.23-Y-B-V) e Pindamonhangaba (SF.23-Y-B-VI).

Conforme Nunes et al. (2020), as duas folhas topográficas estão inseridas numa zona de interferência ou superposição entre a Faixa Brasília Meridional e Faixa Ribeira. De uma maneira geral, a geologia de Campos do Jordão pode ser dividida em dois domínios principais: *Nappe Socorro* e Complexo Embu. O primeiro, configura um antigo arco magmático formado durante o primeiro evento colisional na região (630 a 600 Ma), representado pelo Batólito Serra da Água Limpa e corpos correlatos (TROUW et al. 2013; VINAGRE et al. 2014). O Complexo Embu, é definido como um compartimento tectônico caracterizado por uma sucessão metassedimentar com embasamento Paleoproterozoico, intrudido por rochas de composição granítica. Entretanto, estudos recentes levantam uma nova hipótese para essa região, que consideram o Complexo Embu como a continuação da *Nappe Socorro*, visto que os mapeamentos atuais revelam uma transição gradual de um domínio para o outro, com rochas similares em ambas as localidades, incluindo o Batólito Serra da Água Limpa (TROUW et al., 2013). As unidades litoestratigráficas inseridas em cada domínio são descritas no Quadro 8 a seguir e representadas espacialmente no mapa geológico na Figura 12. A Bacia do Córrego Piracuama abrange as unidades Complexo Gnaíssico Serra Pedra do Baú, Complexo Gnaíssico da Água Limpa e Unidade Itererê.

Apesar de toda a variedade descrita nas unidades litoestratigráficas, há outro mapa geológico para a área de estudo realizado pela DAEE e UNESP em 1984 e numa escala de 1:250.000, que indica apenas duas unidades litológicas para o município – gnaisses e migmatitos (Figura 13). De acordo com Azevedo (2019), os migmatitos são rochas metamórficas heterogêneas formadas pela fusão parcial em auréolas de contato ou no interior de orógenos colisionais. Além de serem formados em temperaturas muito elevadas, os migmatitos são constituídos por duas partes petrograficamente distintas em uma ou mais características (cor, granulação, composição mineralógica, deformação etc.) chamadas fácies. Estas, caracterizadas pela alternância de bandas, são separadas em paleossoma e neossoma,

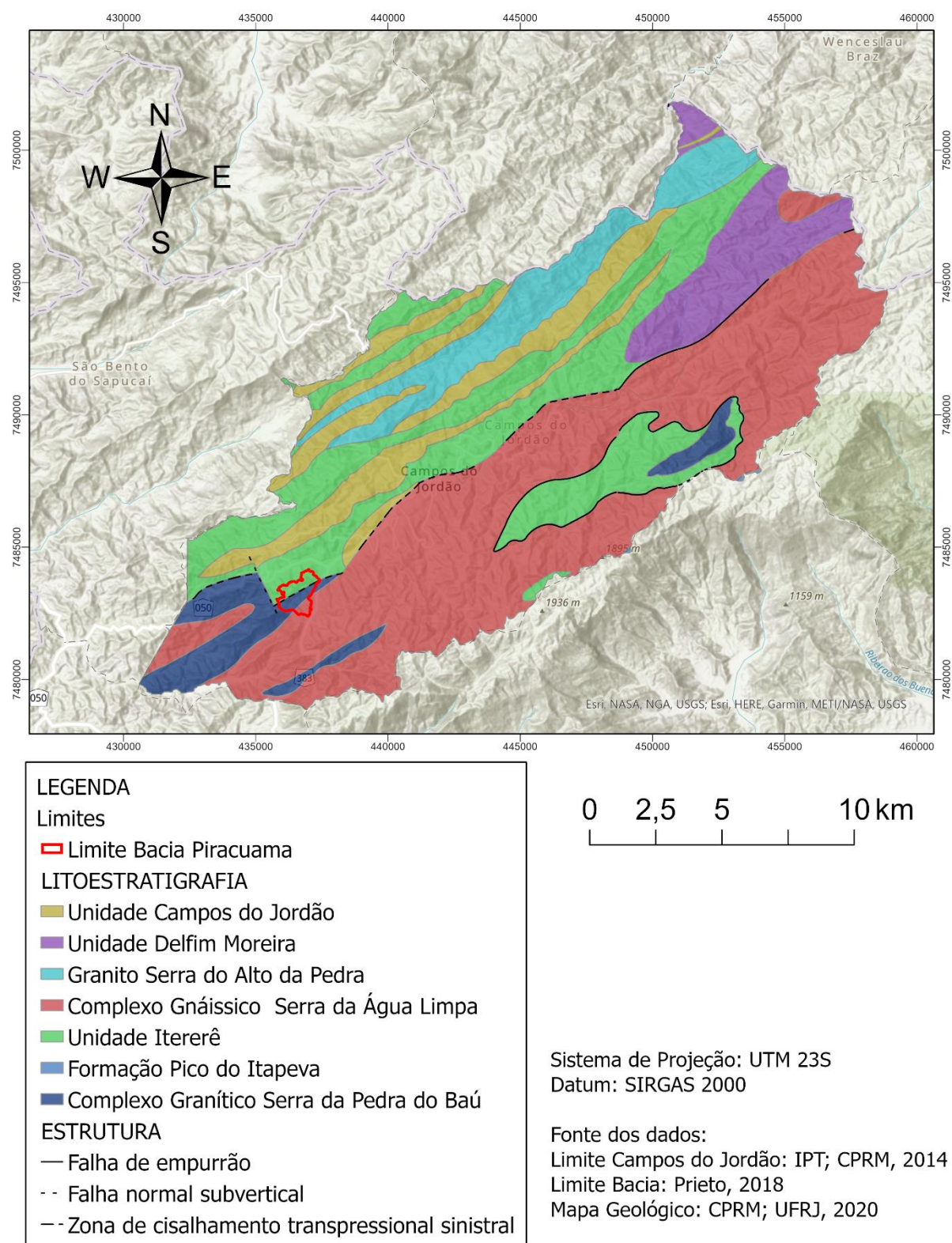
sendo que o primeiro possui natureza gnáissica e estrutura mais antiga e o segundo, mais jovem, apresenta origem granítica.

Quadro 8. Unidades litoestratigráficas que abrangem o município Campos do Jordão

Domínios Geológicos	Litoestratigrafia	Período	Classe	Descrição
<i>Nappe Socorro</i>	Complexo Gnaissico Serra Pedra do Baú	Criogeniano	Metamórfica	Ortognaisses graníticos e granodioríticos
<i>Nappe Socorro e Complexo Embu</i>	Complexo Gnaissico da Água Limpa	Criogeniano	Metamórfica	Ortognaisse e metagranito
<i>Complexo Embu</i>	Unidade Delfim Moreira	Criogeniano	Metamórfica	Biotita gnaisse migmatítico com intercalações de anfibolito e quartzito
	Unidade Campos do Jordão	Toniano	Metamórfica	Quartzito puro a micáceo e feldspático, com intercalações de (sillimanita)-biotita e/ou muscovita xisto
	Unidade Itererê	Toniano	Metamórfica	Rocha metaultramáfica, quartzito feldspático, rocha calcissilicática, xisto, anfibolito
	Formação Pico do Itapeva	Ediacarano	Metamórfica	Riolito, metarritmito, metarenito arcoseano, pelito, ritmito, conglomerado, arenito
<i>Corpos Intrusivos</i>	Granito Serra do Alto da Pedra	Ediacarano	Ígnea	Biotita granito intrudindo rochas das unidades Campos do Jordão e Itererê

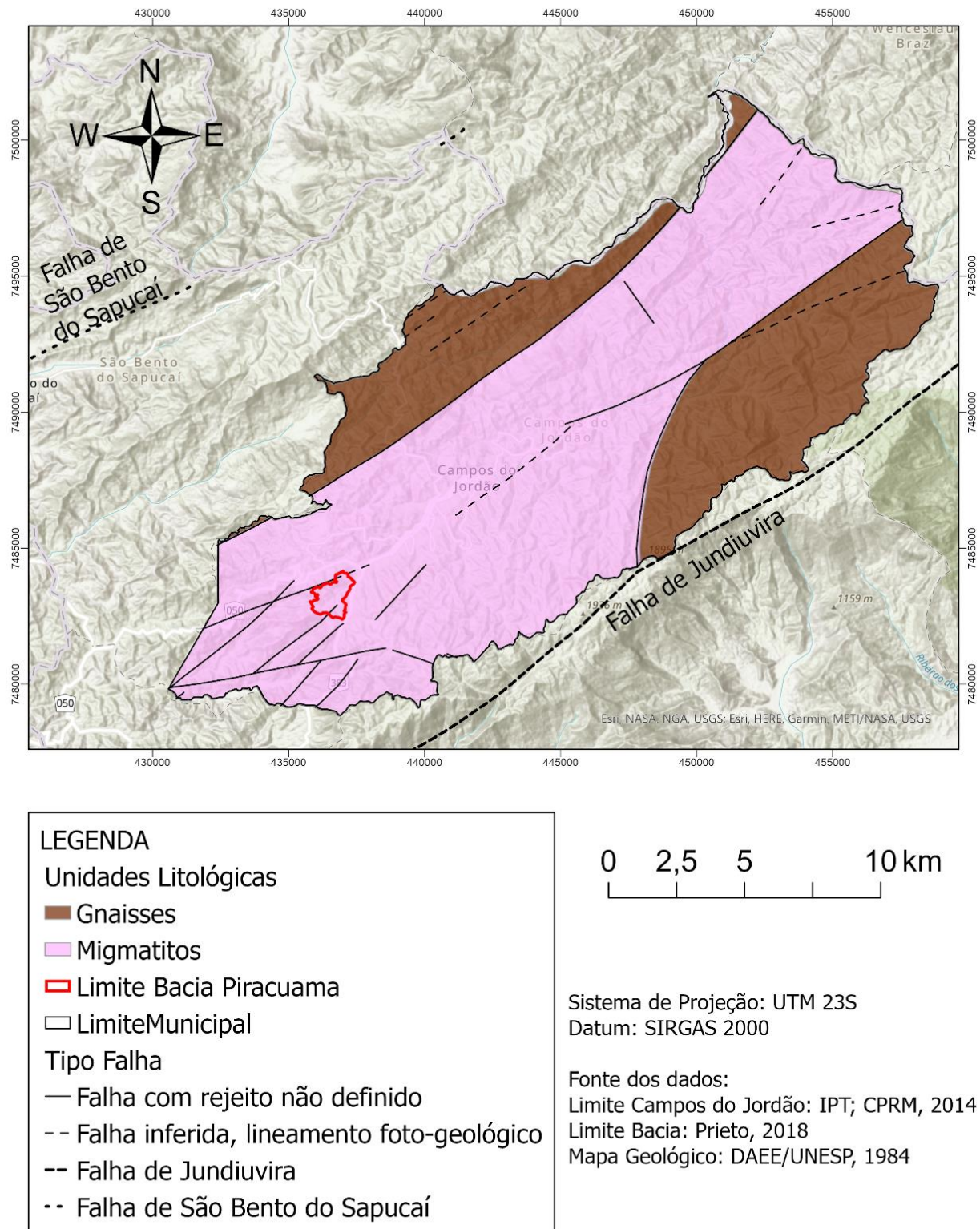
Fonte: produção do autor com base nos dados de CPRM; UFRJ (2020).

Figura 12. Mapa geológico de Campos de Jordão



Fonte: produção do autor.

Figura 13. Mapa geológico de Campos de Jordão



Fonte: produção do autor.

Em termos tectônicos, o planalto é delimitado por duas grandes falhas transcorrentes reativadas – a Falha de Jundiuvira e Falha do Paiol Grande ou São Bento do Sapucaí – que apresentam direção NE e idade pré-cambriana a eopaleozóica. Além disso, Hiruma et al. (2001) descrevem eventos neotectônicos ocorridos na região, que configuraram um papel decisivo na formação da topografia contemporânea. A reativação da Zona de Cisalhamento Jundiuvira, por exemplo, provocou falhamentos recentes (Neopleistoceno e Holoceno) e o desenvolvimento de feições morfotectônicas como anfiteatros suspensos, vales assimétricos e escarpas retilíneas. Feições como cristas truncadas, divisores de água, capturas de drenagem e rios em gancho também podem estar relacionadas a falhamentos. Os lineamentos estruturais, resultantes da ação tectônica atuante em períodos distintos, podem estar associados a deflagração de processos geomorfológicos como a erosão e os movimentos de massa.

Em relação à geomorfologia, o Planalto de Campos de Jordão, localizado no bloco principal da Serra da Mantiqueira e inserido no contexto da Província Geomorfológica do Planalto Atlântico, é definido como uma superfície tectonicamente elevada, de estrutura cristalina complexa e com altitude superior a 2.000 metros (HIRUMA; RICCOMINI; MODENESI-GAUTTIERI, 2001). O Planalto é limitado por escarpas abruptas com cerca de 1.500 metros que se erguem sobre as colinas do médio Vale do Paraíba. De acordo com a classificação proposta por Ross (1992), nesse contexto geomorfológico predominam relevos de morros e serras, caracterizando-se pelas altas amplitudes e declividades, além da dissecação generalizadas com desgaste nas vertentes.

Modenesi-Gauttieri e Hiruma (2004) classificaram a região em unidades, de acordo com um critério de combinações específicas de topografia e formas de relevo, substrato, processos atuantes e cobertura vegetal. Dentre esses compartimentos, destacam-se as depressões turfosas, localizadas nas bases de anfiteatros de erosão e caracterizadas por constituírem depósitos de argila orgânica que variam de espessura. Apesar da alta sensibilidade a intervenções antrópicas, esses depósitos muitas vezes encontram-se aterrados embasando edificações no município de Campos do Jordão, resultando em processos de instabilização e situações de risco (MENDES et al., 2018a).

Segundo IG-SMA (2014), o perfil geotécnico da região é bastante semelhante, apresentando variações texturais que acompanham a própria heterogeneidade do maciço rochoso migmatítico. Os solos residuais são encontrados predominantemente nas encostas e são considerados rasos, com espessuras que variam de 0,5 a 2 metros. Os horizontes do solo residual, que apresentam uma evolução pedológica pouco desenvolvida, são formados

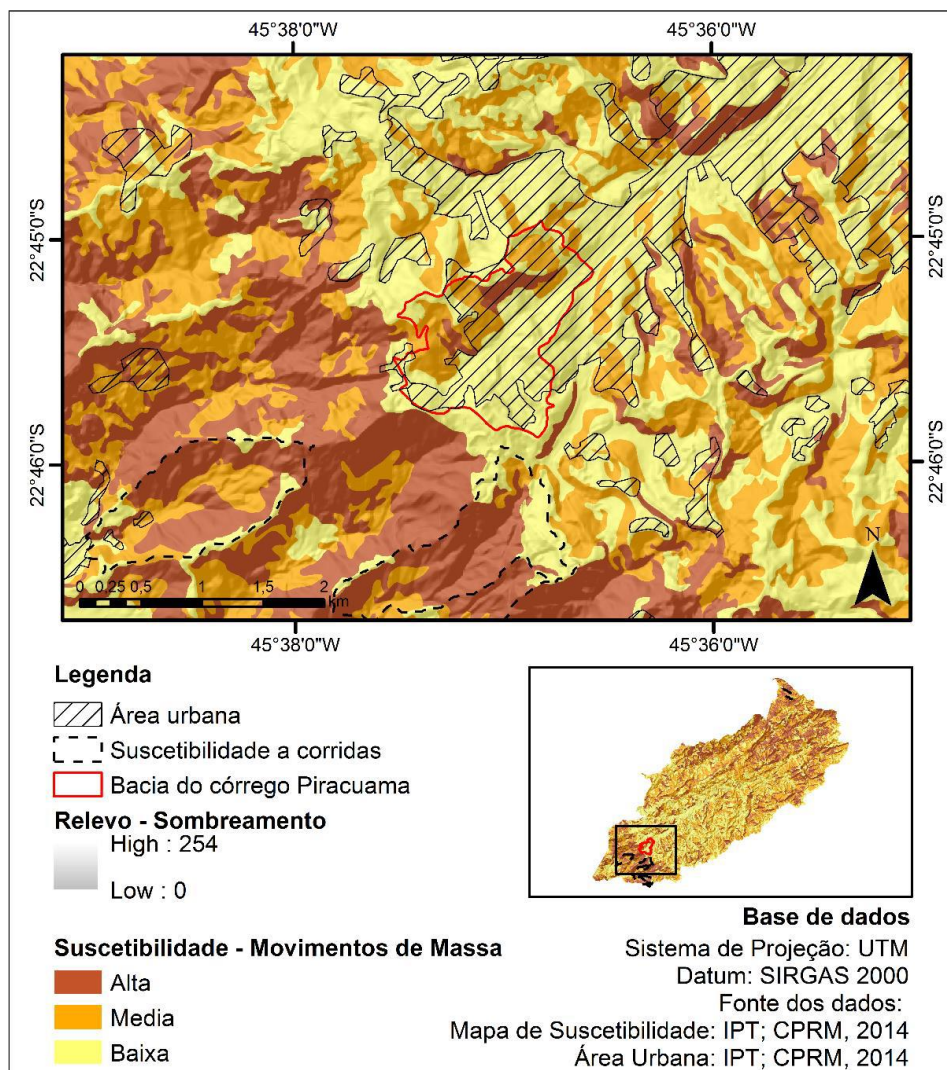
principalmente por silte e areia fina, com possíveis ocorrências de argila, matéria orgânica e raízes próximo à superfície. O material saprolítico, formado subjacente aos horizontes residuais, possui textura predominantemente areno argilosa, que varia de proporção conforme a profundidade. O solo saprolítico é resultante da decomposição da rocha matriz, geralmente preservando as características texturais e estruturais de sua origem como as estruturas reliquias de planos de foliação oriundos das rochas gnáissicas. Na região, esses materiais variam de 1 a 5 metros de profundidade, exceto em porções isoladas em que as espessuras encontradas são maiores (IPT, 2002).

4.1.2.6 Instrumentações de gestão de risco

Campos do Jordão possui diversos instrumentos de gestão - elaborados pelas esferas municipais, estaduais e federais - que apoiam na prevenção e mitigação de riscos decorrentes de desastres naturais. Segundo Vieira et al. (2015), esses principais instrumentos, que embasam a tomada de decisão do município, são:

- a) Mapeamento de perigos e riscos, realizados pelo Instituto Geológico e Instituto de Pesquisas Tecnológicas: abrange estudos que apoiam a elaboração de estratégias e implementação de ações para redução de riscos de desastres, são eles: “Avaliação de riscos de escorregamentos, inundação, erosão e solapamento de margens de drenagens” representado parcialmente na Figura 11, e a “Carta de suscetibilidade a movimentos gravitacionais de massa e inundações”, representada na Figura 14.
- b) Plano Municipal de Redução de Riscos (PMRR): elaborado em 2006 pelo extinto Ministério das Cidades, o relatório visa subsidiar os poderes públicos na gestão de riscos associados a escorregamentos e solapamentos de margens de córregos, apoiar a implementação de ações estruturais e não estruturais e apresentar a demanda orçamentária para a realização das atividades descritas.
- c) Monitoramento Pluviométrico: com o objetivo de monitorar os índices pluviométricos do município, foram instalados cinco pluviômetros semiautomáticos e treze pluviômetros automáticos. O local escolhido para a instalação baseou-se no mapeamento de áreas de risco feito pelo IG-SMA (2014). A alocação dos pluviômetros foi realizada por meio de cooperação técnica-institucional firmada entre a Prefeitura Municipal, Coordenadoria Estadual de Defesa Civil (Cedec) e Cemaden.

Figura 14. Carta de suscetibilidade a movimentos gravitacionais de massa



Fonte: Prieto (2018).

- d) Monitoramento de Encostas: por meio de uma parceria firmada entre a prefeitura e o Cemaden, foi instalada em 2014 uma Estação Total Robotizada (ETR) na região da Vila Albertina. Essa ferramenta é capaz de fornecer dados em tempo real, permitindo uma melhor compreensão da dinâmica de movimentação das encostas, dos mecanismos de ruptura, da quantidade de chuva acumulada e da umidade necessária para a deflagração de deslizamentos em cada localidade.
- e) Estação Hidrológica: equipamento que monitora os níveis de água do rio principal do município – o Capivari – que atravessa a área urbana. Essa estação realiza a medida, o armazenamento e a transmissão de dados hidrológicos, apoiando na tomada de decisão

para o envio de alertas. Essa ferramenta também foi adquirida por meio da parceria institucional entre a prefeitura e o Cemaden.

- f) Participação no Projeto Cidades Resilientes: o município aderiu ao Programa em 2014 e atualmente possui um comitê focado nessa temática. Ao todo, o programa prevê dez objetivos a serem alcançados que são acompanhados por meio de indicadores urbanos locais para que as cidades avaliem sua capacidade de resiliência. De acordo com o relatório da UNISDR (2018), Campos do Jordão encontrava-se entre os dez municípios mais avançados nesse contexto no Brasil, situado no terceiro ciclo da campanha.

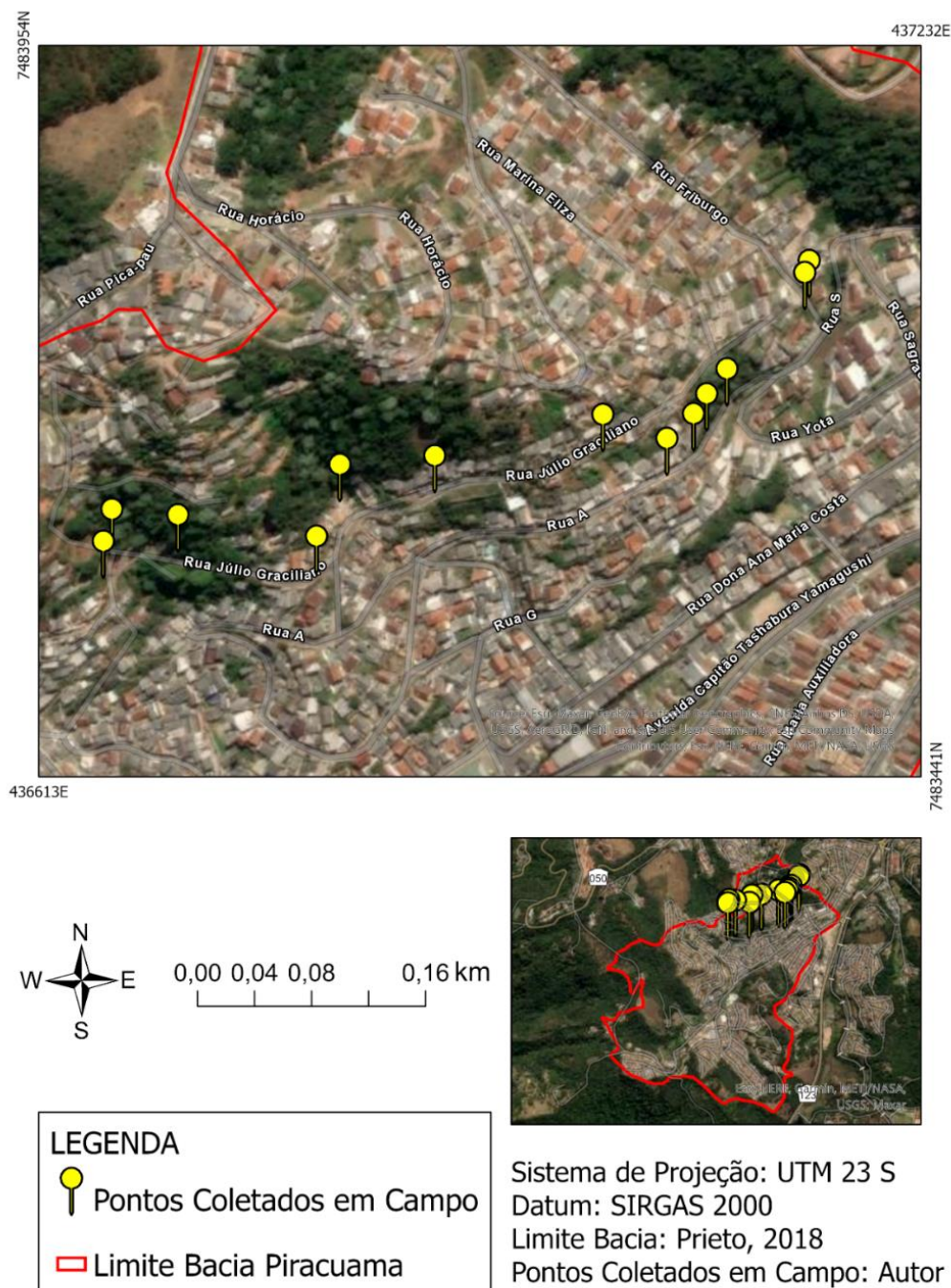
4.1.3 Trabalho de campo

Com o objetivo de caracterizar a área de estudo em escala local, foi realizado um trabalho de campo no dia 26 de junho de 2020. A atividade contou com o apoio da defesa civil de Campos do Jordão no deslocamento e reconhecimento das áreas classificadas como grau de risco muito alto, conforme o mapeamento IG-SMA (2014). Dessa forma, foram visitadas as ruas A e F, localizadas na Vila Albertina, visando mapear as variáveis envolvidas no cenário e associadas aos movimentos gravitacionais de massa. Nesse contexto, foram coletados 13 pontos (Figura 15) que permitiram o mapeamento de informações ligadas à litologia, estruturas geológicas, presença de vegetação e drenagem, identificação de obras estruturais, cicatrizes de escorregamentos, declividade, padrões de ocupação e ações antrópicas. Todo o mapeamento foi realizado com uma das aplicações do Sistema ArcGIS – o Survey123 – um formulário digital georreferenciado que permite a coleta de dados via celular, integrando de forma automática as informações mapeadas em campo com as outras feições geoespaciais utilizadas no projeto.

Em relação às obras estruturais realizadas pela prefeitura, foram identificadas canaletas de drenagem (Figura 16 (a)), escadas hidráulicas (Figura 16 (b)) e muros de contenção (Figura 16 (c)). Essas obras estão ligadas principalmente a localidades de ocorrência de escorregamentos e tem como objetivo reduzir e/ou dizimar o risco a esses processos. Quanto às litologias, encontrou-se as duas fácies do migmatito - o paleossoma e o neossoma - na área mapeada. Em ambos, foi observado o produto de alteração e identificado as relações com os solos residuais. O paleossoma (Figura 17) encontrou-se por vezes alterado, resultando em solos argilosos de coloração marrom-avermelhada que estão associados a presença de biotita na composição dessa fácies. Já o neossoma (Figura 18) apresentou um produto de alteração de coloração clara, relacionados à composição quartzo-feldspática. A rocha matriz preservou

estruturas tectônicas, provavelmente relacionadas às zonas de cisalhamento localizadas na região. No migmatito, foi possível observar uma estrutura sigmóide (Figura 19), feição resultante de regimes tectônicos dúcteis que indicam o sentido da cinemática na região.

Figura 15. Pontos coletados em campo



Fonte: produção do autor.

Figura 16. Obras estruturais observadas em campo. a) Canaleta de drenagem. b) Escada hidráulica. c) Muro de contenção com canaletas de drenagem, que auxiliam no escoamento das águas infiltradas no terreno



Fonte: produção do autor.

Figura 17. Paleossoma do migmatito observado em campo. Observa-se alteração de coloração avermelhada e presença de foliações, derivadas da natureza gnaissíca



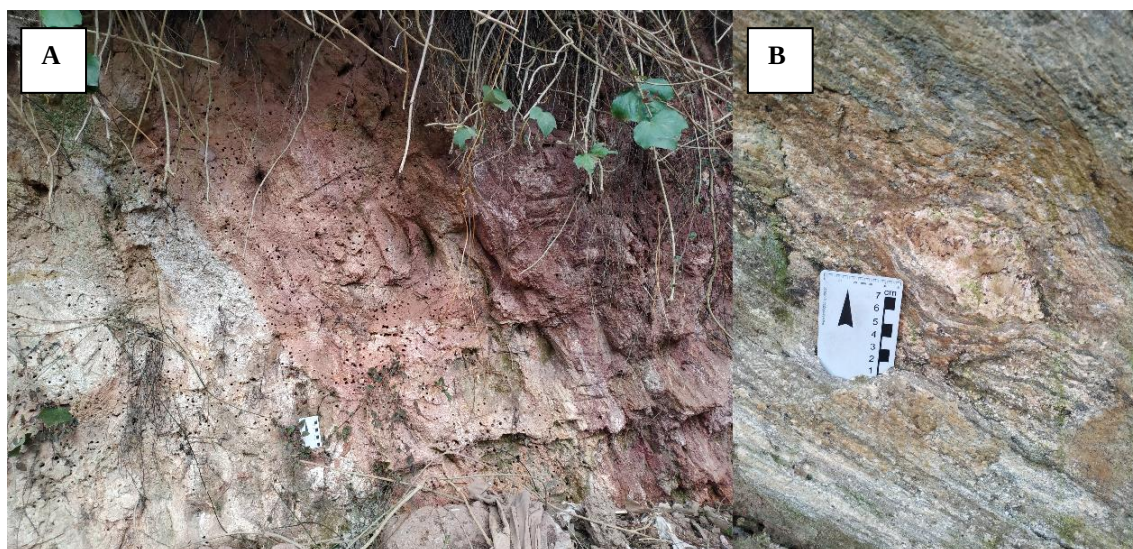
Fonte: produção do autor.

Figura 18. Neossoma do migmatito observado em campo. Nota-se o produto de alteração de coloração clara, associado à origem granítica



Fonte: produção do autor.

Figura 19. A) Superfície de contato entre neossoma e paleossoma. B) Estrutura sigmóide, relacionada às zonas de cisalhamento e que indica a cinemática do movimento



Fonte: produção do autor.

Durante o trabalho de campo, também foi identificada uma cicatriz de escorregamento associada ao evento de 2000. Nesse ponto, observou-se um grande anfiteatro de erosão e vegetações estabelecidas pós-evento (Figura 20 (a)). Em relação às atividades antrópicas, observou-se a presença de lixos domésticos dispostos em terrenos baldios ao ar livre, que

algumas vezes se encontravam situados em áreas de encostas. Além disso, identificou-se terrenos com cortes e aterros irregulares, bem como moradias de baixo padrão construtivo (Figura 20 (b)), que não foram autorizadas nem regularizadas pela prefeitura.

Figura 20. a) Cicatriz de deslizamento observada em campo. Nota-se um anfiteatro na localidade, bem como vegetações pós-evento. b) Moradias de baixo padrão construtivo observadas em campo, apresentando fundação irregular



(a)



(b)

Fonte: produção do autor.

4.2 Etapa 2

Essa etapa consistiu na construção do banco de dados utilizado nos modelos de predição. Para tal, foi realizado o levantamento e a extração das variáveis dependentes – representadas pelas cicatrizes de escorregamentos encontradas na área de estudo – e das variáveis independentes, que compreendem os fatores condicionantes. Toda a etapa foi realizada em ambiente ArcGIS Pro.

4.2.1 Levantamento de cicatrizes

De acordo com Youssef et al. (2016), os escorregamentos são mais prováveis de ocorrer sob as mesmas condições que foram encontradas em escorregamentos anteriores. Dessa forma, um mapa de inventário de ocorrências para a área de estudo auxilia a entender qual o padrão dos atributos do terreno que está relacionado às ocorrências. As cicatrizes foram levantadas no artigo de Mendes et al. (2018) e no banco de dados do Instituto Geológico, sendo que ambas as fontes registram informações do ano e tipo de processo envolvido (Quadro 9 e Figura 21). Ao todo, foram levantadas seis feições georreferenciadas, entretanto, visto que uma delas é referente a fluxos de lama e outra foi disponibilizada em formato de ponto, optou-se por trabalhar somente com quatro registros. Os tipos de escorregamento descritos em todas as cicatrizes restantes referem-se a translacionais rasos.

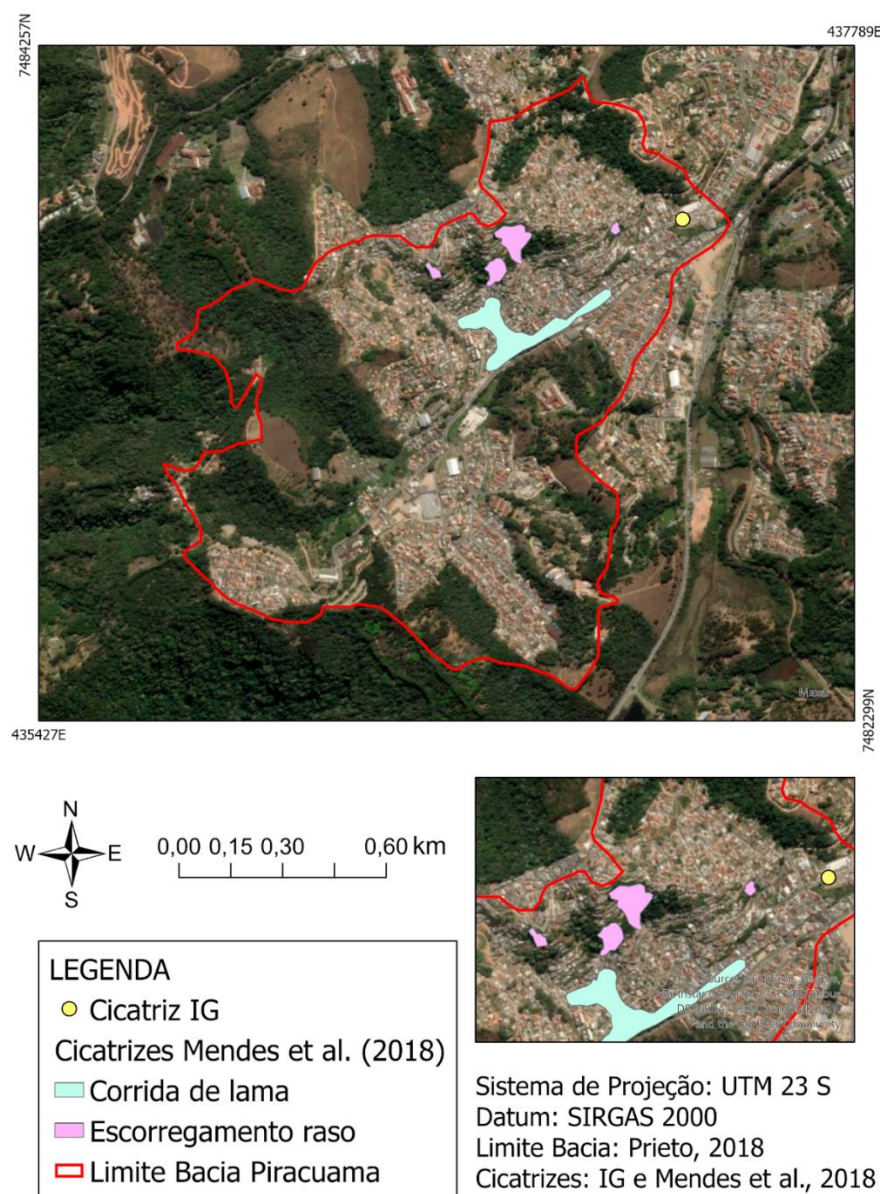
Quadro 9. Levantamento das cicatrizes na área de estudo

Número da Cicatriz	Evento	Ano	Fonte	Tipo de Feição
1	Escorregamento raso	2000	Mendes et al. (2018)	Polígono
2	Escorregamento raso	2000	Mendes et al. (2018)	Polígono
3	Escorregamento raso	2000	Mendes et al. (2018)	Polígono
4	Escorregamento raso	2000	Mendes et al. (2018)	Polígono
5	Corrida de terra	1972	Mendes et al. (2018)	Polígono
6	Escorregamento	1999	Instituto Geológico	Ponto

Fonte: produção do autor.

Com o auxílio do levantamento bibliográfico de ocorrências realizado na primeira etapa, buscou-se em plataformas especializadas – Google Earth, Landsat e Sentinel Explorer da Esri, Geócatálogo mantido pela Secretaria de Meio Ambiente e Catálogo de Imagens do Instituto Nacional de Pesquisas Espaciais (INPE) - imagens de satélite de alta resolução para a vetorização de possíveis cicatrizes. Entretanto, não foi possível vetorizar novas feições devido à densa urbanização da área de estudo e à baixa disponibilidade de imagens de satélite associadas às datas das ocorrências de escorregamentos e com resoluções adequadas para a visualização e delimitação desses processos.

Figura 21. Localização das cicatrizes na área de estudo



Fonte: produção do autor.

4.2.2 Levantamento dos fatores condicionantes

As variáveis independentes foram definidas conforme diversos trabalhos que aplicaram o algoritmo Floresta Aleatória no mapeamento de suscetibilidade a escorregamentos: Youssef et al. (2016), Catani et al. (2013), Segoni et al. (2020), Park e Kim et al. (2019), Sahin et al. (2020) e Chen et al. (2017). Há uma infinidade de fatores condicionantes de escorregamentos que poderiam ser utilizadas no modelo, conforme apontado no Quadro 2 e

descrito no capítulo anterior. Entretanto, a disponibilidade de dados geográficos para a área de estudo ainda é baixa e/ou possui uma escala pouco detalhada, dificultando a inserção de diversos fatores condicionantes na construção dos modelos de predição.

A variável geológica, por exemplo, apesar de prevista no início do trabalho, não foi utilizada na construção do modelo. Conforme descrito acima, o mapeamento geológico realizado pelo DAAE (1984) possui uma escala muito pequena e descreve para a área de estudo apenas uma litologia, sendo impossível trabalhar com diferentes pesos na análise dessa variável. Já no mapeamento realizado por CPRM e UFRJ (2020), as unidades litoestratigráficas são amplas e, em geral, são constituídas por uma associação de litologias com diferentes vulnerabilidades, sendo que nem sempre é possível discernir qual delas é a predominante e/ou apresentou maior importância na deflagração dos escorregamentos. No total, foram utilizadas 14 variáveis explicativas, divididas em parâmetros morfométricos e temáticos:

4.2.2.1 Parâmetros morfométricos

Os parâmetros morfométricos trazem uma visão detalhada do relevo e são essenciais no mapeamento de suscetibilidade a escorregamentos (Youssef et al., 2016). Tais fatores foram construídos a partir de imagens provenientes de radar com resolução de 5 metros, datada de 2014 e fornecida pelo Serviço Geológico do Brasil (CPRM). Posteriormente, a partir desse dado, foi extraído o Modelo Digital de Elevação (MDE), que serviu como base para a construção de todos os fatores inseridos nessa categoria.

As classes de certos parâmetros que serão descritos abaixo foram definidas por meio de um método de classificação chamado quebras naturais (*natural breaks*). Ele é utilizado para identificar, dentro de um conjunto de dados, quais os intervalos que melhor agrupam valores semelhantes e maximizam as diferenças entre as classes. Sendo assim, os fatores são divididos em classes cujos limites são definidos onde existem diferenças relativamente grandes nos valores dos dados (ESRI, 2021). Derivados dos modelos, foram construídos ao todo 11 parâmetros morfométricos, descritos a seguir:

- a) Altimetria (Figura 22 (a)): reflete a superfície de elevação do terreno, em metros, contada a partir do nível médio do mar. De acordo com Catani et al. (2013), as variações das elevações podem estar relacionadas a diferentes configurações de ambiente

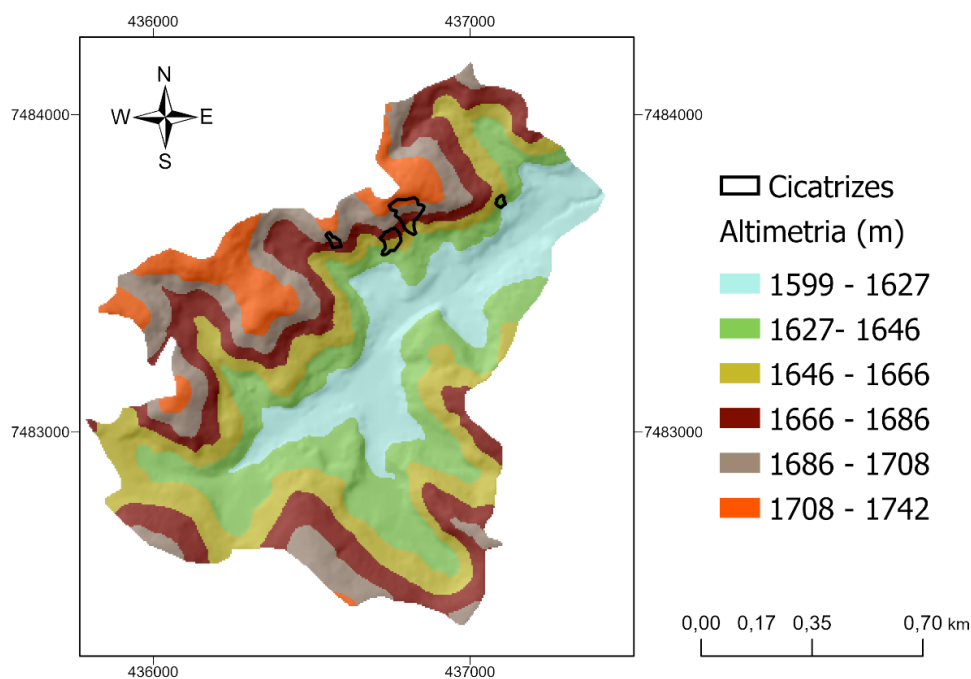
(vegetação, temperatura, regime de chuvas etc.). Neste trabalho utilizamos seis classes conforme o método de classificação quebras naturais.

- b) Aspecto (Figura 22 (b)): configura a orientação do relevo, variando de 0° a 360°. De acordo com Demir (2016), esse fator pode influenciar as quantidades de chuva e radiação solar, bem como a umidade do terreno e crescimento da vegetação. Hiruma (2004) descreve que em Campos do Jordão as encostas da face norte costumam ter maior incidência do sol e são relativamente mais secas. Em contrapartida, as encostas da face sul-sudeste sofrem exposição direta às frentes polares, sendo caracterizadas por serem sombrias e úmidas. Esse fator condicionante foi dividido em dez classes: Plano (-1°), Norte (0°-22,5°), Nordeste (22,5°-67,5°), Leste (67,5°-112,5°), Sudeste (112,5°-157,5°), Sul (157,5°-202,5°), Sudoeste (202,5°-247,5°), Oeste (247,5°-292,5°), Noroeste (292,5°-337,5°) e Norte (337,5°-360°).
- c) Curvatura Combinada (Figura 22 (c)): esse fator combina as curvaturas de perfil e planta. Considerar a forma planta e perfil juntas permite entender com mais precisão o fluxo em uma superfície. Este fator condicionante foi dividido em três classes para representar as áreas côncavas (< -2,23), retilíneas (-2,23 a 1,13) e convexas (> 1,13).
- d) Curvatura em Perfil (Figura 22 (d)): está relacionada a aceleração e desaceleração de um fluxo ao longo de uma superfície e, portanto, influencia a erosão e a deposição em uma determinada área. Um valor positivo indica que a superfície é côncava e que o fluxo será acelerado. Já um valor negativo indica que a superfície é convexa e o fluxo será desacelerado. Valores próximos a zero indicam que a superfície é retilínea. Esse fator foi dividido em três classes, conforme Sánchez (2016): < -0,3 (convexa), -0,3 a 0,4 (retilínea) e > 0,4 (côncava).
- e) Curvatura em Planta (Figura 22 (e)): está relacionada a convergência ou divergência de um fluxo em uma superfície, possibilitando entender a quantidade de água no solo. Valores positivos indicam que a superfície é convexa e apresenta um fluxo divergente. Já os valores negativos indicam uma superfície côncava e fluxos convergentes. Os valores próximos a zero indicam uma superfície retilínea. Essa curvatura foi dividida em três classes, conforme Sánchez (2016): < -0,4 (côncava), -0,4 a 0,1 (retilínea) e > 0,1 (convexa).
- f) Declividade (Figura 22 (f)): representa a inclinação da superfície do terreno e é um dos fatores condicionantes mais importantes, pois controla fortemente as forças de cisalhamento atuantes nas encostas (Catani et al., 2013). Além disso, altas declividades

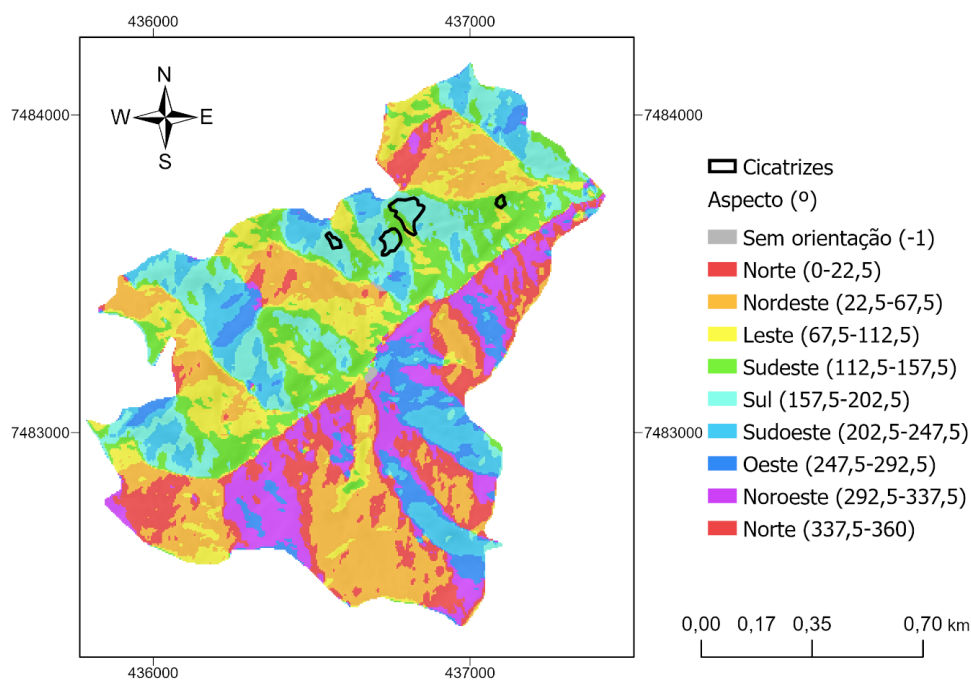
também podem estar relacionadas a horizontes de solos menos desenvolvidos, favorecendo a ocorrência de escorregamentos translacionais rasos. Esse fator condicionante foi dividido em cinco classes, conforme o proposto por De Biasi (1970): < 5%, 5-12%, 12-30%, 30-47% e > 47%.

- g) SPI (*Stream Power Index* ou Índice de Potência de Escoamento) (Figura 22 (g)): é um atributo topográfico secundário que descreve o potencial da topografia para concentrar grandes quantidade de água superficial em uma determinada área (CAPOANE, 2015). À medida que a área de captação e a declividade aumentam, a velocidade do fluxo e o poder erosivo do escoamento superficial se intensificam. Este fator condicionante foi dividido em cinco classes conforme o método quebras naturais.
- h) STI (*Sediment Transport Index* ou Índice de Transporte de Sedimentos (Figura 22 (h)): está diretamente relacionado ao transporte de sedimentos do terreno para o canal durante eventos de deslizamento, representando os procedimentos de deposição (MOORE; WILSON, 1992). Este fator condicionante foi dividido em cinco classes através do método quebras naturais.
- i) TPI (*Topographic Position Index* ou Índice de Posição Topográfica) (Figura 22 (i)): mede a diferença entre a elevação de um ponto central contra a elevação média de um raio especificado em torno desse ponto. TPI positivos são locais maiores do que média de seu ambiente, enquanto TPI negativos representam locais inferiores aos seus arredores (WEISS, 2001). Esta ferramenta pode ser usada para isolar características específicas da paisagem como encostas íngremes. Este fator condicionante foi dividido em cinco classes conforme o método quebras naturais.
- j) TRI (*Topographic Ruggedness Index* ou Índice de Rugosidade do Terreno) (Figura 22 (j)): é definido como a variabilidade ou irregularidade na elevação dentro de uma unidade espacial de terreno amostrada, desenvolvido por Riley et al. (1999). De acordo com Nakileza e Nedala (2020), o TRI descreve a heterogeneidade de uma superfície de um terreno, caracterizando as formas do relevo como planas ou acidentadas. Este fator condicionante foi dividido em cinco classes conforme o método quebras naturais.
- k) TWI (*Topographic Wetness Index* ou Índice de Umidade Topográfico) (Figura 22 (k)): o índice é utilizado para caracterizar espacialmente as zonas de saturação da água superficial e o teor de água no solo (COSTA; TAVARES; OLIVEIRA, 2019). Este fator condicionante foi dividido em cinco classes através do método quebras naturais.

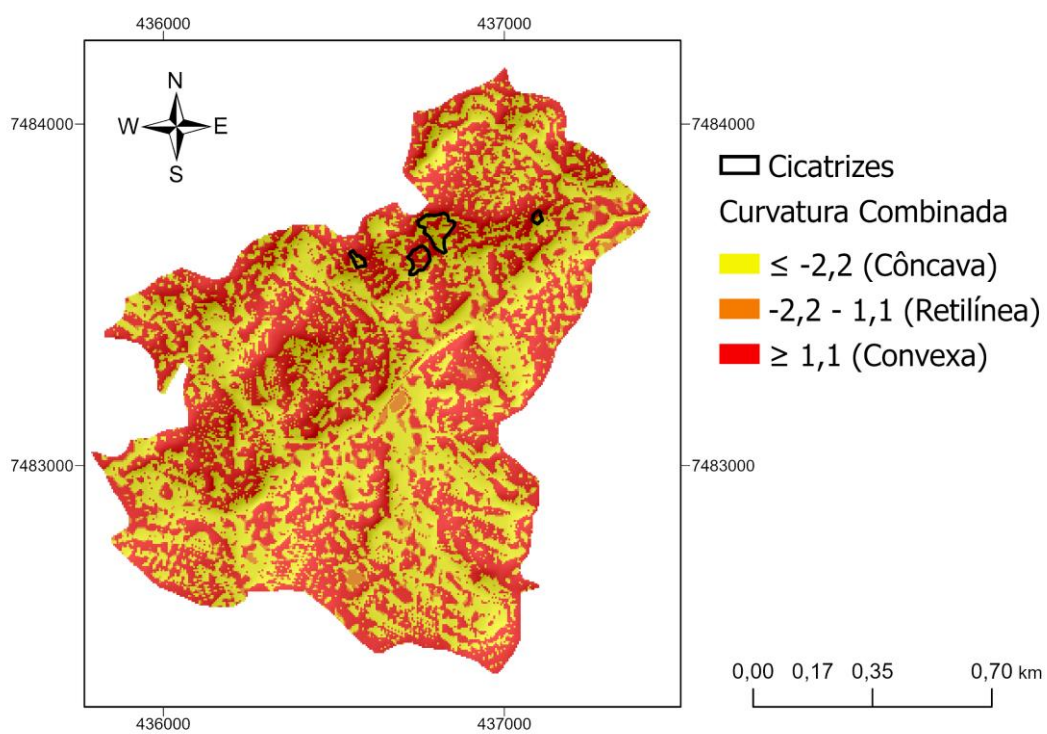
Figura 22. Parâmetros morfométricos utilizados no modelo de acordo com (a) Altimetria; (b) Aspecto; (c) Curvatura Combinada; (d) Curvatura em Perfil; (e) Curvatura em Planta; (f) Declividade; (g) SPI; (h) STI; (i) TPI; (j) TRI; (k) TWI



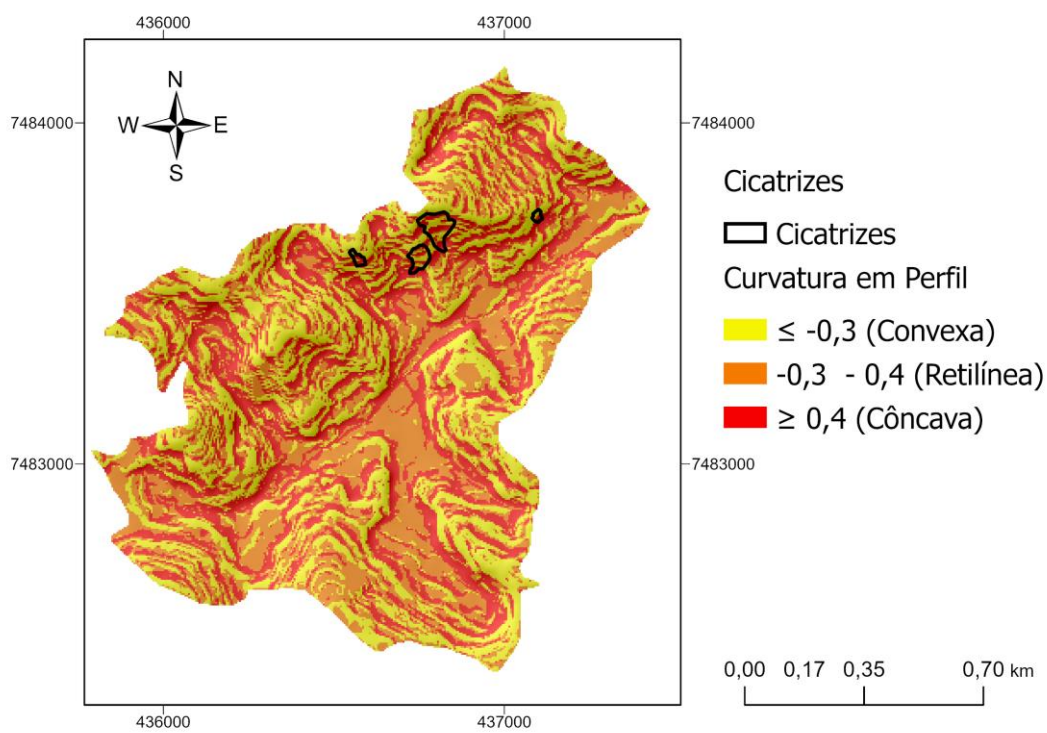
(a) Altimetria



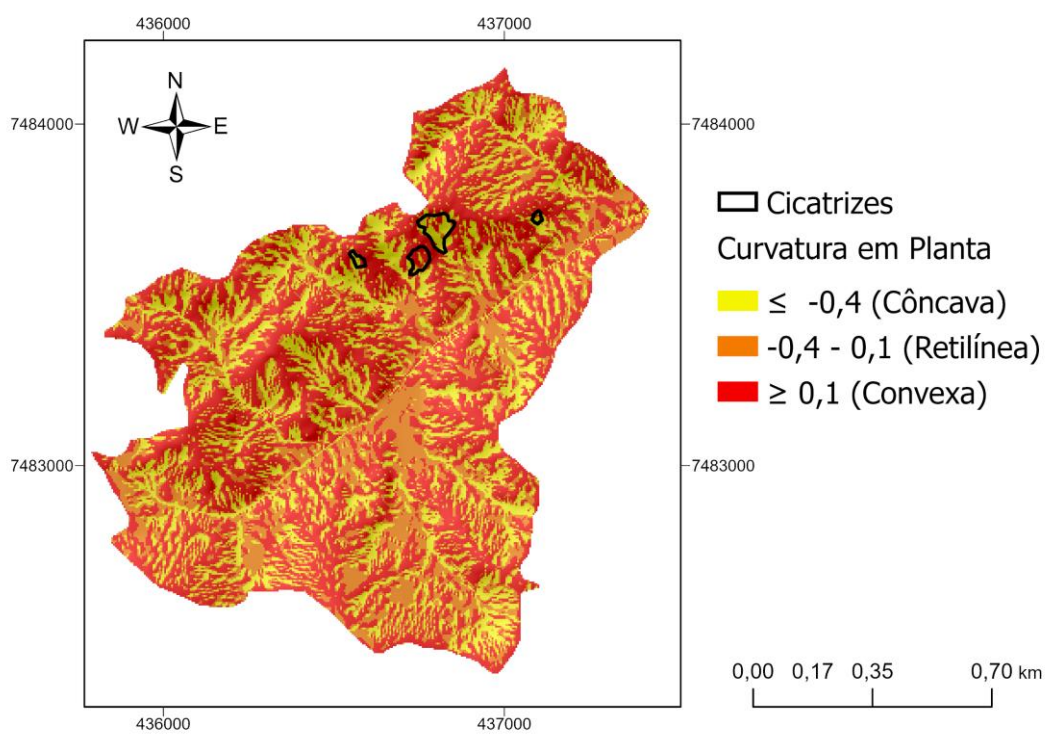
(b) Aspecto



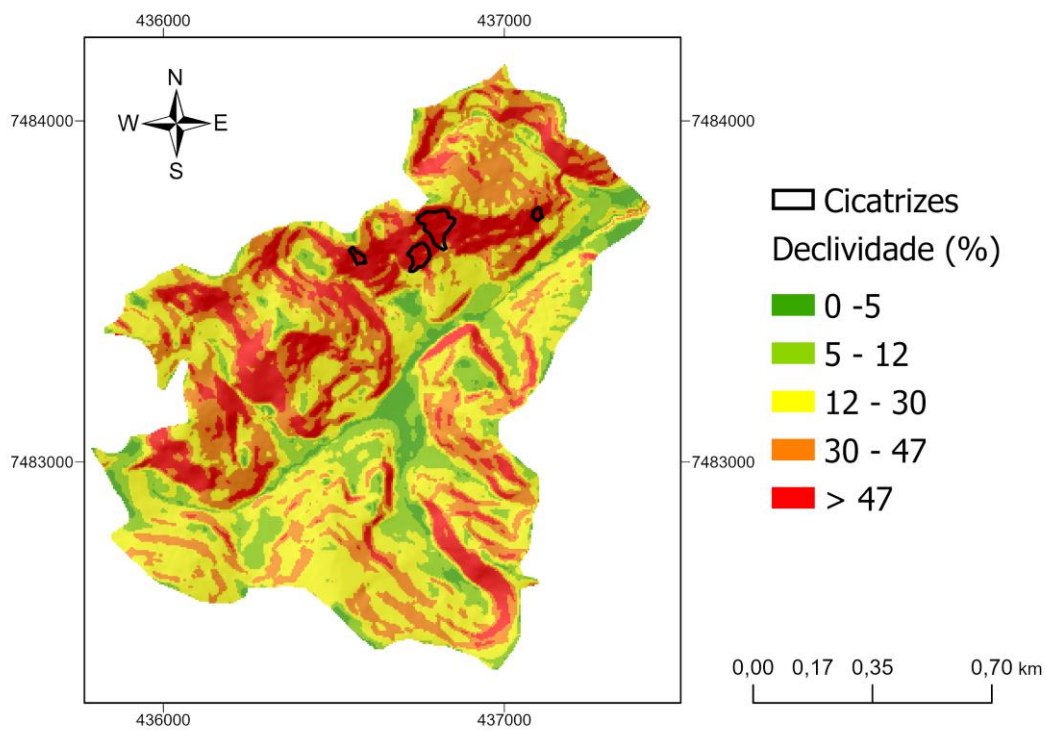
(c) Curvatura Combinada



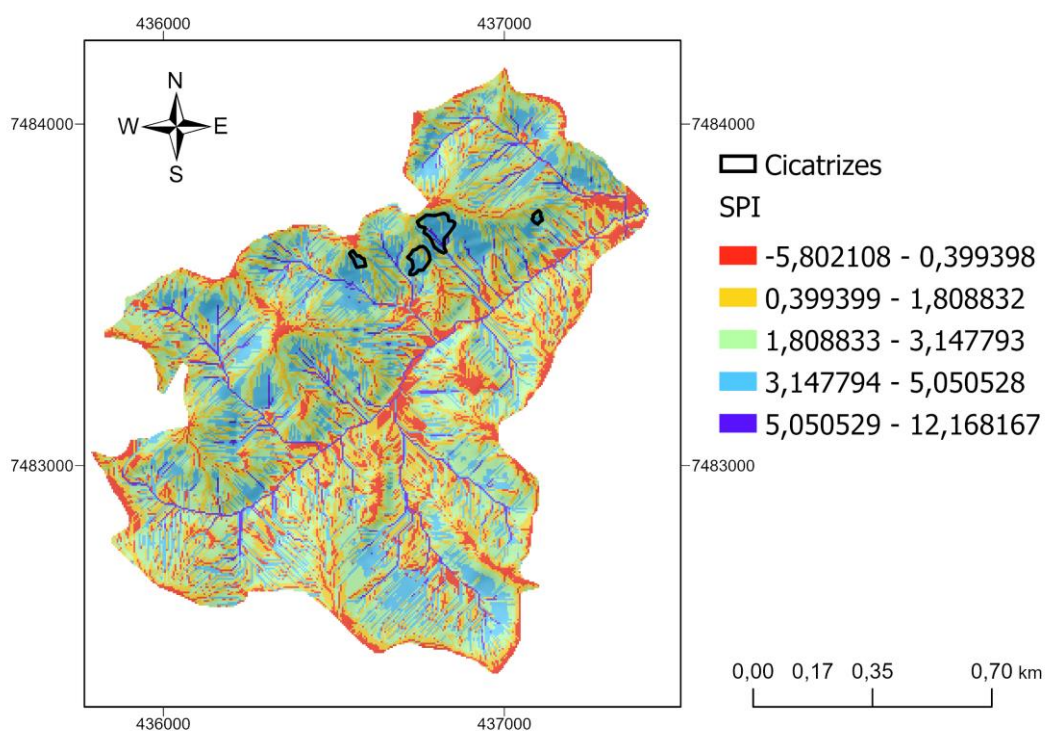
(d) Curvatura em Perfil



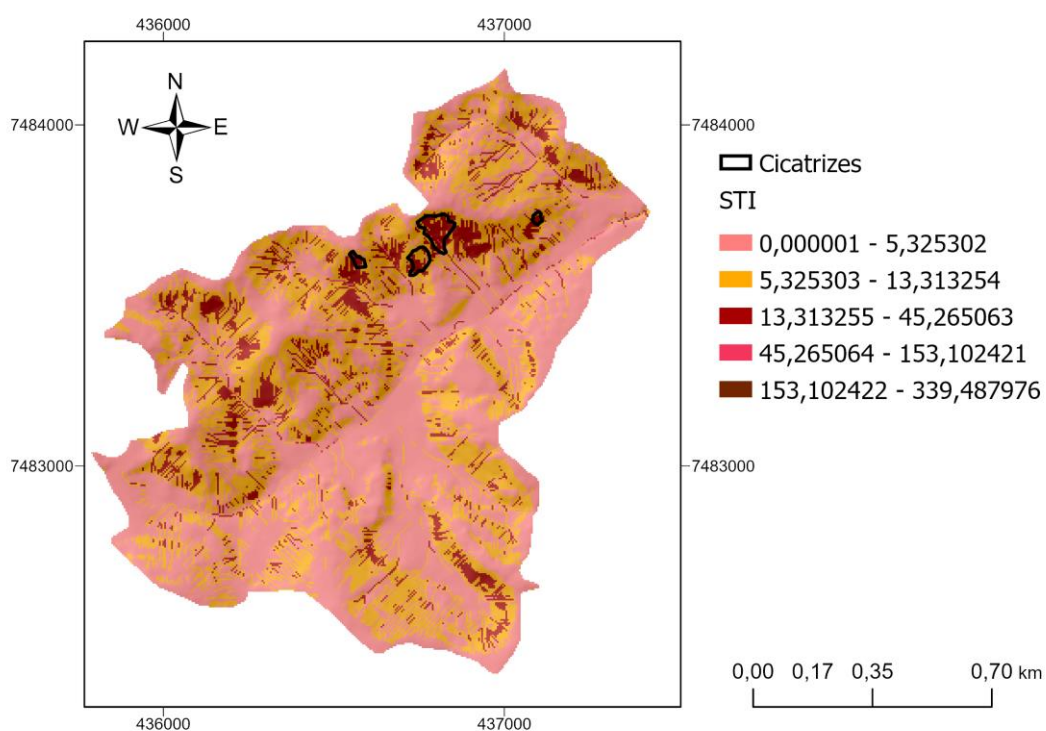
(e) Curvatura em Planta



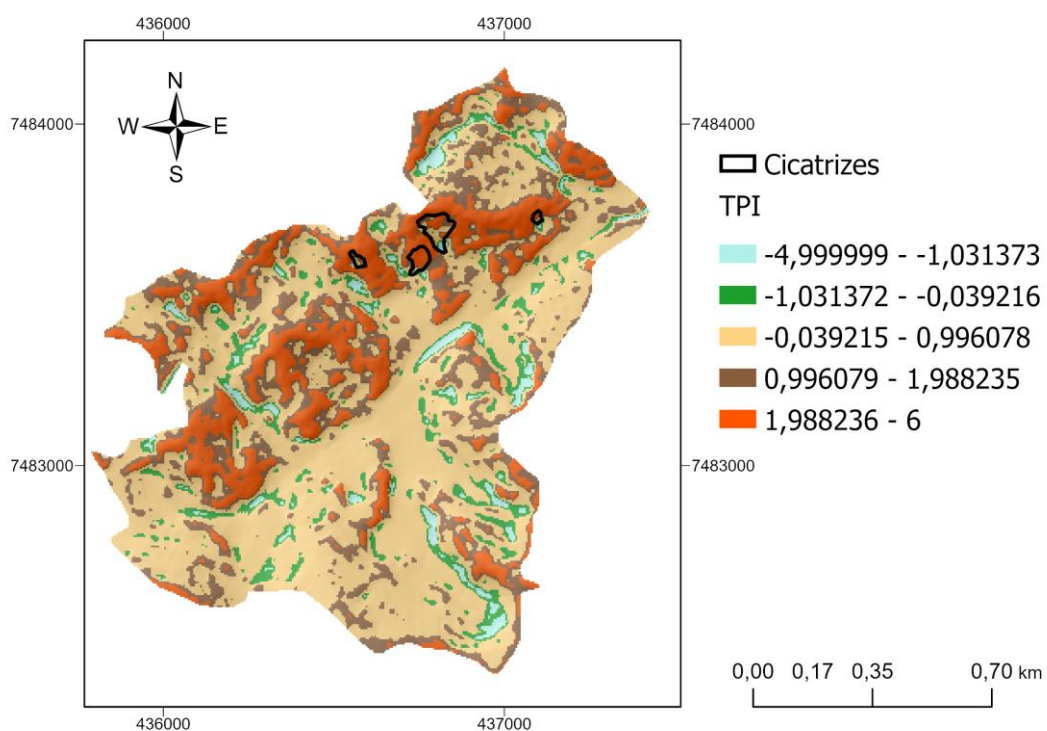
(f) Declividade



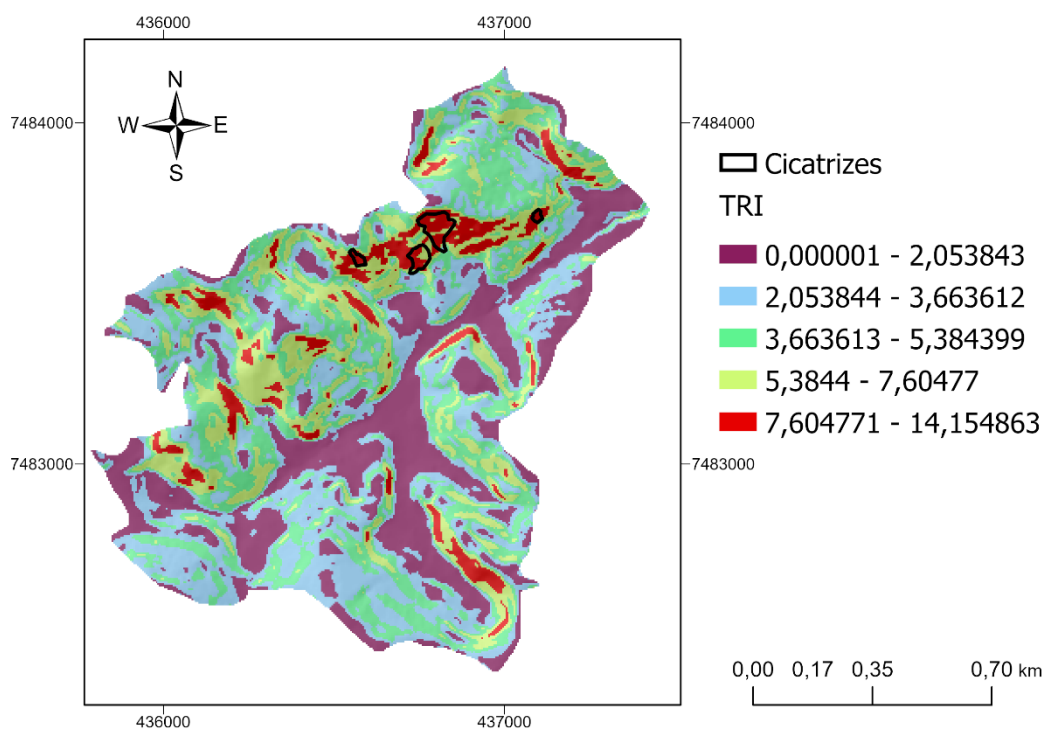
(g) SPI ou Índice de Potência de Escoamento



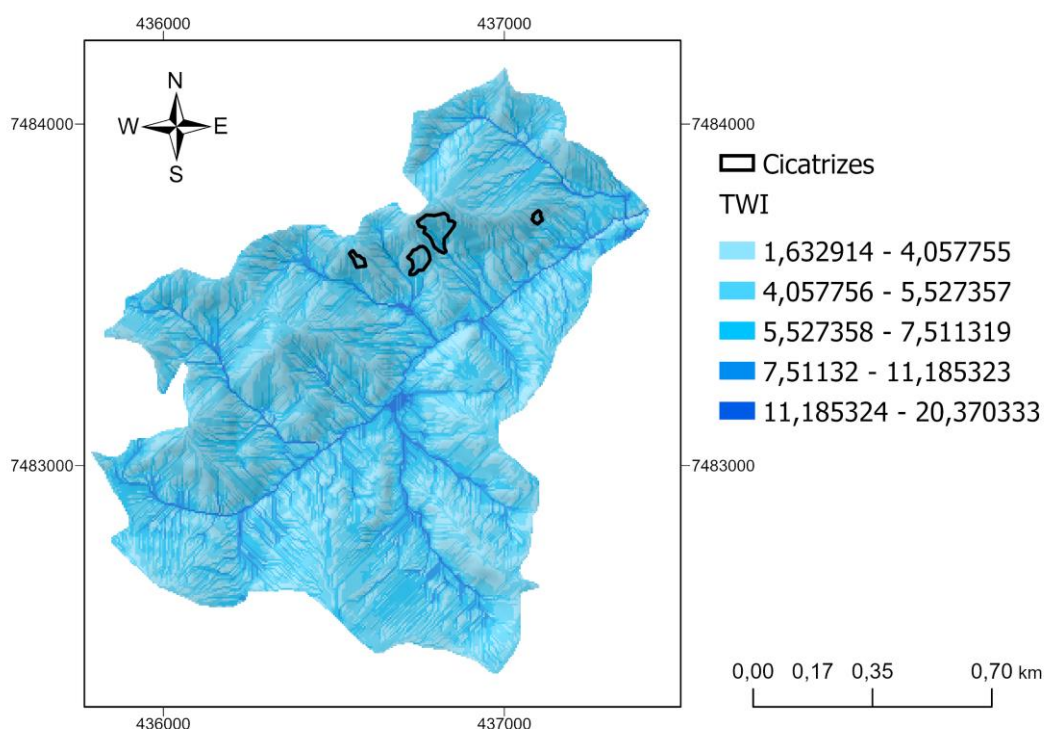
(h) STI ou Índice de Transporte de Sedimentos



(i) TPI ou índice de Posição Topográfica



(j) TRI ou Índice de Rugosidade do Terreno



(k) TWI ou Índice Topográfico de Umidade

Fonte: produção do autor.

4.2.2.2 Parâmetros Temáticos

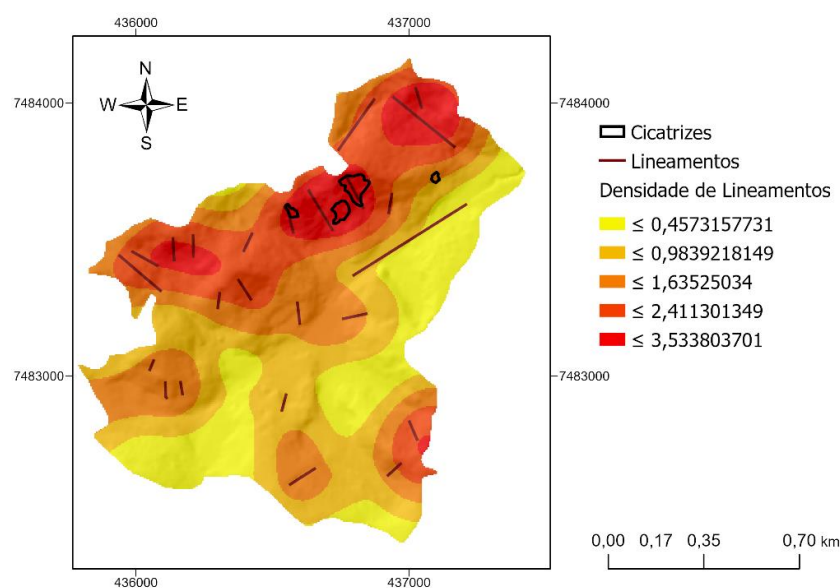
Os mapas temáticos envolveram a densidade de lineamentos, o uso e cobertura do solo e a vegetação. Esses fatores, apesar de não estarem intrinsecamente associados ao relevo, também configuram importantes aspectos a serem analisados, pois estão frequentemente associados às instabilizações nas encostas. Ao todo, foram levantados três parâmetros temáticos:

- a) Densidade de Lineamentos (Figura 23 (a)): foi realizado com base no relevo sombreado derivado do MDE e curvas de nível. Dessa forma, foi possível vetorizar as feições de lineamentos e identificar possíveis zonas de instabilidade de encosta. De acordo com Avelar et al. (1992), o mapeamento dessas estruturas possibilita a análise do papel do substrato geológico no controle do desenvolvimento de formas concavas do relevo, isto é, dos vales canalizados e não-canalizados.
- b) Uso e Cobertura da Terra (Figura 23 (b)): realizado através de ferramentas de classificação de imagem supervisionada baseada em pixel em ambiente ArcGIS Pro.

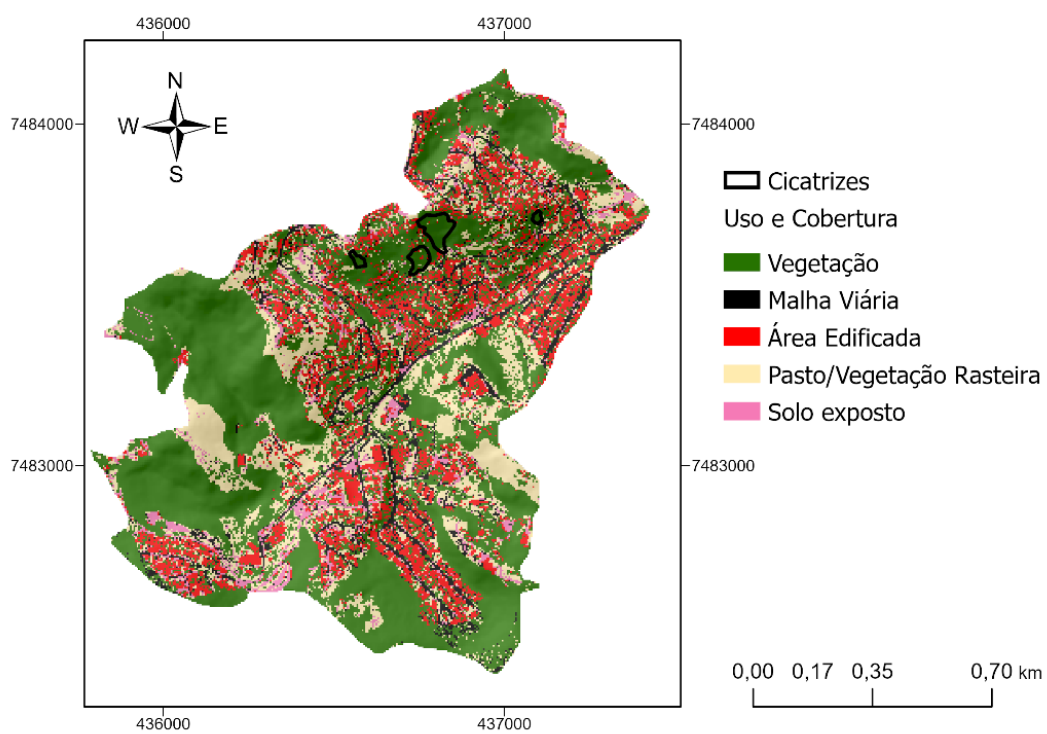
Esse processo consistiu na extração de cinco classes - vegetação, vias pavimentadas/pavimentos, áreas edificadas, pasto/vegetação rasteira e solo exposto/vias não pavimentadas – a partir de imagens de sensoriamento remoto multibanda. O fluxo da classificação compreendeu, primeiramente, a coleta de amostras de treinamento, com o objetivo de rotular grupos de pixels semelhantes e pertencentes às classes já estabelecidas, resultando em um conjunto de dados de referência. Em seguida, utilizou-se o classificador *Random Trees* que permitiu, com base nas amostras de treinamento, a extrapolação das classes de uso da terra para toda a área. Nesse tipo de método, os pixels são analisados individualmente de acordo com a resposta espectral e então classificados. Como métrica de avaliação do classificador, obteve-se um índice Kappa de 0,93, indicando bons resultados para o mapa gerado.

- c) Vegetação Figura 23 (c)): O dado de vegetação do município foi retirado do banco de dados de dados ambientais do IBGE. O mapeamento de vegetação, que possui escala 1:250.000, abrange a cobertura da terra do território nacional, conforme a classificação da vegetação brasileira utilizada pelo IBGE, sendo que as informações foram produzidas em formato digital no período de 1997 a 2017. A área de estudo contém as classes Floresta Ombrófila Mista e Vegetação com influência urbana.

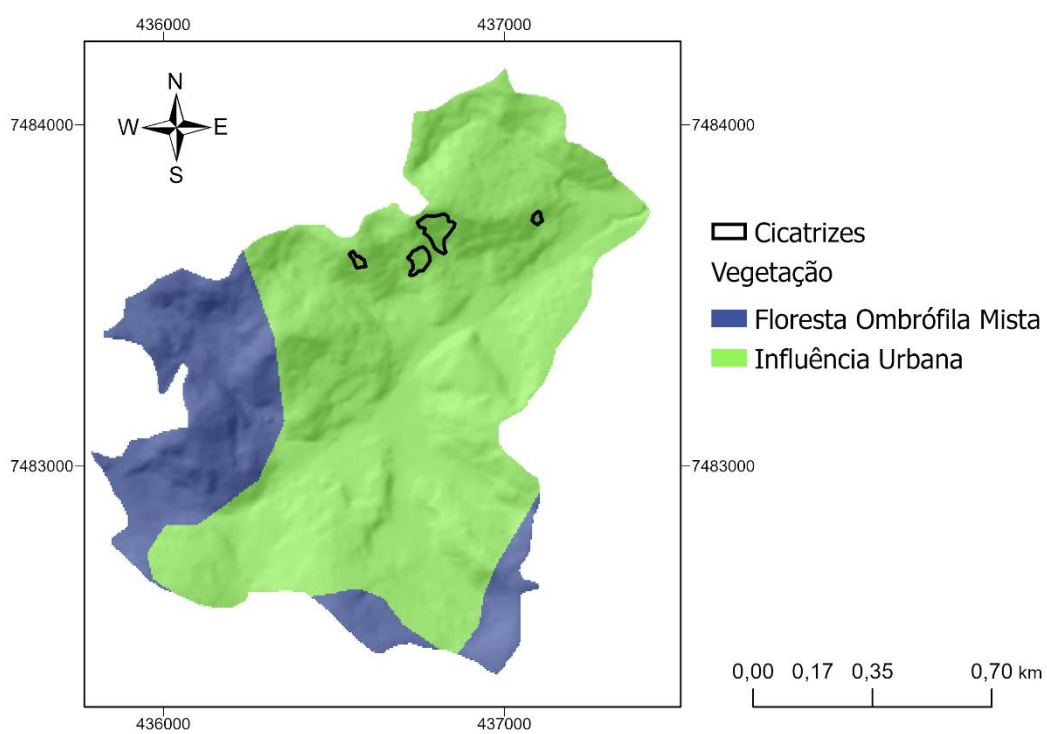
Figura 23. Parâmetros temáticos utilizados no modelo de acordo com: (a) Densidade de Lineamentos; (b) Uso e Cobertura da Terra; (c) Vegetação



(a) Densidade de Lineamentos



(b) Uso e Cobertura da Terra



(c) Vegetação

Fonte: produção do autor.

4.2.3 Razão de Frequência

O passo final de preparação do dado compreendeu a análise da relação entre os fatores dependentes (cicatrizes) e independentes (variáveis explicativas) através da técnica Razão de Frequência (*Frequency Ratio* - FR). O FR é a relação entre a área onde ocorreram os escorregamentos e a área total de estudo, conforme a Equação 1. Se a probabilidade bivariada for maior que 1, mais forte é a correlação entre a incidência de escorregamentos e os fatores desencadeantes e, se a probabilidade for menor que 1, mais fraca será a correlação (AGHDAM; VARZANDEH; PRADHAN, 2016).

$$FR = \frac{N_{pix(Si)} / \sum N_{pix(Si)}}{N_{pix(Ni)} / \sum N_{pix(Ni)}} \quad (1)$$

Onde $N_{pix(Si)}$ é o número de pixels de escorregamentos em um fator de classe, $\sum N_{pix(Si)}$ é o total de pixels de ocorrência da área de estudo, $N_{pix(Ni)}$ é o número total de pixels de um fator de classe e $\sum N_{pix(Ni)}$ é o total de pixels da área de estudo. O FR também pode ser determinado pela porcentagem de pixels de ocorrência de escorregamento em um fator de classe dividido pela porcentagem de pixels total desse fator na área de estudo.

O cálculo foi feito para cada variável explicativa que, posteriormente, foram normalizadas dentro de um intervalo de 0 a 1. Os FRs normalizados foram aplicados ao conjunto de dados inicial e reescalados de 0 a 100 quando implementados no modelo de predição. A Tabela 1 apresenta os cálculos realizados.

Tabela 1. FR calculado para cada variável utilizada no modelo de predição

(continua)

FATOR	CLASSE	% CLASSE	% OCORRÊNCIA	FR	FR NORMALIZADO
ALTIMETRIA (m)	1599 - 1627	18	0	0,00	0,00
	1627- 1646	23	9	0,38	0,06
	1646 - 1666	21	22	1,05	0,15
	1666 - 1686	18	31	1,68	0,25
	1686 - 1708	12	26	2,10	0,31

	1708 - 1742	8	12	1,60	0,23
ASPECTO (°)	Sem orientação (-1)	0	0	0,00	0,00
	Norte (0-22,5)	7	0	0,00	0,00
	Nordeste (22,5-67,5)	20	0	0,00	0,00
	Leste (67,5-112,5)	13	4	0,30	0,04
	Sudeste (112,5-157,5)	14	24	1,78	0,22
	Sul (157,5-202,5)	12	72	5,91	0,74
	Sudoeste (202,5-247,5)	10	0	0,02	0,00
	Oeste (247,5-292,5)	7	0	0,00	0,00
	Noroeste (292,5-337,5)	11	0	0,00	0,00
	Norte (337,5-360)	6	0	0,00	0,00
CURVATURA COMBINADA	≤ -2,2 (Côncava)	12	31	2,56	0,59
	-2,2 a 1,1 (Retilínea)	63	40	0,65	0,15
	≥ 1,1 (Convexa)	25	29	1,13	0,26
CURVATURA EM PERFIL	≤ -0,3 (Convexa)	16	41	2,61	0,45
	-0,3 a 0,4 (Retilínea)	69	15	0,22	0,04
	≥ 0,4 (Côncava)	15	44	2,92	0,51
CURVATURA EM PLANTA	≤ -0,4 (Côncava)	13	56	4,28	0,73
	-0,4 a 0,1 (Retilínea)	62	10	0,16	0,03
	≥ 0,1 (Convexa)	24	34	1,40	0,24
DECLIVIDADE (%)	0 - 5%	4	0	0,00	0,00
	5 - 12%	13	0	0,00	0,00
	12 - 30%	37	1	0,03	0,01
	30 - 47%	28	3	0,09	0,02
	> 47%	18	96	5,30	0,98
DENSIDADE DE LINEAMENTOS	0 - 0,457316	22	0	0,00	0,00
	0,457316 - 0,983922	27	1	0,02	0,00
	0,983922 - 1,63525	24	5	0,23	0,02
	1,63525 - 2,411301	18	0	0,00	0,00
	2,411301 - 3,533804	10	94	9,88	0,97

Tabela 1. FR calculado para cada variável utilizada no modelo de predição

(conclusão)					
SPI	-5,802109 - 0,399398	13	0	0,00	0,00
	0,399398 - 1,808832	28	5	0,18	0,03
	1,808832 - 3,147793	34	22	0,65	0,12
	3,147793 - 5,050528	22	67	3,03	0,55
	5,050528 - 12,168167	4	6	1,68	0,30

STI	0 - 5,325302	54	1	0,03	0,00
	5,325302 - 13,313254	36	29	0,80	0,10
	13,313254 - 45,265063	10	69	6,96	0,89
	45,265063 - 153,102421	0	0	0,06	0,01
	153,102421 - 339,487976	0	0	0,00	0,00
TPI	-5 - -1,031373	3	3	1,33	0,21
	-1,031373 - -0,039216	7	4	0,59	0,09
	-0,039216 - 0,996078	54	22	0,42	0,07
	0,996078 - 1,988235	20	23	1,16	0,19
	1,988235 - 6	17	47	2,76	0,44
TRI	0 - 2,053843	25	0	0,00	0,00
	2,053843 - 3,663612	28	1	0,04	0,00
	3,663612 - 5,384399	25	1	0,05	0,00
	5,384399 - 7,60477	16	21	1,28	0,09
	7,60477 - 14,154863	6	77	12,21	0,90
TWI	1,632913 - 4,057755	30	40	1,32	0,43
	4,057755 - 5,527357	40	52	1,28	0,41
	5,527357 - 7,511319	22	7	0,33	0,11
	7,511319 - 11,185323	6	1	0,17	0,06
	11,185323 - 20,37033	1	0	0,00	0,00
USO E COBERTURA	Vegetação	51	93	1,83	0,80
	Malha Viária	8	0	0,03	0,01
	Área Edificada	16	3	0,21	0,09
	Pasto/Vegetação	19	4	0,19	0,08
	Rasteira				
	Solo Exposto	5	0	0,04	0,02
VEGETAÇÃO	Floresta Ombrófila	23	0	0,00	0,00
	Mista				
	Influência Urbana	77	100	1,30	1,00

Fonte: produção do autor.

4.3 Etapa 3

A construção de modelos de classificação consiste em uma série de etapas complexas. Geralmente, para produzir mapas de suscetibilidade a escorregamentos através de métodos de aprendizado de máquina, é preciso que haja uma combinação entre softwares de análise espacial

e linguagem de programação. Em muitos casos, para desenvolver esses algoritmos, o usuário precisar ter um certo nível de especialização, dificultando o acesso por pessoas que não possuem esse nível de expertise. Dessa forma, é possível usar bibliotecas com códigos já desenvolvidos que facilitam a criação de modelos de classificação, como a *Scikit-Learn*. Essa biblioteca é voltada para a linguagem de programação *Python*, de código aberto, que integra diversos algoritmos de aprendizado de máquina para problemas supervisionados e não supervisionados. Esse pacote concentra-se em levar o aprendizado de máquina para não especialistas usando uma linguagem de alto nível e de uso geral, incorporando uma gama de métodos, algoritmos e técnicas que simplificam as atividades de programação (TEMBHEKAR et al., 2021).

Essa etapa consiste na construção dos modelos de predição, através da combinação entre as ferramentas ArcGIS Pro e linguagem *Python* utilizando a biblioteca *Scikit-Learn*. Portanto, após a construção e adequação do banco de dados realizado na etapa anterior, foram aplicados os seguintes processos para a geração dos mapas de suscetibilidade: definições de amostras de treino, validação e teste, identificação da correlação entre as variáveis, identificação da importância dos fatores condicionantes, aplicação do algoritmo Floresta Aleatória, geração dos mapas de suscetibilidade e avaliação da performance do modelo final.

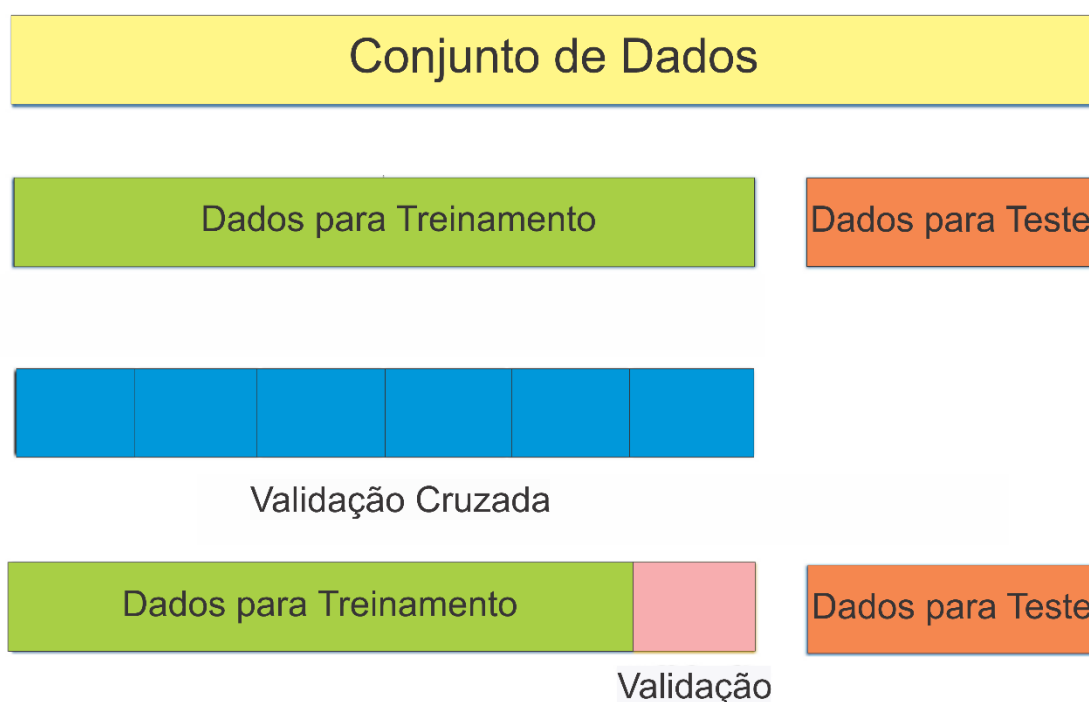
4.3.1 Definição das amostras de treino e teste

O principal objetivo desse passo é dividir os dados de inventário de entrada em um conjunto de dados de treinamento e teste, procedimento padrão na aplicação de algoritmos de aprendizado de máquina (Figura 24). O conjunto de amostras de treinamento é utilizado para construir os modelos de predição, avaliá-los e selecionar a melhor opção. Já o conjunto de teste é usado para analisar desempenho do modelo final. Caso o mesmo conjunto de dados de treinamento for usado para todas as etapas, podem ocorrer problemas de sobreajuste (SAHIN et al., 2020). Durante essa etapa, ainda é necessário considerar os dados de validação, que são utilizados para analisar o desempenho do conjunto de dados de treinamento, conforme a Figura 24.

Desse modo, as cicatrizes levantadas no formato vetorial de polígonos foram convertidas para um arquivo vetorial de pontos, totalizando 100 amostras de ocorrência distribuídas aleatoriamente. O mesmo processo foi realizado com as áreas de não ocorrência, no qual foram geradas aleatoriamente o mesmo número de pontos do conjunto das amostras de

escorregamento. Com o objetivo de garantir uma distância mínima entre as amostras de ocorrência e não ocorrência, foram gerados *buffers* de distância das cicatrizes de 100 metros. Após essa etapa, o conjunto de amostras foi classificado de forma binária: pontos de escorregamento receberam o valor “1”, enquanto para pontos de não escorregamento foram atribuídos valor “0”.

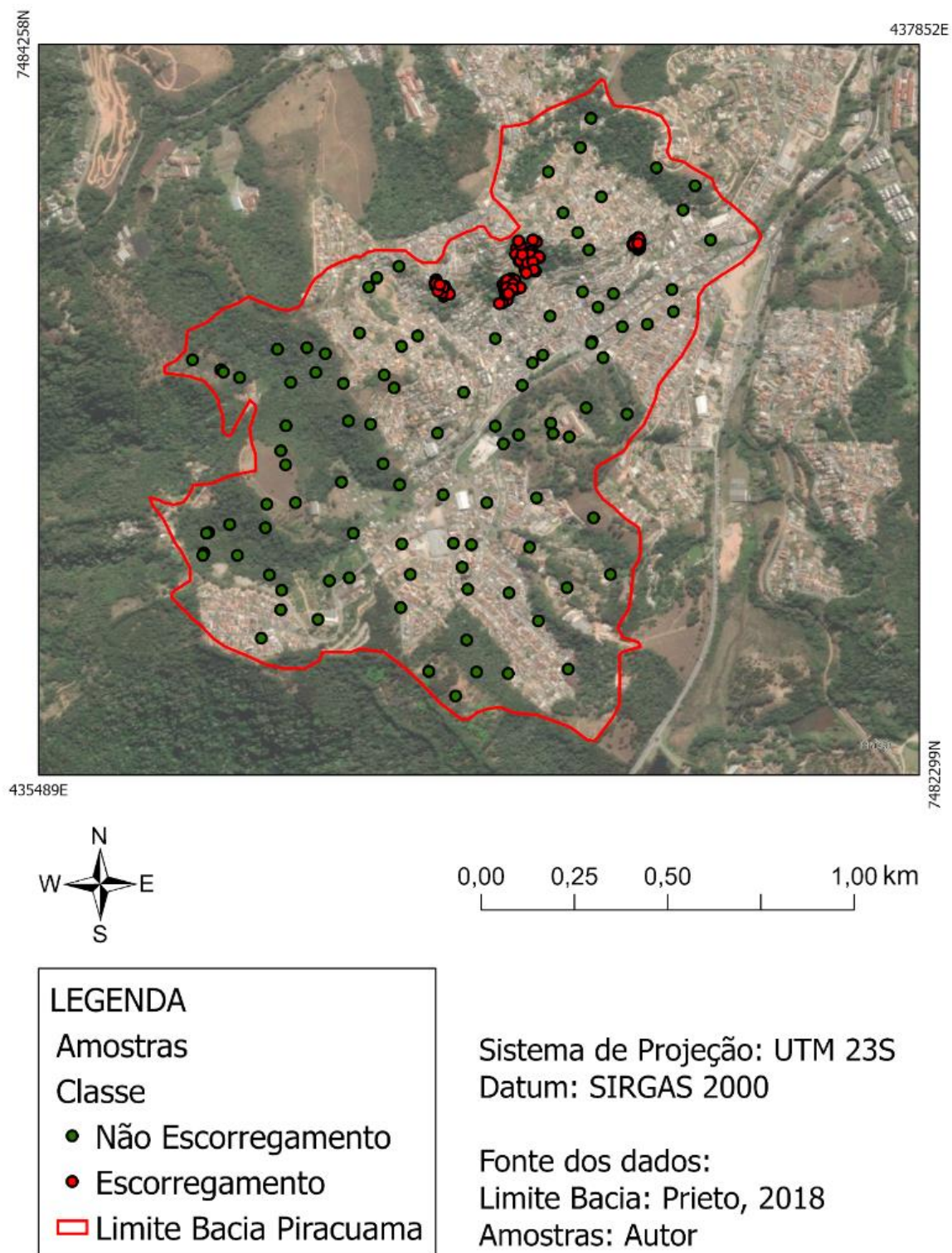
Figura 24. Separação do conjunto de dados para treinamento, validação e teste



Fonte: produção do autor.

Posteriormente, de acordo com o proposto por Bui et al. (2015), Segoni et al. (2020), Dou et al. (2020), Sahin et al. (2020), Dang et al. (2020), as amostras foram divididas aleatoriamente em duas séries: 70% foram destinadas ao treinamento e 30% ao teste. Sendo assim, do total de 200 amostras (100 de ocorrência e 100 de não ocorrência), 140 foram utilizados para o treino (70 de ocorrência e 70 de não ocorrência) e 60 para teste (30 de ocorrência e 30 de não ocorrência). O conjunto de dados de treinamento e validação foi usado para construir e selecionar o modelo de suscetibilidade a escorregamentos, enquanto o conjunto de dados de validação foi utilizado para verificar o desempenho do modelo final. O conjunto de amostra total é representado na Figura 25, classificado em “escorregamentos” e “não escorregamentos”.

Figura 25. Amostras de “escorregamentos” e “não escorregamentos” utilizadas no modelo de predição



Fonte: produção do autor.

4.3.2 Identificação da correlação entre as variáveis

Um passo importante na modelagem de escorregamentos é a identificação da multicolinearidade no conjunto de dados, ou seja, com o objetivo de evitar erros e tendências no modelo, é feito uma análise para identificar o quanto as variáveis independentes estão correlacionadas (TIEN BUI et al., 2016). Para quantificar a correlação entre os fatores condicionantes utilizou-se o coeficiente de correlação Pearson.

Essa técnica permite identificar o quanto os atributos estão associados, possibilitando uma análise de cada par de variáveis. Valores de correlação de Pearson maiores que 0,7 indicam alta colinearidade (BOOTH et al. 1994; TIEN BUI et al., 2016). Quanto menos dispersos estiverem os dados, mais forte será a dependência, isto é, a associação entre as variáveis. O coeficiente de correlação assume um valor entre -1 e 1 , sendo que se essa relação for igual a 1 a correlação é positiva perfeita, se igual a -1 é considerada negativa perfeita e igual a 0 significa que a correlação é nula.

4.3.3 Análise da importância das variáveis condicionantes

Na modelagem de escorregamentos, todos os fatores de condicionantes inseridos no conjunto inicial podem não ter uma capacidade preditiva igual e, até mesmo, alguns deles podem causar um ruído que reduz a capacidade de previsão dos modelos resultantes. Portanto, habilidades preditivas dos fatores condicionantes devem ser quantificadas e posteriormente analisadas (MARTÍNEZ-ÁLVAREZ et al., 2013). De acordo com Sahin et al. (2020), infelizmente, não há um procedimento global bem definido na definição das variáveis condicionantes que devem ser utilizadas nos modelos de predição. Esses fatores são geralmente escolhidos de acordo com a área de estudo, tipo de escorregamento encontrado, disponibilidade de dados e estudos anteriores encontrados na literatura (AYALEW; YAMAGISHI, 2005). Uma alternativa para essa questão é utilizar métodos estatísticos que apoiem na identificação da importância dos atributos e propiciem uma melhor utilização dos fatores condicionantes dentro dos fatores previamente selecionados.

Durante o processo de análise da importância dos fatores condicionantes utilizados na modelagem da suscetibilidade, foi realizado uma classificação ordenada dos atributos a partir do conjunto de dados de treinamento, determinando quais os fatores condicionantes são os mais influentes para a construção do modelo de predição. Essa classificação configura o processo de ordenamento dos atributos através do valor de uma função de pontuação selecionada, muitas vezes baseada em testes estatísticos, que geralmente medem a relevância do atributo (SAHIN et al., 2020). Para identificar a importância das variáveis no modelo de predição, utilizou-se no presente trabalho a técnica Ganho de Informação (*Information Gain*), utilizada também por Park e Kim (2019).

Essa técnica é baseada no conceito de entropia, que é normalmente utilizada para medir a pureza ou impureza de um determinado conjunto de dados, quantificando o grau de incerteza de uma informação ou de uma variável envolvida no processo aleatório. A entropia mede a quantidade de informação contida nos dados e quantifica até que ponto estão espalhados sobre seus valores possíveis. Assim, alta entropia significa que os dados estão espalhados tanto quanto possível, enquanto baixa entropia significa que os dados estão quase todos concentrados em um valor. Se a entropia for baixa, portanto, teremos alto conteúdo de informação e provavelmente apresentaremos um forte padrão (GALAR; KUMAR, 2017). A entropia quantifica informações como a incerteza média sobre os resultados da variável: se todas as observações da variável levam o mesmo valor, não há incerteza e a entropia é zero; se, ao contrário, a variável assume valores diferentes, todas com a mesma probabilidade de ocorrência, a entropia é máxima e reflete incerteza máxima (XIONG; FAES; IVANOV, 2017).

Para diminuir o grau de incerteza das informações, é utilizado o ganho de informação, que rastreia a diminuição da entropia para quantificar a importância dos fatores condicionantes e é considerado a técnica padrão para medir as habilidades preditivas de fatores em mineração de dados (TIEN BUI et al., 2016). Um fator condicionante com um valor de ganho de informação mais alto tem uma classificação mais alta, o que significa que é mais importante para o modelo de escorregamento e possui menor entropia. Em contraste, um fator condicionante com um valor ganho de informação igual a zero deve ser removida do conjunto de dados, visto que ele não faz nenhuma contribuição efetiva no modelo.

4.3.4 Aplicação do método Floresta Aleatória

A precisão do modelo de predição depende não apenas do algoritmo de aprendizagem, mas também dos parâmetros estabelecidos para construir as árvores de decisão, que são definidos antes de iniciar o processo de aprendizagem. Os parâmetros na Floresta Aleatória são utilizados para aumentar o poder preditivo do modelo e/ou para torná-lo mais rápido. Existem diversos parâmetros que podem ser considerados na calibração do algoritmo Floresta Aleatória, fazendo com que o processo de identificar qual a melhor combinação das diversas variáveis seja complexo.

Uma maneira de otimizar esse processo é criar uma grade desses parâmetros e testar exaustivamente todas as combinações possíveis, buscando os melhores valores de acurácia. Este método é referido como busca em grade (*Grid Search*), que seleciona de forma automática os parâmetros que obtiveram os menores erros (DUBEY et al., 2020). De modo geral, utilizando os parâmetros padrão, os modelos baseados na Floresta Aleatória produzem bons resultados, todavia a otimização de parâmetros tem uma grande influência na precisão do modelo de predição, visto que seleciona a melhor combinação que, conseqüentemente, resultará nos melhores índices de acurácia.

Usualmente essa técnica incorpora o procedimento de validação cruzada (*cross-validation*), método estatístico usado para verificar o desempenho do classificador e calcular a acurácia associada a cada configuração de parâmetro. Essa técnica define, a partir dos dados originais, uma parte como o conjunto de treinamento e a outra parte como o conjunto de validação. O conjunto de treinamento é usado para treinar o classificador e o conjunto de validação é usado para testar a acurácia do modelo de treinamento e avaliar o desempenho do classificador.

Em geral, os dados originais são divididos em subconjuntos, sendo que para cada iteração um conjunto formado por $N-1$ é utilizado como subconjunto de treinamento e o subconjunto restante será utilizado para validação, gerando um valor de acurácia. Desta forma, N modelos são obtidos e a média do valor de acurácia obtido para cada iteração é usada como índice final de performance do classificador (YAO et al., 2021).

A biblioteca *Scikit-Learn* oferece uma diversidade de possibilidades para a calibração dessas variáveis, porém, optou-se por analisar somente os parâmetros número de árvores e número máximo de atributos testados em cada nó, conforme o Quadro 10. Park e Kim (2019) e Taalab (2018) descrevem que esses são os parâmetros mais críticos na construção dos

modelos. Os demais foram utilizados de acordo com os valores padrões já estabelecidos na biblioteca.

Quadro 10. Parâmetros de entrada utilizados na construção do algoritmo Floresta Aleatória

Parâmetros	Descrição
Número de Árvores da Floresta (<i>ntree</i>)	Número de árvores de decisão utilizada no modelo. Com o objetivo de avaliar se o número de árvores impacta significativamente nos resultados das predições, testou-se o algoritmo com 50, 100, 200, 300, 400 e 500 árvores
Número de variáveis testadas em cada nó (<i>Mtry</i>)	Número de variáveis testadas em cada nó de forma aleatória. Geralmente, esse valor é representado pela $\sqrt{\text{número de atributos}}$. Durante a calibração do modelo, optou-se por trabalhar com os valores 4 e 3

Fonte: produção do autor.

4.3.5 Métricas de performance

As métricas de desempenho utilizadas – acurácia global, precisão, Curva de Característica de Operação do Receptor (ROC) e Área Sob a Curva (AUC) - foram obtidas através da matriz de confusão, que segundo Park e Kim (2019) é uma tabela que mostra as frequências de classificação para cada classe do modelo e inclui os valores de verdadeiro positivo (VP), falso positivo (FP), verdadeiro negativo (VN) e falso negativo (FN), conforme a Figura 26.

A acurácia global (Equação 2) é uma medida que expressa a proporção da quantidade de predições corretas em relação a quantidade total de observações consideradas. Dessa forma, essa medida irá analisar a quantidade de pixels de escorregamentos ou não escorregamentos classificados corretamente no modelo (TIEN BUI et al., 2016). Os valores próximos de 1 indicam os melhores modelos.

$$\text{Acurácia Global} = (\text{VP} + \text{VN}) / (\text{VP} + \text{FP} + \text{VN} + \text{FN}) \quad (2)$$

A precisão (Equação 3) representa a proporção de instâncias que são classificadas corretamente em uma classe dividida pelo total de instâncias classificadas como positivas (DOU et al., 2019). Em outras palavras, é a proporção de VP em relação ao número total de amostras classificadas como positivas (PRADO et al., 2020). Segundo Sahin et al. (2020), os

resultados variam entre 0 e 1 e quanto mais próximo de 1 mais corretas as instâncias foram classificadas.

$$\text{Precisão} = \text{VP} / (\text{VP} + \text{FP}) \quad (3)$$

Figura 26. Matriz de Confusão

OBSERVAÇÃO	Não Escorregamento	Verdadeiro Negativo (TN)	Falso Positivo (FP)
	Escorregamento	Falso Negativo (FN)	Verdadeiro Positivo (TP)
		Estável	Instável
		PREDIÇÃO	

Fonte: produção do autor.

A curva ROC tem sido amplamente utilizada para validar o desempenho de modelos de suscetibilidade a escorregamentos. É um método útil para retratar de forma bidimensional o desempenho dos modelos de predição, representado através de uma curva de probabilidade obtida através de um gráfico que plota a sensibilidade (Equação 5) ou taxa de verdadeiro positivo (eixo x) versus a especificidade (Equação 6) ou taxa de falso positivo (eixo y) do modelo resultante (TIEN BUI et al., 2015).

A sensibilidade (Equação 4) (*recall* ou *sensitivity*) diz respeito a porcentagem dos pixels de ocorrência de escorregamento que foram previstos corretamente, ou seja, a razão de VP para o número total de amostras de classe positiva (PRADO et al., 2020)

$$\text{Sensibilidade} = \text{VP} / (\text{VP} + \text{FN}) \quad (4)$$

Já a especificidade (Equação 5) refere-se a porcentagem dos pixels previstos de escorregamento de toda a área de estudo:

$$\text{Especificidade} = \text{VN}/(\text{VN}+\text{FP}) \quad (5)$$

Park e Kim (2019) abordam que a curva ROC pode ser classificada como uma curva de taxa de sucesso ou curva de taxa de predição, dependendo do conjunto de dados usado. A curva de taxa de sucesso, calculada usando o conjunto de dados de treinamento representa o quão bem os modelos de suscetibilidade se ajustam aos dados. A curva da taxa de predição, calculada usando o conjunto de dados de teste, representa o quanto o modelo e os fatores condicionantes preveem um escorregamento

A AUC estima a eficácia da predição realizada e é equivalente a probabilidade de que um classificador irá escolher aleatoriamente um registro verdadeiro positivo maior do que um falso negativo. Quanto mais próximo de 1 o valor de AUC, melhor a acurácia do modelo (SEGONI et al., 2019). Por outro lado, um valor de AUC próximo a 0,5 significa resultados ruins. Em geral, considera-se que se a AUC do modelo for maior que 0,7, o modelo tem alta precisão (RAJA et al., 2017).

4.3.6 Geração dos mapas de suscetibilidade

Após a construção e seleção do modelo, o algoritmo foi aplicado para a área inteira com a finalidade de gerar o mapa de suscetibilidade a escorregamentos. O modelo gerado permite a visualização binária da suscetibilidade a escorregamento (classificada em apenas 0 e 1) e a visibilidade em probabilidades de ocorrência (variando entre 0 e 1 ou 0 a 100%). Os mapas de suscetibilidade gerados foram expressos quanto ao índice de escorregamentos em termos de probabilidade de ocorrência, divididos em cinco graus de suscetibilidade: muito alta, alta, média, baixa e muito baixa. Essas classes foram definidas pelos intervalos de quebras naturais, mesmo processo adotado por Park e Kim (2019).

5 RESULTADOS

A seguir são apresentados os resultados e discussões do presente trabalho, envolvendo a relação entre as cicatrizes de escorregamento e as classes de cada fator condicionante determinada através do método razão de frequência, a análise da correlação das variáveis, a identificação dos atributos mais importante para a construção do modelo de acordo com o ganho de informação, as etapas de treinamento e construção do modelo a partir da técnica de busca em grade e seleção do melhor modelo de predição, a aplicação do mesmo para toda a área de estudo, a análise da performance do classificador e do modelo final e a geração dos mapas de suscetibilidade.

5.1 Razão da Frequência

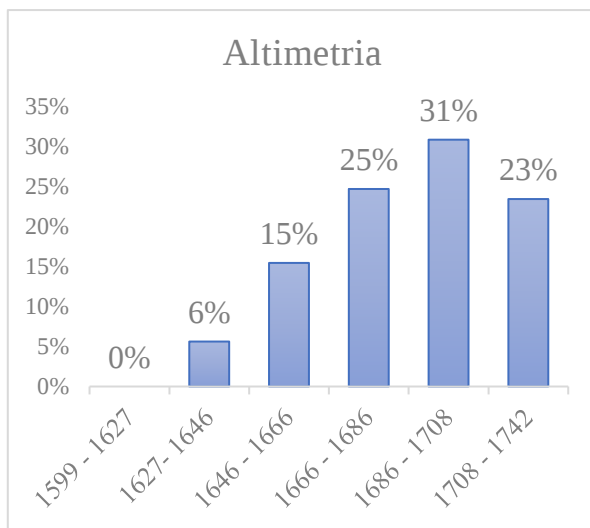
Através da técnica FR foi possível identificar a distribuição da ocorrência das cicatrizes nos intervalos de classes de cada fator condicionante utilizado no modelo de predição. Dessa forma, essa etapa permitiu analisar, para cada fator, quais os intervalos de classes que mais se associavam à deflagração dos escorregamentos e quais seriam os limiares prováveis para cada variável. Outra análise que o FR possibilitou diz respeito à comparação entre os fatores condicionantes em relação à distribuição dos escorregamentos: houve fatores em que os escorregamentos se relacionavam em sua grande maioria a apenas uma classe. Em contrapartida, para outros atributos, a distribuição desses processos se vinculava a diversos intervalos, não permitindo definir com clareza qual seria a classe de maior probabilidade.

Em relação aos parâmetros morfométricos, a altimetria (Figura 27 (a)), apresentou uma distribuição ampla de ocorrência de escorregamentos em todas as classes relacionadas a esse atributo, exceto para as elevações mais baixas (1599-1627 metros). O intervalo de elevação 1689-1708 metros foi a classe que apresentou o valor FR mais alto, perfazendo 31% da área de estudo. Já o aspecto (Figura 27 (b)) concentrou as cicatrizes em apenas três intervalos: Leste (4%), Sudeste (22%) e Sul (74%). Esse padrão é reforçado por Hiruma (2004), que descreve que as encostas da face Sul-Sudeste sofrem exposição direta às frentes polares, sendo caracterizadas por serem menos expostas a incidências solares e úmidas.

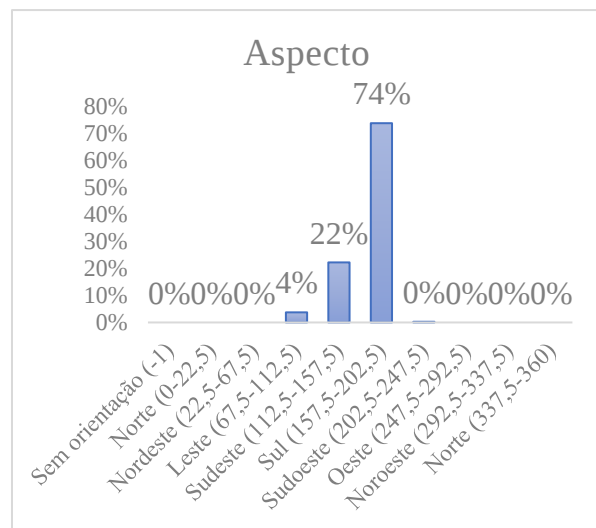
Em relação às curvaturas, observa-se que nos três tipos analisados, a côncava é a que apresenta as maiores relações com a deflagração de escorregamentos. Na curvatura combinada (Figura 27 (c)), as cicatrizes ocorrem nas três classes com a seguinte divisão: côncava (59%),

retilínea (15%) e convexa (26%). Na curvatura em perfil (Figura 27 (d)) as ocorrências se situam em 45% nas curvaturas convexas, 4% nas retilíneas e 51% nas côncavas. Nesse tipo de curvatura, em específico, as classes côncava e convexa apresentam valores muito próximos, o que pode prejudicar a definição de um limiar para esse atributo. Na curvatura em planta (Figura 27 (e)), 73% das cicatrizes ocorrem nas curvaturas côncavas, 3% nas retilíneas e 26% nas convexas.

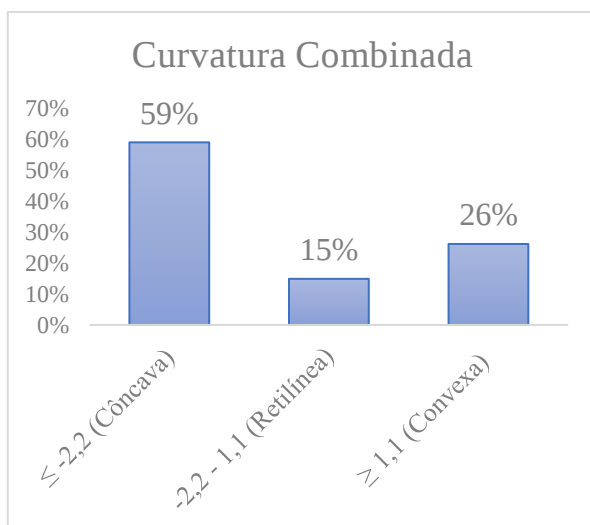
Figura 27. Distribuição da RF normalizada dos parâmetros morfométricos de acordo com (a) Altimetria; (b) Aspecto; (c) Curvatura Combinada; (d) Curvatura em Perfil; (e) Curvatura em Planta; (f) Declividade; (g) SPI; (h) STI; (i) TPI; (j) TRI; (k) TWI



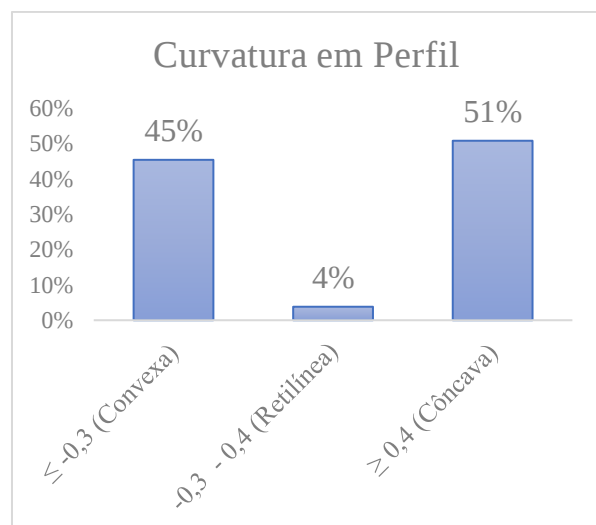
(a)



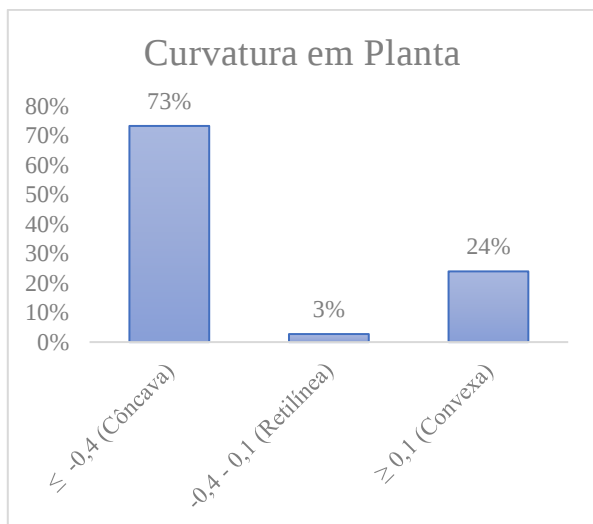
(b)



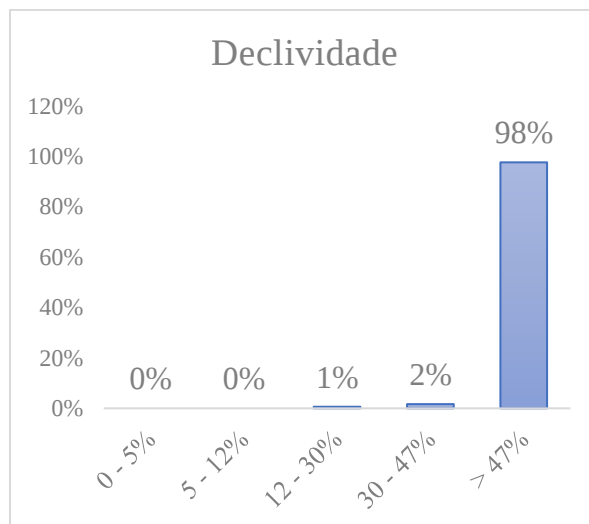
(c)



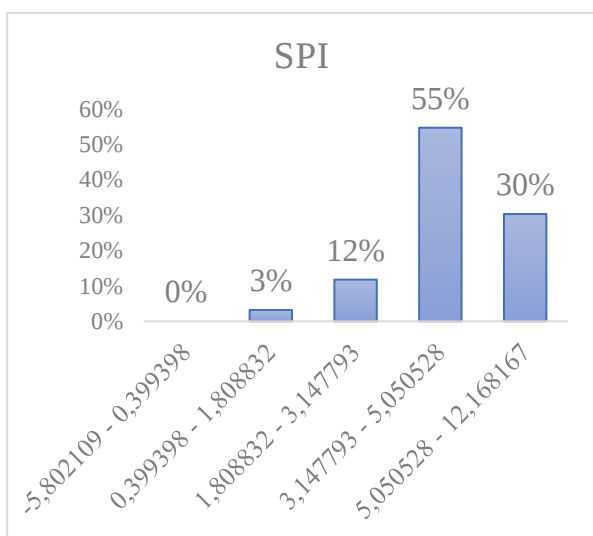
(d)



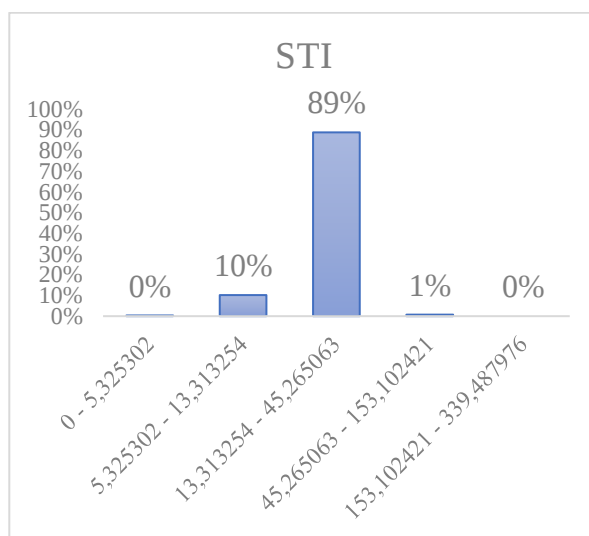
(e)



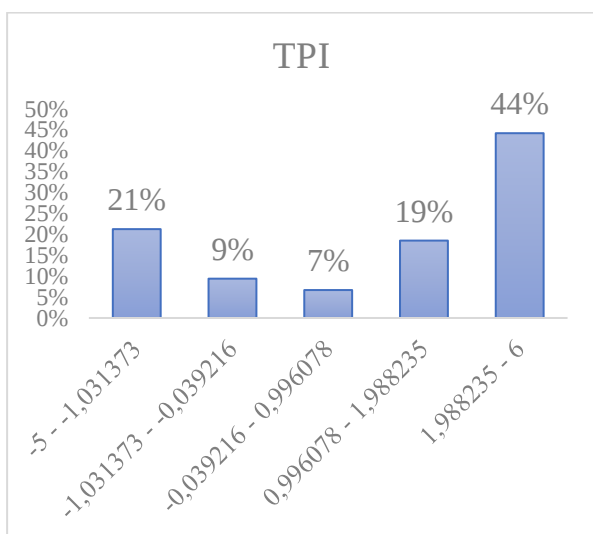
(f)



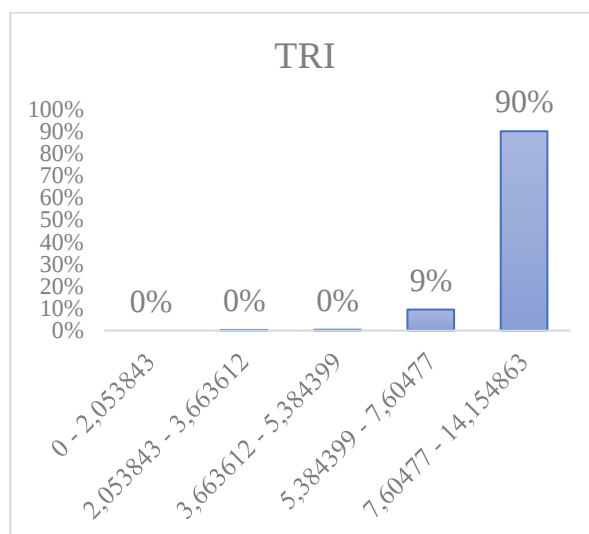
(g)



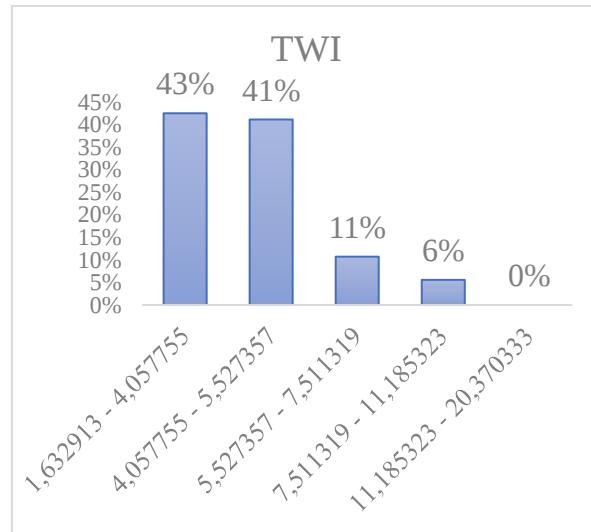
(h)



(i)



(j)



(k)

Fonte: produção do autor.

De acordo com Vieira (2007) as porções côncavas, em planta e em perfil, concentram maior volume de água e sedimentos, permitem a subida rápida da pressão e consequentemente a redução da estabilidade da encosta, principalmente durante eventos pluviométricos intensos. Além disso, muitas áreas côncavas na paisagem podem registrar a presença de material coluvionar, cuja origem está associada a processos de movimentos gravitacionais pretéritos.

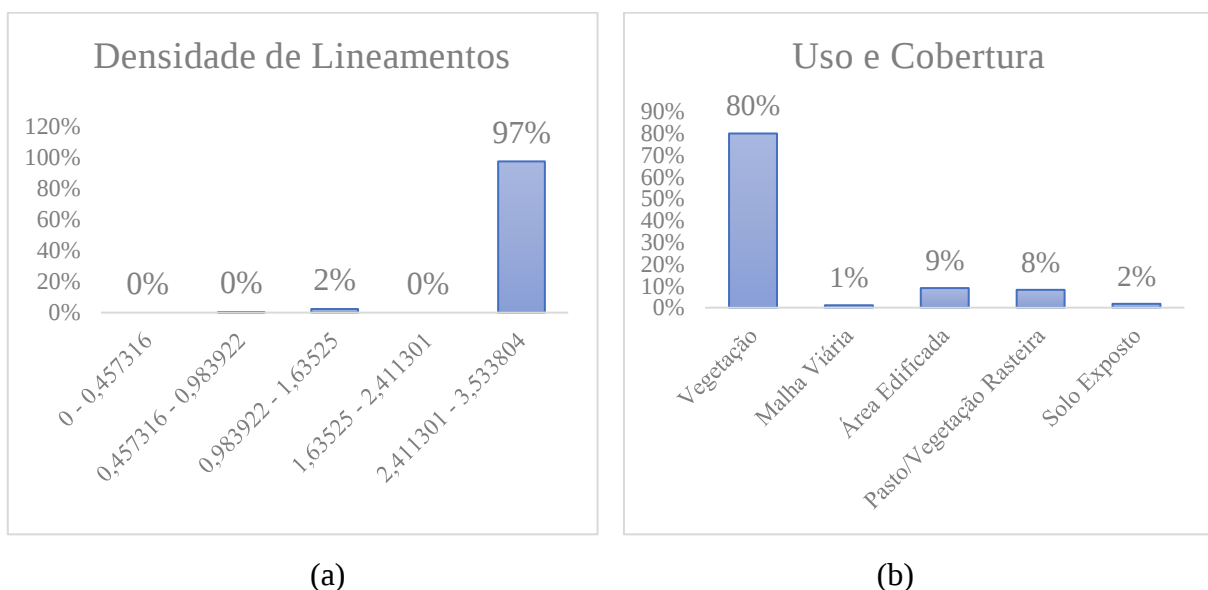
Praticamente todas as cicatrizes de escorregamentos se encontram no intervalo de declividade $>47\%$, concentrando 98% desses processos nessa classe (Figura 27 (f)). Os intervalos 12-30% e 30-47% compreendem, respectivamente, 1% a 2% das ocorrências. Dessa maneira, para a classe $>47\%$, que é a mais representativa em porcentagem de área na bacia, é esperado encontrar um maior número de cicatrizes de escorregamentos situadas nesse intervalo. Para fins de parcelamento urbano, não é recomendado a ocupação em terrenos com declividade igual ou superior a 30%. O intervalo acima de 47% compreende vertentes que apresentam inclinações maiores que 25° e, de acordo com o Código Florestal Brasileiro nessas porções é proibida a derrubada de florestas.

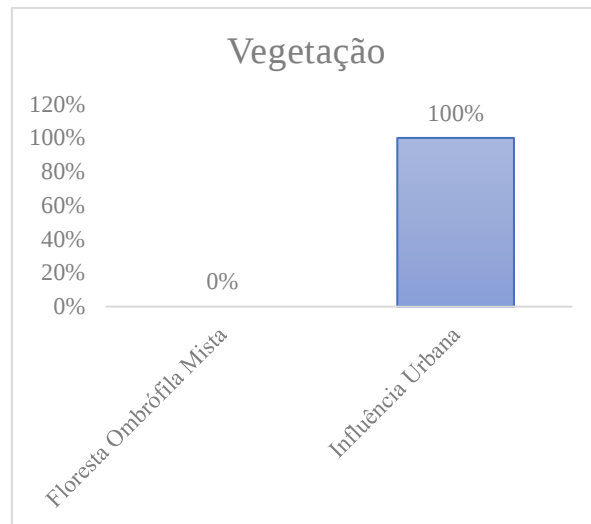
Em referência aos fatores morfométricos secundários, o SPI (Figura 27 (g)) apresentou ocorrências em quatro das cinco classes, sendo que o intervalo 3,147793 - 5,050528 apresentou as maiores concentrações (55%). Já o STI (Figura 27 (h)) possui uma concentração predominantemente na classe intermediária 13,313254 - 45,265063 (89%). O TPI (Figura 27 (i)) está amplamente distribuído em todas os intervalos de classe, sendo que a classe mais elevada (1,988235 – 6) concentra a maioria das cicatrizes (44%). O TRI (Figura 27 (j)) define

claramente a classe em que os escorregamentos apresentam maior frequência, cujo intervalo 7,60477 - 14,154863 compreende (90%) desses processos. O parâmetro TWI (Figura 27 (k)) concentra as cicatrizes principalmente nos dois intervalos mais baixos para o índice: 1,632913 - 4,057755 (43%) e 4,057755 - 5,527357 (41%).

No cenário dos fatores temáticos, a densidade de lineamentos (Figura 28 (a)) concentra as cicatrizes no intervalo que apresenta valores maiores que 2,41 (98%), que corresponde as áreas que apresentam a maior concentração dessas feições. Dessa forma, diante dos valores de FR, quanto maior a presença de lineamentos, mais suscetíveis as áreas estão à deflagração de escorregamentos. Essas feições muitas vezes podem estar preservadas no solo saprolítico, favorecendo os planos de ruptura e, conseqüentemente, a ocorrência desses processos. O uso e cobertura da terra (Figura 28 (b)) concentra as cicatrizes majoritariamente nas áreas de vegetação (80%), seguidas das áreas edificadas (9%), pasto/vegetação rasteira (8%), solo exposto (2%) e malha viária (1%). Apesar das cicatrizes de escorregamentos estarem concentradas nas áreas de vegetação, é importante ressaltar que na área de estudo essas classes estão associadas a áreas declivosas e localizadas próximas a zonas urbanas. Essas porções do território podem ter sido modificadas devido a ações antrópicas e/ou deslizadas e posteriormente preenchidas por vegetação. Essa afirmação é associada ao último atributo utilizado, a vegetação, que mostra que 100% das áreas que concentram as cicatrizes de escorregamentos possuem influência urbana (Figura 28 (c)).

Figura 28. Distribuição da RF normalizada para os parâmetros temáticos: (a) Densidade de Lineamentos; (b) Uso e Cobertura; (c) Vegetação





(c)

Fonte: produção do autor.

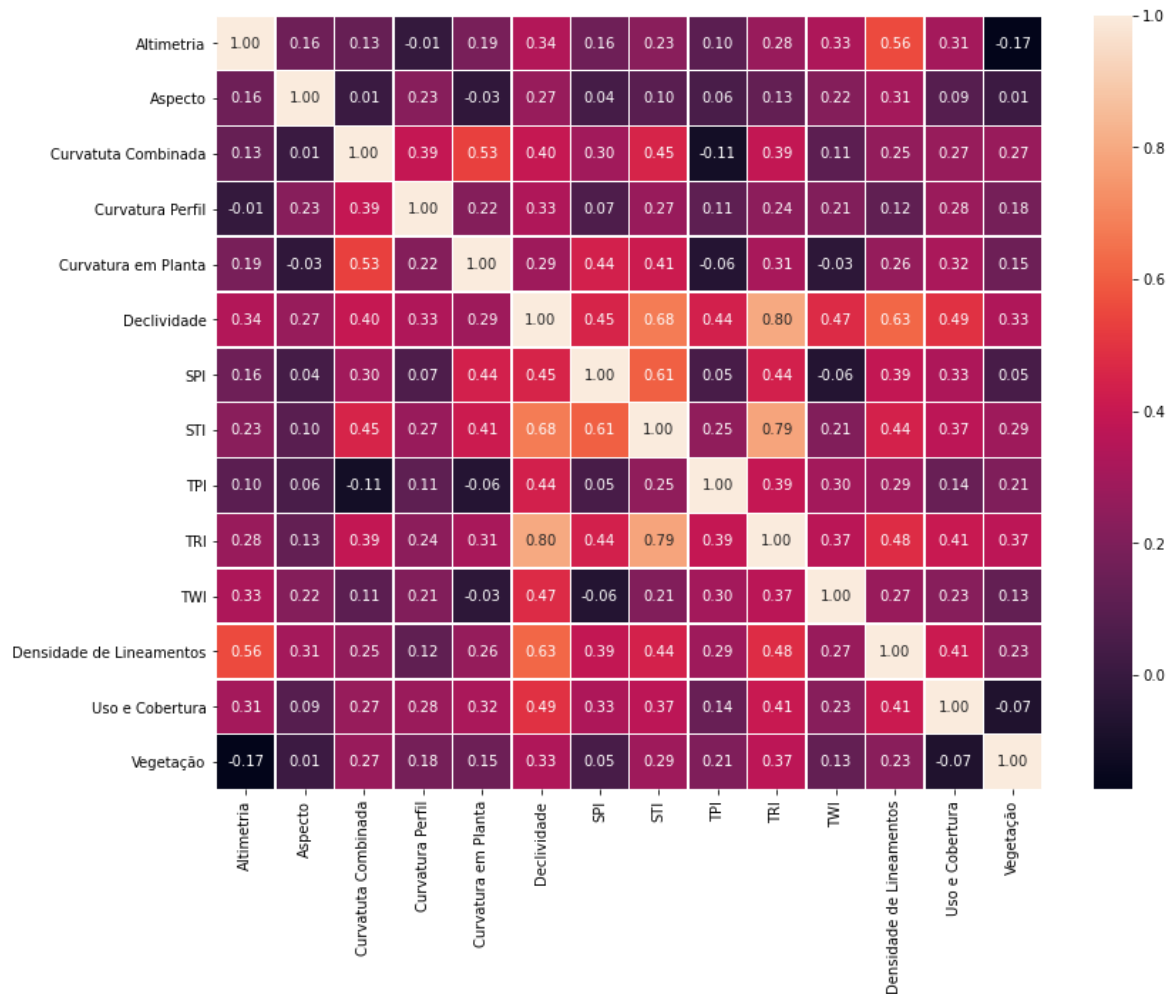
5.2 Identificação da correlação entre as variáveis

A análise da correlação linear de Pearson aplicada aos fatores condicionantes permitiu a identificação da multicolinearidade entre as variáveis selecionadas para a construção do modelo. Pares de atributos que apresentaram índices maiores que 0,7 indicam uma alta correlação e devem ser observados com maior atenção (TIEN BUI et al., 2016).

Os resultados dessa etapa indicaram que as correlações de diferentes fatores condicionantes tomados par a par variaram entre -0,17 a 0,8, conforme a matriz de distância ilustrada na Figura 29. Os maiores valores de correlação se situaram nas relações entre a declividade e o TRI (0,8) e o TRI e o STI (0,79). De acordo com Riley et al. (1999), o TRI é um parâmetro secundário que indica as variações do relevo e pode ser calculado como a variação da declividade dentro de uma determinada área. Dessa forma, o TRI pode apresentar forte correlação com a declividade por utilizar a mesma como base da construção desse índice. Quando comparados os mapas dessas duas variáveis, apresentado no capítulo anterior, percebe-se que eles configuram certa semelhança.

Mesmo com os valores de correlação acima de 0,7 citados acima, optou-se por manter todas as variáveis nessa etapa, visto que os parâmetros TRI, declividade e STI são derivados do mesmo MDE. As demais variáveis independentes não apresentaram valores altos de correlação.

Figura 29. Matriz de distância com os valores de correlação para os fatores condicionantes



Fonte: produção do autor.

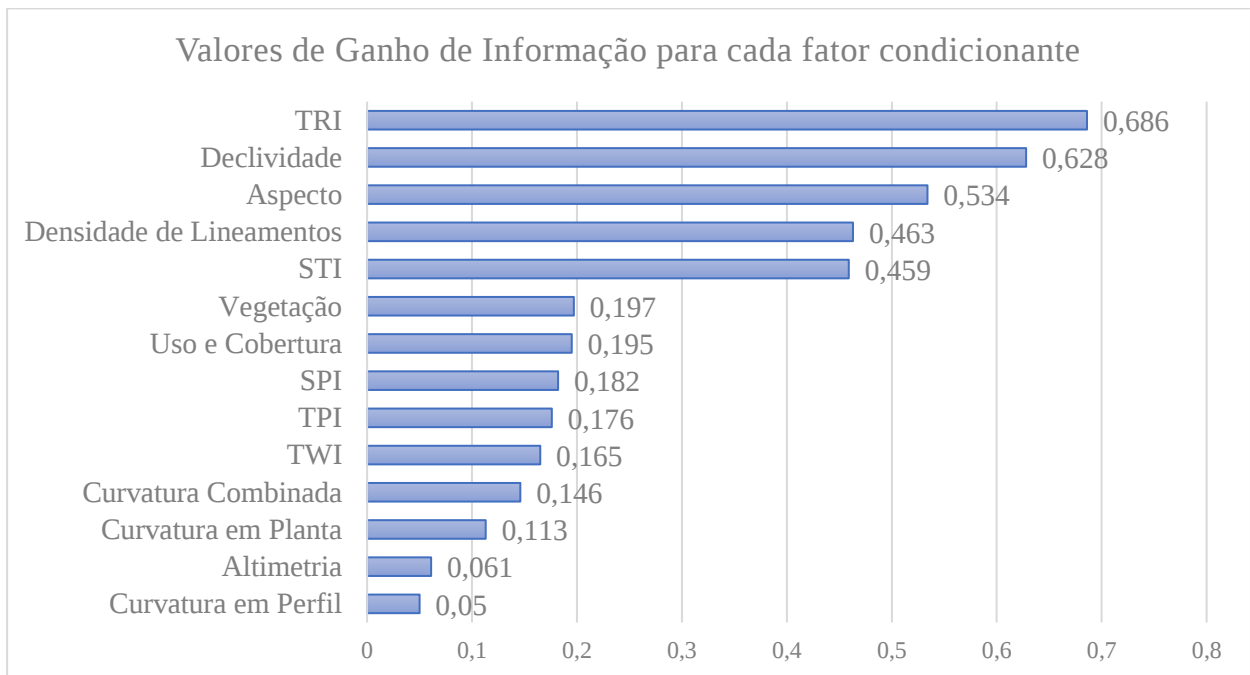
5.3 Importância dos atributos

O ganho de informação foi usado para determinar os melhores fatores condicionantes que processam o máximo de informações sobre as classes “escorregamento” ou não “escorregamento”. Algoritmos baseados em árvores de decisão sempre buscam maximizar o ganho de informação e um atributo com maior ganho de informação será testado ou dividido primeiro na raiz ou nó das folhas. Essa etapa foi utilizada com as variáveis já ajustadas pelo método de FR.

O valor do ganho de informação foi calculado para cada variável explicativa e, de acordo com os resultados mostrados pela Figura 30, os fatores que apresentam maiores importâncias na área de estudo foram o TRI e declividade com valores de ganho de informação de 0,658 e 0,632, respectivamente. Em contrapartida, a altimetria e curvatura em perfil configuraram as

variáveis menos importantes, conforme indicado pelos valores 0,061 e 0,05. Sahin et al. (2020) obteve o TRI seguido da declividade como atributos com maiores valores de ganho de informação. Park e Kim (2019) e Shirvani (2020), também apresentou o TRI como a variável com a maior importância.

Figura 30. Importância dos fatores condicionantes de acordo com o Ganho de Informação



Fonte: produção do autor.

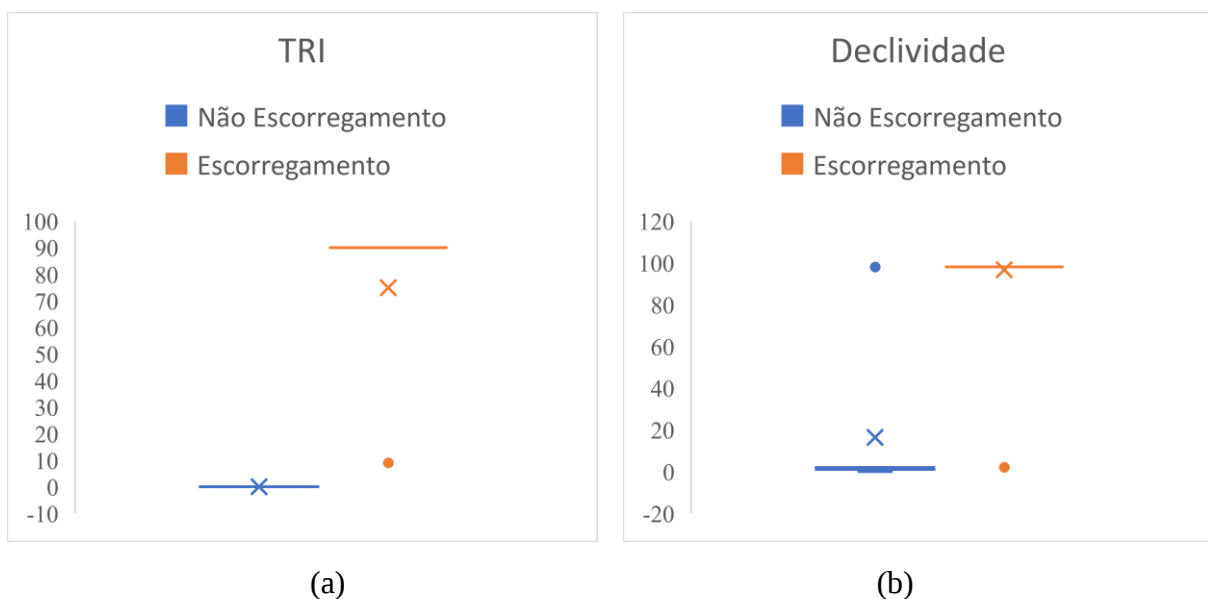
Em geral, quando se analisa o diagrama de caixa para verificar a relação entre os intervalos de classe do TRI e a ocorrência ou não ocorrência de escorregamentos (Figura 31 (a)), percebe-se que as divisões são bem separadas e há um padrão bem definido. É possível identificar que há uma relação entre os valores mais altos de TRI e a deflagração desses processos e, em contrapartida, os valores mais baixos desse atributo concentram as não ocorrências de escorregamento.

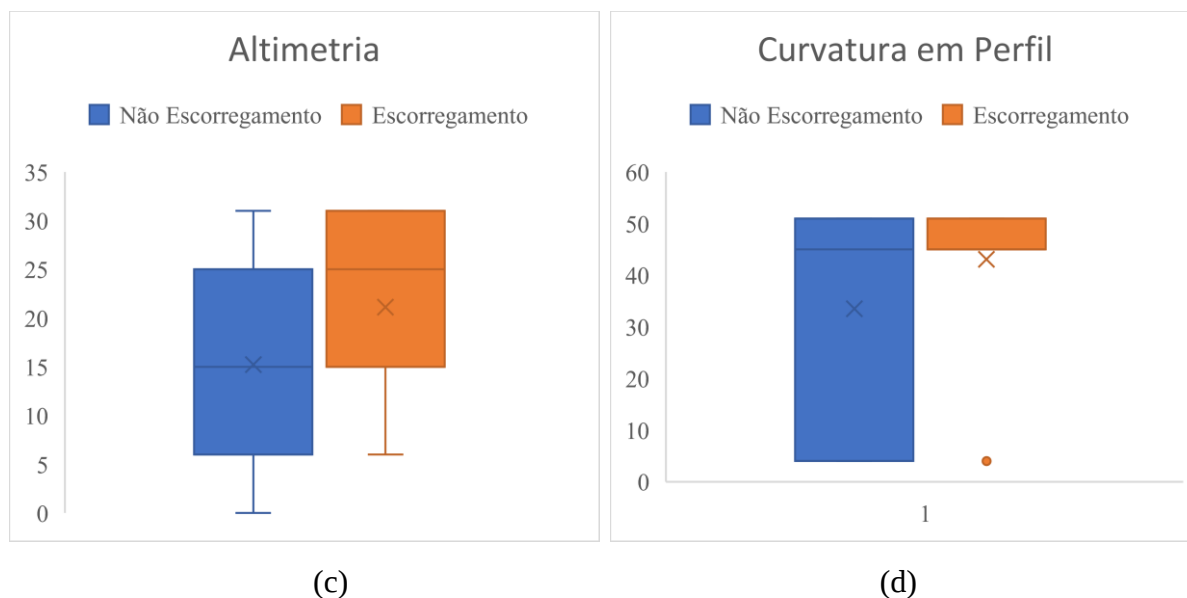
Já a declividade, amplamente utilizada nos modelos de suscetibilidade, configura um dos fatores principais associados a esses processos. Assim como a variável TRI, na área de estudo há uma distinção bem definida das classes em que ocorrem os escorregamentos, representados pelas áreas mais declivosas, enquanto as classes em que esses fenômenos não são deflagrados são aquelas representadas pelas áreas com menor declividade. Essa distribuição fica visível na Figura 31 (b)).

A altimetria, atributo usado na maioria dos modelos, resultou no penúltimo atributo de importância na influência a escorregamentos. Isso se deve ao fato de terem sido deflagrados eventos de escorregamentos em diferentes altimetrias na Bacia do Córrego Piracuama, não havendo concentrações em poucas classes. Os valores de elevação da escarpa do planalto apresentam uma alta variabilidade ao longo da área de estudo, fazendo com que os pontos de ruptura dos escorregamentos apresentem diferenças de elevação para praticamente todas as classes de altimetria da área de estudo, exceto para a classe que concentra os valores mais baixos, distribuição evidenciada pela Figura 31 (c)).

As curvaturas em geral, apesar de serem constantemente empregadas, não configuram os atributos mais importantes nos mapeamentos de suscetibilidade. Sahin et al. (2020) e Park e Kim (2019) ao aplicarem o método de ganho de informação também obtiveram baixa importância quando analisaram esses fatores condicionantes. Na área de estudo, a curvatura em perfil representou uma concentração de ocorrência de escorregamentos nas classes côncava e convexa muito semelhante, fazendo com que esse atributo se torne menos efetivo na diferenciação de quais intervalos de curvatura são mais suscetíveis a escorregamentos (Figura 31 (d)). De todo modo, todas as variáveis explicativas de escorregamento utilizadas nesse estudo contribuíram para os modelos de escorregamento, pois os valores de ganho de dessas variáveis foram maiores que zero.

Figura 31. Diagrama de caixa para os atributos mais relevantes para o modelo: (a) TRI e (b) Declividade e menos relevantes: (c) Altimetria e (d) Curvatura em Perfil





Fonte: produção do autor.

5.4 Construção e aplicação do modelo de predição

5.4.1. Busca em grade

O conjunto de dados de treinamento foi usado para treinar e construir o modelo de suscetibilidade a escorregamentos utilizando o algoritmo Floresta Aleatória. A partir das amostras de treino, aplicou-se o método busca em grade, que identificou qual a melhor combinação dos parâmetros (número de árvores e número de atributos para cada nó) para a seleção do melhor modelo de predição. Ao todo, foram testadas 12 combinações que são descritas no Quadro 11. Durante o processo de busca em grade foram realizadas ao todo 10 iterações, através do método de validação cruzada, que analisaram todas as combinações possíveis entre os parâmetros pré-estabelecidos, resultando em médias de valores de acurácia e precisão (Quadro 10).

Diante dos resultados, todas as combinações apresentaram altos valores de acurácia, porém as combinações 4, 6, 9, 10, 11 e 12 registram os maiores valores médios de acurácia (0,97). Apesar dos valores semelhantes, o método busca em grade selecionou a Combinação 4 ($mtry = 4$; $ntree = 300$), visto que, mesmo que o valor médio de acurácia seja o principal critério de seleção dos melhores parâmetros, há outras variáveis envolvidas nesse processo de escolha como a média do tempo de processamento, que nesse caso foi o menor dentre todas as combinações.

Quadro 11. Combinações realizadas durante a busca em grade e seus respectivos valores médios de acurácia e precisão

Combinação	Número de Atributos	Número de Árvores	Média da Acurácia	Média da Precisão
1	4	50	0,96	0,954762
2	4	100	0,96	0,954762
3	4	200	0,96	0,954762
4	4	300	0,97	0,954762
5	4	400	0,96	0,954762
6	4	500	0,97	0,954762
7	3	50	0,96	0,954762
8	3	100	0,96	0,954762
9	3	200	0,97	0,954762
10	3	300	0,97	0,954762
11	3	400	0,97	0,954762
12	3	500	0,97	0,954762

Fonte: produção do autor.

Os valores de precisão foram exatamente iguais para todas as combinações. A calibração dos parâmetros *ntree* e *mtry* não interferiram de forma significativa no modelo de predição, porém, é válido testar a busca em grade com um intervalo maior entre o número de árvores, bem como testar com outros parâmetros possíveis no algoritmo Floresta Aleatória. A Figura 32 mostra os valores médios de acurácia e precisão para cada uma das combinações.

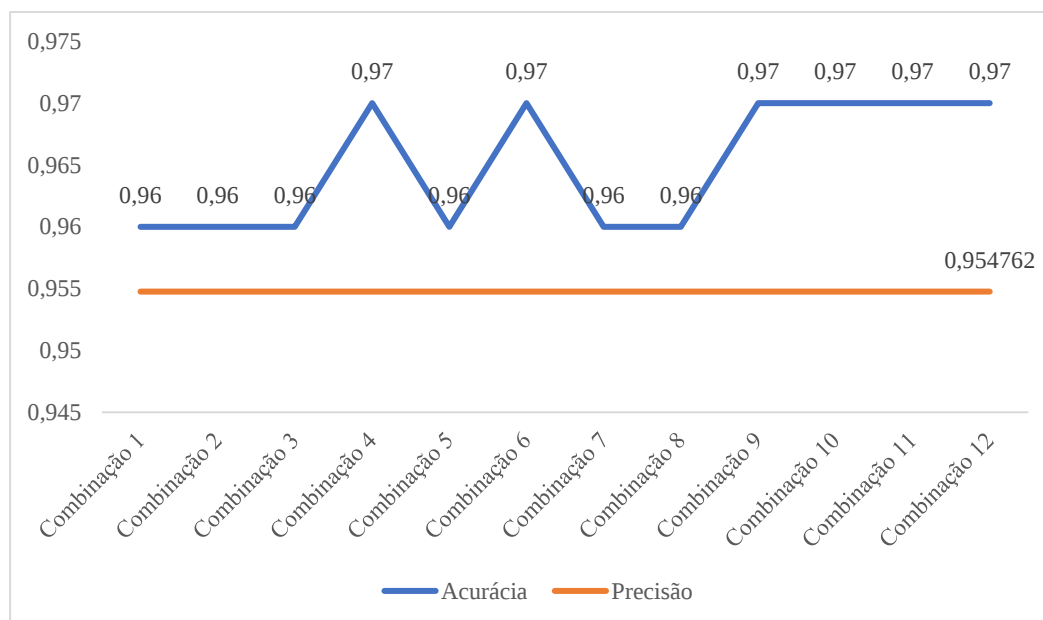
De acordo com os resultados, é possível perceber que o número de árvores não é proporcional à qualidade do modelo, visto que não necessariamente o maior número de árvores configura os melhores resultados. O método de busca em grade permitiu a identificação da melhor combinação de forma otimizada, obtendo a melhor capacidade preditiva do modelo e evitando etapas manuais para a construção e análise de diversos algoritmos de predição.

5.4.2 Avaliação de desempenho

Após a construção e seleção do modelo com base na busca em grade, o melhor modelo de predição foi aplicado para toda a área de estudo para produzir o mapa de suscetibilidade a escorregamentos. A partir do conjunto de dados de teste, foi possível analisar o desempenho do modelo final, através da comparação das amostras de teste contendo as classes

previamente conhecidas com as áreas do mapa de suscetibilidade, em que o modelo de predição classificou como escorregamento e não escorregamento.

Figura 32. Valores médios de acurácia para cada combinação da busca em grade



Fonte: produção do autor.

As métricas de performance foram aplicadas para avaliar o desempenho tanto do modelo de treinamento, com o objetivo de validar se de fato aquele era um bom algoritmo para prosseguir com a classificação da área total, quanto para o modelo final, a partir dos conjuntos de dados destinados para cada etapa. Esses resultados são descritos na Tabela 2 que apresenta as taxas de VP, VN, FP, FN, Acurácia e Precisão. Em geral, todas as métricas resultantes sugerem que os modelos de treinamento e teste apresentaram um ótimo desempenho, visto que, para as métricas de Acurácia e Precisão, quanto mais próximos de 1 mais corretas as instâncias foram classificadas.

As curvas ROC também foram construídas para ambos os modelos, a partir da Taxa de VP e Taxa de FP. A curva ROC pode ser verificada quantitativamente quando a área sob a curva ROC (AUC) é calculada. Segundo Dou et al. (2019), o desempenho de classificação dos modelos pode ser referido como: excelente (0,9-1); bom (0,8-0,9), razoável (0,7-0,8), ruim (0,6 < 0,7) e sem importância (< 0,5). Através dos resultados, foi possível observar que os valores de AUC para o modelo de treinamento resultou em 1 e para o modelo final resultou em 0,969,

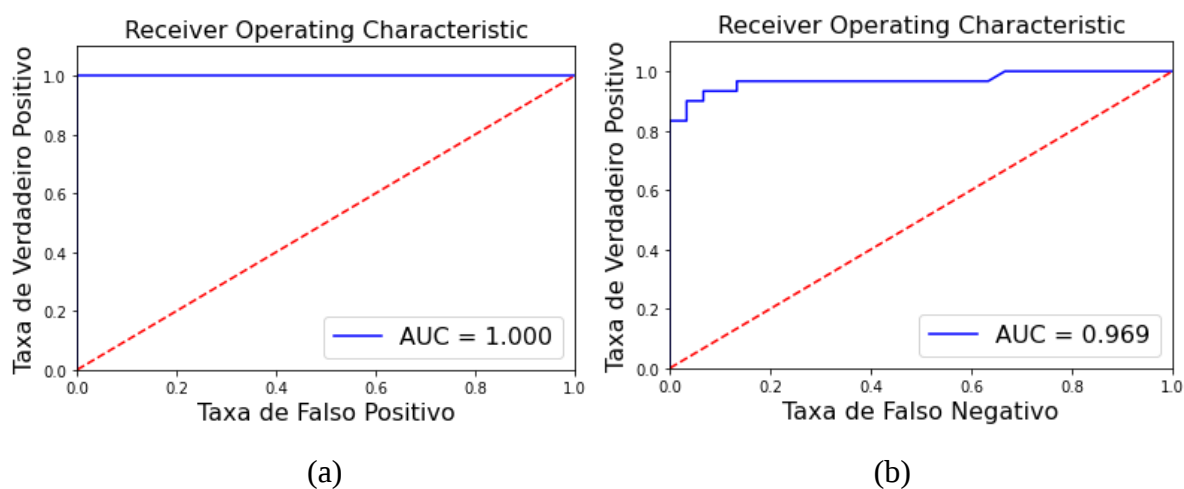
configurando excelentes resultados, como pode ser visto na curva de taxa de sucesso (Figura 33 (a) e (b)).

Tabela 2. Métricas de desempenho para os dados de treinamento/validação e teste

Métrica de Performance	Treinamento/Validação	Teste
VP	17	28
VN	22	28
FP	3	2
FN	0	2
Acurácia	0,928	0,93
Precisão	0,85	0,93

Fonte: produção do autor.

Figura 33. Curva de (a) taxa de sucesso e (b) predição



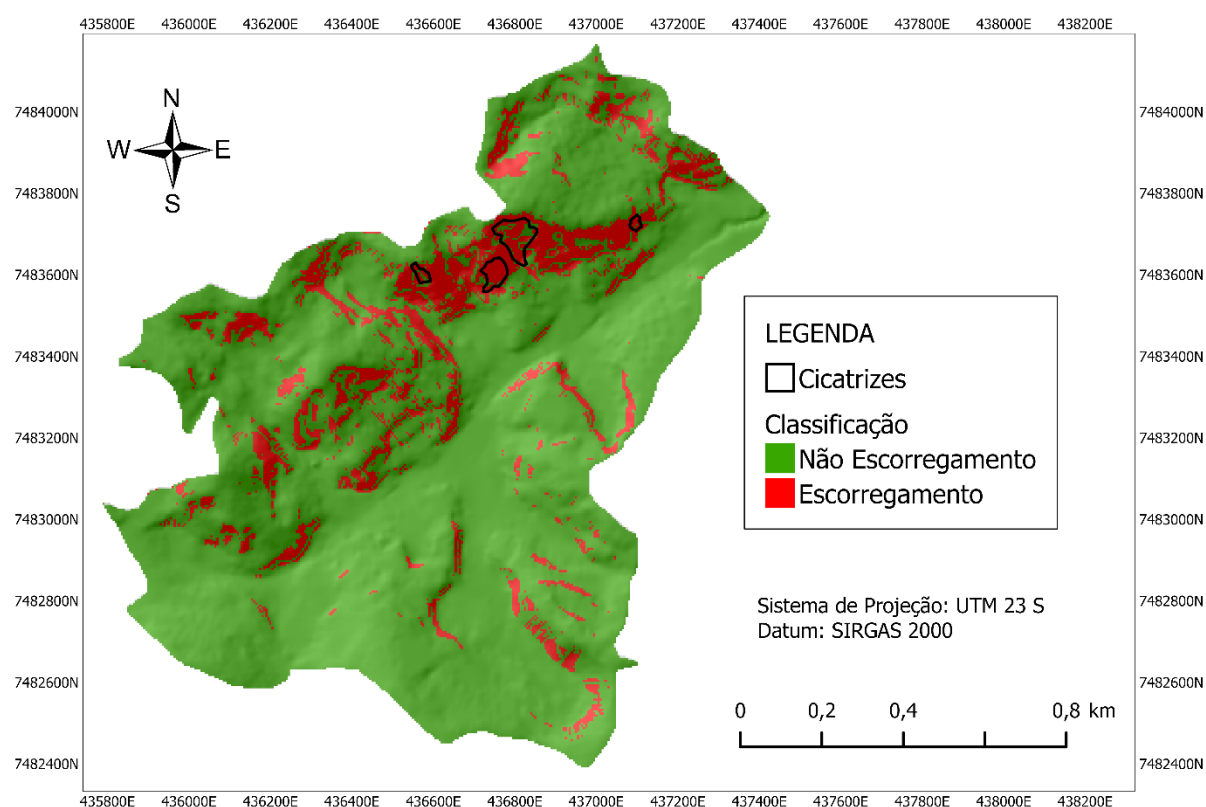
Fonte: produção do autor.

5.5 Geração dos mapas de suscetibilidade

O modelo de predição aplicado na área total de estudo possui como resultado tanto a classificação binária, atribuindo 0 para não escorregamento e 1 para escorregamento (Figura 34), quanto para a classificação de acordo com a probabilidade a escorregamentos (Figura 35), cujos valores de probabilidade foram classificados em cinco graus de suscetibilidade: muito

baixa ($< 0,14$), baixa ($0,14-0,32$), média ($0,32-0,45$), alta ($0,45-0,67$) e muito alta ($0,67-1,0$). Independentemente de ter sido aplicada a metodologia de quebras naturais para a divisão das classes de suscetibilidade, há outras possibilidades de classificação dos dados como a definição por intervalos manuais, intervalos iguais, quantil, intervalos geométricos e desvio padrão. Porém, optou-se por aplicar o método quebras naturais com base em trabalhos anteriores (PARK e KIM, 2019; DOU et al., (2019); POURGHASEMI e KERLE, 2015).

Figura 34. Bacia do Córrego Piracuama classificada em áreas de escorregamento e áreas de não escorregamento

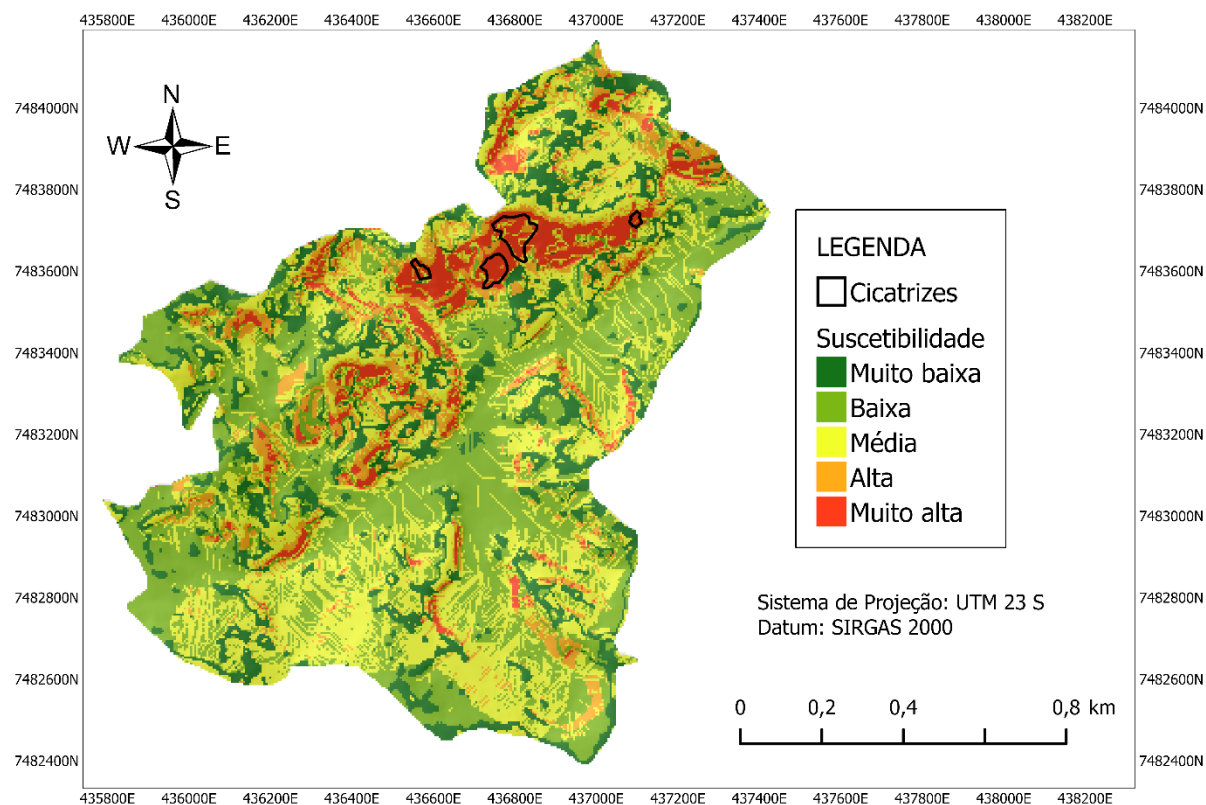


Fonte: produção do autor.

Ao analisar o mapa de suscetibilidade na Figura 35 verifica-se que um total de 16% e 34% da área situou-se sob a classe de zonas de suscetibilidade muito baixa e baixa, 33% da área total na média, enquanto 10% e 7% da área de estudo encontraram-se em zonas de suscetibilidade muito alta e alta (Figura 36). Essas áreas que apresentam maior criticidade concentram-se na porção Norte da Bacia do Córrego Piracuama, principalmente nas áreas

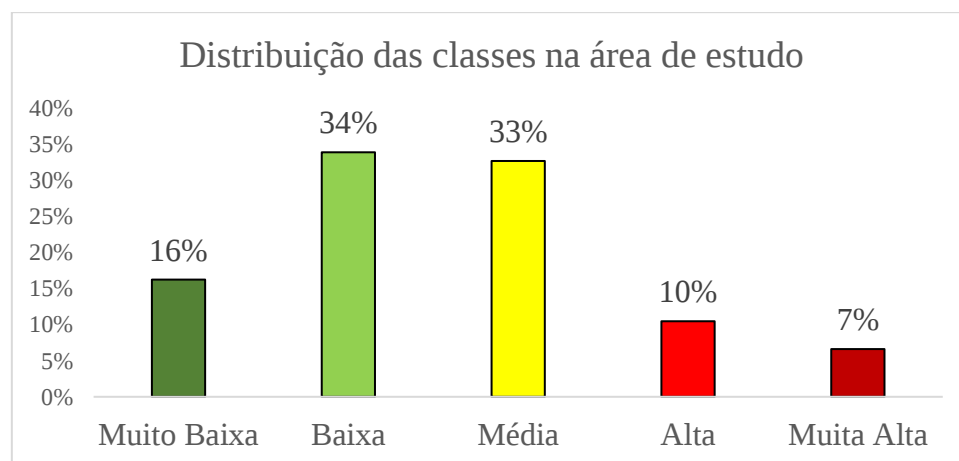
localizadas no setor nordeste. Pontualmente, em trechos localizados na porção sul da bacia, encontra-se também áreas de risco classificadas como muito altas.

Figura 35. Mapa de suscetibilidade a escorregamentos para a área de estudo



Fonte: produção do autor.

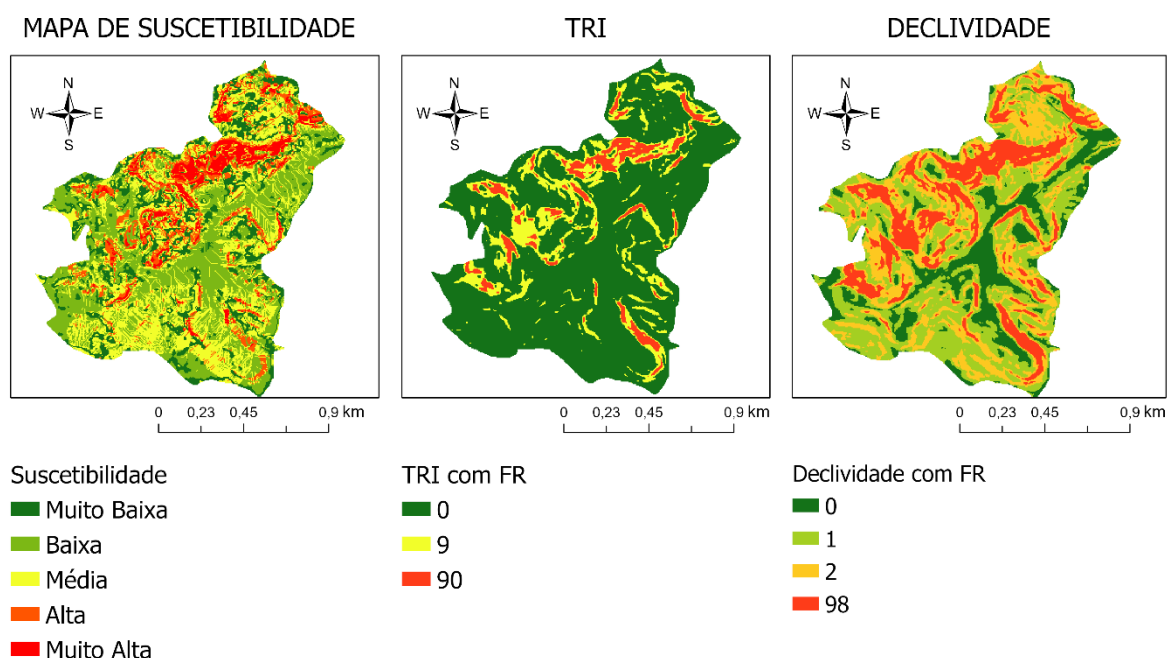
Figura 36. Distribuição das classes de suscetibilidade na Bacia do Córrego Piracuama



Fonte: produção do autor.

De fato, quando se compara os mapas de suscetibilidade com os mapas dos fatores condicionantes normalizados, é possível identificar que as áreas com suscetibilidade muito altas e altas possuem uma grande influência principalmente dos fatores condicionantes TRI e declividade, conforme já comprovado nas etapas anteriores e indicado na Figura 37. As áreas que apresentam altas declividades e altos valores para o índice TRI, em sua grande maioria possuem maiores suscetibilidades a escorregamentos.

Figura 37. Comparação entre o mapa de suscetibilidade e os fatores condicionantes que apresentaram maior influência no modelo de predição



Fonte: produção do autor.

O mapa de suscetibilidade também foi analisado considerando a relação entre os limites de cicatrizes e as classes de suscetibilidade (Tabela 3). Essa relação foi feita sobrepondo o mapa de inventário de cicatrizes e as classes de suscetibilidade. As cicatrizes estão associadas principalmente às classes muito alta (76%) e alta (14%), seguidas da classe média (11%). As classes baixa e muito baixa não apresentaram nenhuma relação com as cicatrizes. Esse resultado é positivo, visto que de fato as cicatrizes se situaram nas classes de maior probabilidade a escorregamentos. As áreas sem a presença de cicatrizes também foram analisadas conforme as classes de suscetibilidade. Dessa forma, observou-se as maiores concentrações nas classes

média (34%), baixa (33%), muito baixa (16%), alta (10%) e muito alta (6%), conforme a Tabela 4.

Tabela 3. Relação das classes de suscetibilidade com o inventário de cicatrizes

Classe	Pixels	Porcentagem
Muito Baixa	0	0%
Baixa	1	0%
Média	50	11%
Alta	65	14%
Muita Alta	358	76%

Fonte: produção do autor.

Tabela 4. Relação das classes de suscetibilidade com o inventário de cicatrizes

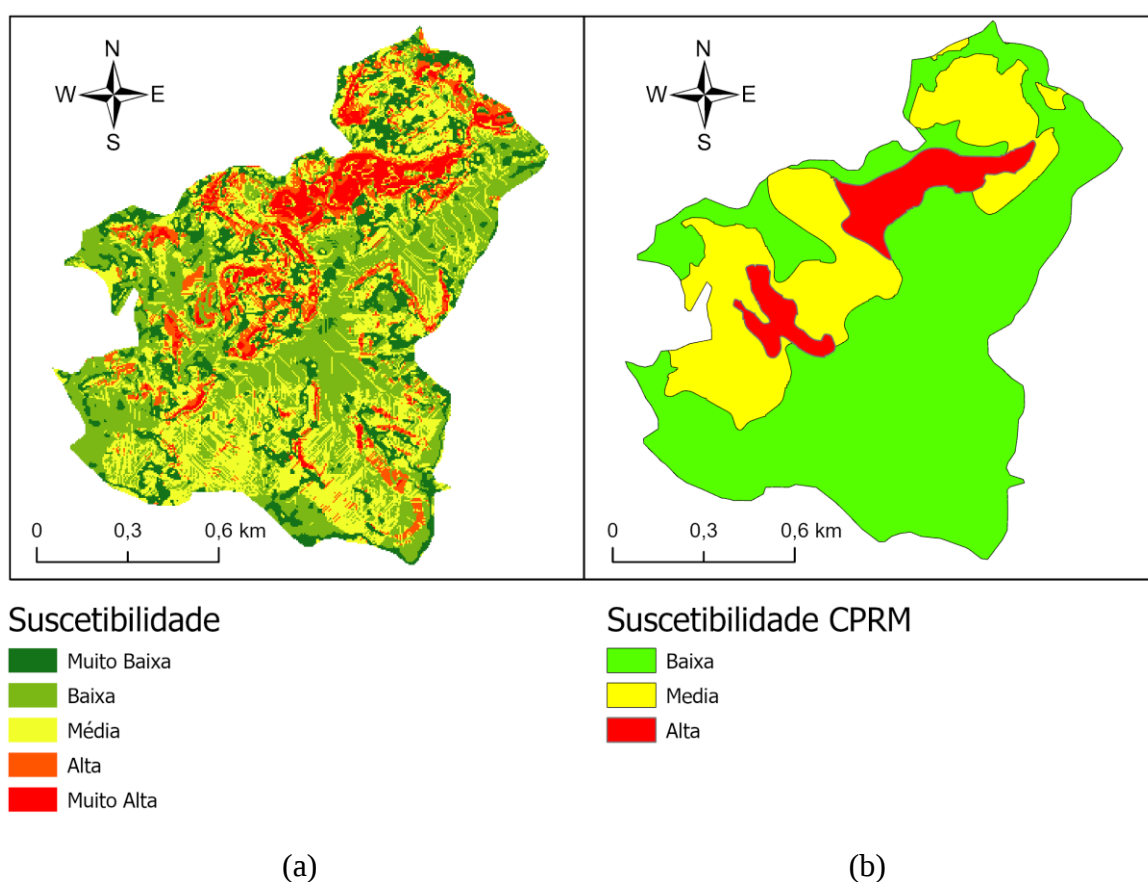
Classe	Pixels	Porcentagem
Muito Baixa	10182	16%
Baixa	21240	34%
Média	20436	33%
Alta	6498	10%
Muita Alta	3788	6%

Fonte: produção do autor.

Com o objetivo de validar o mapa de suscetibilidade de outras formas, além das métricas de avaliação, comparou-se, de forma qualitativa, o resultado obtido com o mapeamento de suscetibilidade a movimentos gravitacionais de massa gerado pela CPRM (Figura 38). Esse mapa foi produzido em 2014, em uma escala de 1:25.000 através de modelagem estatística feita em escritório e posteriormente validada em trabalho de campo. A CPRM trabalhou com o Índice de Suscetibilidade a Deslizamentos (ISD), obtido através da densidade de cicatrizes em cada classe dos fatores condicionantes e o conceito de Unidade Homogênea de Terreno (UHT), método que delimita unidades relativamente homogêneas quanto aos aspectos geológicos, geomorfológicos e hidrológicos. Diferente do mapeamento elaborado pelo método de aprendizado de máquina, as áreas são classificadas em alta, média e baixa suscetibilidade. Ao

confrontar ambos os mapeamentos, observou-se que o realizado no presente trabalho apresentou um maior nível de detalhamento, distinguindo as áreas de suscetibilidade de forma mais precisa. Além disso, o mapa apresenta áreas de muito altas, altas e médias suscetibilidades ao sul da bacia, algo que no mapeamento feito pela CPRM não é indicado.

Figura 38. Comparação entre o (a) modelo final e o (b) mapa de suscetibilidade elaborado por CPRM, 2014

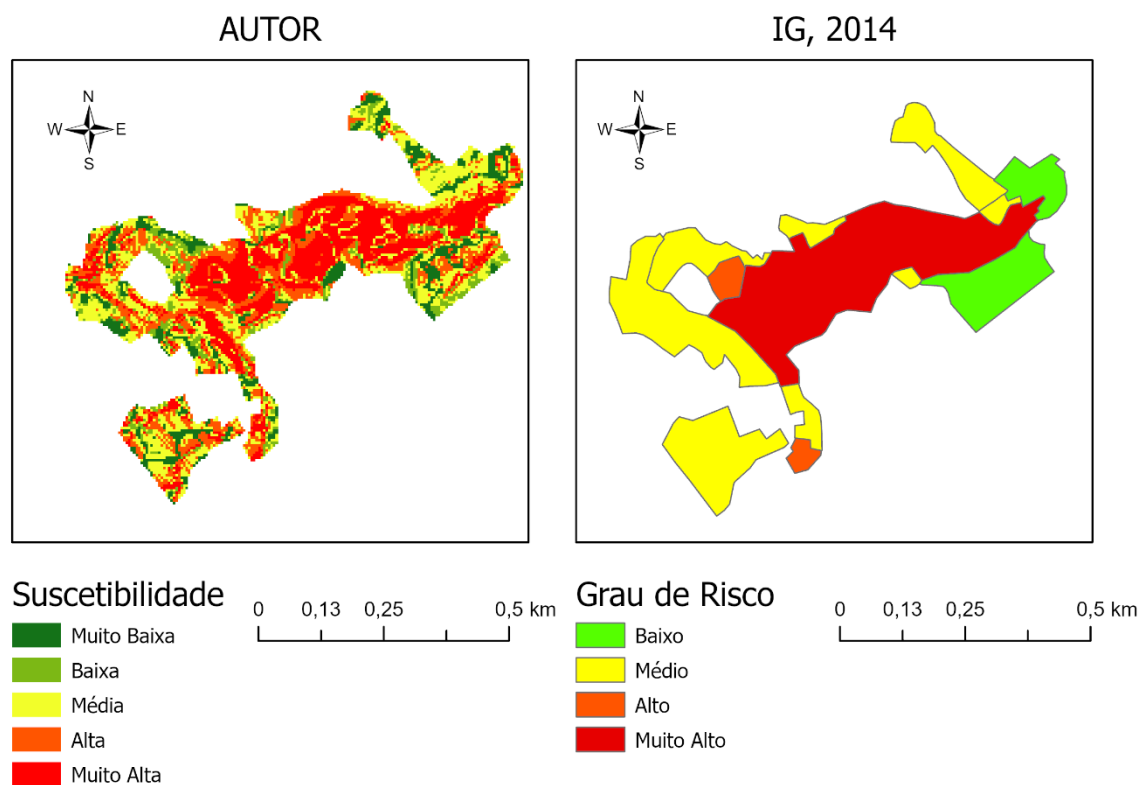


Fonte: produção do autor.

A fim de trazer essa comparação para uma escala ainda mais detalhada, comparou-se o mapa final ao mapeamento de risco elaborado pelo Instituto Geológico (Figura 39). Apesar de serem abordagens e metodologias totalmente distintas, optou-se por prosseguir com a análise, tendo em vista que o mapeamento da suscetibilidade pode nortear a elaboração de mapeamentos de risco e serem bons aliados na identificação de áreas prioritárias. Cabe reiterar a importância da utilização das cartas de suscetibilidade como base à elaboração das cartas geotécnicas de aptidão à urbanização e de risco para o caso específico de escorregamentos. Sendo assim,

identificou-se que os setores de risco muito alto apontados pelo IG são bem parecidos com as classes de suscetibilidade muito altas.

Figura 39. Comparação entre o (a) modelo final e o (b) mapa de risco elaborado por IG, 2014



(a)

(b)

Fonte: produção do autor.

6 DISCUSSÕES

Em geral, o modelo de suscetibilidade apresentou ótimos resultados de desempenho, comprovado pelas métricas de validação e quando comparado aos mapeamentos já existentes para a área de estudo. Entretanto, o número baixo de cicatrizes mapeadas pode ter direcionado os resultados, trazendo uma modelagem concentrada a apenas algumas características pontuais da bacia e determinando somente algumas localidades. Dessa forma, como a área apresenta esse agrupamento de amostras em uma certa porção da bacia, o modelo pode ficar direcionado a não errar. Esse número baixo e concentrado de cicatrizes traz duas discussões à tona. A primeira diz respeito às bases de dados públicos disponíveis no Brasil e a segunda consiste em quais escalas são ideais para que os modelos com base em aprendizado de máquina sejam aplicados.

Visto que a cultura do risco é recente no Brasil, as bases que disponibilizam dados de ocorrência de escorregamentos ainda são muito escassas. Ainda que haja diversas possibilidades de ter acesso às informações relacionadas a esses processos, como mídias impressas e digitais, imagens de sensoriamento remoto, S2ID, bancos de dados públicos, não há uma plataforma unificada que consolide todas essas informações e as atrele à componente espacial. Essa dificuldade também ocorre em relação aos fatores condicionantes, que muitas vezes não são encontrados ou são mapeados em escala pequena, que não favorecem uma análise em áreas menores. A geologia é um exemplo desses fatores para a área de estudo.

O tamanho da área em que será aplicado o modelo com base em aprendizado de máquina também é um ponto importante. A Bacia do Córrego Piracuama pode ser considerada como uma área pequena e, como possui uma ocupação urbana significativa, muitas vezes, as cicatrizes não conseguem ser vistas com base em imagens de satélite. Dessa forma, áreas extensas que tragam um maior número de cicatrizes e que sejam bem distribuídas podem trazer modelos que não concentrem as amostras. Em contrapartida, mesmo com essas concentrações de cicatrizes, o mapa resultante mostrou-se semelhante aos mapeamentos de suscetibilidade e risco existentes para a área, trazendo uma visão mais detalhada e precisa das classes para a área de estudo.

A metodologia proposta, através da Razão da Frequência, permite a identificação de limiares de escorregamentos para cada fator condicionante. Essa visão traz uma possibilidade de extrapolar esses limiares para outras áreas de estudo, que estão situadas no mesmo contexto geomorfológico e geológico.

É importante ressaltar que há outras possibilidades para se trabalhar com o aprendizado de máquina, que podem ser consideradas para trabalhos futuros na região, como por exemplo,

a aplicação de diferentes algoritmos e posterior comparação para identificar os melhores resultados; a análise da resolução do dado de entrada, buscando entender a influência da escala dos dados na análise da suscetibilidade a escorregamentos; o impacto da remoção de variáveis a partir da importância de atributos e correlação, analisando em qual das situações o modelo apresenta melhor desempenho e o estudo do impacto da amostragem, com a finalidade de entender se a distribuição das amostras interfere nos resultados.

Os mapeamentos de suscetibilidade são indispensáveis para a gestão pública e são apontadas inúmeras vezes no Marco de Sendai como ferramentas para a redução dos riscos de desastres. Realizar novas abordagens com base em aprendizado de máquina pode ser um bom caminho para lidar com a falta de mão de obra e otimizar os processos de mapeamento. O uso de SIG nesta abordagem pode fornecer melhorias no manuseio de dados espaciais, além de fornecer o processamento aprimorado dessas informações e auxiliar no processo de tomada de decisão. Outro ponto de destaque é que a modelagem baseada em aprendizado de máquina permite trabalhar com inúmeras variáveis, caracterizando de forma ampla a área de análise.

Mesmo que a temática de modelagens a partir de aprendizado de máquina integrado a ambiente SIG ainda seja recente e ainda há limitações relacionadas a questões computacionais e de capacitação, as bibliotecas disponíveis podem apoiar de forma significativa no desenvolvimento desses processos. Essas modelagens poderão auxiliar na gestão dos riscos de desastres, desempenhando um papel significativo no controle, gestão e elaboração de medidas para mitigar os riscos associados a potenciais escorregamentos de terra futuros.

7 CONCLUSÕES

Utilizar novas abordagens no mapeamento de suscetibilidade é uma questão que deve ser considerada pela administração pública, principalmente quando municípios suscetíveis a esses fenômenos lidam com diversos desafios na gestão de risco de desastres tais como a baixa capacidade operacional, falta de mão-de-obra ou quando o acesso às áreas vulneráveis é difícil. O método Floresta Aleatória apresentou um ótimo desempenho para a Bacia do Córrego Piracuama, mostrando que as técnicas de aprendizado de máquina podem ser ferramentas consideradas na gestão de riscos de desastres.

Para a área de estudo, todas as etapas realizadas – desde a preparação do dado até a avaliação do algoritmo Floresta Aleatória – mostraram:

- Ao aplicar a Razão de Frequência para cada um dos 14 fatores condicionantes, foi possível entender quais as classes dos atributos possuem maior relação com a deflagração dos escorregamentos. Em alguns atributos houve uma concentração em apenas uma classe, como o TRI e a declividade. Em outros, não houve um padrão bem definido e a ocorrência se relacionou de forma ampla a todos os intervalos, como a altitude e as curvaturas. Essa análise permitiu identificar quais são os limiares relacionados a cada fator condicionante, observando quais as classes possuem maior probabilidade à deflagração desses processos
- A análise da multicolinearidade entre as variáveis permitiu identificar altas correlações (maior que 0,7) entre dois pares de variáveis: TRI e declividade e TRI e SPI. Como essas variáveis foram construídas a partir da mesma base de dados, ou seja, do MDE, optou-se por mantê-las no modelo de predição.
- Em relação à importância dos atributos, identificou-se que a declividade e o TRI (variáveis que se destacaram pela alta correlação) apresentaram os melhores valores de ganho de informação. Em contrapartida, os fatores altimetria e curvatura em perfil configuraram as menos importantes no modelo.
- O método busca em grade permitiu, de forma automática, otimizar os parâmetros de entrada do algoritmo. Essa técnica torna-se uma grande aliada nesse campo de estudo, visto que reduz o tempo manual focado em “tentativas e erros” e retorna com o melhor valor de acurácia para diversas combinações. A busca em grade identificou que os melhores parâmetros de calibração foram 300 árvores e 4 atributos testados em cada nó.

- As métricas de avaliação, aplicadas para o conjunto de dados de treinamento e para o conjunto de dados de teste mostraram, respectivamente, ótimos resultados em relação aos índices Acurácia (0,928 e 0,93), Precisão (0,85 e 0,93) e AUC score (1,0 e 0,969).
- O mapa de suscetibilidade a escorregamento foi construído utilizando o método de quebras naturais para classificar os valores de probabilidade em cinco graus de suscetibilidade: muito baixa, baixa, média, alta e muito alta. A relação das cicatrizes com as classes de suscetibilidade identificou que as cicatrizes se situaram em 76% na classe muito alta, 14% na alta e 11% na média. As classes baixa e muito baixa não registraram cicatrizes.
- Ao comparar o mapa final com o mapa de suscetibilidade produzido pela CPRM e o mapa de risco produzido pelo IG identificou-se grandes semelhanças, principalmente no que diz respeito as classes muito alta e alta. O mapeamento feito através do aprendizado de máquina apresentou níveis maiores de detalhamento.

A disponibilidade de dados ainda é um problema enfrentado no Brasil. Conforme apresentado na fundamentação teórica, há uma infinidade de fatores que condicionam os escorregamentos que possibilitam uma modelagem robusta, caso essas informações estejam mapeadas. Ainda que o município de Campos do Jordão seja referência em instrumentações de risco e tenha uma boa infraestrutura quando comparada a municípios do mesmo porte, a falta de dado de qualidade e com a escala adequada ainda é um ponto de atenção. Trabalhar com um banco de dados estruturado e unificado é um grande desafio para a administração pública, porém é um caminho sem volta que precisa ser tratado como prioridade. Os mapeamentos de suscetibilidade realizados através de métodos de aprendizado de máquina podem ser grandes aliados na tomada de decisão e nas estratégias de planejamento urbano, identificando as áreas suscetíveis a escorregamentos e auxiliando na elaboração de medidas que reduzam o risco de desastres.

REFERÊNCIAS

- AGHDAM, I. N.; VARZANDEH, M. H. M.; PRADHAN, B. Landslide susceptibility mapping using an ensemble statistical index (Wi) and adaptive neuro-fuzzy inference system (ANFIS) model at Alborz Mountains (Iran). **Environmental Earth Sciences**, v. 75, n. 7, 2016.
- AHRENDT, A. **Movimentos de massa gravitacionais**: proposta de um sistema de previsão: aplicação na área urbana de Campos do Jordão - SP. 2005. 364 f. Tese (Doutorado em Geotecnia) – Universidade de São Paulo (USP), São Paulo, 2005.
- ANDRADE, E.; BROLLO, M. J.; SILVA, P. C. F.; PENTEADO, D. R.; SANTORO, J.; RIBEIRO, F. S.; GUEDES, A. C. M.; BRAGA, E. S. Campos do Jordão (SP): mapeamento de perigo e risco de escorregamentos e inundação no bairro de Vila Albertina, como subsídio à gestão de risco. In: SIMPÓSIO DE GEOLOGIA DO SUDESTE, 14.; SIMPÓSIO DO CRETÁCEO DO BRASIL, 8., 2015, Campos do Jordão. **Anais [...]**. Campos do Jordão: SBG, 2015.
- AVELAR, A. S.; COELHO NETTO, A. L. Fraturas e desenvolvimento de unidades geomorfológicas côncavas no médio vale do Rio Paraíba do Sul. **Revista Brasileira de Geociências**, v. 22, n. 2, p. 222–227, 1992. Disponível em: <http://bjg.siteoficial.ws/1992/n.2/11.pdf>. Acesso em: 25 ago. 2021.
- AYALEW, L.; YAMAGISHI, H. The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, Central Japan. **Engineering Geology**, v. 65, p.15-31, 2005. Disponível em: www.elsevier.com/locate/geomorph. Acesso em: 24 ago. 2021.
- AZEVEDO, R. A. Conceitos fundamentais sobre migmatitos. **Terra e Didática**, v. 15, p. 7, 2019. Disponível em: <https://periodicos.sbu.unicamp.br/ojs/index.php/td/article/view/8653435>. Acesso em: 10 nov. 2020.
- BRASIL. **Lei 12.608, de 10 de abril de 2012**. Dispõe sobre o Sistema Nacional de Proteção e Defesa Civil - SINPDEC e o Conselho Nacional de Proteção e Defesa Civil – CONPDEC. Institui a Política Nacional de Proteção e Defesa Civil – PNPDEC. Brasil, 2012. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2012/lei/l12608.htm. Acesso em: 10 jul. 2020.
- BRASIL – Ministério da Integração Nacional. Secretaria Nacional de Defesa Civil. Centro Nacional de Gerenciamento e Riscos e Desastres. **Anuário brasileiro de desastres naturais**: 2012. Brasília: CENAD, 2012. 84 p. Disponível em: https://antigo.mdr.gov.br/images/stories/ArquivosDefesaCivil/ArquivosPDF/publicacoes/AnuariodeDesastresNaturais_2013.pdf. Acesso em: 15 jun. 2020.
- BRASIL. Ministério das Cidades. Instituto de Pesquisas Tecnológicas (IPT). **Mapeamento de riscos em encostas e margem de rios**. Brasília: Ministério das Cidades; Instituto de Pesquisas Tecnológicas (IPT), 2007. 176 p.

BREIMAN, L. Random forests. **Machine Learning**, v. 45, n. 1, p. 5–32, 2001. Disponível em: <https://link.springer.com/article/10.1023/A:1010933404324>. Acesso em: 4 jul. 2021.

BROLLO, M.J. & FERREIRA, C.J. **Gestão de riscos de desastres devido a fenômenos geodinâmicos no Estado de São Paulo: cenário 2000-2015**. Boletim do Instituto Geológico nº 67. São Paulo: IG/SMA, 2016. 72p. Disponível em: http://igeologico.sp.gov.br/files/2016/10/boletim_IG_vol_67.pdf. Acesso em: 13 out. 2020.

BUI, T. D.; TUAN, T. A.; KLEMPE, H.; PRADHAN, B.; REVHAUG, I. Spatial prediction models for shallow landslide hazards: a comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. **Landslides**, v. 13, n. 2, p. 361–378, 2016. Disponível em: <http://dx.doi.org/10.1007/s10346-015-0557-6>. Acesso em: 17 set. 2020.

CAPOANE, V. Determinação do índice de potência de escoamento para o município de Palmitinho/RS. **Estudos geográficos**, v. 13, n. 2, p. 106–117, 2015. Disponível em: <http://www.periodicos.rc.biblioteca.unesp.br/index.php/estgeo/article/view/10926/7359>. Acesso em: 10 dez. 2020.

CARVALHO, C. S.; GALVÃO, T. Prevenção de riscos de deslizamentos em encostas em áreas urbanas. In: MORAIS, M. P., KRAUSE, C. NETO, V. C. L. **Caracterização e tipologia de assentamentos precários: estudos de caso brasileiros**. Brasília: IPEA, 2016. p. 169-185.

CARVALHO, L. M. V.; JONES, C.; LIEBMANN, B. Extreme precipitation events in southeastern South America and large-scale convective patterns in the South Atlantic convergence zone. **Journal of Climate**, v. 15, n. 17, p. 2377–2394, 2002.

CATANI, F.; LAGOMARSINO, D.; SEGONI, S.; TOFANI, V. Landslide susceptibility estimation by random forests technique: sensitivity and scaling issues. **Natural Hazards and Earth System Sciences**, v. 13, n. 11, p. 2815–2831, 2013.

CENTRO UNIVERSITÁRIO DE ESTUDOS E PESQUISAS SOBRE DESASTRES NATURAIS (CEPED) ; UNIVERSIDADE FEDERAL DE SANTA CATARINA (UFSC). **Atlas brasileiro de desastres naturais 1991 a 2012**. 2.ed. ver. ampl. Florianópolis: CEPED; UFSC, 2013. 126 p. Disponível em: <https://s2id.mi.gov.br/paginas/atlas/>. Acesso em: 17 abr. 2020.

CENTRE FOR RESEARCH ON THE EPIDEMIOLOGY OF DISASTERS; UNITED NATIONS OFFICE FOR DISASTER RISK REDUCTION. **Economic losses, poverty and disasters**. Brussels: EM-DAT, 2018.

CHEN, W.; SHIRZADI, A.; SHAHABI, H.; AHMAD, B. Bin; ZHANG, S.; HONG, H.; ZHANG, N. A novel hybrid artificial intelligence approach based on the rotation forest ensemble and naïve bayes tree classifiers for a landslide susceptibility assessment in Langao County, China. **Geomatics, Natural Hazards and Risk**, v. 8, n. 2, p. 1955–1977, 2017. a. Disponível em: <https://doi.org/10.1080/19475705.2017.1401560>. Acesso em: 18 nov. 2020.

CHEN, W.; XIE, X.; WANG, J.; PRADHAN, B.; HONG, H.; BUI, D. T.; DUAN, Z.; MA, J.

A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility. **Catena**, v. 151, p. 147–160, 2017. b.

COROMINAS, J.; VAN WESTEN, C.; FRATTINI, P.; CASCINI, L.; MALET, J. P.; FOTOPOULOU, S.; CATANI, F.; VAN DEN EECKHAUT, M.; MAVROULI, O.; AGLIARDI, F.; PITILAKIS, K.; WINTER, M. G.; PASTOR, M.; FERLISI, S.; TOFANI, V.; HERVÁS, J.; SMITH, J. T. Recommendations for the quantitative analysis of landslide risk. **Bulletin of Engineering Geology and the Environment**, v. 73, n. 2, p. 209–263, 2014. Disponível em: <https://link.springer.com/article/10.1007/s10064-013-0538-8>. Acesso em: 1 jul. 2021.

COSTA, I.; TAVARES, F.; OLIVEIRA, J. Predictive lithological mapping through machine learning methods: a case study in the Cinzento Lineament, Carajás Province, Brazil. **Journal of the Geological Survey of Brazil**, v. 2, n. 1, p. 26–36, 2019.

DANG, V. H.; HOANG, N. D.; NGUYEN, L. M. D.; BUI, D. T.; SAMUI, P. A novel GIS-based random forest machine algorithm for the spatial prediction of shallow landslide susceptibility. **Forests**, v. 11, n. 1, 2020.

DEMIR, G. Landslide susceptibility assessment of the part of the north Anatolian fault zone (Turkey) by GIS-based frequency ratio and index of entropy models. **Natural Hazards and Earth System Sciences Discussions**, p. 1–17, 2016.

DIAS, F. P.; HERMANN, M. L. P. **Suscetibilidade a deslizamentos**: estudo de caso no bairro de Saco Grande, Florianópolis-SC. Florianópolis: Caminhos de Geografia, 2002.

DOU, J.; YUNUS, A. P.; BUI, D. T.; MERGHADI, A.; SAHANA, M.; ZHU, Z.; CHEN, C. W.; HAN, Z.; PHAM, B. T. Improved landslide assessment using support vector machine with bagging, boosting, and stacking ensemble machine learning framework in a mountainous watershed, Japan. **Landslides**, v. 17, n. 3, p. 641–658, 2020.

DOU, J.; YUNUS, A. P.; TIEN BUI, D.; MERGHADI, A.; SAHANA, M.; ZHU, Z.; CHEN, C. W.; KHOSRAVI, K.; YANG, Y.; PHAM, B. T. Assessment of advanced random forest and decision tree algorithms for modeling rainfall-induced landslide susceptibility in the Izu-Oshima Volcanic Island, Japan. **Science of The Total Environment**, v. 662, p. 332–346, 2019.

DUBEY, A. K.; NARANG, S.; KUMAR, A.; SASUBILLI, S. M.; GARCÍA-DÍAZ, V. Performance estimation of machine learning algorithms in the factor analysis of COVID-19 dataset. **Computers, Materials and Continua**, v. 66, n. 2, p. 1921–1936, 2020.

GALAR, D.; KUMAR, U. **eMaintenance**: essential eletronic tools for efficiency. Londres: Elsevier, 2017.

GAMEIRO, S. **Capacidade de generalização e extrapolação de redes neurais artificiais no mapeamento a suscetibilidade de deslizamentos**. Tese (Mestrado) – Universidade Federal do Rio Grande do Sul, Programa de Pós-Graduação em Sensoriamento Remoto, Porto Alegre, 2020.

GERSCOVICH, D. M. **Estabilidade de taludes**. 2. ed. Porto Alegre: Oficina de Textos, 2016.

GUERRA, A. J. T. A erosão dos solos no contexto social. **Anuário do Instituto de Geociências**, v.17, 1994.

GUIDICINI, G.; NIEBLE, C. M. **Estabilidade de taludes naturais e de escavação**. São Paulo: USP, 1984.

HARMENDANI, P.; ZANETTI, M. Aprendizado de máquina por reforço aplicado no jogo de cartas Uno. **Revista de Sistemas e Computação (RSC)**, v. 9, n. 2, p. 245–251, 2019. Disponível em: <https://revistas.unifacs.br/index.php/rsc/article/view/6011>. Acesso em: 4 jul. 2021.

HASUI, Y. A grande colisão pré-cambriana do sudeste brasileiro e a estruturação regional. **Geociências**, v. 29, n. 2, p. 141–169, 2010.

HEILBRON, M.; PEDROSA-SOARES, A. C.; CAMPOS-NETO, M.; SILVA, L. C.; TROUW, R.; JANASI, V. Província Mantiqueira. In: MANTESSO-NETO, V.; BARTORELLI, A.; CARNEIRO, C.; BRITO NEVES, B. **Geologia do continente Sul-Americano**: evolução da obra de Fernando Flávio Marques de Almeida. São Paulo: Beca, 2004. p. 203–235.

HIGHLAND, L. M.; BOBROWSKY, P. **The landslide handbook**: a guide to understanding landslides. Reston: U.S. Geological Survey Circular, 2008. 129 p.

HIRUMA, S. T.; RICCOMINI, C.; MODENESI-GAUTTIERI, M. C. Neotectônica do planalto de Campos do Jordão (SP). **Revista Brasileira de Geociências**, v. 31, n. 3, p. 375–384, 2001.

HONG, H.; PRADHAN, B.; XU, C.; TIEN BUI, D. Spatial prediction of landslide hazard at the Yihuang area (China) using two-class kernel logistic regression, alternating decision tree and support vector machines. **Catena**, v. 133, p. 266–281, 2015.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). **Classificação e caracterização dos espaços rurais e urbanos do Brasil**: uma primeira aproximação. Rio de Janeiro: IBGE, 2017. Disponível em: <http://biblioteca.ibge.gov.br/visualizacao/livros/liv100643.pdf> http://www.ibge.gov.br/home/geociencias/geografia/espacos_rurais_e_urbanos/default.shtm. Acesso em: 01 ago. 2020.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). Campos do Jordão. **IBGE**, 2020. Disponível em: <https://cidades.ibge.gov.br/brasil/sp/campos-do-jordao>. Acesso em: 9 out. 2020.

INSTITUTO DE PESQUISAS TECNOLÓGICAS (IPT). Acidentes associados a movimentos gravitacionais de massa ocorridos no município de Campos do Jordão, SP, em janeiro de 2000: ações técnicas após o desastre. In: CONGRESSO BRASILEIRO DE GEOLOGIA DE

ENGENHARIA E AMBIENTAL, 10., 2002, Ouro Preto. **Anais [...]**. Ouro Preto: ABGE, 2002.

INSTITUTO DE PESQUISAS TECNOLÓGICAS (IPT); SERVIÇO GEOLÓGICO DO BRASIL (CPRM). **Cartas de suscetibilidade a movimentos de gravitacionais de massa e inundações**: 1:25.000: nota técnica explicativa. São Paulo: Publicação IPT n. 3016, 2014. 42p.

INSTITUTO GEOLÓGICO (IG). SECRETARIA DE MEIO AMBIENTE DO ESTADO DE SÃO PAULO (SMA). **Relatório de atendimento técnico**: PPDC operação 2012/2013. n. 01/2012-2013. São Paulo: IG-SMA, 2013. 4 p.

INSTITUTO GEOLÓGICO (IG)/SECRETARIA DE MEIO AMBIENTE DO ESTADO DE SÃO PAULO (SMA). **Mapeamento de riscos associados a escorregamentos, inundações, erosões e solapamentos de margens de drenagens do município de Campos do Jordão (SP)**. São Paulo: Instituto Geológico, 2014.

INSTITUTO DE PESQUISAS TECNOLÓGICAS DO ESTADO DE SÃO PAULO (IPT). **Ocupação de encostas**: manual. São Paulo: IPT, 1991. 216 p.

INSTITUTO DE PESQUISAS TECNOLÓGICAS (IPT). **Banco de dados de mortes por escorregamentos no Brasil**. Banco de Dados Digital. São Paulo: IPT, 2009.

JORDAN, M. I.; MITCHELL, T. M. Machine learning: trends, perspectives, and prospects. **American Association for the Advancement of Science**, v. 349, n. 6245, p. 255-260, 2015. Disponível em: <https://science.sciencemag.org/content/349/6245/255>. Acesso em: 19 jun. 2021.

KARANTANELIS, E.; MARINOS, V.; VASSILAKIS, E.; HÖLBLING, D. Evaluation of machine learning algorithms for object-based mapping of landslide zones using UAV data. **Geosciences**, v. 11, n. 8, p. 305, 2021. Disponível em: <https://www.mdpi.com/2076-3263/11/8/305/htm>. Acesso em: 12 fev. 2021.

KÖNIG, T.; MORAES, M. A. E.; MENDES, R. M.; KUX, H. J. H.; MÁRCIO, R. M. Análise da variação do fator de segurança durante evento pluviométrico em áreas de encostas. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 19., 2019, Santos. **Anais [...]**. São José dos Campos: INPE. 2019.

MARTÍNEZ-ÁLVAREZ, F.; REYES, J.; MORALES-ESTEBAN, A.; RUBIO-ESCUADERO, C. Determining the best set of seismicity indicators to predict earthquakes. Two case studies: Chile and the Iberian Peninsula. **Knowledge-Based Systems**, v. 50, p. 198–210, 2013. Disponível em: <http://dx.doi.org/10.1016/j.knosys.2013.06.011>. Acesso em 21 out. 2020.

MENDES, R. **Estudo das propriedades geotécnicas de solos residuais não saturados de Ubatuba (SP)**. 2008. Tese (Doutorado em Engenharia Geotécnica) - Escola Politécnica da Universidade de São Paulo, 2008.

MENDES, R. M.; ANDRADE, M. R. M. D.; TOMASELLA, J.; MORAES, M. A. E. D.; SCOFIELD, G. B. Understanding shallow landslides in Campos do Jordão municipality -

Brazil: disentangling the anthropic effects from natural causes in the disaster of 2000. **Natural Hazards and Earth System Sciences**, v. 18, n. 1, p. 15–30, 2018.

MERGHADI, A.; YUNUS, A. P.; DOU, J.; WHITELEY, J.; THAIPHAM, B.; BUI, D. T.; AVTAR, R.; ABDERRAHMANE, B. Machine learning methods for landslide susceptibility studies: a comparative overview of algorithm performance. **Earth-Science Reviews**, v. 207, p. 103225, 2020. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0012825220302713>. Acesso em: 24 ago. 2021.

MODENESI-GAUTTIERI, M. C.; HIRUMA, S. T. A expansão urbana no planalto de Campos do Jordão: diagnóstico geomorfológico para fins de planejamento. **Revista do Instituto Geológico**, n. 25, v.1/2, p. 1-28, 2004.

MOHRI, M.; ROSTAMIZADEH, A.; TALWALKAR, A. **Foundations of machine learning**. 2. ed. Cambridge: MIT, 2018.

MONARD, M. C.; BARANAUSKAS, J. A. Conceitos sobre aprendizado de máquina. In: REZENDE, S. O. **Sistemas inteligentes: fundamentos e aplicações**. São Carlos: Manole, 2003. p. 89–114.

MOORE, I. D.; WILSON, J. P. Length-slope factors for the revised universal soil loss equation: simplified method of estimation. **Journal of Soil and Water Conservation**, v.47, n.5, p. 423-428, 1992. Disponível em: www.swcs.org. Acesso em: 1 dez. 2020.

NAKILEZA, B. R.; NEDALA, S. Topographic influence on landslides characteristics and implication for risk management in upper Manafwa catchment, Mt Elgon Uganda. **Geoenviron Disasters**, v. 7, n. 27, 2020. Disponível em: <https://doi.org/10.1186/s40677-020-00160-0>. Acesso em: 10 fev. 2021.

NUNES, R. P. M.; TROUW, R. A. J.; COSTA, R. V. C.; CABRAL, R. R. **Geologia e recursos minerais da folha Campos do Jordão SF.23-Y-B-V: escala 1:100.000**, estados de Minas Gerais e São Paulo. Rio de Janeiro: CPRM, 2020.

PARK, S.; KIM, J. Landslide susceptibility mapping based on random forest and boosted regression tree models, and a comparison of their performance. **Applied Sciences**, v. 9, n. 5, 2019. Disponível em: <https://doi.org/10.3390/app9050942>. Acesso em: 21 mar. 2020.

POURGHASEMI, H.R., KERLE, N. Random forests and evidential belief function-based landslide susceptibility assessment in Western Mazandaran Province, Iran. **Environ Earth Sciences**, v. 75, n. 185, 2016. Disponível em: <https://doi.org/10.1007/s12665-015-4950-1>. Acesso em: 29 abr. 2021.

PRADO, E. M. G.; DE SOUZA FILHO, C. R.; CARRANZA, E. J. M.; MOTTA, J. G. Modeling of Cu-Au prospectivity in the Carajás mineral province (Brazil) through machine learning: dealing with imbalanced training data. **Ore Geology Reviews**, v. 124, n. 103611, 2020. Disponível em: <https://doi.org/10.1016/j.oregeorev.2020.103611>. Acesso em: 20 abr. 2021.

PRIETO, C. C. **Previsão de deslizamentos em encostas por meio de modelagem numérica: estudo de caso na Bacia Piracuama, município de Campos do Jordão, SP**. Tese

(Doutorado em Ciência do Sistema Terrestre) - Instituto de Pesquisas Espaciais. São José dos Campos, p. 187, 2018.

RAJA, N. B.; ÇIÇEK, I.; TÜRKOĞLU, N.; AYDIN, O.; KAWASAKI, A. Correction to: landslide susceptibility mapping of the Sera River Basin using logistic regression model. **Natural Hazards** 2017, v. 91, n. 3, p. 1423–1423, 2017. Disponível em: <https://link.springer.com/article/10.1007/s11069-017-3145-3>. Acesso em: 25 ago. 2020.

REICHENBACH, P.; ROSSI, M.; MALAMUD, B. D.; MIHIR, M.; GUZZETTI, F. A review of statistically-based landslide susceptibility models. **Earth-Science Reviews**, v. 180, n. March, p. 60–91, 2018. Disponível em: <https://doi.org/10.1016/j.earscirev.2018.03.001>. Acesso em: 18 abr. 2020.

RILEY, S. J.; DEGLORIA, S. D.; ELLIOT, S. D. A terrain ruggedness index that quantifies topographic heterogeneity. **Intermountain Journal of Sciences**, v. 5, p. 23–27, 1999. Disponível em: https://www.researchgate.net/publication/259011943_A_Terrain_Ruggedness_Index_that_Quantifies_Topographic_Heterogeneity. Acesso em: 7 ago. 2021.

ROKACH, L.; MAIMON, O. Top-down induction of decision trees classifiers: a survey. **IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews**, v. 35, n. 4, p. 476–487, 2005.

ROSA FILHO, A. Percepção geográfica de escorregamentos de encostas em favelas nas áreas de risco: Campos do Jordão-SP. **Aleph**, v. 15, p. 190 f., 2006.

SAHIN, E. K.; COLKESEN, I.; ACMALI, S. S.; AKGUN, A.; AYDINOGLU, A. C. Developing comprehensive geocomputation tools for landslide susceptibility mapping: LSM tool pack. **Computers and Geosciences**, v. 144, n. August, p. 104592, 2020. Disponível em: <https://doi.org/10.1016/j.cageo.2020.104592>. Acesso em: 23 set. 2020.

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. Disponível em: <http://ieeexplore.ieee.org/document/5392560/>. Acesso em: 4 jul. 2021.

SÁNCHEZ, I. F. D. P. Mapeamento semi-automático de feições deposicionais quaternárias associadas a fundos de vale e reentrâncias de cabeceiras de drenagem em anfiteatro. In: SIMPÓSIO NACIONAL DE GEOMORFOLOGIA, 9., 2012, Rio de Janeiro. **Anais [...]**. Rio de Janeiro: UFRJ, 2012.

SANTOS, A. R. **A grande barreira da Serra do Mar**: da trilha dos Tupiniquins à rodovia dos Imigrantes. São Paulo: O Nome da Rosa, 2004.

SANTOS, A. R. **Cidades e geologia**: discussão técnica e proposição de projetos de lei de grande interesse para as populações urbanas. São Paulo: Rudder, 2017.

SATO, L.; SHIMABUKURO, Y.; KUPLICH, T.; GOMES, V. Análise comparativa de algoritmos de árvore de decisão do sistema WEKA para classificação do uso e cobertura da terra. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 15., 2013, Foz do Iguaçu. **Anais [...]**. Foz do Iguaçu: INPE, 2013.

SAUSEN, T. M.; LACRUZ, M. S. P. **Sensoriamento remoto para desastres**. São Paulo: Oficina de Textos, 2015.

SEGONI, S.; PAPPAFICO, G.; LUTI, T.; CATANI, F. Landslide susceptibility assessment in complex geological settings: sensitivity to geological information and insights on its parameterization. **Landslides**, v. 17, P. 2443-2453, 2020.

SHIRVANI, Z. A holistic analysis for landslide susceptibility mapping applying geographic object-based random forest: a comparison between protected and non-protected forests. **Remote Sensing**, v. 12, n. 3, p. 1–22, 2020.

SOBRAL, A. P. B. **Previsão de carga horária uma nova abordagem por árvore de decisão**. Tese (Doutorado em Engenharia Elétrica) - Pontifícia Universidade Católica do Rio de Janeiro. Rio de Janeiro, 2003.

SOUSA, M. C. C. S. **Uma análise do algoritmo k-means como introdução ao aprendizado de máquinas**. Monografia Graduação - Universidade Federal do Tocantins - Curso de Matemática. Araguaína, p.74, 2019.

TAALAB, K.; CHENG, T.; ZHANG, Y. Mapping landslide susceptibility and types using random forest. **Big Earth Data**, v. 2, n. 2, p. 159–178, 2018. Disponível em: <https://www.tandfonline.com/doi/full/10.1080/20964471.2018.1472392>. Acesso em: 20 abr. 2021.

TEMBHEKAR, J.; KATKAR, J.; DAHEKAR, D.; DAF, P. Building and training a supervised machine learning model using scikit-learn for elevating business sales. **International Research Journal of Engineering and Technology**, 2021. Disponível em: www.irjet.net. Acesso em: 2 jul. 2021.

TOMINAGA, L. K. **Avaliação de metodologias de análise de risco a escorregamentos: aplicação de um ensaio em Ubatuba, SP.** (2007). p. 207. Tese (Doutorado em Geografia) - Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2007.

TOMINAGA, L. K.; SANTORO, J.; AMARAL, R. **Desastres naturais: conhecer para prevenir**. São Paulo: Instituto Geológico, 2009. 196p.

TROUW R.A.J., PETERNEL R., RIBEIRO A., HEILBRON M., VINAGRE R., DUFFLES P., TROUW C.C., FONTAINHA M., KUSSAMA H.H. 2013. A new interpretation for the interference zone between the Southern Brasília Belt and the Central Ribeira Belt, SE Brazil. **Journal of South American Earth Sciences**, v. 48, p. 43-57, 2013.

UNDESA. **World urbanization prospects: the 2018 revision demographic research**. Nova York: United Nations, 2019. Disponível em: <https://population.un.org/wup/Publications/Files/WUP2018-Report.pdf>. Acesso em: 13 ago. 2020.

UNISDR, 2009. **UNISDR Terminology on disaster risk reduction**. Geneva: United Nations International Strategy for Disaster Reduction, 2009. Disponível em: https://www.preventionweb.net/files/7817_UNISDRTerminologyEnglish.pdf. Acesso em: 10 jul. 2020.

UNISDR – AM. The ‘silent disaster of local losses’. **United nations office for disaster risk reduction**, 2013. Disponível em: <http://www.unisdr.org/archive/35664>. Acesso em: 21 ago. 2020.

VIEIRA, B. C. **Previsão de escorregamentos translacionais rasos na Serra do Mar (SP) a partir de modelos matemáticos em bases físicas**. Tese (Doutorado em Geografia) - Departamento de Geografia, Instituto de Geociências, Universidade Federal do Rio de Janeiro. Rio de Janeiro, p.193, 2007.

VIEIRA, M. I.; KAZMIERCZAK, M. L.; MALTA, F. J. N. C. Proposta metodológica para identificação de áreas de risco de movimentos de massa em áreas de ocupação urbana. Estudo de caso: Campos do Jordão, SP. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 12., 2015, Goiânia. **Anais [...]**. Goiânia: INPE, 2005.

VIEIRA, W.; OLIVEIRA, C. A.; ANDRADE, E.; BROLLO, M. Campos do Jordão (SP): utilização de instrumentos técnicos na gestão municipal de riscos de escorregamentos e inundações. In: SIMPÓSIO DE GEOLOGIA DO SUDESTE, 14., 2015, Campos do Jordão, SP. **Anais [...]** Campos do Jordão: SBG, 2015.

VINAGRE R, TROUW RAJ, MENDES JC, LUDKA I. 2014B. Química mineral e geotermobarometria do Batólito Serra da Água Limpa, Nappe Socorro-Guaxupé, faixa Brasília Meridional, sudeste do Brasil. **Brazilian Journal of Geology**, v.44, n.3, p. 387-414, 2014.

VON ZUBEN, F. J. **Máquinas de vetores suporte**. Campinas: Unicamp, 2013.

WEISS, A. Topographic position and landforms analysis. In: PROCEEDINGS OF THE ESRI USER CONFERENCE 2001, San Diego, CA, 2001.

XIONG, W.; FAES, L.; IVANOV, P. C. Entropy measures, entropy estimators, and their performance in quantifying complex dynamics: effects of artifacts, nonstationarity, and long-range correlations. **Physical Review E**, v. 95, n. 6, 2017.

YAO, L.; FANG, Z.; XIAO, Y.; HOU, J.; FU, Z. An intelligent fault diagnosis method for lithium battery systems based on grid search support vector machine. **Energy**, v. 214, p. 118866, 2021. Disponível em: <https://sci-hub.se/https://doi.org/10.1016/j.energy.2020.118866>. Acesso em: 6 ago. 2021.

YOUSSEF, A. M.; POURGHASEMI, H. R.; POURTAGHI, Z. S.; AL-KATHEERI, M. M. Landslide susceptibility mapping using random forest, boosted regression tree, classification and regression tree, and general linear models and comparison of their performance at Wadi Tayyah Basin, Asir Region, Saudi Arabia. **Landslides**, v. 13, n. 5, p. 839–856, 2016. b. Disponível em: <http://dx.doi.org/10.1007/s10346-015-0614-1>. Acesso em: 18 mar. 2020.

ZHANG, K.; WU, X.; NIU, R.; YANG, K.; ZHAO, L. The assessment of landslide susceptibility mapping using random forest and decision tree methods in the Three Gorges Reservoir area, China. **Environmental Earth Sciences**, v. 76, n. 11, p. 405, 2017. Disponível em: <https://link.springer.com/article/10.1007/s12665-017-6731-5>. Acesso em: 4 jul. 2021.

ZHOU, X.; WEN, H.; ZHANG, Y.; XU, J.; ZHANG, W. Landslide susceptibility mapping using hybrid random forest with GeoDetector and RFE for factor optimization. **Geoscience Frontiers**, v. 12, n. 5, p. 101211, 2021.