



**UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE TECNOLOGIA E GEOCIÊNCIAS
DEPARTAMENTO DE ENGENHARIA CIVIL E AMBIENTAL
CURSO DE GRADUAÇÃO EM ENGENHARIA CIVIL**

ALAN LINS ALVES

**APLICAÇÃO DE REDES NEURAIIS COM VARIÁVEIS
GEOMORFOLÓGICAS PARA MAPEAMENTO DA
SUSCEPTIBILIDADE A DESLIZAMENTOS DE TERRA EM LAGOA
ENCANTADA - PE**

RECIFE, 2024

ALAN LINS ALVES

**APLICAÇÃO DE REDES NEURAIIS COM VARIÁVEIS
GEOMORFOLÓGICAS PARA MAPEAR A SUSCEPTIBILIDADE A
DESLIZAMENTOS DE TERRA EM LAGOA ENCANTADA - PE**

Monografia apresentada à Universidade Federal de Pernambuco como parte dos requisitos para a obtenção do título de Engenheiro Civil.

Área de concentração: Engenharia Civil.

Orientador: Prof. Dr. Samuel França Amorim

RECIFE, 2024

Ficha de identificação da obra elaborada pelo autor,
através do programa de geração automática do SIB/UFPE

Alves, Alan Lins.

Aplicação de redes neurais com variáveis geomorfológicas para mapeamento da susceptibilidade a deslizamentos de terra em Lagoa Encantada - PE / Alan Lins Alves. - Recife, 2024.

107 p. : il., tab.

Orientador(a): Samuel França Amorim

Trabalho de Conclusão de Curso (Graduação) - Universidade Federal de Pernambuco, Centro de Tecnologia e Geociências, Engenharia Civil - Bacharelado, 2024.

8,3.

Inclui referências.

1. RNA. 2. QGIS. 3. Python. 4. Susceptibilidade. 5. Deslizamentos. I. Amorim, Samuel França. (Orientação). II. Título.

620 CDD (22.ed.)



UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE TECNOLOGIA E GEOCIÊNCIAS
DEPARTAMENTO DE ENGENHARIA CIVIL E AMBIENTAL
CURSO DE GRADUAÇÃO EM ENGENHARIA CIVIL

PARECER DA COMISSÃO DE
TRABALHO DE CONCLUSÃO DE CURSO I

ALAN LINS ALVES

**APLICAÇÃO DE REDES NEURAIS COM VARIÁVEIS
GEOMORFOLÓGICAS PARA MAPEAR A SUSCEPTIBILIDADE A
DESLIZAMENTOS DE TERRA EM LAGOA ENCANTADA - PE**

ÁREA DE CONCENTRAÇÃO: Engenharia Civil

A comissão examinadora composta pelos professores abaixo, sob a presidência do primeiro, considera o candidato ALAN LINS ALVES **APROVADO COM NOTA 8,3**.

Recife, 01 de novembro de 2024.

Prof. Dr. Samuel França Amorim (Orientador)

Prof. Dr. Silvio Romero de Melo Ferreira (Integrante Interno da Banca Examinadora)

Prof. Dr. Pedro Eugênio Silva de Oliveira (Integrante Externo da Banca Examinadora)

Dedico este trabalho a minha família, em especial minha mãe Marcia Lins e meu pai Almir Alves por toda a base moral cedida.

A minha noiva Karolaine Santos pelo apoio incondicional aos meus objetivos.

AGRADECIMENTOS

Agradeço aos meus pais Almir e Marcia, por serem meus maiores incentivadores na vida educacional, por não medirem esforços para me apoiar em cada novo desafio.

Agradeço a minha noiva, Karolaine Santos, pela torcida, incentivo, paciência, por acompanhar o passo a passo em revisões das ideias e do texto, por sempre trazer soluções nos momentos em que me vi sem saída.

Agradeço à Universidade Federal de Pernambuco – UFPE e todo seu corpo docente e administrativo, por todos os aprendizados e oportunidades, e por ter me possibilitado a chance de contribuir com a evolução da Engenharia.

Aos meus colegas de Universidade, que se mobilizaram e me incentivaram na organização e concretização deste curso.

Ao meu orientador Samuel França Amorim, pelas dicas, conselhos e pelas impecáveis correções.

A todos citados acima e aos que mesmo não tendo seus nomes explicitados contribuíram de alguma forma, meus sinceros agradecimentos.

RESUMO

Este trabalho aborda a aplicação de redes neurais artificiais (RNAs) na criação de mapas de susceptibilidade a deslizamentos de terra para a comunidade de Lagoa Encantada, localizada em Recife/PE, destacando sua eficácia e rapidez em comparação com métodos tradicionais. A crescente incidência de deslizamentos de terra e a necessidade de métodos precisos para avaliar e mitigar esses riscos motivaram esta pesquisa, que visa contribuir para a segurança da população e a gestão de desastres naturais. O objetivo principal deste estudo é desenvolver um mapa de susceptibilidade a deslizamentos de terra da comunidade de Lagoa Encantada, utilizando técnicas de inteligência artificial, especificamente RNAs, juntamente a programas de código aberto como QGIS e Python. Para alcançar este objetivo, foram geradas as variáveis geomorfológicas do local de estudo com o QGIS e foram utilizadas bibliotecas como Pandas para análise e tratamento de dados, Scikit-learn para normalização, pré-processamento, treinamento e validação dos modelos utilizando o Python. Os resultados demonstraram que as Redes Neurais Artificiais são altamente eficazes na geração de mapas de suscetibilidade com boa precisão. As principais métricas de avaliação, incluindo acurácia, precisão, revocação e F1 Score, todas superaram a marca de 90%. Além disso, a área sob a curva ROC (AUC) também apresentou valores acima de 0,9, tanto no conjunto de treino quanto no conjunto de teste. Esses resultados confirmam a robustez e a confiabilidade das RNAs na previsão de áreas suscetíveis a deslizamentos. Uma avaliação dos erros do resultado também foi realizada a fim de identificar a causa principal, permitindo que, no futuro, os fatores mapeados sejam levados em consideração para aprimorar o modelo e reduzir a ocorrência de erros similares em análises subsequentes. Conclui-se que a utilização de RNAs representa uma abordagem promissora para a avaliação de susceptibilidade a deslizamentos de terra, oferecendo uma alternativa rápida e eficiente aos métodos convencionais. Este estudo contribui para o avanço das técnicas de mapeamento de riscos geotécnicos e reforça a importância da integração de inteligência artificial em estudos de engenharia civil.

Palavras-chave: RNA, QGIS, Python, Susceptibilidade, Deslizamentos.

ABSTRACT

This work addresses the application of artificial neural networks (ANNs) in creating susceptibility maps for landslides in the community of Lagoa Encantada, located in Recife/PE, highlighting their efficiency and speed compared to traditional methods. The increasing incidence of landslides and the need for accurate methods to assess and mitigate these risks motivated this research, which aims to contribute to the safety of the population and the management of natural disasters. The main objective of this study is to develop a landslide susceptibility map for the community of Lagoa Encantada using artificial intelligence techniques, specifically ANNs, along with open-source programs such as QGIS and Python. To achieve this objective, geomorphological variables of the study area were generated using QGIS, and libraries such as Pandas for data analysis and processing, and Scikit-learn for normalization, preprocessing, training, and validation of models using Python were utilized. The results showed that Artificial Neural Networks are highly effective in generating susceptibility maps with good accuracy. The main evaluation metrics, including accuracy, precision, recall, and F1 Score, all surpassed the 90% mark. Additionally, the area under the ROC curve (AUC) also presented values above 0.9, both in the training and test sets. These results confirm the robustness and reliability of ANNs in predicting susceptible areas for landslides. A evaluation of the errors in the results was also conducted to identify the main cause, allowing the mapped factors to be considered in the future to improve the model and reduce the occurrence of similar errors in subsequent analyses. It is concluded that the use of ANNs represents a promising approach for assessing landslide susceptibility, offering a fast and efficient alternative to conventional methods. This study contributes to the advancement of geotechnical risk mapping techniques and reinforces the importance of integrating artificial intelligence in civil engineering studies.

Keywords: ANN, QGIS, Python, Susceptibility, Landslides.

LISTA DE FIGURAS

Figura 1 - Precipitação média mensal para a cidade de Recife-PE (1961-2019).....	24
Figura 2 - Representação do comportamento do movimento de queda.....	28
Figura 3 - Representação do comportamento do movimento de tombamento	28
Figura 4 - Da esquerda para a direita, representação do comportamento do movimento de deslizamento rotacional e translacional	29
Figura 5 - Representação do comportamento do movimento de expansão lateral	30
Figura 6 - Representação do comportamento do movimento de corridas	30
Figura 7 - Municípios monitorados da região Nordeste.....	32
Figura 8 - Modelo não linear de um neurônio	37
Figura 9 - Função de ativação ReLU	39
Figura 10 - RNA perceptron multicamadas com apenas uma camada oculta	40
Figura 11 - Divisão do conjunto de dados em conjuntos de treino, validação e teste	42
Figura 12 - Propagação direta	43
Figura 13 - Cálculo do erro	43
Figura 14 - Retropropagação	44
Figura 15 - Otimizador iterando até o ponto de mínimo local da função de custo.....	45
Figura 16 - Modelos e possíveis ajustes	47
Figura 17 - Exemplo de divisão dos dados em 5 dobras para a validação cruzada	48
Figura 18 - Matriz de confusão	50
Figura 19 - Curva ROC.....	51
Figura 20 - Mapa de localização da área de estudo	54
Figura 21 - Localização dos pontos instáveis	56
Figura 22 - Área de influência dos pontos instáveis	57
Figura 23 - Pontos estáveis fora da área de influência.....	58
Figura 24 - Modelo digital de terreno	59
Figura 25 - Fluxograma da dependência de variáveis	60
Figura 26 - Declividade do terreno	61
Figura 27 - Ilustração diagramática da curvatura transversal do terreno	62
Figura 28 - Curvatura transversal do terreno	63
Figura 29 - Ilustração diagramática da curvatura longitudinal do terreno.....	63
Figura 30 - Curvatura longitudinal do terreno	64

Figura 31 - Aspecto.....	65
Figura 32 - Distância da bacia de contribuição sob a superfície do terreno	66
Figura 33 - Inclinação média da bacia de contribuição	67
Figura 34 - Ilustração da extração dos dados do QGIS	68
Figura 35 - Logo do QGIS.....	69
Figura 36 - Logo do Python.....	70
Figura 37 - Índice de uso das linguagens de programação ao longo dos anos	71
Figura 38 - Logo do scikit-learn	71
Figura 39 - Histogramas das variáveis.....	73
Figura 40 - Diagrama de caixa das variáveis	74
Figura 41 - Matriz de correlação das variáveis.....	75
Figura 42 - Código: Divisão em treino e teste	77
Figura 43 - Divisão espacial dos dados em treino e teste	77
Figura 44 - Código: Definição do fluxo de dados.....	78
Figura 45 - Código: Definição dos valores da busca em grade	78
Figura 46 - Em destaque: Falsos negativos na matriz de confusão	79
Figura 47 - Código: Busca da melhor rede neural	80
Figura 48 - Rede neural treinada.....	81
Figura 49 - Matriz de confusão: Conjunto de treino.....	82
Figura 50 - Matriz de confusão: Conjunto de teste.....	83
Figura 51 - Curva ROC: Conjunto de treino.....	84
Figura 52 - Curva ROC: Conjunto de teste.....	85
Figura 53 - Fluxo de obtenção do mapa de susceptibilidade	86
Figura 54 - Mapa de susceptibilidade de Lagoa Encantada.....	87
Figura 55 - Mapa de susceptibilidade: Falsos positivos	88
Figura 56 - Falsos positivos: Área de vegetação	90
Figura 57 - Mapa de susceptibilidade: Falsos negativos	91
Figura 58 - Mapa de susceptibilidade a deslizamentos de terra gerado por RNA.....	93
Figura 59 - Mapa de susceptibilidade a deslizamentos de terra criado por Coutinho	94

LISTA DE TABELAS

Tabela 1 - Classificação dos movimentos de massa	26
Tabela 2 - Classificação dos movimentos de massa de acordo com Hungr	27
Tabela 3 - Graus de riscos de deslizamentos	34
Tabela 4 - Funções de ativação usuais	38
Tabela 5 - Nomes simplificados das variáveis.....	60
Tabela 6 - Estatísticas dos dados	72
Tabela 7 - Melhores hiperparâmetros encontrados	80
Tabela 8 - Métricas no conjunto de treino	82
Tabela 9 - Métricas no conjunto de teste	83
Tabela 10 - Falsos Positivos e suas características	89
Tabela 11 - Falsos Negativos e suas características.....	91
Tabela 12 - Indicadores e subindicadores de susceptibilidade a movimento e transporte de massa.....	93

SUMÁRIO

1	INTRODUÇÃO.....	16
1.1	JUSTIFICATIVA E MOTIVAÇÃO	17
2	OBJETIVOS	22
2.1	OBJETIVO GERAL	22
2.1.1	Objetivos específicos	22
3	FUNDAMENTAÇÃO TEÓRICA.....	23
3.1	DESASTRES NATURAIS.....	23
3.1.1	Influência da Chuva nos Movimentos de Massa.....	24
3.1.2	Interferência da Ação Antrópica nos Movimentos de Massa	24
3.2	TIPOS DE MOVIMENTOS DE MASSA.....	25
3.2.1	Quedas	28
3.2.2	Tombamentos.....	28
3.2.3	Deslizamentos.....	29
3.2.4	Expansões laterais.....	29
3.2.5	Corridas	30
3.3	GERENCIAMENTO DE RISCOS EM DESLIZAMENTOS	31
3.3.1	Monitoramento de Desabamento de Terra em Pernambuco	31
3.4	ANÁLISE DE RISCO DE DESLIZAMENTO	33
3.5	APRENDIZADO DE MÁQUINA	34
3.5.1	Redes Neurais Artificiais.....	36
3.5.2	Neurônio Artificial e Funções de Ativação	36
3.5.3	Perceptron Multicamadas.....	39
3.5.4	Treino da Rede Neural Artificial.....	41
3.5.4.1	Divisão dos Dados em Treino, Validação e Teste	41
3.5.4.2	Propagação Direta.....	42
3.5.4.3	Cálculo do Erro.....	43
3.5.4.4	Retropropagação do Erro	44
3.5.4.5	Atualização dos Pesos.....	45
3.5.4.6	Épocas.....	46
3.5.5	O Problema de Subajuste (<i>underfit</i>) e Sobreajuste (<i>overfit</i>).....	46
3.5.6	Fluxo de Dados.....	47

3.5.7	Validação Cruzada	47
3.5.8	Hiperparâmetros de uma Rede Neural.....	48
3.5.9	Busca em Grade	49
3.5.10	Validação dos Resultados.....	49
3.5.10.1	Matriz de Confusão.....	49
3.5.10.2	Curva ROC	51
3.5.11	Desafios da Inteligência Artificial	52
4	MATERIAIS E MÉTODOS	54
4.1	ÁREA DE ESTUDO	54
4.2	OBTENÇÃO DOS DADOS.....	55
4.2.1	Pontos Instáveis.....	55
4.2.2	Pontos Estáveis.....	56
4.2.3	Modelo Digital de Terreno	58
4.2.4	Declividade do Terreno	60
4.2.5	Curvatura Transversal do Terreno	61
4.2.6	Curvatura Longitudinal do Terreno.....	63
4.2.7	Aspecto.....	65
4.2.8	Distância da Bacia de Contribuição Sob a Superfície do Terreno	66
4.2.9	Inclinação Média da Bacia de Contribuição	67
4.2.10	Base de Dados Compilada.....	68
4.3	PROGRAMAS UTILIZADOS.....	69
4.3.1	Quantum Geographic Information System - QGIS	69
4.3.2	Python	70
5	RESULTADOS E DISCUSSÕES.....	72
5.1	ANÁLISE DOS DADOS	72
5.1.1	Estatísticas Básicas	72
5.1.2	Histogramas.....	73
5.1.3	Diagramas de Caixa.....	74
5.1.4	Matriz de Correlação	75
5.2	TREINAMENTO E VALIDAÇÃO DA REDE NEURAL.....	76
5.2.1	Divisão dos Dados em Treino e Teste.....	76
5.2.2	Criação do Fluxo de Dados	77
5.2.3	Definição dos Conjuntos para a Busca em Grade	78

5.2.4	Treino do Modelo de RN - Realização da Busca em Grade com Validação Cruzada.....	79
5.2.5	Validação do Modelo – Matriz de Confusão e Métricas Derivadas.....	81
5.2.6	Validação do Modelo – Curva ROC	84
5.2.7	Mapa de Susceptibilidade	85
5.2.8	Análise dos Falsos Positivos	88
5.2.9	Análise dos Falsos Negativos.....	90
5.2.10	Comparação Entre Mapas de Susceptibilidade	92
6	CONCLUSÕES.....	96
7	REFERÊNCIAS BIBLIOGRÁFICAS.....	98

1 INTRODUÇÃO

Desastres naturais ocorrem em diversos níveis de intensidade e gravidade, dependendo de elementos próprios da dinâmica da natureza e de intervenções humanas, podendo resultar em danos materiais, econômicos, sociais e ambientais. Segundo Moreira e Barbieri (2020), desastres relacionados a inundações, deslizamentos de terra e outros eventos relacionados aos padrões climáticos e pluviométricos colocam o Brasil como um dos mais vulneráveis às mudanças climáticas na América Latina.

O Instituto Brasileiro de Geografia e Estatística (IBGE, 2019) enfatiza por meio do estudo “Suscetibilidade a Deslizamentos do Brasil: Primeira aproximação” que o conjunto de fatores envolvendo geologia, geomorfologia, pedologia, vegetação e uso da terra, declividade e pluviosidade quando somados favorecem em algumas regiões do Brasil o desencadeamento de fenômenos de movimentos gravitacionais de massa, nesse quesito a região Nordeste apresentou taxas de 3,8% para áreas com suscetibilidade muito alta e 10,1% para áreas de alta suscetibilidade a deslizamentos.

Como agravante, conforme observado por Almeida *et al.* (2014), a população mais carente tende a ocupar áreas naturalmente inadequadas e/ou suscetíveis a riscos naturais e que são ocupadas sem os mínimos preceitos técnicos. Isso ocorre, pois espaços não expostos aos desastres são majoritariamente ocupados por classes sociais com maior poder econômico (Santana, 2019). Essa situação força a população menos favorecida a recorrer às moradias que se encontram em locais de risco, muitas delas são oriundas de desocupações devido ao decreto de desabitação por incompatibilidade de condições de permanência, de modo que, estas habitações são invadidas apesar de suas inadequações e tais comunidades permanecem expostas de forma iminente ao risco de catástrofes naturais como deslizamentos de terra.

No Brasil, entre os anos de 1991 e 2012 a ocorrência de movimentos de massa resultou no segundo maior número de óbitos em desastres decorrentes de fenômenos naturais (CEPED UFSC, 2013). De acordo com um levantamento do Instituto de Pesquisas Tecnológicas, entre 1988 e junho de 2022, Recife está entre os 10 municípios com maior número de mortes por deslizamentos (Macedo e Sandre, 2022). Além disso, Recife ocupa a quinta posição entre as cidades brasileiras com o maior número de habitantes vivendo em áreas de morro propensas a deslizamentos de terra (IBGE, 2019).

Diante de tais fatos torna-se imprescindível a aplicação de medidas baseadas em critérios técnicos objetivos, dentre os quais destacam-se os mapeamentos de suscetibilidade a movimentos gravitacionais de massa. De acordo com Bragagnolo et al. (2019) existem limitações referentes às metodologias tradicionalmente aplicadas, como a demora na geração de resultados e, em alguns casos, alto nível de detalhamento técnico, exigindo um grau de conhecimento elevado e, portanto, reduzindo suas aplicabilidades. Desta forma, as tecnologias de Sistema de Informações Geográficas (SIG) são identificadas como ferramentas favoráveis para a otimização de análises referentes aos movimentos de massa em todas as etapas de estudo, incluindo a detecção, o monitoramento, a espacialização das áreas suscetíveis e até mesmo a previsão de recorrência da ameaça (Tralli et al., 2005).

O método heurístico é tradicionalmente utilizado para avaliação de susceptibilidade a deslizamentos de terra, entretanto, de acordo com Lucchese (2022), nos últimos anos, tem havido um interesse crescente em aplicar metodologias de Inteligência Artificial para a geração de mapas de suscetibilidade a movimentos de massa. Metodologias de mineração de dados (MD) e de aprendizado de máquina (AM) são consideradas apropriadas para aplicações em grandes áreas, e especialmente se as características a serem estudadas possuem relações complexas não lineares, como é o caso da identificação de áreas suscetíveis a deslizamentos de terra, além disso esses algoritmos podem ser facilmente integrados em Sistemas de Informações Geográficas (Bragagnolo et al., 2019).

Entre as diversas técnicas de aprendizado de máquina que são atualmente utilizadas para mapeamento da suscetibilidade a deslizamentos de terra, as Redes Neurais Artificiais (RNA) destacam-se, sendo consideradas um dos métodos mais precisos para essa finalidade, por sua elevada acurácia, capacidade de aprendizagem e generalização de resultados (Gameiro, 2020; Azevedo, 2022). Trabalhos como o de Amorim (2012), Zare et al. (2013) e Braun et al. (2019) utilizam-se dessa técnica para classificação e identificação dos locais com alto risco de deslizamento.

1.1 JUSTIFICATIVA E MOTIVAÇÃO

De acordo com o Instituto Nacional de Pesquisas Espaciais (INPE), um desastre é o resultado de eventos adversos que causam grandes impactos na sociedade, sendo distinguidos principalmente em função de sua origem, isto é, da natureza do fenômeno que o desencadeia (INPE, 2008). Sua ocorrência causa altos danos humanos, materiais, econômicos e/ou

ambientais, as quais excedem a habilidade da comunidade afetada de se reestabelecer com seus próprios recursos.

Apesar da atribuição de diferentes conceitos, desastres podem ser classificados quanto à sua tipologia, sendo divididos em: desastres de origem natural, quando provocados por fenômenos e desequilíbrios da natureza, desastres humanos ou antropogênicos, quando provocados pelas ações e/ou omissões humanas e desastres mistos, quando atividades humanas contribuem para intensificar, agravar, modificar ou complicar fenômenos naturais, geradores de desastres. Desastres são frequentemente o resultado de vulnerabilidades sociais e ambientais que são agravadas por fatores como a mudança climática, o crescimento populacional e a urbanização (Souza et al., 2009).

Nos últimos anos houve aumento na frequência e intensidade dos desastres naturais, sendo eles eventos catastróficos com origens atmosféricas, geológicas e hidrológicas que podem provocar mortes, danos à infraestrutura e interrupção do funcionamento social, nessa categoria estão, por exemplo, as inundações, deslizamentos de terras, terremotos, entre outros (XU et al., 2016).

Segundo o relatório publicado no Intergovernmental Panel on Climate Change - IPCC (2012), é esperado um aumento significativo em perdas e danos causados por desastres naturais devido a mudanças climáticas e o aumento da exposição e vulnerabilidade da sociedade, visto que alterações climatológicas podem intensificar o ciclo hidrológico, acarretando em maiores precipitações, levando a mudanças na intensidade, frequência e severidade dos eventos desastrosos (Apurv et al., 2015). Assim sendo, a possibilidade de ocorrência de desastres naturais vem assumindo notoriedade como uma importante pauta para os governantes, organizações internacionais, pesquisadores e agências de gerenciamento de emergências que objetivam mapear maneiras eficientes em resposta às consequências de tais eventos (Nascimento e Alencar, 2016).

A intensidade das chuvas sempre foi objeto de estudo no meio científico, justamente pelos impactos e modificações que causam nas bacias hidrográficas e nas cidades (Da Silva et al., 2021). No Brasil, os deslizamentos de terra, inundações e enxurradas são responsáveis por parte dos prejuízos ocasionados por desastres, provocando a destruição de moradias, o aumento da vulnerabilidade social e a perda de bens essenciais. De acordo com Azevedo (2022), apenas nos primeiros meses do ano de 2024 os desastres naturais no Brasil causaram prejuízos financeiros estimados em R\$ 32 bilhões.

Segundo Alves (2022), em Pernambuco (PE), o desastre das chuvas ocorridas em 2022 ficou marcado como a maior tragédia natural do século XXI no Estado. Estudo do

Centro Nacional de Monitoramento e Alertas de Desastres Naturais (CEMADEN), aponta que são 207 mil pessoas vivendo em área de risco, o que significa 13,4% da população sendo impactadas pelas enxurradas e deslizamentos de barreiras. No município de Jaboatão dos Guararapes (PE), a situação é ainda mais crítica, já que quase um terço da população vive em área de risco (29,2%), totalizando 188 mil pessoas.

Deslizamento de terra é caracterizado como um evento natural cujo desencadeamento é provocado por um conjunto de fatores como geomorfologia do solo, clima, vegetação, alta pluviosidade e a ação antrópica (Rocha e Körting, 2023). Além desses fatores, os deslizamentos são potencializados, principalmente, pela ação das águas das chuvas, que infiltram o solo das encostas, provocando a redução da sua estabilidade e, consequentemente, induzindo os movimentos de massa. Associado aos problemas ambientais, existe a questão social e econômica, relacionada com a ocupação desordenada de terrenos em áreas de risco, essa questão surge principalmente devido ao aumento dos preços dos terrenos encontrados em áreas mais estáveis, como consequência dessa problemática, famílias inteiras acabam residindo em locais considerados com maior propensão a ocorrência de desastres como movimentação de terra (Pereira *et al.*, 2020).

Os movimentos gravitacionais de massa representam um dos principais desafios enfrentados todos os anos no Brasil (IBGE, 2019). Estreitando um pouco mais essa perspectiva para Pernambuco, principalmente durante os períodos chuvosos, o Estado enfrenta dificuldades com as inundações e os deslizamentos de terra (De Souza *et al.*, 2014). Por esse motivo, é fundamental conhecer os limites da relação entre as chuvas e a ocorrência de deslizamentos de terra, uma vez que desse modo, é possível ajudar na prevenção e reparação de danos oriundos desses eventos (Da Silva *et al.*, 2022).

A Cidade do Recife, localizada no Estado de Pernambuco, como outros centros urbanos localizados no litoral brasileiro, apresenta 23 bairros que concentram parte de suas habitações em áreas cuja topografia favorece, frente ao somatório dos eventos já anteriormente mencionados, a ocorrência de deslizamentos de massa. Em maio de 2022, o bairro Jardim Monte Verde, localizado na divisa entre Recife e Jaboatão dos Guararapes, foi palco de uma tragédia devastadora devido a deslizamentos de terra causados por chuvas intensas resultando em mais de 20 mortes.

Na comunidade de Lagoa Encantada, localizada no bairro do Ibura, deslizamentos de terra têm sido uma preocupação constante, especialmente durante períodos de chuvas intensas. Essa localidade apresenta forte alteração da paisagem natural por meio da utilização do espaço pela sociedade para a construção de habitações de forma desordenada. A escolha

da comunidade de Lagoa Encantada como área de estudo para avaliar a suscetibilidade a deslizamentos é motivada pela alta frequência desses eventos na Região Metropolitana do Recife (RMR).

Vários pesquisadores têm concentrado esforços para desenvolver modelos e metodologias para gerenciar o risco destes eventos. Para realizar análises quantitativas do perigo de deslizamentos e inundações são necessários dados históricos em quantidade e qualidade suficientes. Entretanto, a falta de dados com as características citadas torna a tarefa de predição de deslizamentos desafiadora.

Graças aos avanços recentes no poder computacional, popularização da internet, redes de telefonia móvel, avanços tecnológicos de sensores de monitoramento remoto e aumento da quantidade de dados compartilhados, o volume de informação disponível teve aumento significativo. Porém, tais dados necessitam passar por um processo de tratamento e interpretação, para só então, serem usadas na tomada de decisão. Tal processo, quando realizado manualmente torna-se custoso e demorado, causando uma desmotivação e consequentemente um abandono de tais dados, levando a perda de informações valiosas que poderiam auxiliar a população.

O desenvolvimento de abordagens que utilizam como fonte inicial dados semiestruturados para o mapeamento do perigo de desastres naturais tem como principal vantagem superar o problema de indisponibilidade de dados amplamente relatado na literatura. Além disso, ao usar algoritmos de aprendizado de máquina, o processo de análise e classificação pode ser automatizado o que proporciona maior velocidade e precisão a baixo custo. Nesse cenário, uma das abordagens inovadoras utilizadas são redes neurais. Essas estruturas são capazes de aprender com exemplos e identificar padrões complexos em grandes conjuntos de dados (Modesto et al., 2023). Quando aplicadas ao estudo de deslizamentos de terra, as redes neurais são uma ferramenta poderosa para analisar e prever esses eventos, auxiliando na identificação de riscos e na tomada de decisões mais embasadas.

Ao utilizar redes neurais treinadas com dados que envolvem aspectos geotécnicos, geomorfológicos, hidrológicos, meteorológicos e outras variáveis pertinentes, é possível revelar padrões e identificar relações complexas entre os fatores que contribuem para o desencadeamento de deslizamentos de terra. Com isso, a rede neural pode fornecer uma análise detalhada da vulnerabilidade de uma área, destacando regiões com maior risco de ocorrência. Essa informação é essencial para a aplicação de estratégias de prevenção e mitigação de desastres. Além disso, esses modelos preditivos têm a capacidade de estimar,

com maior precisão, a probabilidade de deslizamentos em pontos específicos, proporcionando um suporte eficaz para a tomada de decisões voltadas à segurança pública e ambiental.

Embora as redes neurais alcancem inúmeras possibilidades e vantagens na análise de características de terra, é importante destacar que o sucesso desses modelos depende da qualidade e representatividade dos conjuntos de dados utilizados, bem como da expertise dos especialistas na interpretação dos resultados obtidos, identificando erros e comportamentos anômalos. A combinação do conhecimento humano e da capacidade de processamento das redes neurais abre novas perspectivas na compreensão e no enfrentamento dos desafios relacionados aos deslizamentos de massa, tornando-se uma alternativa para a proteção das comunidades e preservação do meio ambiente em áreas de risco.

2 OBJETIVOS

2.1 OBJETIVO GERAL

Elaborar um mapa de suscetibilidade a deslizamentos de terra na comunidade de Lagoa Encantada, localizada em Recife-PE, utilizando inteligência artificial por meio de redes neurais artificiais. O projeto fará uso de programas gratuitos e será baseado na análise detalhada da geomorfologia local, visando identificar áreas de risco e contribuir para a prevenção de desastres naturais na região.

2.1.1 Objetivos específicos

Nesse contexto, o presente trabalho tem como objetivos específicos:

- Utilizar programas como QGIS, Python e algumas bibliotecas de código aberto como Pandas para análise e tratamento de dados, avaliando a biblioteca scikit-learn para normalização dos dados e treino na RN;
- Calcular variáveis geomorfológicas a partir do Modelo Digital de Elevação do local procedente do levantamento LIDAR do PE3D usando o QGIS;
- Obter os melhores hiperparâmetros para a rede neural artificial visando identificar o modelo mais adequado para o mapa de susceptibilidade.
- Avaliar o Mapa de susceptibilidade a deslizamento de terra.

3 FUNDAMENTAÇÃO TEÓRICA

3.1 DESASTRES NATURAIS

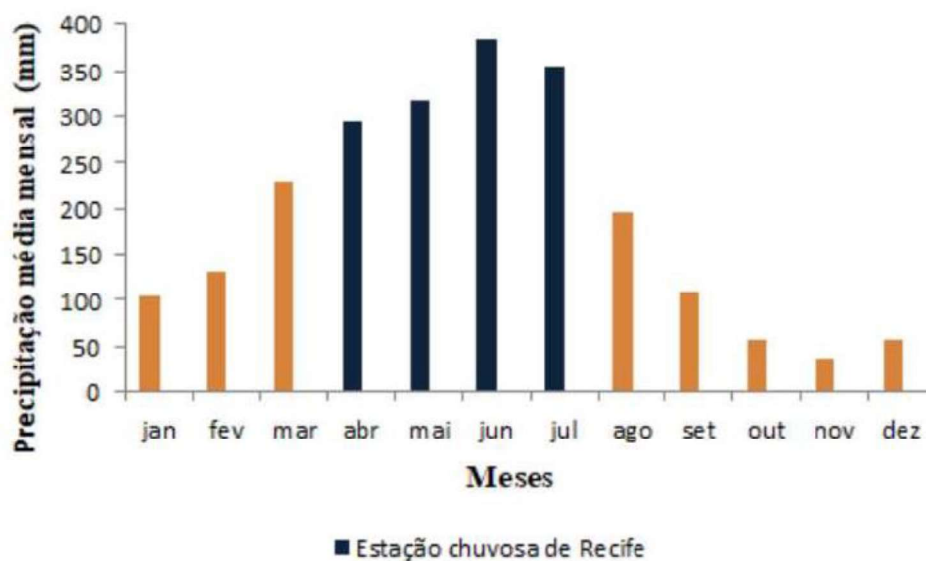
Os Desastres Naturais representam uma preocupação constante, especialmente para aquelas pessoas que residem em áreas de risco. Em primeiro momento, o termo nos faz associá-lo com terremotos, tsunamis, erupções vulcânicas, ciclones e furacões, entretanto, desastres naturais também engloba processos e fenômenos como deslizamentos, inundações, subsidências e erosão do solo, que podem ocorrer naturalmente ou induzidos pelo homem (Goswami et al., 2018).

Devido à recorrência destes eventos e impactos cada vez mais intensos, suas consequências tornam-se expressivas em danos e perdas, de caráter social, econômico e ambiental, de maneira que os cientistas sugerem já ser resultado das mudanças climáticas globais. Estas alterações são as principais alavancas para a ocorrência de desastres naturais de origem hidrometeorológica e climática, tais como inundações, enxurradas, deslizamentos de terra, destruição por vendavais, colapsos de safras e de sistemas de abastecimento de água por secas, entre outros (Sulaiman e Aledo, 2016).

No Brasil, tem-se observado o crescimento de ocorrências de desastres naturais, seja em virtude da intensificação de eventos hidrometeorológicos e climáticos em algumas regiões do país, seja devido ao aumento da exposição ao risco pela ocupação de áreas suscetíveis a desastres naturais e/ ou suas ações desordenadas. Nesse panorama de magnitude e de frequência de eventos climáticos extremos, observa-se a possibilidade que no futuro alguns lugares no Brasil tornem-se cada vez mais difíceis para se viver e trabalhar (Freitas et al., 2014).

A ocorrência de deslizamentos na cidade do Recife são mais frequentes nos meses de abril a julho, tendo sua frequência aumentada pelos eventos climáticos que ocorrem de modo natural e pela ocupação desordenada da terra devido às dinâmicas antrópicas (Da Silva et al., 2015). Entretanto, é importante salientar que a mitigação, prevenção e remediação, com ações de forma ordenada, são importantes para evitar ou minimizar os eventuais danos. A Figura 1 mostra as médias mensais de precipitação de 1961 a 2019, destacando os meses de abril a julho como os períodos com maior volume de chuvas.

Figura 1 - Precipitação média mensal para a cidade de Recife-PE (1961-2019)



Fonte: Da Silva *et al.* (2021).

3.1.1 Influência da Chuva nos Movimentos de Massa

Movimentos de massa em encostas podem ocorrer devido ao excesso de água no solo, geralmente devido a fortes precipitações ou mesmo por chuvas de média ou baixa intensidade precipitadas em solos já saturados ou quase saturados. Os eventos de deslocamentos de massa relacionam-se, no Brasil, diretamente com a ocorrência das chuvas, o fator determinante baseia-se justamente nos aspectos extrínsecos das regiões de risco (De Lima Santos *et al.*, 2019). Em países de clima tropical, como no Brasil, o volume de água no solo demonstra-se como uma das variáveis mais importantes para a ocorrência de deslizamentos de massa (Molina *et al.*, 2015).

As precipitações promovem a instabilidade das encostas, devido ao aumento do nível de água e da geração de forças de percolação, do preenchimento temporário de fendas e/ou estruturas no solo e rochas, atrelado a geração de pressões hidrostáticas e da formação de frentes de saturação, reduzindo a resistência dos solos pela perda da coesão (Highland e Bobrowsky, 2008).

3.1.2 Interferência da Ação Antrópica nos Movimentos de Massa

Terrenos que apresentam uma predisposição natural ao deslizamento quando adicionado às alterações realizadas pelo homem como mudança no relevo, retirada da

vegetação, mudança da drenagem entre outras ações, e somados à fatores climáticos como a chuva, são caracterizados como área de maior suscetibilidade aos movimentos de massa (Souza et al., 2009).

A ação do homem, nas áreas de morros e encostas, é um grande indutor de escorregamentos, que podem ser agravados com a ação da precipitação. No entanto, segundo Santos (2012), havendo interferência humana em determinada área, não é necessária uma intensidade elevada no nível pluviométrico para ocorrência de deslocamento de terra, quando comparado a situações em que o desencadeamento de escorregamentos ocorre em terrenos cujas condições são naturais, ou seja, sem ação antrópica.

3.2 TIPOS DE MOVIMENTOS DE MASSA

Movimentos de massa podem descrever uma ampla variedade de processos, descendente e ao longo da transversal, dos materiais de formação de uma encosta sob a ação da gravidade, que ocorre quando a tensão cisalhante atuante é maior que a tensão resistente da massa deslizando (Souza et al., 2009).

De acordo com Tominaga, Santoro e Amaral (2009), dentre as classificações mais adotadas internacionalmente é a proposta por Varnes (1978), que classifica os movimentos de massa com base em duas variáveis principais, o tipo de movimento e o material. Eles propõem cinco tipos de movimentos, sendo eles quedas, tombamentos, deslizamentos, expansões laterais e escorregamentos, e duas classes de material sendo rochas e solo que pode ser subdividido em detritos e terra, como demonstrado na Tabela 1. Para eles, qualquer movimento de massa pode ser classificado e descrito por dois nomes, o primeiro descreve o tipo de movimento e o segundo descreve o material.

Tabela 1 - Classificação dos movimentos de massa

Tipo de Movimento		Tipo de material		
		Rocha	Solo (engenharia)	
			de detrito	fino
Quedas		de rocha	de detritos	de terra
Tombamentos		de rocha	de detritos	de terra
Deslizamentos	Rotacional	abatimento de rocha	abatimanto de detritos	abatimento de terra
		de blocos rochosos	de blocos de detritos	de blocos de terra
	Translacional	de rocha	de detritos	de terra
Expansões laterais		de rocha	de detritos	de terra
Escorregamentos		de rocha	de detritos	de terra
		(rastejo profundo)	restejo de solo	
Complexos: Combinação de 2 ou mais dos principais tipos de movimentos				

Fonte: Adaptado de Cruden e Varnes (1993).

Em 2014, Hungr, Leroueil e Picarelli publicaram uma atualização da classificação proposta por Varnes (1978). O principal objetivo dessa atualização, segundo os autores, foi substituir os termos utilizados para descrever os materiais (rocha, detrito e terra) por uma terminologia mais aceita nas áreas de geologia e geotecnia. A nova classificação de Hungr, Leroueil e Picarelli (2014) resulta em 32 tipos de movimentos de massa e não inclui uma categoria separada para movimentos complexos; no entanto, os usuários podem criar tipos compostos ao combinar dois ou mais tipos. A tabela 2 apresenta essa classificação.

Tabela 2 - Classificação dos movimentos de massa de acordo com Hungr

Tipo de movimento	Rocha	Solo
Queda	1. Queda de rocha/gelo	2. Queda de pedregulho/ detrito/
Tombamento	3. Tombamento de bloco de rocha	5. Tombamento de cascalho/ areia/ silte
	4. Tombamento a flexão de rocha	
Escorregamentos	6. Escorregamento rotacional de rocha	11. Escorregamento rotacional de argila/silte
	7. Escorregamento planar de rocha	12. Escorregamento planar de argila/silte
	8. Escorregamento em cunha de rocha	13. Escorregamento de pedregulho/areia/detrito
	9. Escorregamento composto de rocha	14. Escorregamento composto de argila/silte
	10. Escorregamento irregular de rocha	
Expansão lateral	15. Espalhamento de talude de rocha	16. Espalhamento de areia/ silte liquefeito
		17. Espalhamento de argila sensível
Fluxo	18. Avalanche de rocha/gelo	19. Fluxo seco de areia/ silte/ detrito
		20. Fluxo de areia/silte/detrimento
		21. Corrida úmida de argila sensível
		22. Fluxo de detrito
		23. Fluxo de lama
		24. Inundação de detritos
		25. Avalanche de detritos
		26. Fluxo de terra
		27. Fluxo de turfa
Deformação de talude	28. Deformação de talude de montanha	30. Deformação de talude de solo
	29. Deformação de talude de rocha	31. Rastejo de solo
		32. Solifluxão

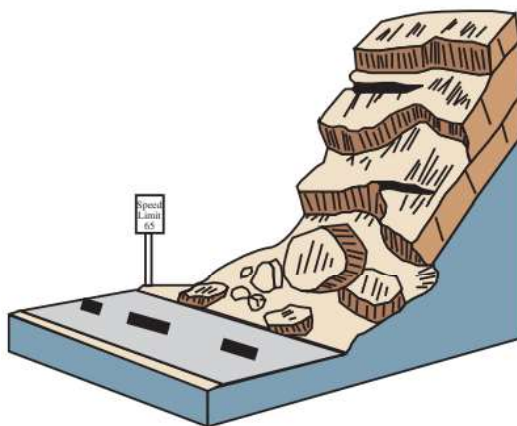
Fonte: Hungr, Leroueil e Picarelli (2014).

A seguir temos uma breve descrição sobre cada tipo de movimento.

3.2.1 Quedas

Quedas são movimentos abruptos de massas de materiais geológicos, como rochas e pedregulhos, que se desprendem de encostas íngremes ou penhascos, conforme demonstrado na Figura 2. A separação ocorre ao longo de discontinuidades, como fraturas, juntas e planos de acabamento, e o movimento ocorre por queda livre sendo fortemente influenciadas pela gravidade, intemperismo mecânico e presença de água intersticial (Varnes, 1978).

Figura 2 - Representação do comportamento do movimento de queda

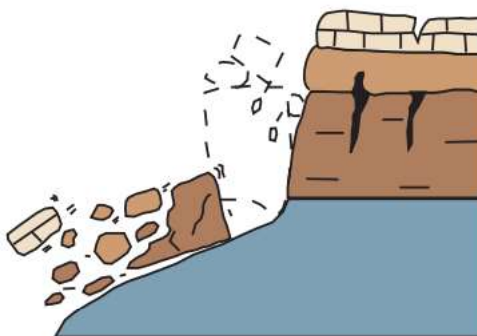


Fonte: Highland, 2004.

3.2.2 Tombamentos

As falhas por tombamento são distinguidas pela rotação para frente de uma unidade em torno de algum ponto pivô, sob as ações da gravidade e forças exercidas por unidades adjacentes ou por fluidos em rachaduras (Highland, 2004), conforme demonstrado na Figura 3.

Figura 3 - Representação do comportamento do movimento de tombamento



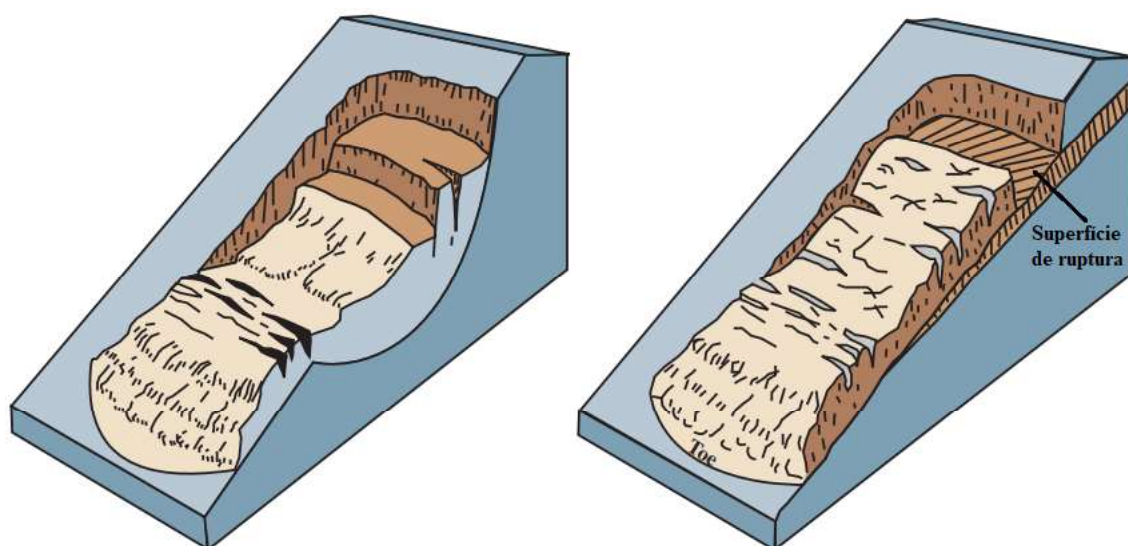
Fonte: Highland, 2004.

3.2.3 Deslizamentos

De acordo com Varnes (1978), os deslizamentos são categorizados em rotacionais e translacionais, dependendo da forma da superfície de falha.

O deslizamento rotacional ocorre quando a superfície de ruptura é curvada de forma côncava para cima e o movimento da massa é aproximadamente rotacional em torno de um eixo paralelo à superfície do solo e transversal ao deslizamento. Já o deslizamento translacional, ocorre quando a massa de deslizamento se move ao longo de uma superfície aproximadamente plana com pouca rotação ou inclinação para trás, conforme demonstrado na Figura 4.

Figura 4 - Da esquerda para a direita, representação do comportamento do movimento de deslizamento rotacional e translacional

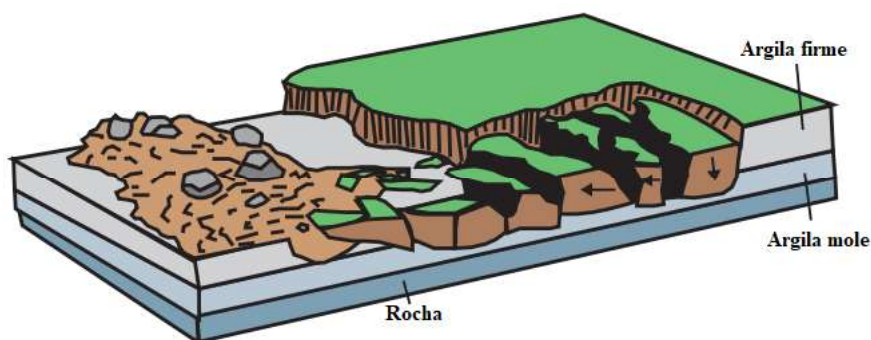


Fonte: Highland, 2004.

3.2.4 Expansões laterais

As expansões laterais geralmente ocorrem em declives muito suaves ou terreno plano. O modo dominante de movimento é a extensão lateral acompanhada por cisalhamento ou fraturas por tração, conforme demonstrado na Figura 5. A falha é causada pela liquefação, o processo pelo qual sedimentos saturados, soltos e sem coesão são transformados de um estado sólido em um estado liquefeito. A falha começa repentinamente em uma pequena área e se espalha rapidamente (Varnes, 1978).

Figura 5 - Representação do comportamento do movimento de expansão lateral



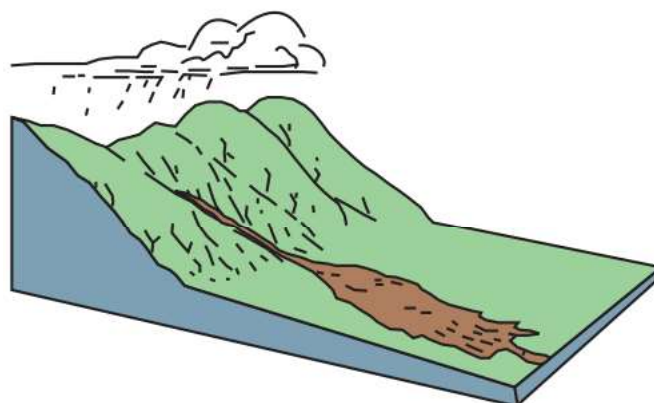
Fonte: Highland, 2004.

3.2.5 Corridas

As corridas de detritos são movimentos gravitacionais de massa complexos, ligados a eventos pluviométricos excepcionais. Ocorrem a partir de deslizamentos nas encostas e mobilizam grandes volumes de material, sendo o seu escoamento ao longo de um ou mais canais de drenagem, tendo comportamento líquido viscoso e alto poder de transporte (Carvalho et al., 2007), conforme demonstrado na Figura 6.

As corridas de detritos e lama estão entre os fenômenos mais destrutivos da natureza, em que o material com alto teor de umidade perde coesão e ganha velocidade no caminho de descida (Cruden e Varnes, 1993).

Figura 6 - Representação do comportamento do movimento de corridas



Fonte: Highland, 2004.

3.3 GERENCIAMENTO DE RISCOS EM DESLIZAMENTOS

A sociedade enfrenta diariamente riscos de diferentes naturezas, tanto que analisá-los tornou-se uma atividade necessária aos seres humanos. O risco pode ser definido como a união de um evento e suas consequências (Aven, 2015).

A criação de um sistema de alerta como ferramenta da gestão de risco é a cada evento mais necessário, fomentando uma base de dados para prover informações técnicas e estratégias a fim de subsidiar resoluções quanto a redução de impactos causados pelos desastres e promover ações, em todos os níveis de governo e nas organizações sociais, servindo de fonte de informação a sociedade. Para tal, são necessárias ações relacionadas a monitoramento, sistema de alerta e análise de risco (mapas de vulnerabilidade de áreas de risco), divulgação de informações e de apoio à tomada de decisões (Adition et al., 2018; Oliveira et al., 2018).

O gerenciamento de risco de desastres naturais pode ser entendido como um processo sistemático de uso de decisões administrativas, organizacionais, habilidades e capacidades operacionais para implementar políticas e estratégias, bem como capacidade de lidar com a sociedade para minimizar os impactos dos perigos naturais e desastres ambientais e tecnológicos relacionados. Isso inclui todas as formas de atividades, podendo ser medidas estruturais e não estruturais para evitar (prevenção) ou diminuir (mitigação) os efeitos adversos dos perigos. No Brasil, em esferas federal, estadual e municipal são realizados estudos e colocados em prática projetos destinados à mitigação e prevenção de deslizamentos.

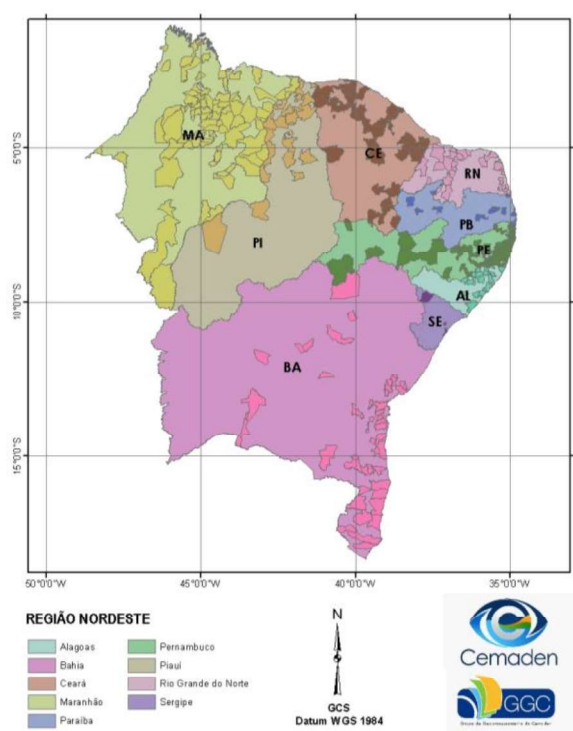
3.3.1 Monitoramento de Desabamento de Terra em Pernambuco

O Centro Nacional de Monitoramento e Alertas de Desastres Naturais - CEMADEN, foi criado em 2011 e vinculado ao Ministério da Ciência, Tecnologia, Inovação e Comunicações (MCTIC), para estabelecer uma atuação coordenada entre os órgãos envolvidos nas questões relativas à gestão de monitoramento e alertas, alarme e articulação, resposta e mobilização. Por sua vez, o CEMADEN emite alertas para o Centro Nacional de Gerenciamento de Riscos e Desastres – CENAD, auxiliando o Sistema Nacional de Defesa Civil.

O CEMADEN tem como objetivo principal desenvolver atividades de pesquisa científica, tecnológica e de inovação nas áreas de desastres naturais, meteorologia, hidrologia, geologia, agrometeorologia e incêndios florestais. Estas atividades têm ênfase na previsão, preparação, prevenção e mitigação do impacto de eventos extremos, em bacias hidrográficas urbanas e rurais, com reduzido tempo de resposta. Tais ações visam reduzir as perdas em termos de vidas humanas, bens materiais e meio ambiente, como decorrência desses eventos.

Atualmente, o CEMADEN monitora 333 municípios na região Nordeste, sendo 66 deles no estado de Pernambuco, como demonstrado na Figura 7, (CEMANDEN, 2023).

Figura 7 – Municípios monitorados da região Nordeste



Fonte: CEMADEN - Centro Nacional de Monitoramento e Alerta de Desastres Naturais, 2023.

O CEMADEN trabalha, em parceria, com outros órgãos, entre eles o Serviço Geológico do Brasil – SGB/CPRM, a Agência Nacional de Águas e Saneamento Básico - ANA e a Agência Pernambucanas de Águas e Clima - APAC. Isto permite que as informações levantadas pela setorização de riscos sejam disponibilizadas para as áreas identificadas como de alto e muito alto risco a deslizamentos e enchentes, para que seja possível emitir alertas, tornando ciente as comunidades sobre a forte possibilidade da ocorrência de desastres (CEMADEN, 2023).

A Prefeitura do Recife, por meio da Coordenadoria de Defesa Civil (Sedec/CODECIR), está desenvolvendo o Plano Municipal de Redução de Riscos (PMRR), com foco na proteção de assentamentos precários e na mitigação de desastres. O principal objetivo do PMRR é mapear e classificar os setores de risco na cidade, identificando áreas vulneráveis a deslizamentos e inundações, e implementando ações preventivas que garantam a segurança das populações mais expostas. Entre as iniciativas planejadas, estão intervenções estruturais nas regiões mais críticas, a criação dos Núcleos Comunitários de Defesa Civil (NUDEC), que envolvem a população na preservação e melhoria do ambiente urbano, e programas educativos voltados para a conscientização dos moradores sobre práticas seguras.

Além de suas ações preventivas, o PMRR busca promover uma integração eficaz entre os diferentes níveis de governo, assegurando que as medidas de redução de riscos sejam coordenadas e abrangentes. A importância desse plano está não apenas em evitar vítimas em situações de desastre, mas também em minimizar os danos materiais e fortalecer a resiliência das comunidades. Ao implementar o PMRR, a Prefeitura visa criar uma estrutura sólida e duradoura que melhore a resposta da cidade frente a desastres naturais, contribuindo para a segurança e a qualidade de vida da população (Brasil, 2023).

Apesar da dificuldade em prever os desastres naturais, a avaliação da vulnerabilidade, mitigação do risco e planos de gerenciamento de emergência podem reduzir os impactos dos eventos e facilitar a recuperação da área atingida (Frigerio et al., 2016).

3.4 ANÁLISE DE RISCO DE DESLIZAMENTO

A análise do risco depende de dois condicionantes: a perigosidade e a vulnerabilidade. A perigosidade indica a potencialidade do terreno de ocorrência de processos destrutivos em uma dada área. Ou seja, expressa a propensão do meio físico para a ocorrência do processo (Dias e Herrmann, 2002). Já a vulnerabilidade é descrita como indicador do grau de perdas sociais, econômicas e ecológicas para um dado elemento, grupo ou comunidade, na área passível de ser afetada por um determinado processo destrutivo (Marandola Jr. e Hogan, 2005).

De acordo com Almeida (2012), o risco é uma medida da ameaça e das consequências que um evento poderá causar num dado intervalo de tempo. Pacheco e Coutinho (2006) afirmam que o risco comumente é definido como o produto entre a probabilidade de um

evento ocorrer e as suas potenciais consequências, onde estas são representadas por quaisquer resultados adversos ou indesejáveis.

O risco pode ser diferenciado ou classificado de diferentes formas ou graus, como categorizados, de forma qualitativa, conforme demonstrado na Tabela 3. A Carta de Riscos serve para delimitar, em uma área, as zonas expostas a um mesmo risco. Esta define os diferentes graus de risco e as medidas a adotar relacionadas com cada um deles.

Tabela 3 - Graus de riscos de deslizamentos

Grau de Risco	Descrição
R1 Baixo	Não há indícios de desenvolvimento de processos destrutivos em encostas e margens de drenagens. Mantidas as condições existentes, não se espera a ocorrência de eventos destrutivos.
R2 Médio	Observa-se a presença de alguma(s) evidência(s) de instabilidade (encostas e margens de drenagens), porém incipiente(s). Mantidas as condições existentes, é reduzida a possibilidade de ocorrência de eventos destrutivos, durante episódios de chuvas intensas e prolongadas.
R3 Alto	Observa-se a presença de significativa(s) evidência(s) de instabilidade (trincas no solo, degraus de abatimento em taludes etc.) Mantidas as condições existentes, é perfeitamente possível a ocorrência de eventos destrutivos, durante episódios de chuvas intensas e prolongadas.
R4 Muito Alto	As evidências de instabilidade são expressivas e estão presentes em grande número e/ou magnitude. Mantidas as condições existentes, é muito provável a ocorrência de eventos destrutivos, durante episódios de chuvas intensas e prolongadas. Exemplos de evidências comuns são: trincas no solo, degraus de abatimento em taludes, trincas em moradias ou em muros de contenção, árvores ou postes inclinados, cicatrizes de escorregamento, feições erosivas, proximidade da moradia em relação ao córrego etc.

Fonte: Almeida, 2012.

3.5 APRENDIZADO DE MÁQUINA

O aprendizado de máquina (do inglês, *machine learning*), é uma área de Inteligência Artificial cujo objetivo é o desenvolvimento de técnicas computacionais sobre o aprendizado bem como a construção de sistemas capazes de adquirir conhecimento de forma automática, podendo aprender com dados a identificar padrões e tomar decisões com o mínimo de

intervenção humana, de modo a automatizar a construção de modelos analíticos (Shalev-Shwartz; Ben-David, 2014).

Um sistema de aprendizado é um programa de computador que toma decisões com base em experiências acumuladas a partir da solução bem-sucedida de problemas anteriores. Segundo Lantz (2019), o processo de aprendizado de máquina pode ser dividido em seis etapas. A primeira etapa consiste na coleta de dados, seguida pela segunda etapa, que envolve a exploração e preparação dos dados coletados. A terceira etapa é o treinamento do modelo, e a quarta etapa refere-se à avaliação de sua performance. Na quinta etapa, podem ser implementadas melhorias no modelo, de acordo com as necessidades específicas de cada problema. Finalmente, a sexta etapa envolve a inferência, onde o modelo realiza previsões para dados desconhecidos, aplicando técnicas de extrapolação e interpolação para prever comportamentos futuros.

Existem três paradigmas principais no aprendizado de máquina: supervisionado, não supervisionado e por reforço (Haykin, 2000). No aprendizado supervisionado, o objetivo é aprender a partir de um mapeamento de entrada, utilizando uma base de dados que fornece os valores corretos das saídas. Assim, o algoritmo aprende com exemplos.

O aprendizado supervisionado é dividido em dois métodos principais: classificação e regressão (Bishop, 2006). A classificação trata de problemas em que os dados de entrada são conhecidos e o algoritmo deve atribuí-los a categorias específicas, ou seja, distribuir esses dados entre uma das classes de saída. Por outro lado, a regressão aborda problemas em que os dados de entrada e saída estão relacionados por uma função contínua, e a saída do modelo é um número (Bishop, 2006).

No aprendizado não supervisionado, não há uma base de dados com "respostas corretas"; o algoritmo trabalha apenas com dados de entrada, buscando identificar padrões e estruturas subjacentes.

Por fim, no aprendizado por reforço, o processo de aprendizado se dá por meio de recompensas recebidas pelo agente. Nesse método, não é necessário um modelo com dados de entrada e saída previamente definidos (Harmendani e Zanetti, 2019). O aprendizado ocorre por meio da exploração de um ambiente desconhecido e da obtenção de recompensas após a execução de ações específicas.

Com o crescente avanço tecnológico e aumento da disponibilidade de grandes volumes de dados em curto tempo em todas as áreas do conhecimento humano, surge o gargalo e a demanda pelo entendimento de algoritmos de aprendizado de máquina (Jordan;

Mitchell, 2015). Uma das técnicas de Aprendizado de Máquina que tem tido sucesso em resolver muitos problemas são as Redes Neurais Artificiais (RNA).

3.5.1 Redes Neurais Artificiais

As RNAs são modelos matemáticos que se inspiram nas estruturas neurais biológicas e que têm a capacidade computacional adquirida por meio de aprendizado (Goodfellow et al., 2016). O primeiro tipo de neurônio utilizado em uma RNA, foi inspirado em um neurônio humano, conhecido como Perceptron, originado por Frank Rosenblatt, em 1957. Ele tem arquitetura simples e se tornou obsoleto para solução de problemas complexos pela sua inflexibilidade e ausência de estabilidade de sua função de ativação, que é a função que analisa se um neurônio pode ser ou não ativado.

O aprendizado da Rede Neural ocorre, de forma resumida, em duas fases: uma fase de treinamento e uma fase de teste do algoritmo. Na fase de treinamento os dados são apresentados ao algoritmo. Os parâmetros da rede são ajustados a cada apresentação de um novo modelo à rede. Depois da adequação dos parâmetros, na fase de teste, o sistema é avaliado (Monard e Baranauskas, 2003).

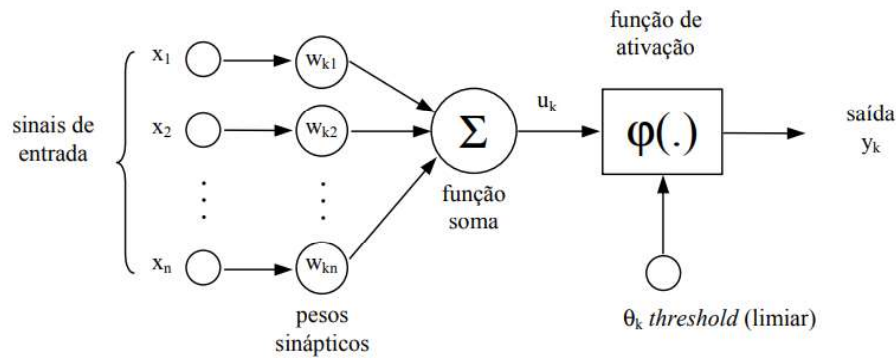
As RNA podem ser empregadas em diversos problemas relacionados às engenharias. Atualmente, o campo de aplicação é muito amplo. Holanda (2022), por exemplo, utilizando RNA identificou o grau de probabilidade e classificou a suscetibilidade de ocorrência de solos colapsíveis e expansivos no Brasil baseadas em variáveis geotécnicas e variáveis pedológicas, geológicas e climatológicas. Magalhães, Ferreira e Soares (2022) buscou construir um método seguro e mais rápido de obtenção do Fator de Segurança (FS) para garantir a estabilidade de taludes, utilizando redes neurais. Sousa, Medeiros e Freire (2024) desenvolveu um modelo para previsão da capacidade de carga em fundações diretas assentadas em diversos tipos de solos, mostrando que os modelos de aprendizado de máquina são capazes de fornecer estimativas precisas e confiáveis, superando as abordagens tradicionais baseadas em métodos empíricos.

3.5.2 Neurônio Artificial e Funções de Ativação

O neurônio artificial é a unidade de processamento de informação fundamental para a operação de uma rede neural que surgiu como uma tentativa de descrever a complexa

estrutura de processamento do neurônio humano. De acordo com Haykin (1998), são identificados cinco elementos básicos do neurônio artificial, são eles conjunto de sinais de entrada e pesos sinápticos, a função soma, função de ativação e saída.

Figura 8 - Modelo não linear de um neurônio



Fonte: Haykin, 1998.

De acordo com Haykin (1998) a resposta y_k de um neurônio artificial k pode ser descrita por meio das equações matemáticas a seguir.

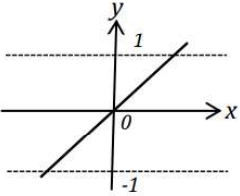
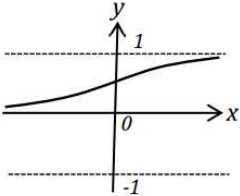
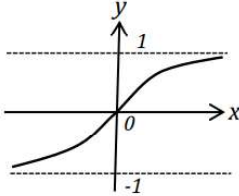
$$y_k = \varphi(u_k + \theta_k) \quad (2.1)$$

$$u_k = \sum_{j=1}^n w_{kj} x_j \quad (2.2)$$

Onde x_j são as variáveis de entrada, w_{kj} são os pesos, θ_k é o limiar e φ é a função de ativação.

Existem vários tipos de função de ativação as quais produzem diferentes saídas e devem ser escolhidas de acordo com o tipo de aplicação, tais funções determinam o nível de transformação de um neurônio perante as suas entradas (Kovács, 2002). Dentre as quais se destacam a linear, logística (ou sigmoide) e tangente hiperbólica, conforme demonstrado na Tabela 4.

Tabela 4 - Funções de ativação usuais

Linear	Logística	Tangente Hiperbólica
$y = a.x$	$y = \frac{1}{1 + e^{-\gamma.x}}$	$y = \frac{1 - e^{-\gamma.x}}{1 + e^{-\gamma.x}}$
$[-\infty, \infty]$	$[0, 1]$	$[-1, 1]$
		

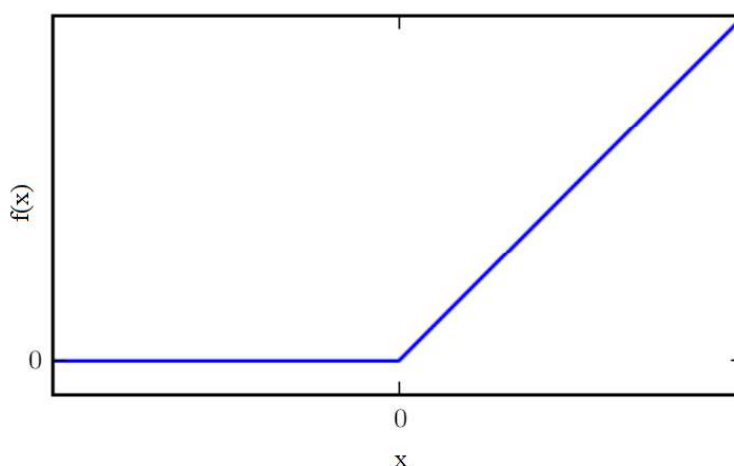
Fonte: Kovács, 2002.

A função de ativação ReLU (unidade linear retificada), introduzida por Hahnloser *et al.* em 2000, é amplamente utilizada em redes neurais artificiais devido às suas características vantajosas. Conforme destacado por Bai (2022), a função ReLU apresenta boa esparsidade, rápida convergência e simplicidade de cálculo, tornando-se uma escolha eficiente para o treinamento de RNAs. Ela é definida pela seguinte fórmula:

$$f(x) = \max\{0, x\} \quad (2.3)$$

Para essa função, qualquer entrada negativa é convertida para zero, enquanto entradas positivas são mantidas inalteradas. Tal característica permite que a ReLU introduza não linearidade no modelo, que segundo Kulathunga *et al.* (2020) é um requisito essencial para a capacidade das redes neurais de aprender e representar relações complexas nos dados.

Figura 9 - Função de ativação ReLU



Fonte: Goodfellow *et al.*, 2016.

Conforme destacado por Goodfellow *et al.* (2016), a função de ativação ReLU é recomendada como a função padrão para redes neurais de propagação direta. Ele explica que aplicar essa função à saída de uma transformação linear resulta em uma transformação não linear em uma rede multicamadas. No entanto, a ReLU permanece próxima ao comportamento linear, pois é uma função linear por partes com duas peças lineares. Esta proximidade à linearidade permite que as unidades ReLU preservem muitas das propriedades que tornam os modelos lineares fáceis de otimizar com métodos baseados em gradiente descendente.

3.5.3 Perceptron Multicamadas

Em geral, as arquiteturas de redes neurais podem ser classificadas em três tipos fundamentais com base no número de camadas, descritas a seguir.

As redes unidirecionais de camada única, todas as conexões entre neurônios seguem uma trajetória de entrada para saída. Essa configuração representa a forma mais básica de uma rede neural, consistindo em uma camada de entrada que se conecta diretamente a uma única camada de saída. Esse tipo de rede também é conhecido como rede feedforward, acíclica ou de camada simples (Haykin, 1998).

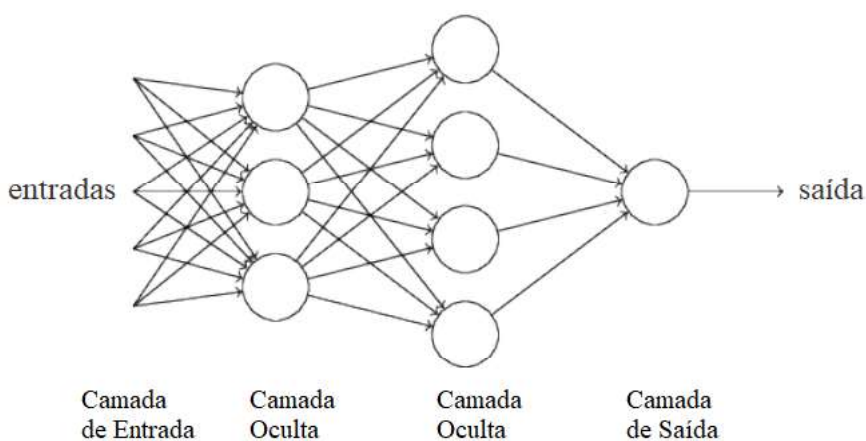
As redes unidirecionais, conhecidas como redes de múltiplas camadas, possuem mais de uma camada de neurônios. Um exemplo são as Redes Perceptron Multicamadas (MLP), que incluem pelo menos uma camada intermediária de neurônios. Nessas redes, as camadas ocultas processam informações entre a camada de entrada e a camada de saída (Haykin,

2000). Os sinais de entrada são apresentados à primeira camada oculta, cujas saídas são usadas como entrada para a próxima camada, continuando até a camada final.

Por fim, as redes recorrentes se distinguem pela presença de realimentações em seu funcionamento, o que afeta significativamente sua capacidade de aprendizado e desempenho (Haykin, 2000). Nesses modelos, a camada de entrada pode enviar sinais de volta para as entradas de outros neurônios, incluindo a auto-realimentação.

Um Perceptron Multicamadas (MLP) é uma arquitetura de rede neural artificial composta por várias camadas de neurônios interconectados. De acordo com Nielsen (2019), um MLP é estruturado em três tipos principais de camadas, a camada de entrada, responsável pela recepção dos dados brutos, camadas ocultas, que processam esses dados por meio de transformações não lineares, e a camada de saída, que gera as previsões finais do modelo.

Figura 10 - RNA perceptron multicamadas com apenas uma camada oculta



Fonte: Adaptado de Nielsen, 2019.

Cada neurônio em uma camada oculta calcula uma combinação linear dos valores de entrada, aplicando em seguida uma função de ativação para introduzir não linearidade (Izbicki e Santos, 2020). Essas funções permitem que o MLP aprenda e represente relações complexas e não lineares entre os dados de entrada.

Cada neurônio em uma camada oculta calcula uma combinação linear dos valores de entrada, ponderada por pesos sinápticos e acrescida de um valor de viés (bias), que é uma constante adicionada ao cálculo para garantir que a rede possa deslocar funções de ativação, permitindo uma melhor adaptação às particularidades dos dados. Após essa combinação linear e adição do viés, aplica-se uma função de ativação para introduzir a não linearidade

necessária ao modelo (Izbicki e Santos, 2020). Essas funções permitem que o MLP aprenda e represente relações complexas e não lineares entre os dados de entrada.

As camadas ocultas desempenham um papel fundamental no aprendizado da rede, pois são responsáveis por processar as informações recebidas da camada anterior e realizar transformações nos dados que ajudam a identificar padrões complexos. Ao passar por múltiplas camadas ocultas, a rede consegue extrair e combinar características cada vez mais abstratas dos dados, o que amplia sua capacidade de generalização e de captura de relações não triviais. Assim, o MLP se torna capaz de realizar tarefas como classificação e regressão de maneira eficaz, modelando padrões que seriam difíceis de detectar por algoritmos lineares simples.

3.5.4 Treino da Rede Neural Artificial

O treinamento de uma rede neural artificial do tipo Perceptron Multicamadas (MLP) pode ser dividido em cinco etapas principais: Propagação Direta, Cálculo do Erro, Retropropagação do Erro, Atualização dos Pesos e Épocas. Cada uma dessas etapas desempenha um papel crucial para garantir que a rede neural aprenda de maneira eficaz a partir dos dados de treino. Nas próximas subseções, cada uma dessas etapas será explicada de forma a proporcionar a compreensão do processo de treinamento, em conformidade com as explicações fornecidas por Aggarwal *et al.* (2018).

3.5.4.1 Divisão dos Dados em Treino, Validação e Teste

De acordo com Santos *et al.* (2019), a divisão da amostra em dados de treinamento e de teste é realizada para verificar se um modelo apresenta boa performance não apenas em dados utilizados para seu ajuste, mas também capacidade de generalização para novas observações. O conjunto de treino, geralmente composto por 70% a 80% dos dados, é utilizado para ajustar os parâmetros do modelo, enquanto o conjunto de teste, constituído pelos 20% a 30% restantes, serve para avaliar o desempenho do modelo em dados não vistos, proporcionando uma medida objetiva de sua capacidade de generalização. De acordo com Bishop (2006), existem algumas regras práticas para essa divisão, em casos com grandes volumes de dados, pode-se reservar uma menor porcentagem para o teste, já em situações

com poucos dados, técnicas como a validação cruzada, onde o conjunto de dados é dividido em múltiplos subconjuntos, são recomendadas para garantir uma avaliação mais robusta.

Além dos dois mencionados anteriormente, o conjunto de validação é frequentemente utilizado durante o treinamento para ajustar hiperparâmetros e tomar decisões sobre a configuração do modelo. Em situações em que o conjunto de dados não é grande o suficiente para ser dividido em três partes, técnicas de reamostragem podem ser utilizadas para aproximar o conjunto de validação por meio da reutilização de observações do conjunto de treinamento (Hastie et al, 2008). Ou seja, os dados de validação são extraídos como um subconjunto separado dos dados de treino, permitindo uma avaliação intermediária do modelo durante o treinamento. A avaliação final do desempenho do modelo é realizada utilizando o conjunto de teste, que contém dados completamente novos e não vistos durante o treinamento e a validação.

Figura 11 - Divisão do conjunto de dados em conjuntos de treino, validação e teste

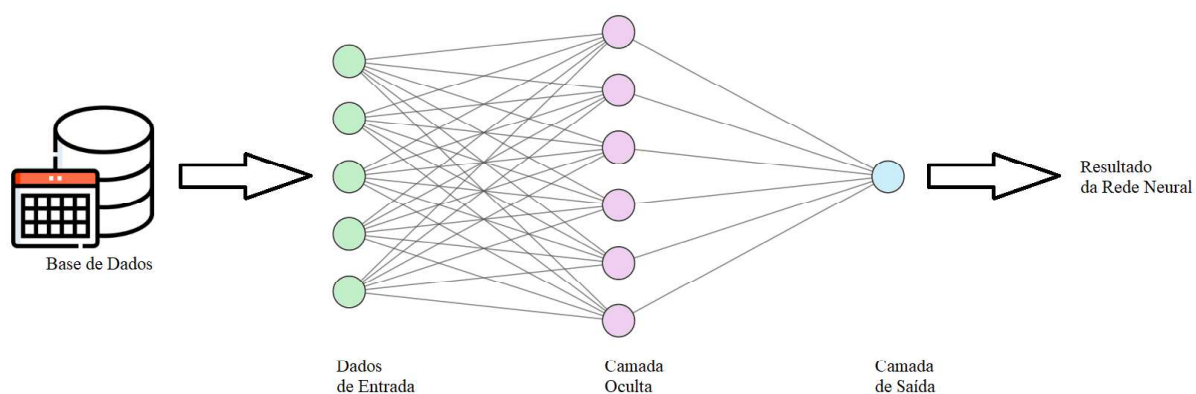


Fonte: Autor.

3.5.4.2 Propagação Direta

Durante a propagação direta de uma RNA feedforward, os dados de entrada são passados através da rede neural de forma sequencial, camada por camada, até alcançar a camada de saída. Em cada camada, cada neurônio recebe os valores de entrada, que são multiplicados pelos respectivos pesos sinápticos. A seguir, esses produtos ponderados são somados e, geralmente, um viés é adicionado para formar uma combinação linear. Essa combinação linear é então transformada por uma função de ativação, que introduz a não linearidade necessária para que a rede possa modelar relações complexas. O resultado desse processo são as ativações dos neurônios, que servem como entradas para a próxima camada, continuando assim até a camada de saída, onde a rede produz a predição final.

Figura 12 - Propagação direta



Fonte: Autor.

3.5.4.3 Cálculo do Erro

Após a propagação direta, o erro entre a predição da rede neural e o valor real é calculado utilizando uma função de custo apropriada. Para problemas de classificação, uma função comum é a entropia cruzada, enquanto para problemas de regressão, o erro quadrático médio é frequentemente utilizado. Este cálculo do erro considera os pesos sinápticos em todas as camadas, pois eles influenciam diretamente as predições da rede. A função de custo quantifica a discrepância entre a saída prevista pela rede e os valores reais, fornecendo uma medida de desempenho que será utilizada para ajustar os pesos sinápticos durante o processo de treinamento, minimizando o erro nas predições futuras.

Figura 13 - Cálculo do erro



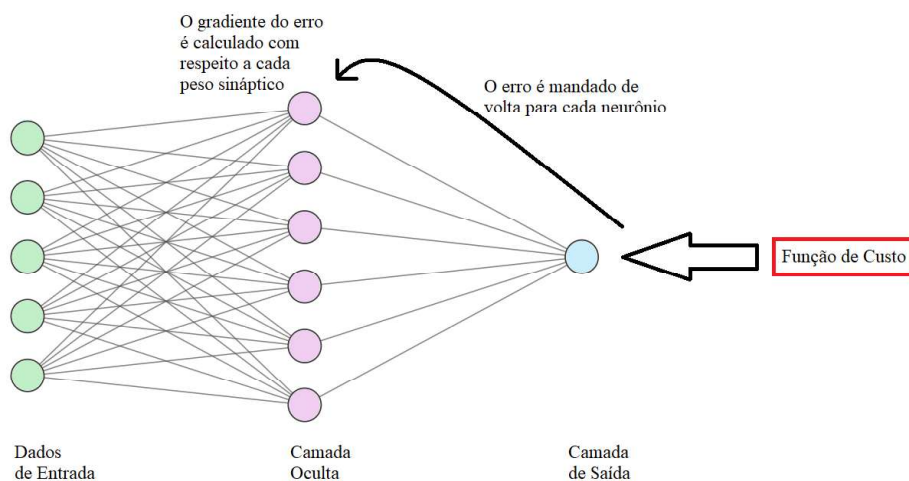
Fonte: Autor.

3.5.4.4 Retropropagação do Erro

A retropropagação é um algoritmo essencial no treinamento de redes neurais artificiais, utilizado para ajustar os pesos sinápticos com o objetivo de minimizar o erro entre as previsões do modelo e os valores reais. Esse ajuste é realizado após o cálculo do erro na fase de propagação direta (forward propagation), que é seguido pelo cálculo dos gradientes durante o processo de retropropagação. Esses gradientes indicam o quanto cada peso influencia o erro total, permitindo que sejam atualizados de forma a reduzir a função de custo da rede.

O processo de retropropagação envolve o cálculo das derivadas parciais da função de custo em relação a cada peso da rede, usando a regra da cadeia. Inicialmente, a derivada da função de custo é calculada em relação à saída da rede, e posteriormente, a função de ativação é derivada para capturar a influência das não linearidades no modelo. A cada peso sináptico é associado um gradiente que, quando calculado, é utilizado para ajustar o peso por meio de algoritmos de otimização como o gradiente descendente. Essa atualização é feita de forma iterativa, refinando os pesos da rede até que o erro seja minimizado. O processo garante que a rede aprenda padrões complexos, melhorando sua capacidade de generalização para novos dados.

Figura 14 - Retropropagação



Fonte: Autor.

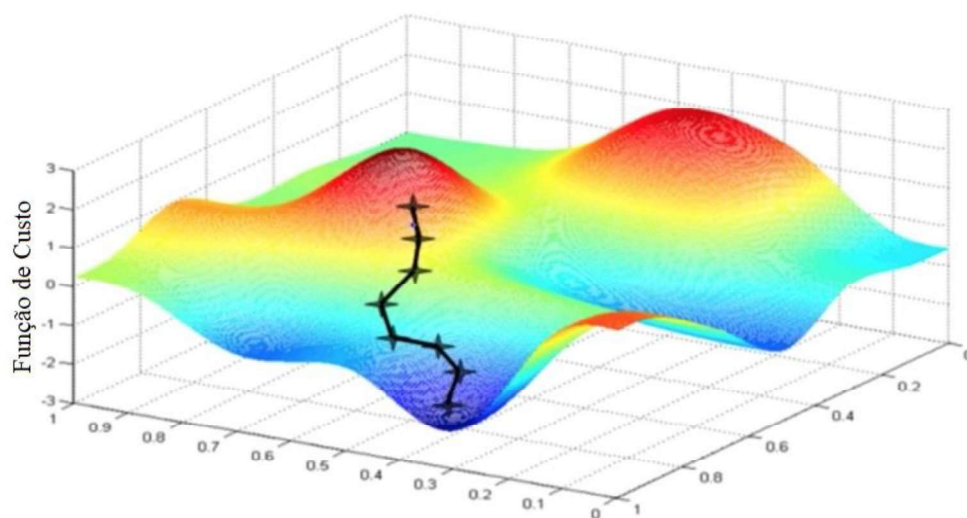
3.5.4.5 Atualização dos Pesos

O ajuste dos pesos sinápticos é essencial para o aprendizado da rede, pois ele refina os parâmetros de modo a melhorar a precisão das predições futuras. Esse ajuste é controlado por hiperparâmetros, que são variáveis definidas antes do treinamento e influenciam diretamente o comportamento do modelo durante o aprendizado. Hiperparâmetros como a taxa de aprendizado, o momentum e o número de épocas são fundamentais para determinar a eficácia e a velocidade do ajuste dos pesos.

Após o cálculo dos gradientes na etapa de retropropagação, um otimizador é aplicado para ajustar os pesos da rede neural minimizando a função de custo. Entre os otimizadores, existem vários, como o gradiente descendente, o Adam (Adaptive Moment Estimation), o RMSProp, o Adagrad e o SGD (Stochastic Gradient Descent). Cada um possui características distintas e seus próprios hiperparâmetros, o que os torna adequados para diferentes tipos de problemas.

O Adam é amplamente utilizado em redes neurais profundas e problemas que envolvem grandes volumes de dados, devido à sua capacidade de ajustar a taxa de aprendizado de forma adaptativa. O SGD, por outro lado, é mais eficiente em problemas onde a generalização é importante, como em redes menos complexas ou com conjuntos de dados menores, já que ele realiza atualizações frequentes com base em subconjuntos aleatórios dos dados. Assim, a escolha do otimizador depende do tipo de problema, da quantidade de dados disponíveis e da arquitetura da rede neural.

Figura 15 - Otimizador iterando até o ponto de mínimo local da função de custo



Fonte: NG (2023).

3.5.4.6 Épocas

Esse processo de propagação direta, cálculo do erro, retropropagação e atualização dos pesos é repetido iterativamente por várias épocas que são iterações completas sobre os dados de treinamento. Cada iteração ajuda a reduzir o erro de predição e ajustar os pesos da rede para melhorar o desempenho do modelo.

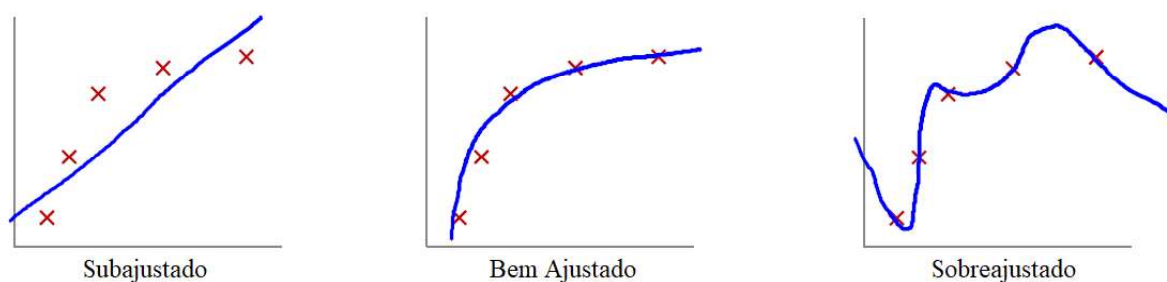
3.5.5 O Problema de Subajuste (*underfit*) e Sobreajuste (*overfit*)

O problema do subajuste ocorre quando um modelo de aprendizado de máquina é demasiado simples para capturar os padrões subjacentes nos dados. Isso geralmente acontece quando o modelo tem uma capacidade muito baixa, como um número insuficiente de parâmetros ou uma estrutura inadequada para a complexidade do problema (Aalst *et al.*, 2010). Consequentemente, o modelo apresenta um desempenho ruim tanto nos dados de treinamento quanto nos dados de teste, resultando em alta variância e baixo viés. Para resolver o subajuste, é possível aumentar a complexidade do modelo, adicionar mais características aos dados ou ajustar melhor os hiperparâmetros para que o modelo possa aprender os padrões mais profundos presentes nos dados.

Por outro lado, o sobre ajuste acontece quando o modelo é excessivamente complexo e se ajusta demasiado aos dados de treinamento, capturando não apenas os padrões subjacentes, mas também o ruído e as peculiaridades dos dados de treinamento (Zhang *et al.*, 2019). Como resultado, o modelo apresenta um excelente desempenho nos dados de treinamento, mas falha em generalizar bem para novos dados, levando a um alto viés e baixa variância.

A seguir um exemplo simples de um modelo de regressão que mostra de forma assertiva quando um modelo está subajustado, bem ajustado e sobreajustado.

Figura 16 - Modelos e possíveis ajustes



Fonte: NG (2023).

3.5.6 Fluxo de Dados

Um fluxo de dados (ou *pipeline*) em aprendizado de máquina é uma sequência estruturada de etapas de processamento que automatiza o fluxo de trabalho desde a preparação dos dados até a modelagem e a avaliação. Segundo Oliveira (2020), essa prática é importante para padronizar a limpeza e preparação dos dados garantindo que o algoritmo de aprendizado receba um pacote de dados adequado, independente da origem ou conteúdo numérico presente no conjunto de dados.

Cada etapa no fluxo de dados pode incluir processos como a limpeza e a transformação dos dados, a seleção de características relevantes, a divisão dos dados em conjuntos de treino e teste, a aplicação de algoritmos de aprendizado de máquina e a validação do modelo. Ao encapsular essas etapas em um pipeline, facilita-se a reprodução e a manutenção do fluxo de trabalho, garantindo que cada passo seja executado de maneira consistente e ordenada.

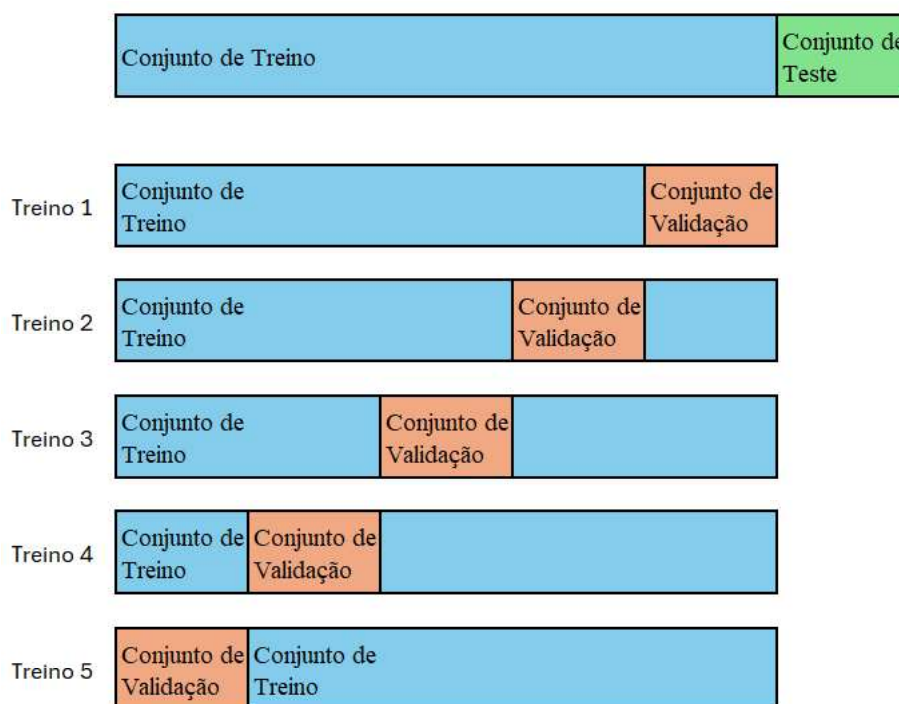
3.5.7 Validação Cruzada

A validação cruzada é uma técnica amplamente utilizada em aprendizado de máquina para avaliar a capacidade de generalização de um modelo de maneira mais robusta e confiável. Em vez de simplesmente dividir os dados em um conjunto de treino e um conjunto de teste, a validação cruzada envolve a divisão dos dados em múltiplos subconjuntos ou dobras. A técnica mais comum, conhecida como validação cruzada k-dobras, divide o conjunto de dados em k partes iguais. O modelo é então treinado k vezes, cada vez utilizando

k-1 dobras para treino e a dobra restante para validação. No presente estudo essa técnica foi utilizada por conta da baixa quantidade de dados, totalizando 442 dados.

Segundo Kuhn *et al.* (2013), toda essa repetição de treino e validação são utilizadas para aumentar a precisão das estimativas, de modo que, a cada repetição, diferentes partições do conjunto de treinamento são consideradas para compor cada uma das k partes do processo de validação cruzada. Ao fim de todo o processo a média dos resultados de todas as iterações é calculada para fornecer uma estimativa mais precisa da performance do modelo.

Figura 17 - Exemplo de divisão dos dados em 5 dobras para a validação cruzada



Fonte: Autor.

3.5.8 Hiperparâmetros de uma Rede Neural

Os hiperparâmetros de uma rede neural são parâmetros definidos antes do treinamento e que controlam aspectos essenciais do modelo e do aprendizado (Diaz *et al.*, 2017). Entre os principais hiperparâmetros da rede neural estão o número de camadas ocultas, o número de neurônios em cada camada, o tipo de função de ativação utilizada, número de épocas. Além desses, os hiperparâmetros do otimizador também desempenham um papel crucial no treinamento do modelo.

Em Goodfellow *et al.* (2016) é visto que os otimizadores como o gradiente descendente, a taxa de aprendizado é um hiperparâmetro fundamental que determina o tamanho dos passos dados na direção do gradiente, já em algoritmos mais avançados, como o Adam, os hiperparâmetros incluem a taxa de aprendizado, bem como os parâmetros beta que controlam as taxas de decaimento das médias móveis dos gradientes e dos quadrados dos gradientes. A escolha adequada desses hiperparâmetros pode significativamente afetar a performance e a eficiência do modelo, exigindo frequentemente uma combinação de experiência e experimentação para otimizar o desempenho da rede neural.

3.5.9 Busca em Grade

A busca em grade (ou *grid search*) é uma técnica de otimização amplamente utilizada para encontrar a melhor combinação de hiperparâmetros em modelos de aprendizado de máquina. Esse método consiste em especificar um conjunto de valores possíveis para cada hiperparâmetro e, em seguida, realizar uma busca exaustiva por todas as combinações possíveis desses valores (Lopes *et al.*, 2023). A busca em grade avalia o desempenho do modelo para cada combinação, geralmente utilizando validação cruzada para garantir que a avaliação seja robusta e generalizável.

Apesar de sua simplicidade e eficácia na identificação de combinações ótimas, o *Grid Search* pode ser computacionalmente caro, especialmente quando o espaço dos hiperparâmetros é grande ou o modelo é complexo (Mendes *et al.*, 2023). No entanto, sua abordagem sistemática e abrangente proporciona uma maneira confiável de encontrar as melhores configurações para maximizar o desempenho do modelo.

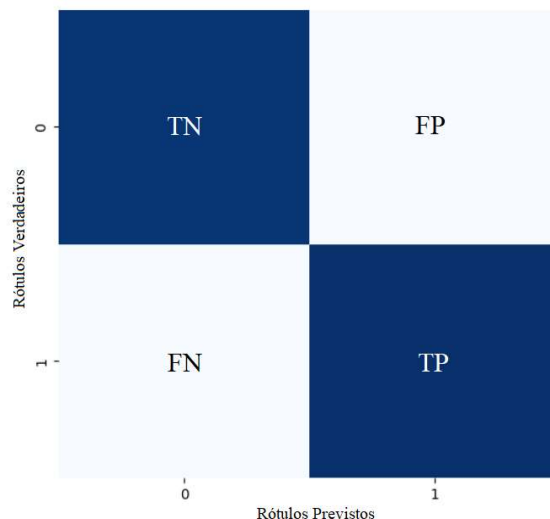
3.5.10 Validação dos Resultados

3.5.10.1 Matriz de Confusão

A matriz de confusão é uma ferramenta crucial na avaliação de modelos de classificação, proporcionando uma visão detalhada da performance do modelo em termos de erros e acertos. Segundo Géron (2022), ela mostra a distribuição das previsões feitas pelo modelo em relação aos valores reais, organizando os resultados em quatro categorias

principais: verdadeiros positivos (VP), falsos positivos (FP), verdadeiros negativos (VN) e falsos negativos (FN).

Figura 18 - Matriz de confusão



Fonte: Autor.

Os verdadeiros positivos são as instâncias corretamente classificadas como pertencentes à classe positiva, enquanto os falsos positivos são aqueles incorretamente classificadas como positivas. Os verdadeiros negativos referem-se às instâncias corretamente identificadas como negativas, e os falsos negativos são as instâncias que foram erroneamente classificadas como negativas, embora pertençam à classe positiva.

Em um contexto de estabilidade de taludes, os verdadeiros positivos seriam as áreas corretamente identificadas como suscetíveis a deslizamentos, resultando em avisos preventivos adequados. Por outro lado, os falsos positivos representam áreas erroneamente classificadas como de risco, levando a evacuações desnecessárias, o que pode gerar desconfiança no modelo. Os verdadeiros negativos são as áreas corretamente identificadas como estáveis, onde não há risco de deslizamento. O maior prejuízo está nos falsos negativos, quando áreas de risco são incorretamente classificadas como seguras, não gerando alertas. Nesse caso, o deslizamento ocorre sem aviso prévio, levando a possíveis perdas de vidas e danos materiais significativos.

De acordo com Fawcett (2006), a matriz de confusão permite calcular diversas métricas de desempenho, como a acurácia, precisão, revocação (ou recall) e F1-score, que são fundamentais para compreender a eficácia do modelo. Ela oferece uma visão clara dos tipos

de erros cometidos pelo modelo e ajuda na identificação de padrões de falhas que podem ser corrigidos ou melhorados. A seguir as fórmulas dessas métricas.

$$Acurácia = \frac{VP+VN}{VP+FP+VN+FN} \quad (2.4)$$

$$Precisão = \frac{VP}{VP+FP} \quad (2.5)$$

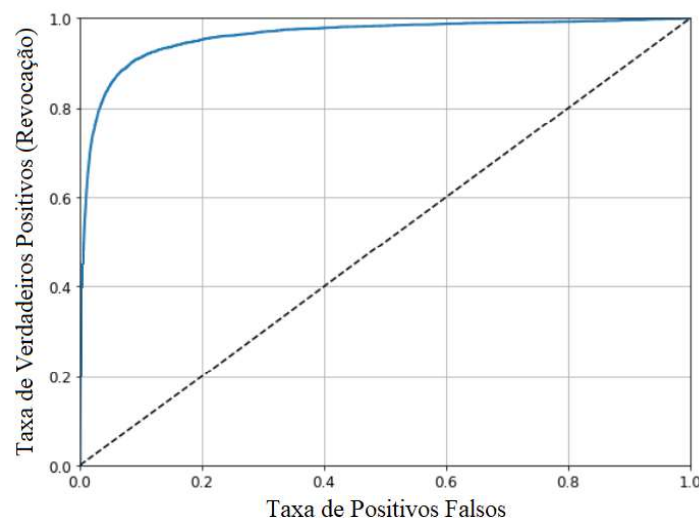
$$Revocação = \frac{VP}{VP+FN} \quad (2.6)$$

$$F1\ Score = 2 * \frac{Precisão * Revocação}{Precisão + Revocação} \quad (2.7)$$

3.5.10.2 Curva ROC

A curva ROC (Receiver Operating Characteristic) é uma ferramenta gráfica utilizada para avaliar o desempenho de modelos de classificação, especialmente em problemas de classificação binária. Ela ilustra a capacidade do modelo de distinguir entre as classes positiva e negativa, variando o limiar de decisão ao longo de todo o intervalo de possíveis valores de classificação (Géron, 2022). A curva é construída ao plotar a taxa de verdadeiros positivos (revocação) contra a taxa de falsos positivos para cada limiar. A taxa de verdadeiros positivos, também conhecida como revocação, representa a proporção de casos positivos corretamente identificados, enquanto a taxa de falsos positivos indica a proporção de casos negativos incorretamente classificados como positivos.

Figura 19 - Curva ROC



Fonte: Autor.

Para construir a curva ROC, o modelo é avaliado em vários limiares de decisão, que determinam quais previsões são consideradas positivas ou negativas. A curva resultante mostra como o equilíbrio entre a taxa de verdadeiros positivos e a taxa de falsos positivos muda com diferentes limiares.

A área sob a curva ROC (AUC) é uma métrica importante derivada dessa análise, que quantifica a capacidade geral do modelo de separar as classes. Um valor de AUC próximo a 1 indica um modelo com excelente capacidade de discriminação, enquanto um valor próximo a 0.5 sugere um desempenho equivalente ao acaso. A curva ROC é especialmente útil para comparar o desempenho de diferentes modelos e selecionar o mais adequado, considerando a taxa de falsos positivos e a taxa de verdadeiros positivos em diversos cenários (Bradley, 1997).

3.5.11 Desafios da Inteligência Artificial

Parte do sucesso da inteligência artificial é oriunda da quantidade de dados disponíveis atualmente e do poder computacional das máquinas modernas. Segundo Coda *et al.* (2020), em relação à quantidade e complexidade dos dados, as máquinas precisam de uma quantidade muito maior de exemplos que os humanos para aprender. A quantidade de exemplos necessários para que uma máquina aprenda uma tarefa complexa varia conforme a natureza do problema e a qualidade das variáveis envolvidas.

Além da necessidade de muitos exemplos no processo de treinamento, alguns exemplos das tarefas reais podem ser limitados e custosos, já que muitos conceitos e conhecimentos sobre o mundo são fornecidos por seres humanos, que rotulam os dados de treinamento. Esse processo de rotulagem pode introduzir incertezas e subjetividade, influenciando a qualidade do aprendizado dos algoritmos de máquina.

Dessa maneira, os conjuntos de treinamento fornecidos aos algoritmos podem não ser uma boa representação do mundo real, e as amostras podem estar enviesadas. Em termos estatísticos, um conjunto de dados é considerado enviesado quando certas características ou classes são desproporcionalmente representadas em relação à população real. Esse viés pode levar o modelo a generalizar mal, pois ele aprende a partir de dados que não refletem corretamente o comportamento ou distribuição do mundo real.

Por exemplo, um sistema pode ser desenvolvido para distinguir cães de gatos, mas no conjunto de treinamento apresentado ao algoritmo, todos os cães são marrons e todos os gatos são amarelos, então o sistema pode aprender a distinguir marrons de amarelos e não cães de

gatos. Nesses casos, é necessária a experiência dos humanos para adicionar mais variáveis para que a IA execute de forma mais assertiva em solucionar o problema (Kaufman, 2021).

Outro aspecto relacionado aos dados é que muitos problemas reais são dinâmicos e a geração de dados ocorre continuamente, como por exemplo no gerenciamento de transportes e monitoramento por redes de sensores. Sendo assim, são necessárias técnicas de aprendizado em ambientes dinâmicos de fluxos contínuos de dados (Gama, 2012).

Em resumo, a resolução de problemas complexos utilizando técnicas de AM necessita de um design eficiente e correto do sistema. Para isso, são necessárias escolhas eficientes de pré-processamento dos dados, seleção de algoritmos apropriados, escolha dos parâmetros dos algoritmos, seleção de atributos e pós-processamento, de modo que possibilite maior aplicabilidade das técnicas de aprendizagem de máquina.

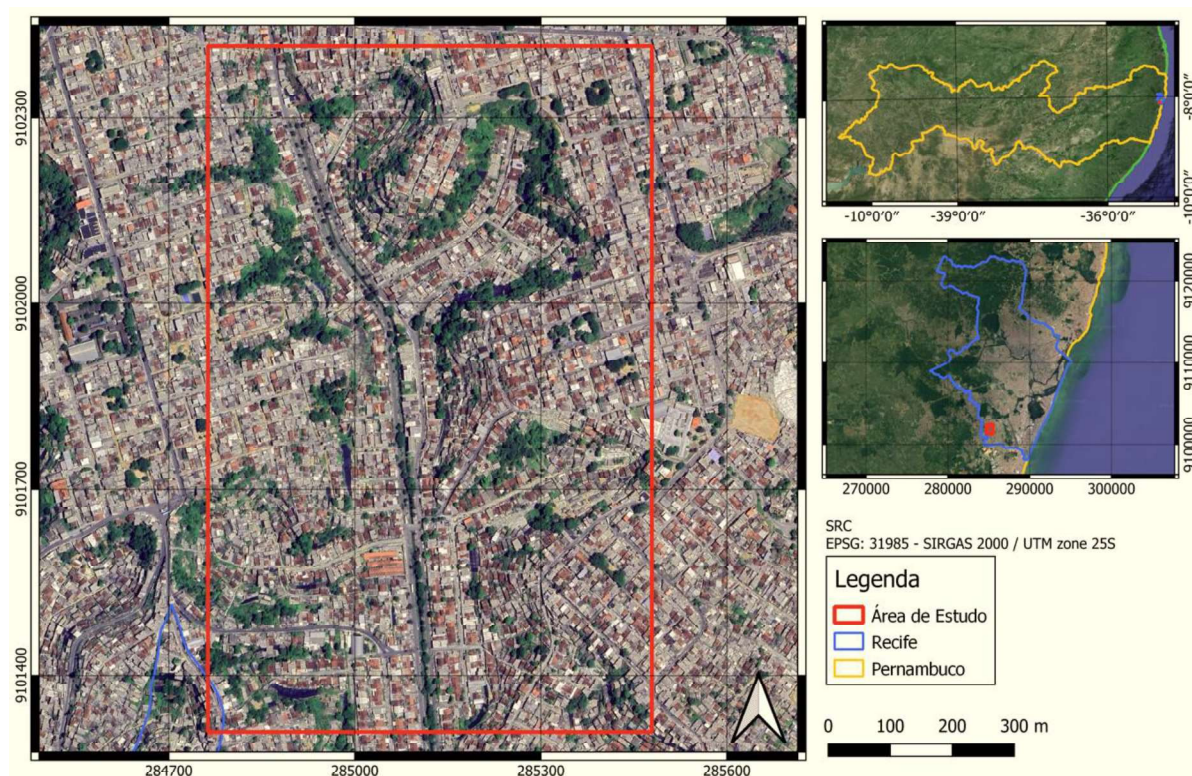
4 MATERIAIS E MÉTODOS

Nos itens a seguir serão descritas as atividades que foram desenvolvidas para obtenção do mapa de susceptibilidade a deslizamentos de terra de Lagoa Encantada.

4.1 ÁREA DE ESTUDO

A comunidade de Lagoa Encantada, em Recife, Pernambuco, está situado em uma região de transição entre o bioma da Mata Atlântica e a vegetação de manguezal característica do litoral. É possível encontrar uma diversidade de vegetação na área. A região apresenta terrenos de altitude de mais de 70 metros acima do nível do mar. De acordo com EMBRAPA (2021), o clima predominante em Pernambuco é o clima tropical úmido, do tipo 'Ams', conforme a classificação de Köppen, o que significa que a região apresenta temperaturas elevadas ao longo do ano, com médias acima de 24 °C, e uma distribuição de chuvas bem definida. A Figura 20 apresenta a localização da área de estudo, que corresponde à comunidade de Lagoa Encantada em Recife – PE.

Figura 20 - Mapa de localização da área de estudo



Fonte: Autor.

No Nordeste do Brasil, os meses mais chuvosos ocorrem geralmente entre abril e julho, correspondendo ao período de inverno. Nesse período, a região é influenciada pela Zona de Convergência Intertropical (ZCIT), que favorece a formação de áreas de instabilidade e chuvas mais intensas. É importante mencionar que a proximidade com o oceano Atlântico também influencia o clima da Lagoa Encantada. Além disso, é relevante destacar que a região está sujeita a eventos climáticos como frentes frias e tempestades, que podem trazer chuvas mais intensas e ventos fortes.

Segundo Cavalcanti et al. (2010), a formação do solo e a geomorfologia de Lagoa Encantada são complexas e influenciadas por vários fatores. O local tem predominância da Formação Barreiras, composta por camadas arenosas na base, que se intercalam com camadas argilosas no topo além de ação antrópica em tabuleiros instáveis e morros com alta declividade (Coutinho *et al.*, 2023).

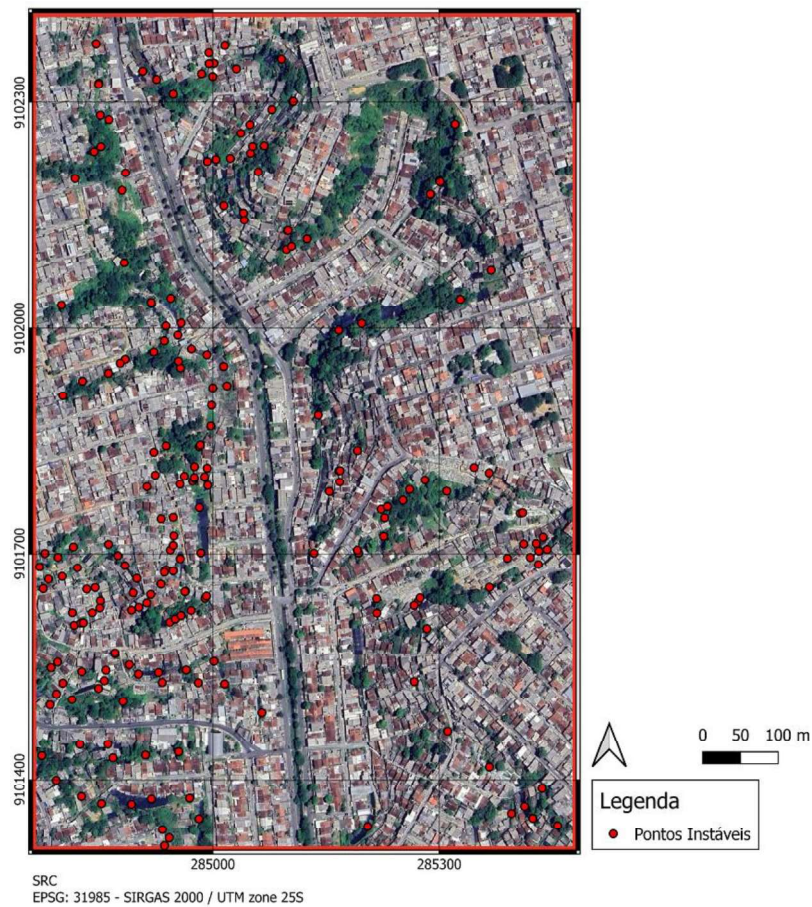
Os movimentos de massa na região ocorrem devido à falta de impermeabilização do terreno, desmatamento, lançamento e concentração de resíduos sólidos e líquidos, além de cortes nos taludes. Em termos de ocorrências de deslizamentos, Melo et al. (2020) afirma que, entre os anos de 2013 e 2017, a Defesa Civil registrou 2.141 chamados relacionados a deslizamentos de terra na comunidade. Essas características geomorfológicas e climáticas moldam a paisagem e afetam a biodiversidade, a infraestrutura e as atividades na região de estudo.

4.2 OBTENÇÃO DOS DADOS

4.2.1 Pontos Instáveis

A base de dados utilizada para identificar pontos instáveis foi a mesma empregada no estudo conduzido por Melo (2021). O georreferenciamento foi realizado a partir do endereço, complementado pela confirmação obtida através da fotointerpretação de imagens de satélite, que permitiu identificar as cicatrizes resultantes dos movimentos de massa. Esse conjunto de dados contempla a localização específica de 221 pontos de instabilidade. A seguir, na Figura 21, é apresentada a distribuição geográfica desses pontos críticos, que servem como referência fundamental para a análise da suscetibilidade a deslizamentos de terra na comunidade de Lagoa Encantada.

Figura 21 - Localização dos pontos instáveis



Fonte: Autor.

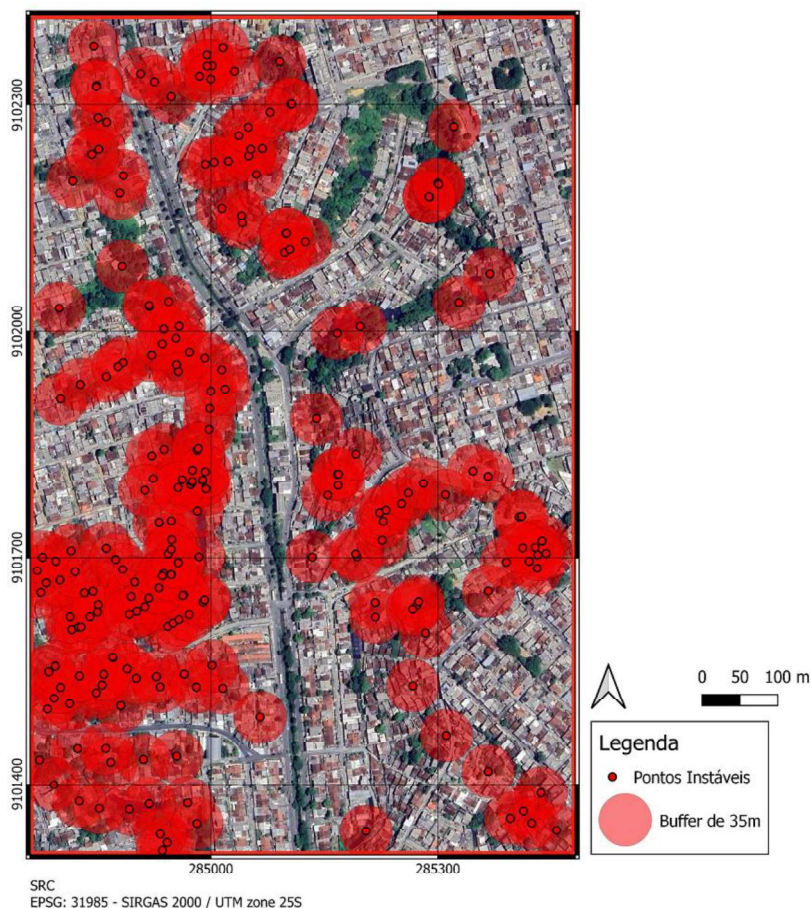
4.2.2 Pontos Estáveis

Para realizar a avaliação de suscetibilidade utilizando técnicas de Aprendizado de Máquina, é indispensável contar também com informações referentes a pontos estáveis. Dessa forma, o modelo de Inteligência Artificial pode diferenciar com mais assertividade a classe em que cada célula do mapa pertence, permitindo uma análise mais robusta e acurada da suscetibilidade a deslizamentos de terra.

Neste estudo, foi estabelecido o critério de 35 metros para classificar uma célula como um potencial ponto estável. Essa distância foi escolhida com base em considerações práticas e análises visuais. Se uma distância muito pequena fosse utilizada, os pontos estáveis poderiam ficar excessivamente próximos dos pontos instáveis, comprometendo a eficácia da classificação. Por outro lado, uma distância excessivamente grande poderia resultar em pontos estáveis muito próximos e concentrados em áreas planas, onde não há risco de deslizamento. Durante os testes, foi observado que 35 metros representava um equilíbrio

adequado. Essa distância assegura que os pontos estáveis estejam afastados de áreas potencialmente perigosas, ao mesmo tempo em que aproveitem a geomorfologia do terreno, evitando que fiquem restritos apenas a locais planos. Com base nesse critério, foi gerada uma área de influência (buffer) ao redor de cada ponto instável, delimitada pela distância especificada, como demonstrado na Figura 22.

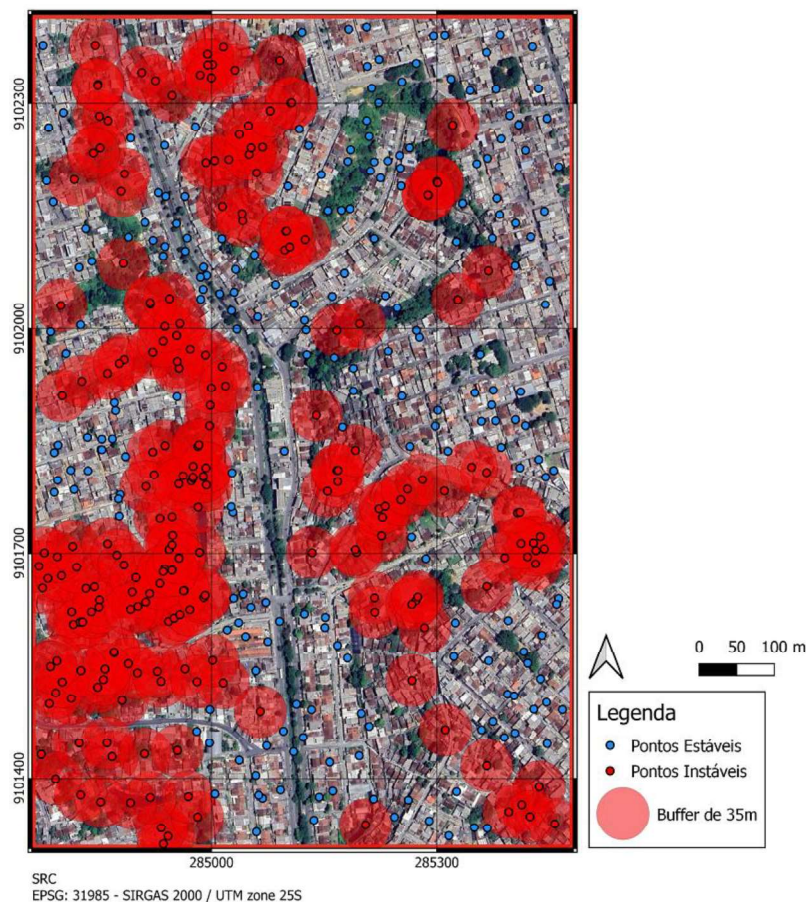
Figura 22 - Área de influência dos pontos instáveis



Fonte: Autor.

Para a seleção dos pontos aleatórios estáveis, foi criteriosamente considerada a exclusão de células localizadas dentro da área de influência dos pontos instáveis. Esse processo assegurou que os 221 pontos aleatórios gerados estivessem estritamente localizados fora das zonas previamente identificadas como instáveis. Ao adotar essa abordagem, priorizou-se a integridade e a confiabilidade dos futuros resultados, garantindo que os pontos selecionados estivessem em áreas consideradas seguras e estáveis para fins de análise geotécnica, como demonstrado na Figura 23.

Figura 23 - Pontos estáveis fora da área de influência



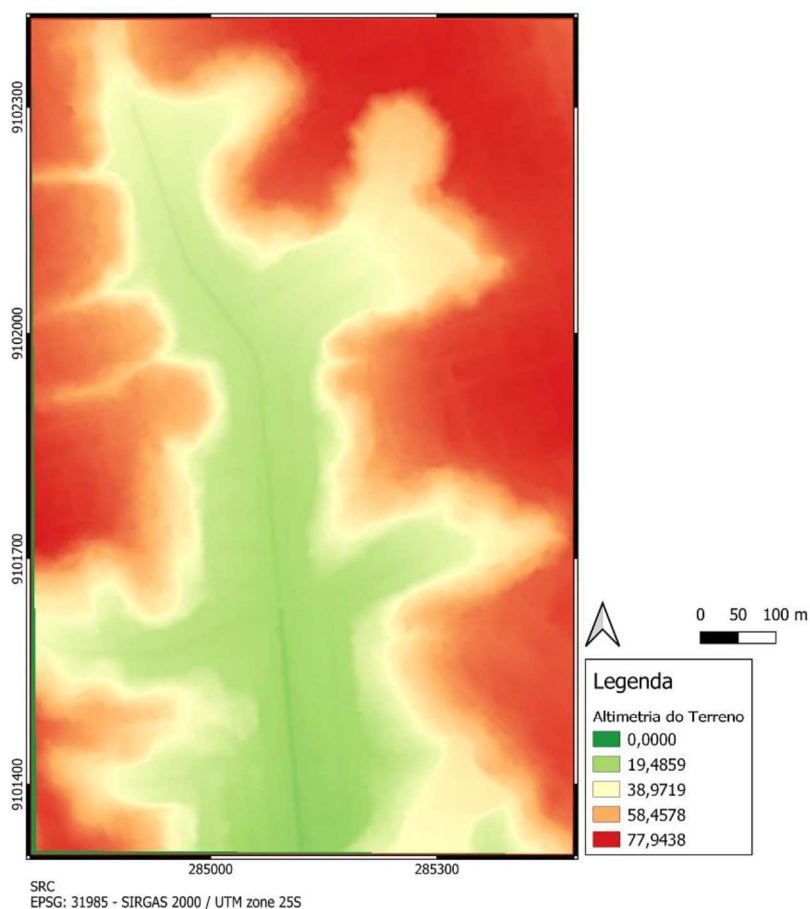
Fonte: Autor.

4.2.3 Modelo Digital de Terreno

É importante destacar que o Modelo Digital de Terreno (MDT) se refere à representação do terreno sem a vegetação e outras estruturas, enquanto o Modelo Digital de Elevação (MDE) inclui a altimetria das superfícies, considerando também a vegetação e construções. Essa diferenciação é crucial para entender as características topográficas da região e suas implicações em análises geoespaciais.

O MDT foi extraído da base de dados do PE3D (<http://www.pe3d.pe.gov.br/>), que oferece um mapeamento digital e tridimensional abrangente do estado de Pernambuco. Essa plataforma, disponível desde o levantamento realizado entre 2010 e 2011, utiliza tecnologia de ponta para gerar representações detalhadas e precisas das características geográficas e urbanas do estado. Com uma resolução plana de 1 metro e uma resolução altimétrica de 25 centímetros, proporcionando uma compreensão precisa da topografia local e facilitando a interpretação dos resultados obtidos, como demonstrado na Figura 24.

Figura 24 - Modelo digital de terreno



Fonte: Autor.

Outras variáveis condicionantes foram derivadas do MDT utilizando o software QGIS, incluindo a Declividade do Terreno, Aspecto, Curvatura Transversal, Curvatura Longitudinal, Distância da Bacia de Contribuição e Inclinação Média da Bacia de Contribuição. Essas variáveis fornecem informações essenciais de geomorfologia, permitindo uma compreensão detalhada das características topográficas da área em estudo.

Para facilitar a compreensão podemos adotar nomes simplificados para essas variáveis, que podem ser adotados ao longo deste documento, como demonstrado na Tabela 5.

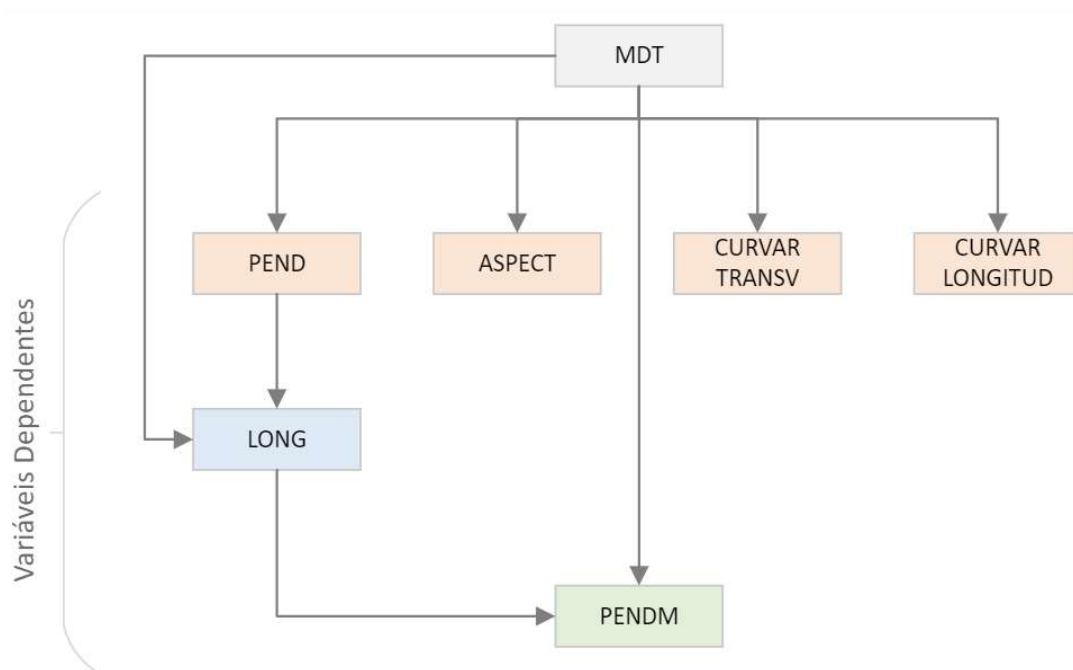
Tabela 5 - Nomes simplificados das variáveis

Nome da variável geomorfológica	Nome simplificado
Modelo Digital de Terreno	MDT
Declividade do Terreno	PEND
Aspecto	ASPECT
Curvatura Transversal	CURVAR TRANSV
Curvatura Longitudinal	CURVAR LONGITUD
Distância da Bacia de Contribuição	LONG
Inclinação Média da Bacia de Contribuição	PENDM

Fonte: Autor.

A seguir, é apresentado na Figura 25 um fluxograma simplificado que ilustra o processo de obtenção das variáveis dependentes e suas interdependências. Este fluxograma oferece uma visão sistemática do método utilizado para derivar as variáveis a partir do MDT utilizando a ferramenta QGIS.

Figura 25 - Fluxograma da dependência de variáveis



Fonte: Autor.

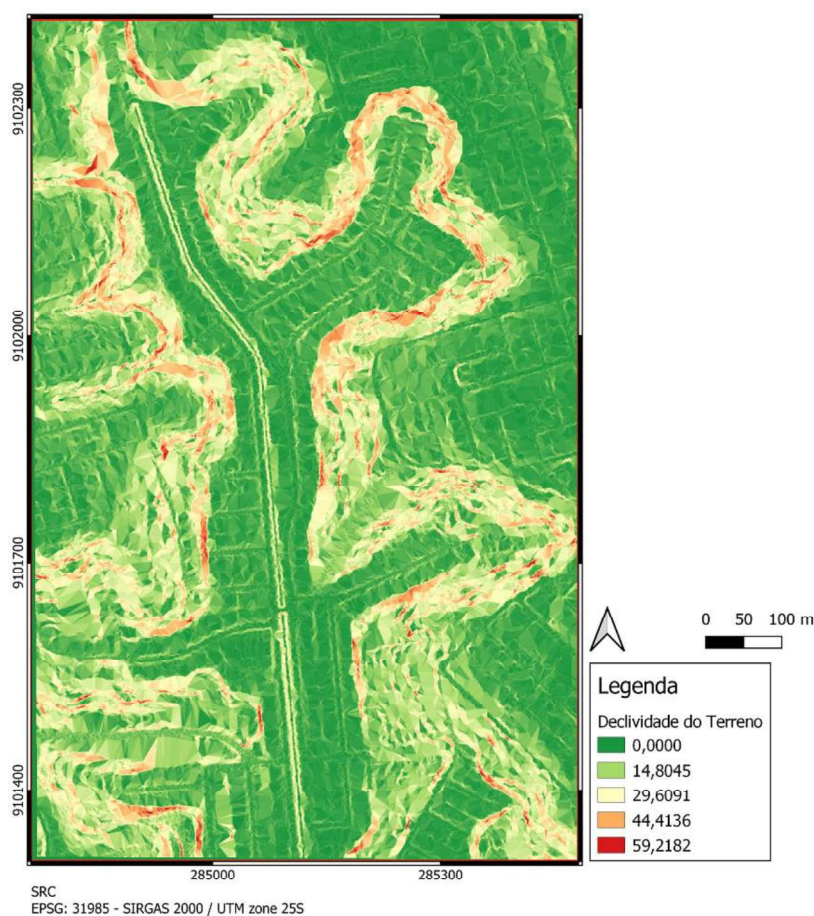
4.2.4 Declividade do Terreno

A declividade do terreno (PEND) é um parâmetro crucial na análise de suscetibilidade a deslizamentos de terra. Este parâmetro define o ângulo de inclinação das superfícies

terrestres em relação à horizontal, podendo variar de 0° a 90° graus, e influencia diretamente a estabilidade e segurança das estruturas construídas sobre ou dentro dessas superfícies. Terrenos com inclinações acentuadas são particularmente mais suscetíveis a deslizamentos, especialmente em áreas com solos argilosos ou altamente saturados, onde a coesão do solo pode ser significativamente comprometida.

Essa variável é uma das saídas da função “*Slope, Aspect, Curvature*” do QGIS. Essa função tem como entrada o MDT. A Figura 26 demonstra a declividade do terreno em estudo.

Figura 26 - Declividade do terreno



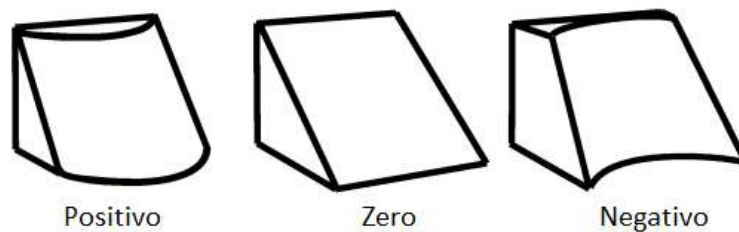
Fonte: Autor.

4.2.5 Curvatura Transversal do Terreno

A curvatura transversal (CURVAR TRANSV) do terreno refere-se à curvatura horizontal da superfície terrestre quando visualizada em planta. Esta característica descreve como o terreno se curva ao longo de uma linha perpendicular à direção principal da inclinação. Seus valores podem variar entre sinais positivos e negativos. Essa variável é uma

das saídas da função “*Slope, Aspect, Curvature*” do QGIS. A Figura 27 demonstra a curvatura transversal do terreno em estudo.

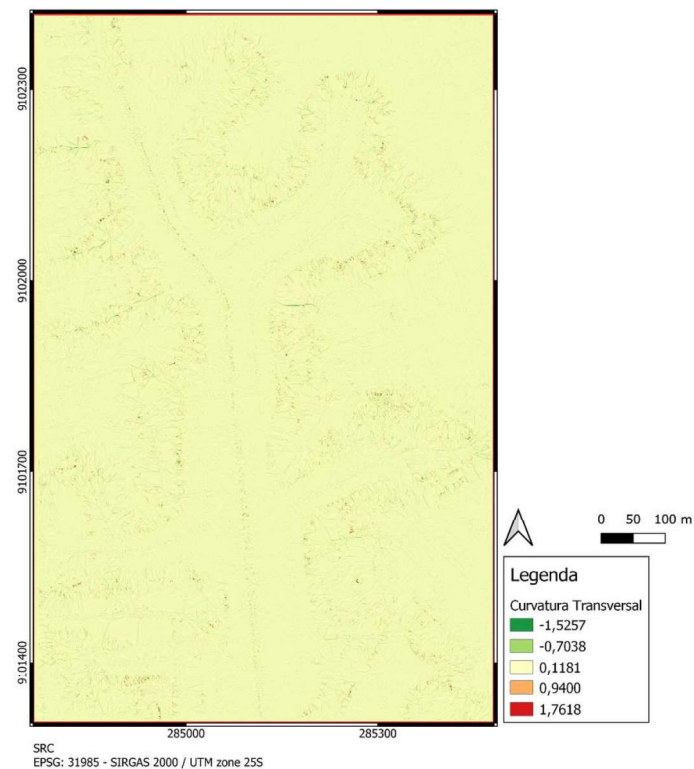
Figura 27 - Ilustração diagramática da curvatura transversal do terreno



Fonte: Adaptado de Heine *et al.* (2004).

- Valores positivos indicam uma curvatura convexa, onde a superfície do terreno se curva para fora. Esta condição é geralmente associada a áreas de encostas ou cumes, onde a água tende a escoar para fora, reduzindo a susceptibilidade a deslizamentos de terra.
- Valores próximos a zero indicam uma superfície plana ou com pouca curvatura, onde o escoamento de água no solo é mais uniforme.
- Valores negativos, indicam uma curvatura côncava, onde a superfície do terreno se curva para dentro. Nessas áreas, a água tende a se acumular, aumentando o risco de saturação do solo e, consequentemente, a susceptibilidade a deslizamentos de terra.

Figura 28 - Curvatura transversal do terreno

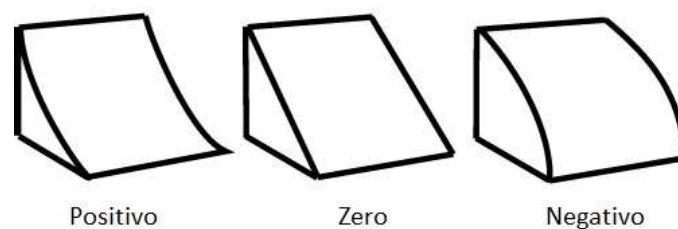


Fonte: Autor.

4.2.6 Curvatura Longitudinal do Terreno

A curvatura longitudinal (CURVAR LONGITUD) do terreno é um parâmetro que descreve a curvatura da superfície terrestre ao longo da direção principal de inclinação. Essa medida revela se o terreno se eleva ou desce ao longo dessa direção específica. Seus valores também podem variar entre sinais positivos e negativos. Essa variável é uma das saídas da função “*Slope, Aspect, Curvature*” do QGIS.

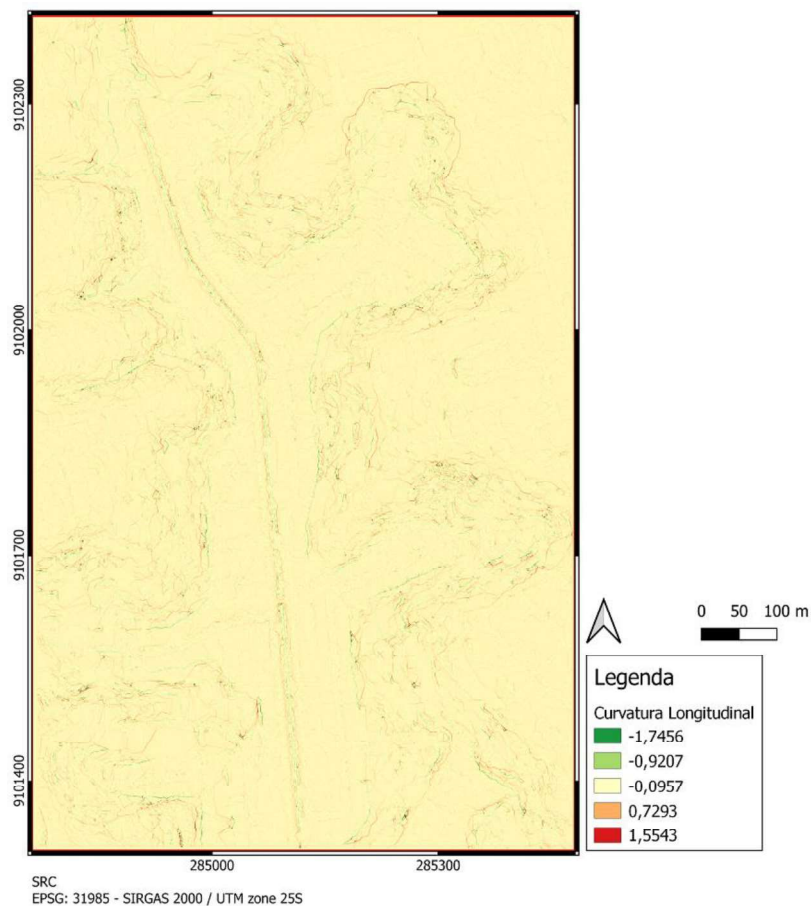
Figura 29 - Ilustração diagramática da curvatura longitudinal do terreno



Fonte: Adaptado de Heine *et al.* (2004).

- Valores positivos indicam uma curvatura côncava, onde a superfície do terreno se curva para baixo ao longo da linha principal de inclinação. Nessas áreas, a água tende a se acumular, aumentando o risco de saturação do solo e, conseqüentemente, a susceptibilidade a deslizamentos de terra.
- Valores próximos a zero indicam uma superfície plana ou com pouca curvatura ao longo do perfil. Nesse caso o escoamento de água no solo é mais uniforme.
- Valores negativos, indicam uma curvatura convexa, onde a superfície do terreno se curva para cima ao longo da linha principal de inclinação. Isso tende a dispersar o escoamento superficial, reduzindo a acumulação de água e, portanto, a susceptibilidade a deslizamentos de terra.

Figura 30 - Curvatura longitudinal do terreno

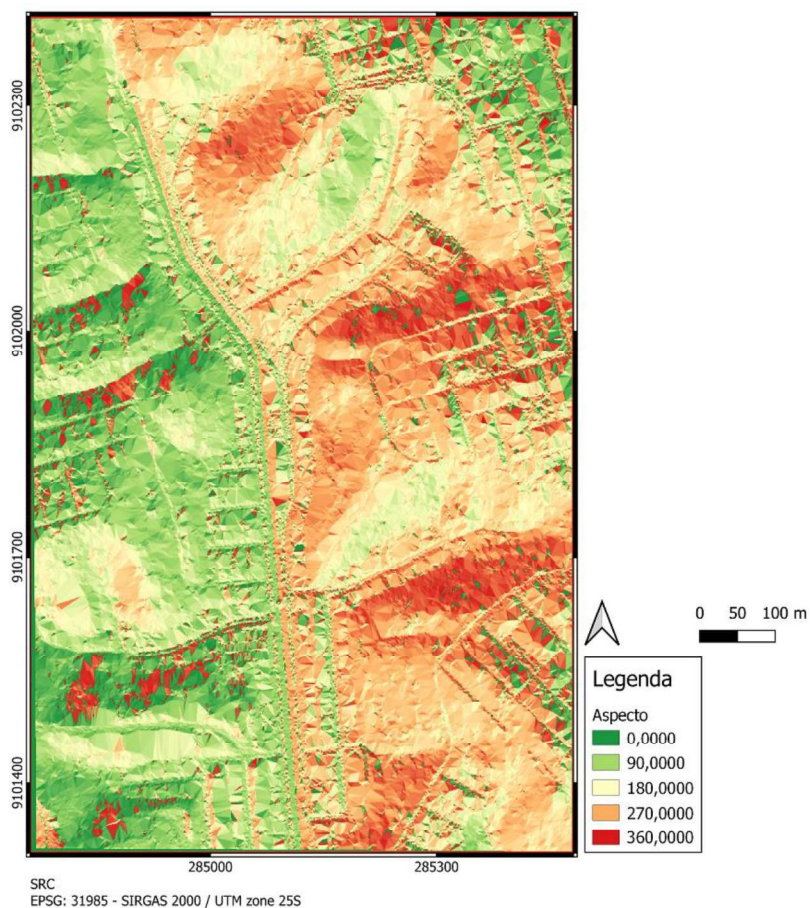


Fonte: Autor.

4.2.7 Aspecto

O aspecto (ASPECT) do terreno é uma medida que indica a direção para a qual uma determinada inclinação está voltada em relação aos pontos cardeais. Essa característica topográfica exerce uma influência direta na exposição solar, na quantidade de precipitação recebida e na umidade do solo em diferentes áreas de uma região. Essas variações são cruciais na determinação da susceptibilidade a deslizamentos de terra. O aspecto do terreno é expresso em uma escala de 0° a 360°, na qual tanto 0° quanto 360° representam o Norte, 90° representa o Leste, 180° representa o Sul e 270° representa o Oeste. Essa variável é uma das saídas da função “*Slope, Aspect, Curvature*” do QGIS.

Figura 31 - Aspecto



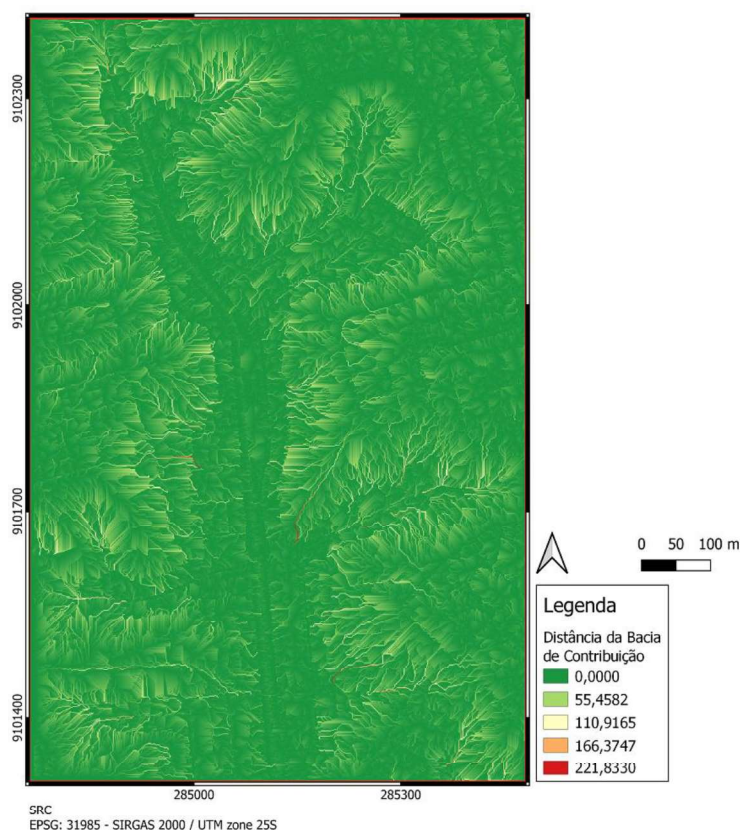
Fonte: Autor.

4.2.8 Distância da Bacia de Contribuição Sob a Superfície do Terreno

A distância da bacia de contribuição sobre a superfície do terreno (LONG) refere-se à distância entre um ponto na encosta e o limite superior da área que contribui para o escoamento de água nesse ponto específico, especificada em metros. Essa característica desempenha um papel fundamental na geotecnia, especialmente na análise da susceptibilidade a deslizamentos de terra. Como é explicado em Amorim (2012), quanto maior for a distância máxima da bacia de contribuição, maior será a quantidade de água que pode acumular-se e escoar pela encosta antes de atingir o ponto de interesse. Esse fenômeno pode resultar em uma maior saturação do solo, aumentando consideravelmente o potencial de instabilidade e ocorrência de deslizamentos de terra.

Essa variável é calculada a partir da função *maximumflowpathlength* do QGIS, onde as entradas são o *weights* e o *Direction_of_Measurement*. O primeiro parâmetro pode ser calculado a partir de $1/\cos(\text{PEND})$ com o PEND convertido em radianos, já o segundo parâmetro apenas deve-se escolher a partir da lista suspensa a opção *upstream*.

Figura 32 - Distância da bacia de contribuição sob a superfície do terreno



Fonte: Autor.

4.2.9 Inclinação Média da Bacia de Contribuição

A inclinação média da bacia de contribuição (PENDM) é uma medida que representa o declive médio da que contribui para o escoamento de água em direção a um ponto específico em uma encosta. Essa medida varia de 0° a 90° graus e desempenha um papel crucial na avaliação da estabilidade das encostas. De acordo com Amorim (2012), a inclinação média da bacia hidrográfica reflete sua capacidade de absorver a água no solo. Valores baixos de inclinação média indicam uma maior capacidade de infiltração.

Essa variável é calculada a partir de ACos ($\text{FlowLenFillUp} / \text{LONG}$), onde o *FlowLenFillUp* é calculado da mesma forma que o LONG mas sem preencher o parâmetro *weights*. O resultado encontrado deve ser convertido de radianos para graus.

Figura 33 - Inclinação média da bacia de contribuição

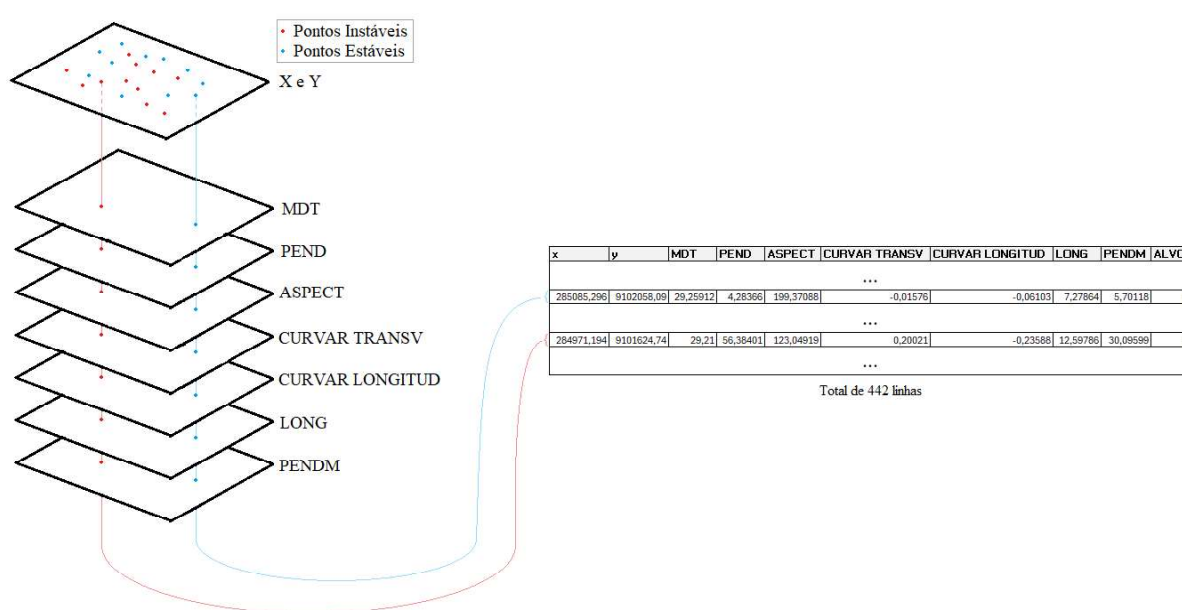


Fonte: Autor.

4.2.10 Base de Dados Compilada

Com base em todas as informações das variáveis geomorfológicas reunidas, procedeu-se à extração dos dados localizados no QGIS com a ferramenta *point sampling tool*, utilizando como referência a localização dos pontos estáveis e instáveis. Esse processo resultou na obtenção de dados tabulados, os quais já estavam categorizados acordo com sua classificação como instáveis ou estáveis. Essa abordagem proporcionou uma estrutura organizada e pronta para análise.

Figura 34 - Ilustração da extração dos dados do QGIS



Fonte: Autor.

Sabendo que a base de dados é composta por 221 pontos instáveis e 221 pontos estáveis, totalizando 442 pontos, o conjunto de dados inclui dez colunas no total. Dessas colunas, sete representam as variáveis geomorfológicas, duas contêm as coordenadas geográficas (X e Y) e uma contém os dados da variável alvo que identifica se um ponto é instável ou estável por meio dos valores 1 ou 0, respectivamente. Essa estrutura tabelada possibilita a análise dos dados e o treinamento de um modelo de Rede Neural, capacitando-o a interpretar e classificar os dados de maneira generalizada, com base nas características da região de Lagoa Encantada.

4.3 PROGRAMAS UTILIZADOS

4.3.1 Quantum Geographic Information System - QGIS

O QGIS (Quantum Geographic Information System) é um software livre e de código aberto para sistemas de informação geográfica (SIG), amplamente utilizado por profissionais da cartografia, geografia, engenharia ambiental, engenharia civil e outras áreas que necessitam de análise espacial e geoprocessamento. Lançado pela primeira vez em 2002 e posteriormente adotado como um projeto incubado pela Open Source Geospatial Foundation em 2007, o QGIS tem se destacado pela sua versatilidade, facilidade de uso e pela extensa gama de ferramentas disponíveis para manipulação e visualização de dados geoespaciais.

Figura 35 - Logo do QGIS



Fonte: Quantum Geographic Information System.

Uma das grandes vantagens do QGIS é a sua capacidade de integração com diversos formatos de dados e sistemas de banco de dados geográficos. Além disso, o QGIS suporta uma ampla variedade de formatos de arquivos vetoriais e raster, permitindo que os usuários importem e exportem dados de diferentes fontes para utilizá-los em seus projetos. A interoperabilidade do QGIS com outras ferramentas de SIG e software de análise estatística torna-o uma escolha robusta para análises complexas e projetos de grande escala.

Em termos de aplicação prática, o QGIS tem sido amplamente utilizado em diversas áreas, como planejamento urbano, gestão de recursos naturais, monitoramento ambiental e monitoramento urbano. A capacidade do QGIS de realizar análises detalhadas e produzir visualizações de alta qualidade o torna uma ferramenta indispensável para profissionais que dependem de informações geoespaciais precisas e atualizadas na tomada de decisões. Sua versatilidade e precisão permitem que especialistas desenvolvam estratégias informadas e

eficazes em contextos variados, reforçando a importância do QGIS no campo das geotecnologias.

4.3.2 Python

Python, uma linguagem de programação de alto nível, foi concebida por Guido van Rossum e lançada pela primeira vez em 1991. Nasceu da necessidade de uma linguagem que fosse acessível para aprendizado e uso, sem comprometer a expressividade para soluções complexas. Desde então, Python tem trilhado um caminho de evolução constante, passando por diversas iterações que agregaram novas funcionalidades e aprimoramentos de desempenho, consolidando sua posição como uma das linguagens mais populares e versáteis na computação moderna. Atualmente, seu desenvolvimento e manutenção estão sob os cuidados da Python Software Foundation (PSF), uma entidade sem fins lucrativos dedicada à promoção, proteção e avanço dessa linguagem de programação.

Figura 36 - Logo do Python



Fonte: python.org.

De acordo com rankings, como o TIOBE e o IEEE Spectrum, Python consistentemente alcança as primeiras posições entre as linguagens de programação mais populares e influentes. Seu alcance é vasto e diversificado que vai desde o desenvolvimento web, até a automação de tarefas administrativas e o desenvolvimento de soluções corporativas robustas. Além disso, Python é uma escolha proeminente em campos emergentes, como ciência de dados, aprendizado de máquina e inteligência artificial. Isso se deve não apenas à sua sintaxe clara e intuitiva, mas também à abundância de bibliotecas especializadas que simplificam tarefas complexas, permitindo aos desenvolvedores concentrarem-se na lógica do problema em vez de se preocuparem com detalhes técnicos.

Figura 37 - Índice de uso das linguagens de programação ao longo dos anos



Fonte: tiobe.com.

No âmbito do aprendizado de máquina, Python se destaca com bibliotecas como TensorFlow, PyTorch e scikit-learn (sklearn). Entre essas, o scikit-learn se destaca como uma ferramenta essencial para a implementação de algoritmos de aprendizado de máquina em Python. Sua abrangência é notável, oferecendo desde recursos para regressão linear até complexas redes neurais.

Figura 38 - Logo do scikit-learn



Fonte: scikit-learn.org.

O que torna o scikit-learn particularmente valioso é sua combinação de simplicidade e eficácia. A biblioteca proporciona ferramentas acessíveis e eficientes para análise de dados e modelagem preditiva, simplificando tarefas que vão desde a pré-processamento de dados até a avaliação de modelos. Essa acessibilidade e robustez fazem do scikit-learn uma ferramenta indispensável para cientistas e engenheiros que buscam construir modelos preditivos precisos e eficientes em Python. Com sua vasta gama de funcionalidades e suporte ativo da comunidade, essa biblioteca continua a ser uma escolha confiável e poderosa para aplicações de aprendizado de máquina em diversos campos, desde pesquisa acadêmica até aplicações comerciais de larga escala.

5 RESULTADOS E DISCUSSÕES

5.1 ANÁLISE DOS DADOS

5.1.1 Estatísticas Básicas

Essa análise inicial serve como um ponto de partida para uma exploração mais profunda dos dados, visando entender melhor suas características. Dentre as variáveis temos as coordenadas geográficas (x, y) e características geomorfológicas como o Modelo Digital de Terreno (MDT), Declividade do Terreno (PEND), Aspecto (ASPECT), Curvatura Transversal (CURVAR TRANSV), Curvatura Longitudinal (CURVAR LONGITUD), Distância da Bacia de Contribuição Sobre a Superfície do Terreno (LONG) e Inclinação Média da Bacia de Contribuição (PENDM). Além dessas, a variável binária META é de particular interesse, pois é o alvo que desejamos que a rede neural classifique corretamente. A seguir na Tabela 6, apresentamos um resumo das estatísticas básicas para cada uma dessas variáveis, proporcionando uma visão geral das suas distribuições e características essenciais.

Tabela 6 - Estatísticas dos dados

	x	y	MDT	PEND	ASPECT	CURVAR TRANSV	CURVAR LONGITUD	LONG	PENDM	META
count	442	442	442	442	442	442	442	442	442	442
mean	285090,597077	9,101851E+06	48,454077	19,655135	163,534238	-0,003368	-0,013463	9,702287	15,417037	0,5
std	202,733787	3,025207E+02	17,344753	14,887024	93,873676	0,156336	0,181015	14,725637	13,958945	0,500567
min	284770,296000	9,101312E+06	16,28	0,28648	0,6782	-0,90093	-1,08061	0	0	0
25%	284931,266250	9,101613E+06	33,965872	5,45275	90,404525	-0,02959	-0,024	1,382955	2,23067	0
50%	285051,122500	9,101815E+06	47,491025	18,113835	157,546755	0	0	5,014675	12,239075	0,5
75%	285263,146750	9,102120E+06	64,257512	31,136717	232,874455	0,02873	0,026673	12,955278	27,166822	1
max	285467,160000	9,102401E+06	77,01623	58,59708	360	0,77453	0,81417	179,13211	54,13013	1

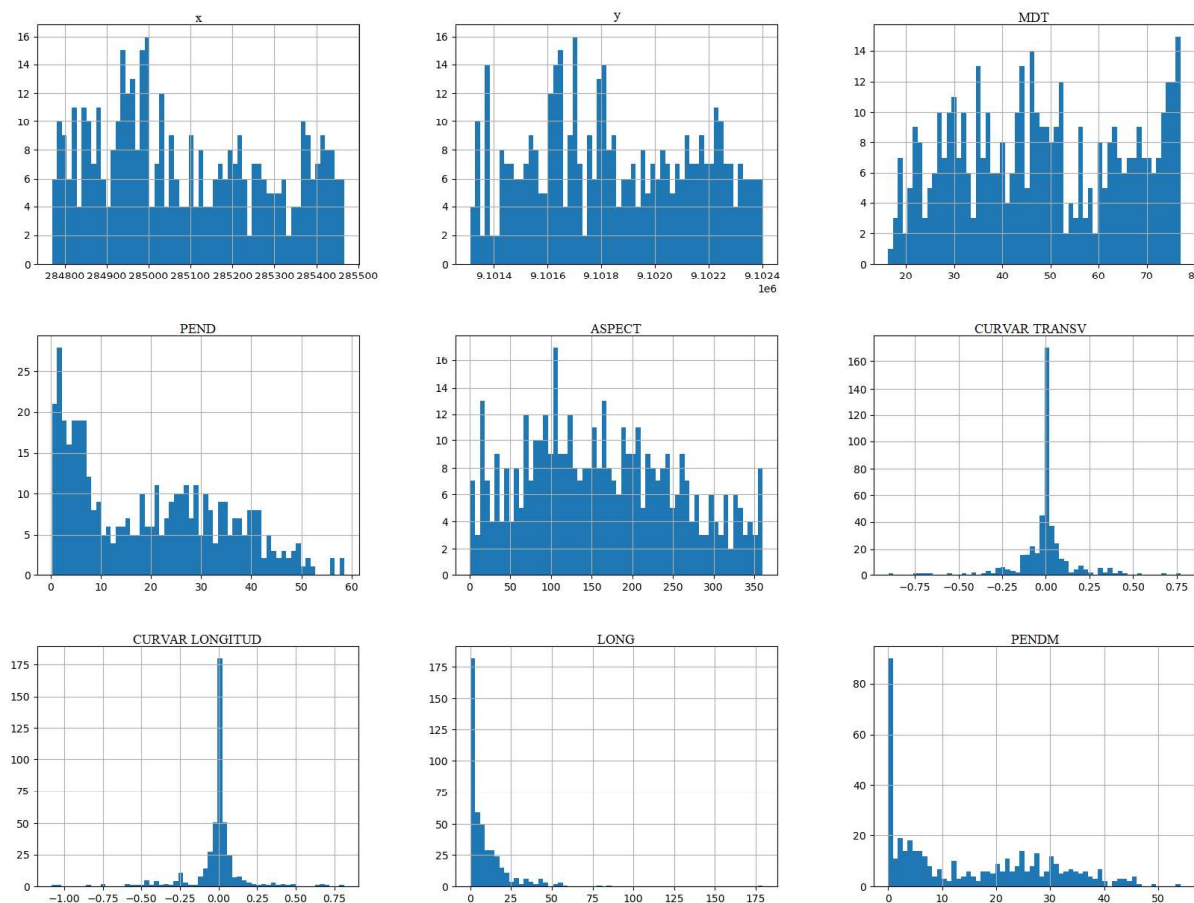
Fonte: Autor.

As variáveis MDT, PEND, ASPECT, LONG e PENDM mostram uma distribuição que pode ser ligeiramente assimétrica, sugerindo a necessidade de uma análise mais detalhada por meio de histogramas e diagramas de caixa.

5.1.2 Histogramas

Histogramas são ferramentas essenciais em análise estatística e exploratória de dados. Eles fornecem uma visão clara e imediata da distribuição dos dados, facilitando a identificação de padrões e características fundamentais da distribuição.

Figura 39 - Histogramas das variáveis



Fonte: Autor.

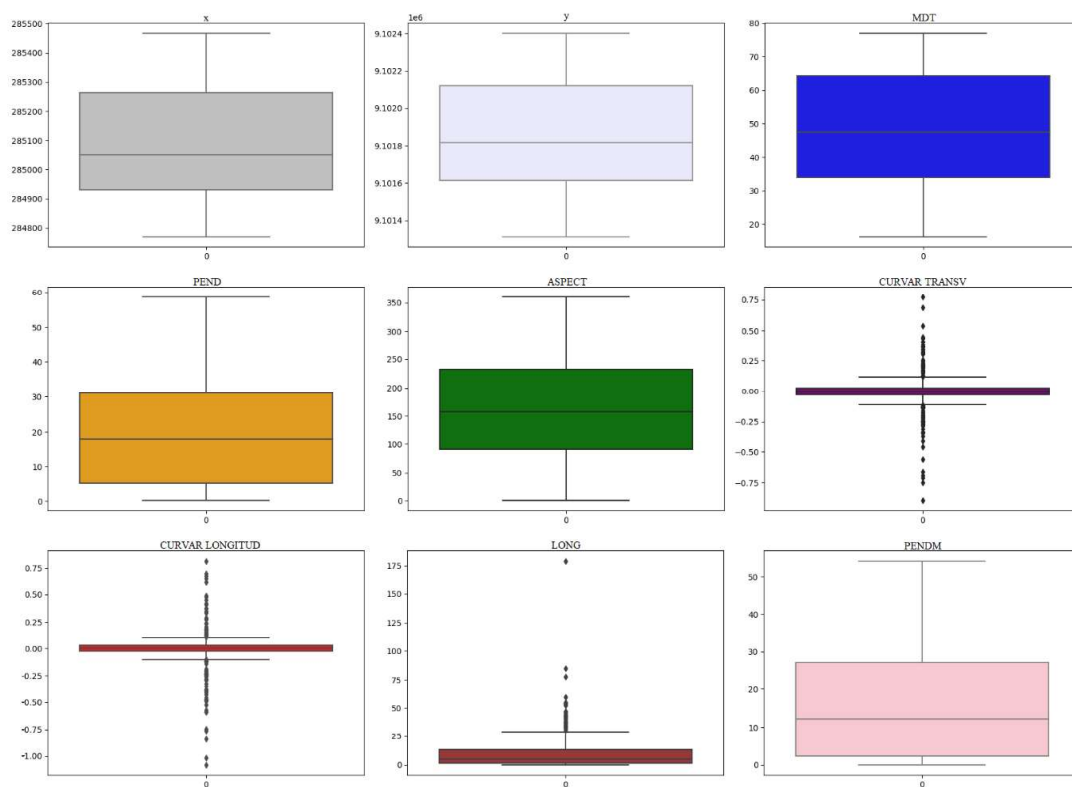
O Modelo Digital de Terreno (MDT) revela que a distribuição das elevações é mais frequente entre 40 e 80, indicando que essas elevações são comuns na região. Em relação à Declividade (PEND), a maioria dos dados concentra-se entre 0 e 10, sugerindo que os terrenos apresentam predominantemente baixa inclinação, mas também contendo declividades entre 20 e 40 presentes. A Orientação (ASPECT) mostra uma distribuição mais dispersa, com picos notáveis entre 100 e 300, o que indica que certas orientações são mais frequentes em superfícies inclinadas.

No que diz respeito à Curvatura Transversal (CURVAT TRANSV), a distribuição é centralizada em torno de 0, apresentando um pico maior entre -0,25 e 0,25, sugerindo a predominância de superfícies quase planas ou levemente curvas. A Curvatura Longitudinal (CURVAR LONGITUD) exibe um comportamento semelhante, com a concentração de valores também próxima de 0, indicando superfícies com pouca curvatura longitudinal. Por sua vez, a Longitude (LONG) apresenta a maior parte dos valores entre 0 e 25, com uma rápida queda na frequência após esse intervalo. Por fim, a Inclinação média da bacia de contribuição (PENDM) mostra uma distribuição com maior frequência de valores ainda se encontra entre 0 e 10, reforçando a ideia de que os terrenos geralmente possuem uma inclinação da bacia não tão elevadas.

5.1.3 Diagramas de Caixa

Os diagramas de caixa fornecem uma visão clara da distribuição das variáveis, destacando a mediana, os quartis e os outliers. Eles ajudam a orientar análises subsequentes e a focar em áreas específicas que podem requerer atenção adicional.

Figura 40 - Diagrama de caixa das variáveis



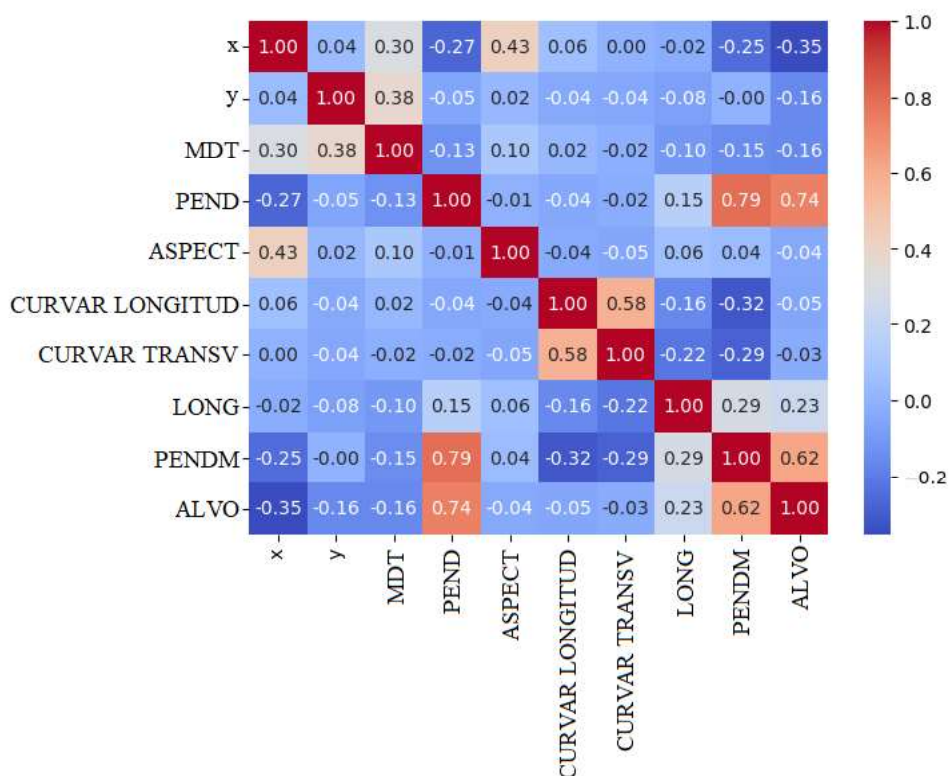
Fonte: Autor.

Variáveis como CURVAR TRANSV, CURVAR LONGITUD e LONG mostram a presença de outliers significativos. Mas devido à escassez de dados disponíveis neste estágio inicial do projeto, foi decidido manter os outliers identificados durante a análise. A decisão de reter esses pontos atípicos se baseia na premissa de que, com um volume limitado de observações, cada ponto de dado desempenha um papel significativo na representação do comportamento geral do sistema estudado.

5.1.4 Matriz de Correlação

A correlação é uma medida estatística que indica a força e a direção da relação linear entre duas variáveis. Os valores de correlação variam de -1 a 1, onde 1 indica uma correlação positiva perfeita, -1 uma correlação negativa perfeita e 0 indica nenhuma correlação. Uma matriz de correlação é uma tabela que mostra as correlações entre variáveis em um conjunto de dados. Ela é frequentemente usada em análises estatísticas e científicas para entender as relações lineares entre diferentes variáveis e como elas se comportam entre si.

Figura 41 - Matriz de correlação das variáveis



Fonte: Autor.

As variáveis PEND, PENDM mostram correlações fortes com ALVO, o que pode indicar que áreas mais íngremes são significativamente associadas com o valor da variável ALVO. Isso é relevante para estudos que possam estar investigando a propensão a deslizamentos ou outros fenômenos relacionados à inclinação do terreno. Mesmo que as correlações lineares entre as demais variáveis em relação a variável ALVO sejam baixas, em muitos casos é estratégico mantê-las durante o treinamento de uma rede neural devido ao potencial significativo que essas variáveis podem ter no contexto do modelo.

De acordo com Goodfellow *et al.* (2016), redes neurais têm a capacidade de capturar padrões complexos e não lineares nos dados. Isso implica que variáveis que apresentam baixas correlações lineares com a variável alvo ainda podem contribuir significativamente para a precisão e eficácia do modelo final, devido à habilidade das redes neurais de explorar interações mais profundas e relações não evidentes em uma análise linear simples.

5.2 TREINAMENTO E VALIDAÇÃO DA REDE NEURAL

5.2.1 Divisão dos Dados em Treino e Teste

A divisão dos dados em conjuntos de treino e teste foi feita na proporção de 80% para treino e 20% para teste. Isso resultou em 353 observações destinadas ao treinamento do modelo e 89 observações reservadas para a etapa de teste. Essa divisão foi cuidadosamente escolhida para garantir que o modelo tenha uma quantidade substancial de dados para aprender, enquanto mantém um conjunto significativo para avaliar seu desempenho e generalização.

A biblioteca scikit-learn (sklearn) oferece funções que realizam a divisão do conjunto de dados em treino e teste de forma a preservar as distribuições das classes ou dos dados. A função mais comum para isso é a `train_test_split`, que pode ser configurada para realizar uma divisão estratificada. Para problemas de classificação, onde você quer garantir que as proporções das classes são mantidas em ambos os conjuntos de treino e teste, você pode usar o parâmetro `stratify` da função `train_test_split`.

Figura 42 - Código: Divisão em treino e teste

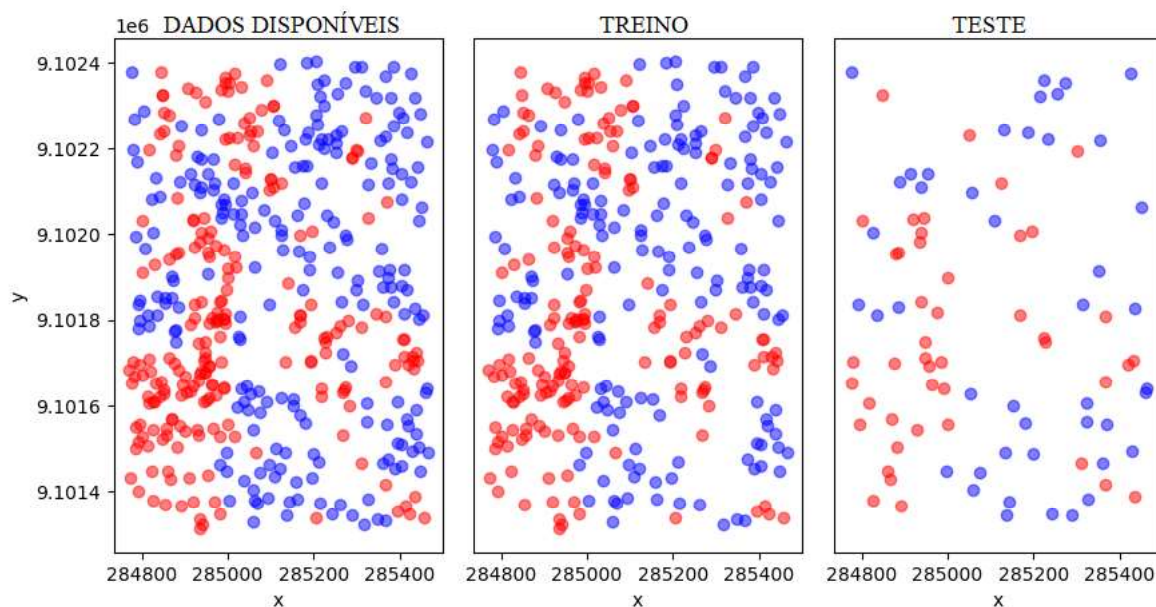
```
from sklearn.model_selection import train_test_split

# Divisão em treino e teste
X_train, X_test, y_train, y_test = train_test_split(df.drop('TARGET', axis = 1),
                                                    df['TARGET'],
                                                    test_size = 0.2,
                                                    stratify = df['TARGET'])
```

Fonte: Autor.

A seguir, podemos observar como foi realizada a divisão entre os conjuntos de treinamento e teste de forma espacialmente distribuída. É evidente que a distribuição geoespacial ficou bem equilibrada.

Figura 43 - Divisão espacial dos dados em treino e teste



Fonte: Autor.

5.2.2 Criação do Fluxo de Dados

O fluxo de dados foi definido usando a biblioteca sklearn em Python, composto por duas etapas principais, sendo elas a normalização dos dados e classificação utilizando uma rede neural MLP.

Figura 44 - Código: Definição do fluxo de dados

```

from sklearn.preprocessing import MinMaxScaler
from sklearn.neural_network import MLPClassifier
from sklearn.pipeline import Pipeline

# Definindo os passos do pipeline (Normalização dos dados e Classificador do tipo Redes Neurais MLP)
steps = [('scaler', MinMaxScaler()), # Normalizar os dados com o normalizador Min-Max
         ('classifierRNA', MLPClassifier(max_iter = 5000, # Número máximo de iterações que o algoritmo adam deve fazer
                                       tol = 0.00001, # Tolerância para convergência do Treinamento
                                       n_iter_no_change = 500, # Número máximo de épocas sem melhoria da função de perda
                                       early_stopping = True, # Permite o a parada prévia
                                       solver = 'adam'))] # Otimizador a ser usado para treinar o modelo

# Criando o pipeline
pipeline = Pipeline(steps)

```

Fonte: Autor.

Primeiro, a etapa de normalização (MinMaxScaler) ajusta os dados para um intervalo entre 0 e 1, assegurando que todas as features estejam na mesma escala. Em seguida, a etapa de classificação (MLPClassifier) treina um modelo de rede neural com parâmetros configurados para otimizar o treinamento, incluindo um máximo de 5000 iterações, uma tolerância de convergência de 0.00001, e parada prévia (*early stopping*) ativado para interromper o treinamento se a função de perda não melhorar após 500 épocas consecutivas.

Para esclarecer, no caso do MLPClassifier, a função de ativação específica para a última camada é determinada automaticamente pelo algoritmo do sklearn com base no tipo de problema de classificação. Em problemas de classificação binária, a função de ativação sigmoide é aplicada.

5.2.3 Definição dos Conjuntos para a Busca em Grade

Para realizar a busca em grade, foi definido um conjunto de valores específicos para cada hiperparâmetro da rede neural. Abaixo, apresentamos o código em Python que estabelece esses conjuntos de valores, a serem utilizados na pesquisa em grade com o classificador MLPClassifier previamente configurado.

Figura 45 - Código: Definição dos valores da busca em grade

```

# Definindo os conjuntos de parâmetros para pesquisa em grade
parameters = {
    'classifierRNA_hidden_layer_sizes': [(i,) for i in range(5, 11)], # Apenas uma camada oculta variando de 5 a 10 neurônios
    'classifierRNA_activation': ['relu','tanh'], # Funções de ativação
    'classifierRNA_validation_fraction': [0.1,0.2,0.3], # Tamanho do conjunto de validação
    'classifierRNA_learning_rate_init': [0.001,0.01,0.1,0.3], # Taxas de aprendizado
    'classifierRNA_alpha': [0.0001,0.001,0.01,0.1], # Valores de alpha
    'classifierRNA_beta_1': [0.7,0.75,0.8,0.85], # Valores de beta 1
    'classifierRNA_beta_2': [0.9,0.95,0.99,0.999], # Valores de beta 2
    'classifierRNA_epsilon': [1e-4,1e-5,1e-6,1e-7] # Valores de epsilon
}

```

Fonte: Autor.

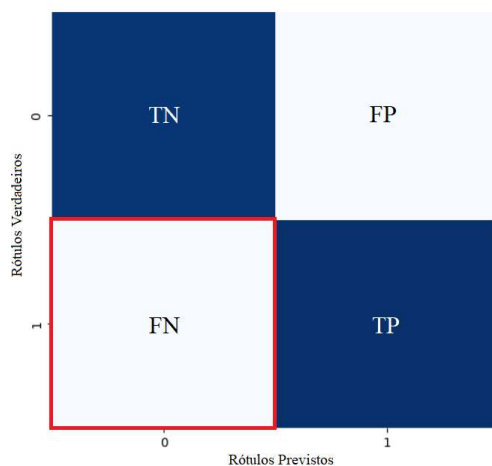
Os parâmetros incluem várias configurações para a estrutura e o treinamento da rede neural, como o número de neurônios em uma camada oculta (*hidden_layer_sizes*), funções de ativação (*activation*), frações de validação (*validation_fraction*), taxas de aprendizado iniciais (*learning_rate_init*), coeficientes de regularização (*alpha*), e hiperparâmetros específicos do otimizador Adam, como *beta_1*, *beta_2*, e *epsilon*. Ao explorar todas as combinações desses parâmetros, a pesquisa em grade busca identificar a configuração que maximiza o desempenho do classificador RNA, garantindo que o modelo final seja bem ajustado aos dados de treinamento e validação.

5.2.4 Treino do Modelo de RN - Realização da Busca em Grade com Validação Cruzada

Antes de iniciar a busca em grade com validação cruzada, é fundamental definir qual métrica de desempenho será avaliada durante o processo de otimização de busca em grade. É possível escolher entre várias métricas válidas para modelos de classificação, como precisão, revocação ou f1-score.

No caso de deslizamentos de terra, é crucial minimizar os erros do tipo Falso Negativo, isto é, quando o modelo prevê que não ocorrerá um deslizamento, mas ele de fato ocorre. Em outras palavras, é essencial otimizar a métrica de revocação. Ela avalia a capacidade do modelo de identificar corretamente todas as instâncias positivas (deslizamentos), garantindo que os deslizamentos de terra sejam previstos de maneira mais abrangente, reduzindo riscos e potencialmente salvando vidas e propriedades.

Figura 46 - Em destaque: Falsos negativos na matriz de confusão



Fonte: Autor.

Para realizar a busca em grade com validação cruzada, foi utilizada a classe `GridSearchCV` da biblioteca `scikit-learn`, permitindo a exploração sistemática de múltiplas combinações de hiperparâmetros por meio de validação cruzada.

Figura 47 - Código: Busca da melhor rede neural

```
from sklearn.model_selection import GridSearchCV
# Realizando a procura em grade com validação cruzada
grid_search = GridSearchCV(pipeline, parameters, cv = 5, scoring = 'recall')
```

Fonte: Autor.

O objeto `grid_search` é configurado para explorar diferentes combinações de hiperparâmetros definidos anteriormente. A busca em grade é realizada com validação cruzada de 5 dobras ($cv = 5$). O parâmetro `scoring = 'recall'` especifica que a métrica de avaliação utilizada durante a busca em grade é a revocação.

O tempo total de execução foi de aproximadamente duas horas. O que torna o processo demorado é o número de redes neurais que precisam ser treinadas. Ao calcular todas as combinações possíveis de parâmetros e multiplicar pelo número de dobras da validação cruzada, resulta em um total de 184.320 redes neurais treinadas. Por fim, a melhor rede neural retornada pelo código possui os seguintes hiperparâmetros.

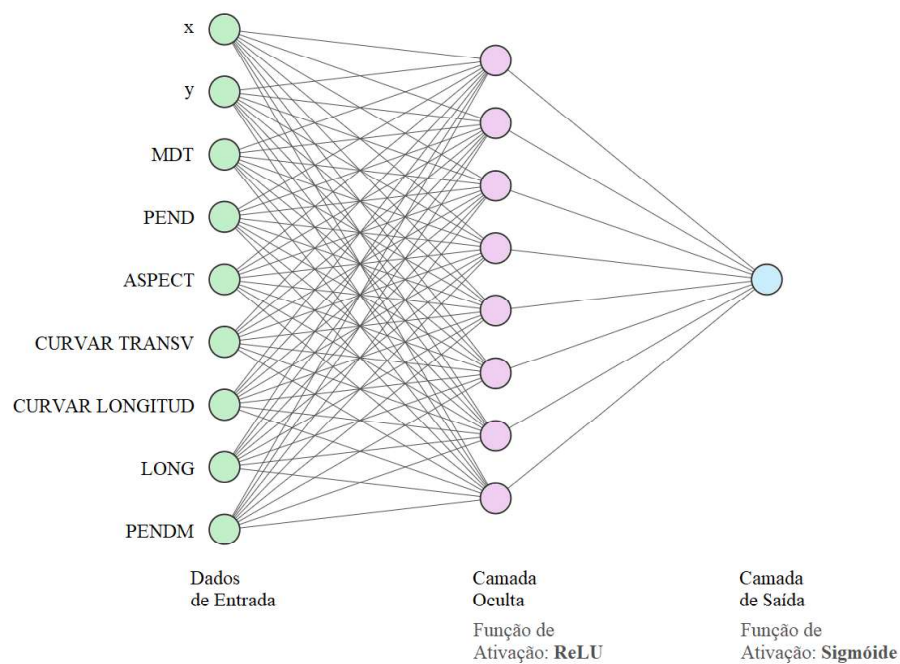
Tabela 7 - Melhores hiperparâmetros encontrados

Hiperparâmetro	Valor
<code>validation_fraction</code>	0.2
<code>hidden_layer_sizes</code>	(8,)
<code>activation</code>	relu
<code>learning_rate_init</code>	0.1
<code>alpha</code>	0.001
<code>beta_1</code>	0.8
<code>beta_2</code>	0.95
<code>epsilon</code>	1e-06

Fonte: Autor.

Visualmente, a RN treinada apresenta nove entradas, oito neurônios na camada oculta utilizando a função de ativação ReLU, e um único neurônio na camada de saída com a função de ativação sigmóide.

Figura 48 - Rede neural treinada

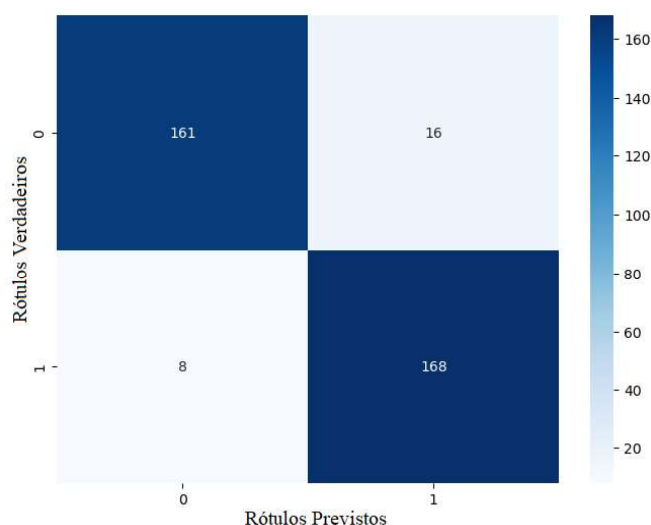


Fonte: Autor.

5.2.5 Validação do Modelo – Matriz de Confusão e Métricas Derivadas

Após identificar a melhor rede neural, otimizada para maximizar a revocação, procedeu-se à verificação da matriz de confusão tanto no conjunto de treino quanto no conjunto de teste. Essa análise abrangente é fundamental para avaliar o desempenho do modelo, verificando se a RN não só se ajustou bem aos dados de treino, mas também generalize efetivamente aos dados de teste. Avaliar a matriz de confusão em ambos os conjuntos permite identificar possíveis problemas de sobreajuste e assegurar que o modelo tem uma capacidade robusta de prever instâncias corretas em cenários não vistos anteriormente no treinamento.

Figura 49 - Matriz de confusão: Conjunto de treino



Fonte: Autor.

Embora a rede neural tenha sido otimizada para minimizar a revocação, ainda ocorreram oito falsos negativos. A ausência de informações como o período do ano e dados hidrológicos locais, além da escassez de dados, pode ter contribuído para a dificuldade do modelo em conseguir atingir uma taxa menor de falsos negativos. As características geomorfológicas, embora fundamentais, podem não capturar as nuances sazonais ou variações hidrológicas que influenciam diretamente na estabilidade do terreno.

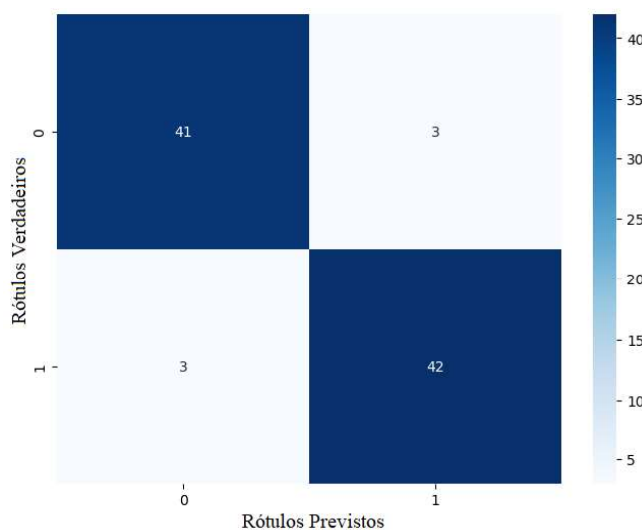
Tabela 8 - Métricas no conjunto de treino

Métrica	Valor
Acurácia	0,931
Precisão	0,913
Revocação	0,955
F1-Score	0,934

Fonte: Autor.

O modelo de classificação apresenta um bom desempenho no conjunto de treino, apresentando alta acurácia, precisão, revocação e F1 Score. Esses resultados são indicativos de que o modelo está aprendendo bem os padrões existentes nos dados de treino, conseguindo distinguir corretamente a maioria dos casos instáveis dos estáveis.

Figura 50 - Matriz de confusão: Conjunto de teste



Fonte: Autor.

Os 3 falsos negativos observados no conjunto de teste indicam que o modelo ainda está cometendo erros ao prever instâncias da classe positiva como negativas em dados não vistos durante o treinamento da rede neural. Avaliar apenas o número de falsos negativos não é suficiente para uma compreensão completa do desempenho do modelo. É fundamental analisar as métricas de desempenho de forma abrangente para avaliar a capacidade de generalização do modelo.

Tabela 9 - Métricas no conjunto de teste

Métrica	Valor
Acurácia	0,933
Precisão	0,933
Revocação	0,933
F1-Score	0,933

Fonte: Autor.

As métricas de desempenho mostram uma consistência notável entre os conjuntos de treino e teste. No conjunto de treino, a precisão foi de 91,3% e a revocação de 95,5%. Já no conjunto de teste, ambas as métricas foram de 93,3%. O F1 Score também é elevado em ambos os conjuntos, assim como a acurácia.

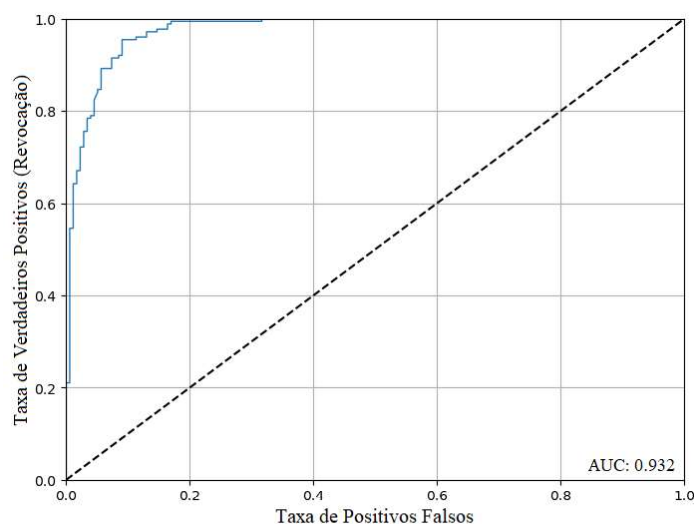
Essa pequena diferença nas métricas entre treino e teste sugere que o modelo não está sobreajustado, o que significa que ele está generalizando bem para novos dados. A

consistência nas métricas de desempenho entre os dois conjuntos indica que o modelo é robusto e confiável para a tarefa de classificação para a qual foi treinado.

5.2.6 Validação do Modelo – Curva ROC

É igualmente valioso examinar a curva ROC (Receiver Operating Characteristic) e a área sob a curva (AUC) tanto nos dados de treino quanto nos dados de teste. Essa análise permite verificar o comportamento do modelo durante o processo de treinamento e garantir que ele esteja aprendendo os padrões corretamente sem sobreajuste, conforme discutido anteriormente.

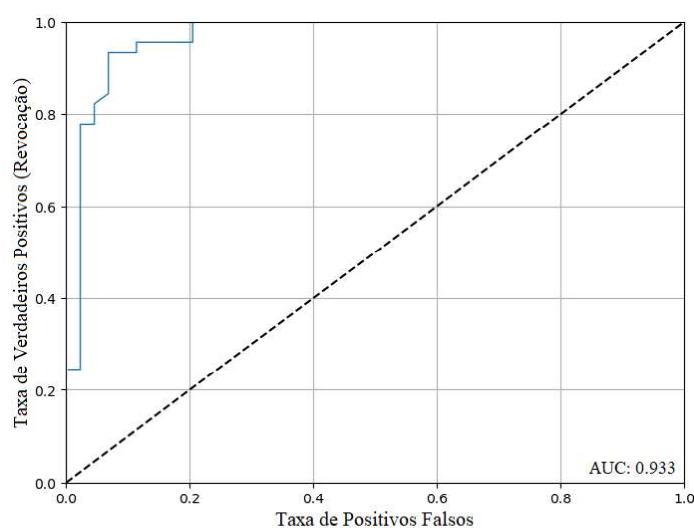
Figura 51 - Curva ROC: Conjunto de treino



Fonte: Autor.

A curva ROC do conjunto de treino mostra uma excelente performance, com a taxa de verdadeiros positivos (revocação) aumentando rapidamente à medida que a taxa de falsos positivos aumenta. A curva está próxima do canto superior esquerdo do gráfico, indicando uma alta capacidade do modelo em discriminar entre as classes. A AUC próxima de 1 indica um bom desempenho do modelo.

Figura 52 - Curva ROC: Conjunto de teste



Fonte: Autor.

A curva ROC do conjunto de teste também mostra um bom desempenho, com um comportamento semelhante ao observado no conjunto de treino. A curva está bem acima da linha diagonal (linha de não-discriminação), o que demonstra que o modelo mantém uma alta taxa de verdadeiros positivos para uma baixa taxa de falsos positivos. Além do mais a AUC também próxima de 1 indica um bom desempenho.

As curvas ROC e as AUCs para os conjuntos de treino e teste indicam um desempenho sólido e consistente do modelo. A similaridade entre as curvas ROC e as altas AUCs próximas de 0.93 em ambos os conjuntos sugerem que o modelo está bem treinado e é capaz de generalizar eficientemente para novos dados. Este tipo de desempenho é altamente desejável em aplicações práticas, pois garante que o modelo mantém sua eficácia mesmo quando aplicado a dados não vistos anteriormente.

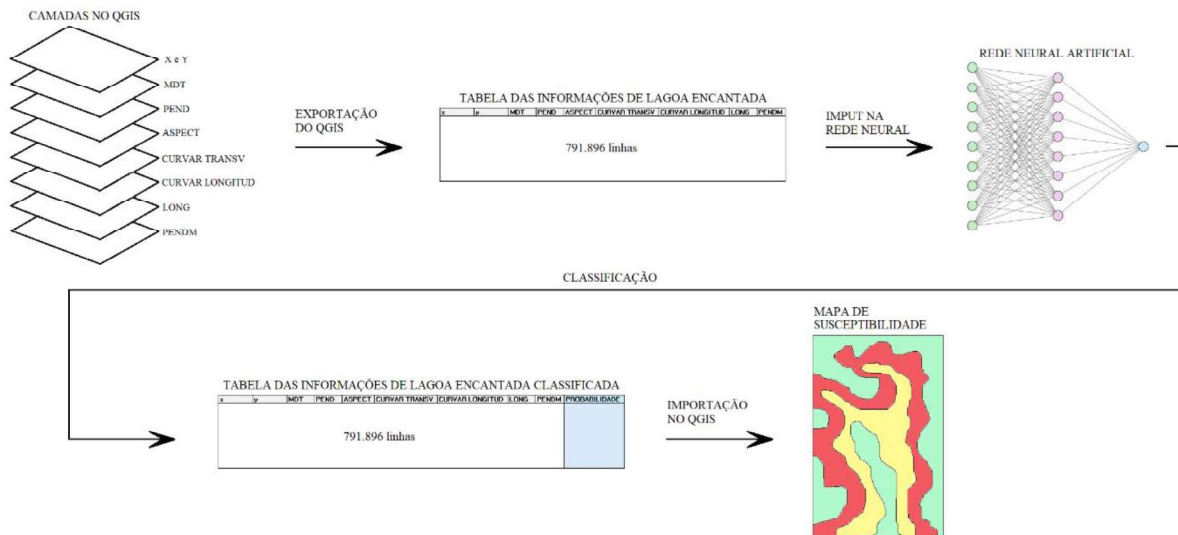
5.2.7 Mapa de Susceptibilidade

Com o modelo de rede neural obtido, avaliado, aprovado e exportado para uso em produção, resta apenas classificar todos os dados disponíveis na região de Lagoa Encantada. Para isso, os valores de todas as variáveis foram exportados do QGIS em formato de tabela, de modo que cada linha dessa tabela corresponde a um pixel da área de estudo.

O próximo passo envolveu o uso do modelo treinado para classificar cada linha (pixel) dessa tabela, gerando previsões para cada célula. Após a classificação, os valores previstos para cada célula foram importados de volta para o QGIS, permitindo uma

visualização geoespacial dos resultados. É importante destacar que a etapa de classificação retorna a probabilidade de um pixel pertencer à classe instável. A seguir o fluxo de trabalho realizado.

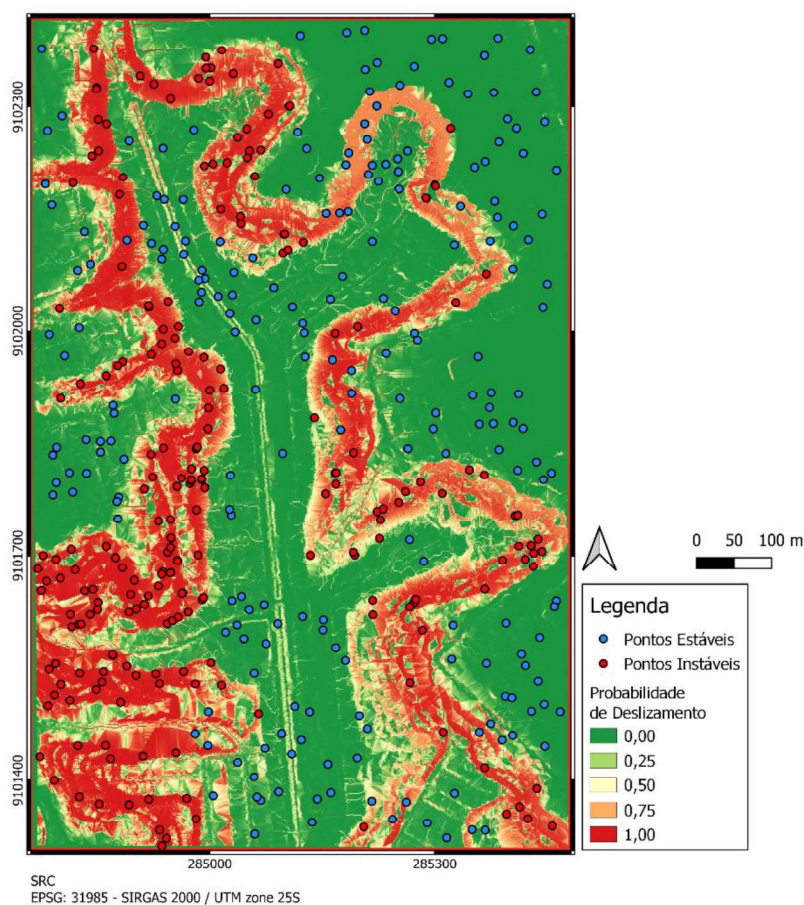
Figura 53 - Fluxo de obtenção do mapa de susceptibilidade



Fonte: Autor.

A figura 54 representa o resultado do mapa de susceptibilidade a deslizamentos de terra na região de Lagoa Encantada, elaborado com base exclusivamente nas características geomorfológicas do terreno. Este mapa proporciona uma visão detalhada das áreas mais propensas a deslizamentos, facilitando a identificação de zonas de risco e contribuindo para a tomada de decisões em estratégias de mitigação e prevenção.

Figura 54 - Mapa de susceptibilidade de Lagoa Encantada



Fonte: Autor.

O mapa utiliza uma escala de cores para representar a probabilidade de deslizamento de terra, proporcionando uma visualização intuitiva e clara das áreas de risco. A legenda de cores é a seguinte:

- Verde a Verde Claro (0 a 0,25): Indica áreas com baixíssima probabilidade de deslizamento.
- Verde Claro a Amarelo (0,25 a 0,50): Indica áreas com probabilidade moderada de deslizamento.
- Amarelo a Laranja (0,50 a 0,75): Indica áreas com alta probabilidade de deslizamento.
- Laranja a Vermelho (0,75 a 1): Indica áreas com altíssima probabilidade de deslizamento.

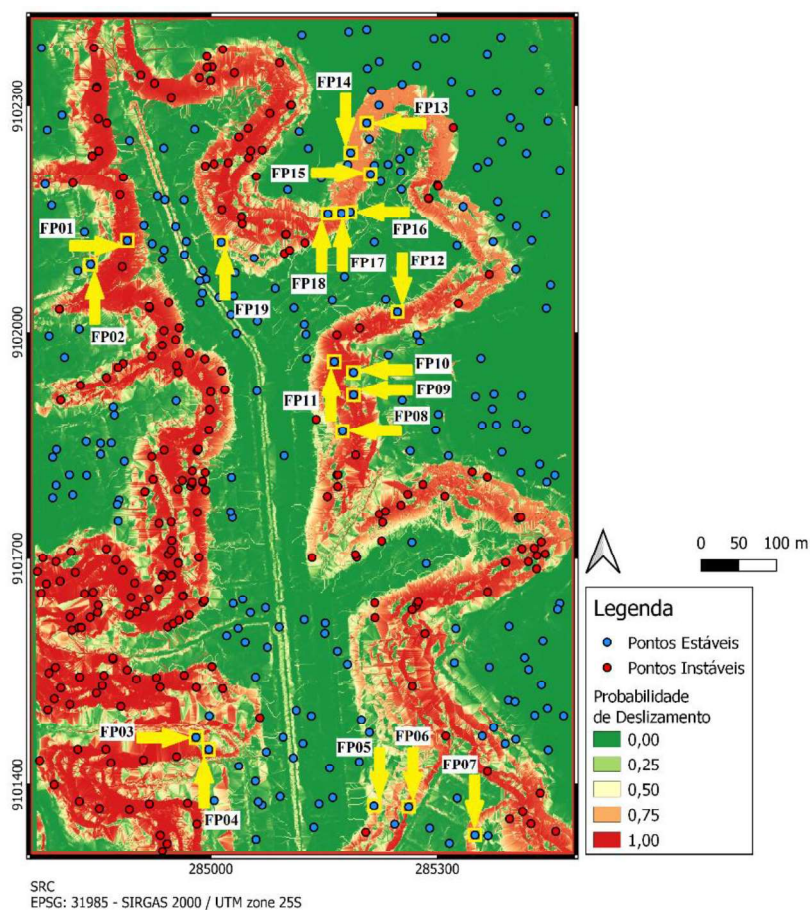
Analisando de forma visual a distribuição das cores no mapa de susceptibilidade a deslizamentos e os mapas das variáveis geomorfológicas, nota-se que as áreas de maior risco estão predominantemente associadas a regiões com maior declividade do terreno e maior

inclinação da bacia de contribuição. A concentração de pontos vermelhos nas áreas destacadas em vermelho reforça a validade do modelo, confirmando que essas zonas são de fato mais propensas a deslizamentos.

5.2.8 Análise dos Falsos Positivos

Considerando os pontos estáveis no mapa, identificamos dezenove erros do tipo falso positivo (figura 55), sendo que dezesseis deles pertencem ao conjunto de treino e três ao conjunto de teste.

Figura 55 - Mapa de susceptibilidade: Falsos positivos



Fonte: Autor.

Tabela 10 - Falsos Positivos e suas características

ID	x	y	MDE	PEND	ASPECT	PROFILE CURVATURE	PLANFORM CURVATURE	LONG	PENDM	CONJUNTO
FP01	284889,378	9102121,49	42,98778	37,35899	93,21537	0,00381	0,03339	2,36963	32,43357	TESTE
FP02	284840,663	9102089,2	60,31866	24,3633	124,32164	0	0	10,23996	14,81612	TREINO
FP03	284980,271	9101462,41	35,9515	30,46024	108,1808	-0,00721	0,07772	1,08715	23,09759	TREINO
FP04	284997,299	9101446,84	32,65622	15,58135	132,7534	0,03513	0,07907	1,44666	12,15684	TESTE
FP05	285215,944	9101371,98	22,98083	22,37522	301,68365	0	0	6,10446	22,0777	TREINO
FP06	285262	9101370,62	35,7469	19,88178	284,72256	-0,00325	0,00324	12,72089	18,39255	TREINO
FP07	285349,788	9101333,95	32,91667	49,29432	109,35722	-0,02745	0,00801	9,63452	35,65519	TREINO
FP08	285174,475	9101868,58	37,33	19,38415	271,22137	-0,40151	-0,27703	46,47328	27,9011	TREINO
FP09	285189,007	9101916,68	51,12067	31,2011	199,16403	-0,08259	-0,0722	18,83198	29,97116	TREINO
FP10	285188,905	9101946,94	57,19937	16,56819	328,5303	0,00336	-0,00337	11,53396	11,21516	TREINO
FP11	285163,538	9101960,94	40,09154	30,94357	339,49567	0	0	6,9986	30,98361	TREINO
FP12	285247,205	9102027,28	43,312	44,84474	349,10431	-0,47951	-0,04182	21,62041	44,52476	TREINO
FP13	285206,575	9102277,38	60,59685	36,89839	119,31545	0	0	14,16124	36,97286	TREINO
FP14	285185,162	9102238,21	65,86714	46,84626	99,83249	-0,06073	-0,02022	2,7492	43,32418	TESTE
FP15	285211,562	9102209,8	47,65333	26,90049	108,18489	-0,05023	-0,0271	29,07394	32,88855	TREINO
FP16	285184,422	9102159,9	44,1775	47,55088	125,72823	0,08635	0,20878	3,93664	44,07023	TREINO
FP17	285172,932	9102157,98	52,03111	43,23889	139,60063	0,02701	-0,02701	27,97663	27,5735	TREINO
FP18	285154,985	9102157,43	60,05333	27,85598	159,01724	0,09935	0,25437	0	0	TREINO
FP19	285013,757	9102119,33	32,64	36,70915	225,78108	0,26902	0,10972	8,15354	29,86062	TREINO

Fonte: Autor.

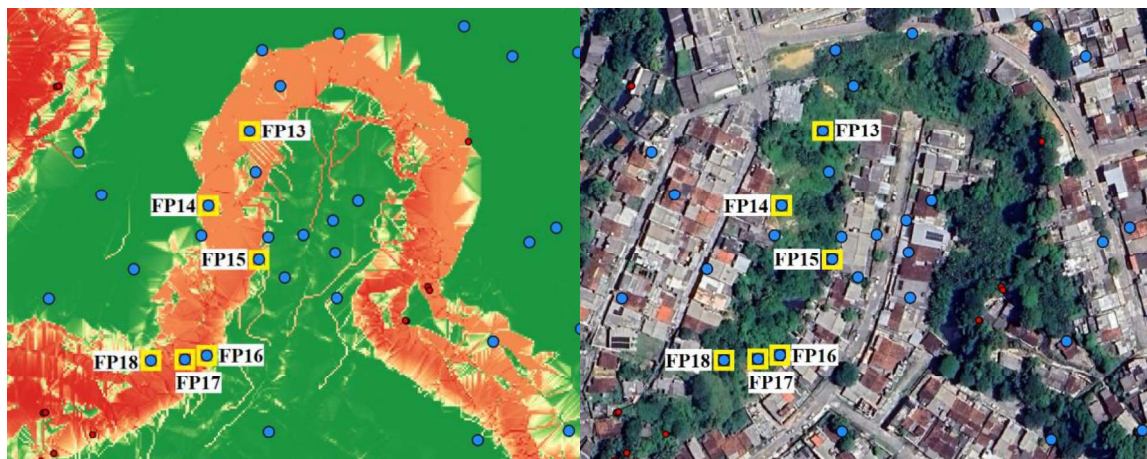
Considerando uma análise entre PENDM e LONG, alguns falsos positivos, como FP05, FP07 e FP16, ocorreram em regiões onde a declividade da bacia é alta (acima de 35°), mas o comprimento da bacia é relativamente curto, inferior a 20 metros. Esse comportamento sugere que o modelo está superestimando o impacto de bacias pequenas e íngremes na geração de deslizamentos, desconsiderando que bacias curtas geralmente acumulam pouca água, o que reduz a probabilidade de instabilidade do solo.

Na análise da variável PEND, observou-se que os falsos positivos ocorrem em uma ampla faixa de inclinações, abrangendo desde declividades moderadas até elevadas, como demonstrado nos pontos FP04 (15,58°), FP02 (24,36°) e FP07 (49,29%). Isso indica que o modelo pode estar superestimando o risco de deslizamento em áreas com declividades mais acentuadas, sem distinguir corretamente entre regiões instáveis e aquelas que, apesar de inclinações semelhantes, permanecem estáveis. Fatores como coesão do solo e a presença de cobertura vegetal podem manter a estabilidade em terrenos moderadamente inclinados, e a falta de uma análise mais integrada dessas variáveis pode estar resultando em falsos positivos em áreas de declividade moderada.

Há uma particularidade nos falsos positivos localizados no canto superior direito da área de estudo (figura 56). O mapa indica que essa área coberta por vegetação é instável, enquanto o histórico de deslizamentos demonstra que é uma área estável, caracterizada por densa vegetação, ausência de uso e ocupação do solo, e falta de ocorrências de deslizamentos.

Essa discrepância ressalta que, com os dados atualmente utilizados, tais erros são inevitáveis, carecendo de mais informações que caracterizem o local. Como a utilização da variável vegetação não utilizada por indisponibilidade.

Figura 56 - Falsos positivos: Área de vegetação

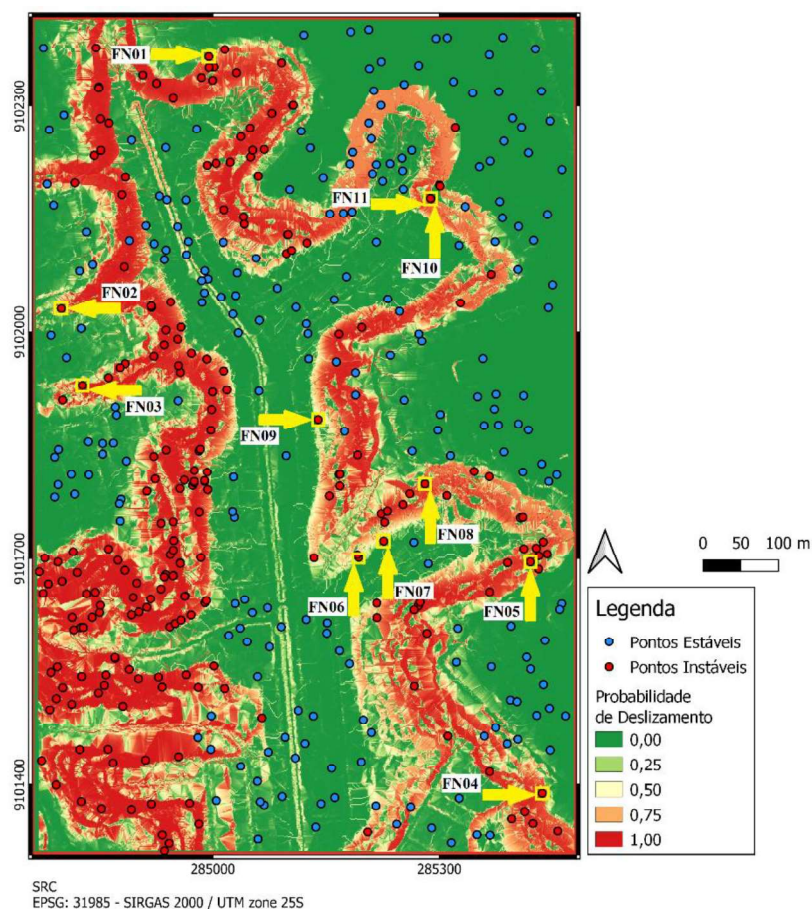


Fonte: Autor.

5.2.9 Análise dos Falsos Negativos

Em relação aos pontos instáveis, observamos onze erros do tipo falso negativo (figura 57), dos quais oito são do conjunto de treino e três do conjunto de teste.

Figura 57 - Mapa de susceptibilidade: Falsos negativos



Fonte: Autor.

Tabela 11 - Falsos Negativos e suas características

ID	x	y	MDE	PEND	ASPECT	PROFILE CURVATURE	PLANFORM CURVATURE	LONG	PENDM	CONJUNTO
FN01	284994,636	9102365,65	63,33017	15,95413	152,11676	0	0	10,28527	15,74217	TREINO
FN02	284799,374	9102030,76	61,03416	17,87519	21,12806	0,03768	0,04315	0	0	TESTE
FN03	284861,54	9101940,13	59,755	12,80927	141,89359	0,33201	0,22392	0	0	TREINO
FN04	285436,178	9101388,59	63,05119	5,64539	259,37695	-0,03354	-0,09099	28,38366	18,6104	TESTE
FN05	285420,631	9101695,47	58,615	9,79829	337,89038	0,12699	-0,067	5,43183	4,61509	TESTE
FN06	285193,16	9101701,47	27,53404	20,65628	162,72963	0,06588	-0,06587	12,18883	26,59158	TREINO
FN07	285226,234	9101723,42	28,14333	15,10783	142,20627	-0,00689	0,02777	8,21137	30,5561	TREINO
FN08	285280,879	9101798,35	56,03	5,30403	187,02032	0,1473	0,07573	0	0	TREINO
FN09	285139,48	9101884,28	25,52	22,53083	229,84779	-0,20197	-0,09329	16,18111	27,79843	TREINO
FN10	285288,793	9102177,55	60,28	20,71089	243,20926	0,02538	0,06379	0	0	TREINO
FN11	285288,18	9102177,86	60,28	20,71089	243,20926	0,02538	0,06379	0	0	TREINO

Fonte: Autor.

A variável PEND é amplamente reconhecida como uma das variáveis mais importantes para a previsão de deslizamentos, uma vez que terrenos íngremes estão diretamente associados a um maior risco de instabilidade gravitacional. No entanto, ao analisar os falsos negativos identificados, observamos que muitos dos pontos estão

localizados em áreas de baixa declividade. Pontos como FN03, FN04, FN05 e FN08 apresentam valores de declividade relativamente baixos, sugerindo que o modelo pode estar subestimando o risco de deslizamento em áreas com essa característica

Embora deslizamentos de terra possam ocorrer em locais com baixa declividade, a rede neural foi predominantemente treinada com pontos instáveis situados em áreas de declividade considerável. Para contornar esses tipos de classificações errôneas, seria benéfico adicionar mais características ao modelo, como espessura e formação do solo, uso e ocupação do solo, e vegetação. Essas informações adicionais podem fornecer mais pistas para que a rede neural faça classificações mais precisas e assertivas, melhorando assim a capacidade de prever deslizamentos em diversas condições.

5.2.10 Comparação Entre Mapas de Suscetibilidade

O mapa de suscetibilidade gerado neste trabalho foi desenvolvido com base em variáveis estritamente geomorfológicas, incluindo o Modelo Digital de Elevação (MDE), declividade do terreno, aspecto, curvaturas longitudinal e transversal além do comprimento e da declividade da bacia de contribuição a montante. Essas variáveis processadas por uma rede neural artificial, resultando em um mapa contínuo de probabilidades de deslizamento, variando de 0 a 1. Esse mapa oferece uma representação do risco de deslizamento para a comunidade de Lagoa Encantada, permitindo uma análise mais precisa das áreas suscetíveis a desastres naturais, com base em características do terreno.

Por outro lado, o mapa de risco apresentado no estudo de Coutinho *et al.* (2023) foi elaborado por meio de uma análise multicritério, integrando diversos indicadores geológico-geotécnicos, geomorfológicos, e de uso e ocupação do solo (ver Tabela 12). Entre os indicadores utilizados estão a formação geológica, as características pedológicas, o tipo e morfologia do relevo, a extensão e a inclinação das encostas, além do uso do solo, que considera aspectos como a categoria de ocupação, densidade populacional e a presença de sistemas de drenagem. Esse conjunto de variáveis permite uma avaliação mais abrangente do risco de deslizamento para a comunidade de Lagoa Encantada, levando em conta não apenas as características físicas do terreno, mas também a forma como ele é utilizado e alterado pela atividade humana, ampliando a compreensão dos fatores que influenciam a suscetibilidade da área.

Tabela 12 - Indicadores e subindicadores de susceptibilidade a movimento e transporte de massa

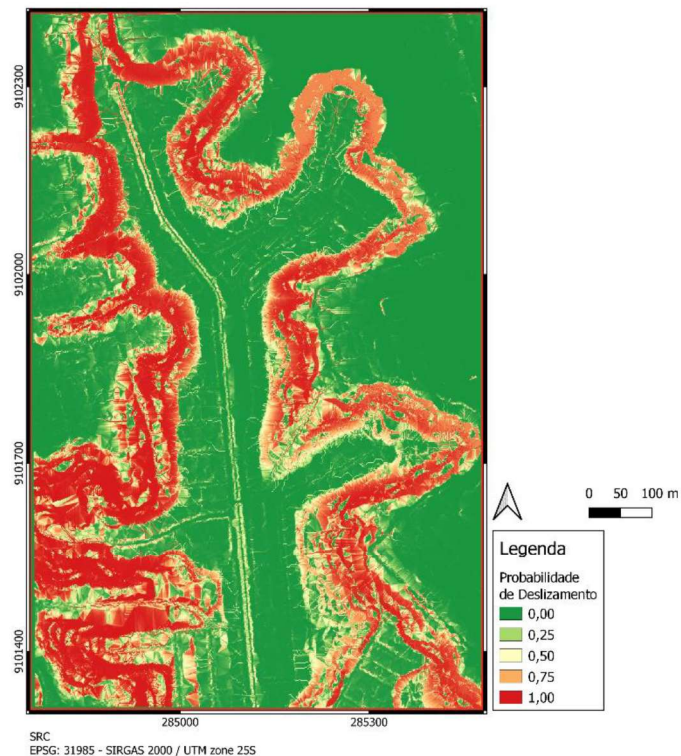
Geológico-Geotécnicas
Formação geológica
Feições Estruturais
Perfil geotécnico (estrutura, litologia e textura)
Características pedológicas
Evidências de movimentos
Geomorfologia
-Tipo relevo
-Visão geral encosta: altura (m), extensão (m) e inclinação (°);
-Classe de curvatura da encosta: vertical e horizontal;
-Visão talude de corte/aterro da encosta: altura (m), extensão (m), largura (m) e inclinação (°).
Uso e Ocupação do solo
-Categoria de ocupação: estágio de ocupação, modo de ocupação, predominância das edificações, padrão das edificações;
-Cobertura superficial: tipo de cobertura superficial, percentual de desmatamento;
-Condição de tratamento: condição do revestimento da encosta, percentual da encosta que apresenta revestimento, condição da estrutura de contenção e percentual, condição do sistema de drenagem e percentual;
-Fatores antrópicos: sistema de drenagem superficial, direcionamento do sistema de coleta de água de chuva dos telhados, destino do esgoto, abastecimento de água e fatores agravantes.

Fonte: Coutinho *et al.* (2023).

A seguir pode ser verificado os mapas em questão nas figuras 58 e 59.

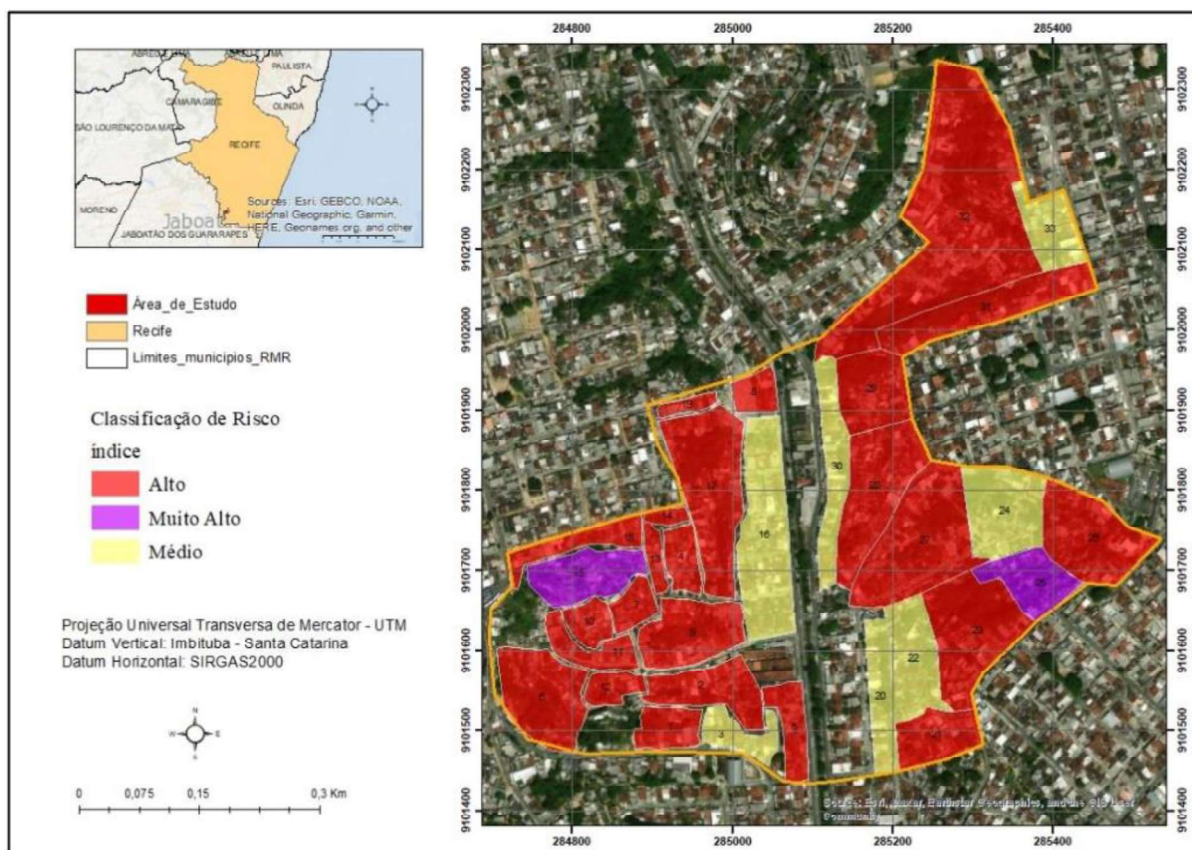
Figura 58 - Mapa de susceptibilidade a deslizamentos de terra gerado por RNA

Mapa de Susceptibilidade a Deslizamento de Terra



Fonte: Autor.

Figura 59 - Mapa de susceptibilidade a deslizamentos de terra criado por Coutinho



Fonte: Coutinho *et al.* (2023).

A análise espacial dos mapas evidencia a concentração de áreas de alto risco e suscetibilidade em regiões de relevo acidentado e encostas íngremes. No mapa de suscetibilidade deste estudo, as áreas com maior probabilidade de deslizamento, representadas em vermelho, coincidem em grande parte com as zonas classificadas como "Alto" e "Muito Alto" risco no trabalho de Coutinho *et al.* (2023).

Entretanto, observa-se que algumas áreas classificadas como de alto risco no mapa de Coutinho não apresentam a mesma intensidade de probabilidade de deslizamento no mapa de suscetibilidade geomorfológica. Essa diferença pode ser atribuída à diversidade de fatores considerados em cada abordagem. Enquanto o mapa de Coutinho *et al.* (2023) incorpora aspectos como o uso e ocupação do solo, a condição das estruturas de contenção e o estado dos sistemas de drenagem, o mapa de suscetibilidade baseia-se exclusivamente em variáveis geomorfológicas, como declividade e curvaturas do terreno.

A convergência entre os dois mapas é notável em áreas de relevo acidentado, onde ambos identificam elevado risco de deslizamento, refletindo as características de encostas íngremes e variações topográficas que favorecem a instabilidade. No entanto, em regiões

urbanizadas ou com intervenções humanas, o mapa de Coutinho *et al.* (2023) frequentemente indica um risco mais elevado, destacando a relevância de fatores antrópicos na avaliação de risco de deslizamento, além das condições naturais do terreno.

6 CONCLUSÕES

O estudo desenvolvido abordou a problemática dos deslizamentos de terra, fenômeno natural que impacta significativamente diversas regiões do Brasil, especialmente áreas urbanizadas e vulneráveis. Através de uma análise minuciosa da literatura e da aplicação de metodologias de Inteligência Artificial, este trabalho propôs um protótipo de solução para o mapeamento de áreas suscetíveis a esses eventos.

A implementação computacional deste estudo foi realizada utilizando o QGIS e a linguagem de programação Python, ambos amplamente reconhecidos por sua versatilidade e robustez. O QGIS, uma plataforma de código aberto, destacou-se no processamento de dados geoespaciais, permitindo a manipulação, cálculo e análise detalhada da geomorfologia de Lagoa Encantada. Por outro lado, o Python, também de código aberto, foi essencial na implementação da Rede Neural, oferecendo bibliotecas gratuitas permitindo a criação de um fluxo de trabalho eficiente e integrado.

A biblioteca Scikit-learn desempenhou um papel crucial neste estudo, fornecendo ferramentas robustas para o treinamento e validação dos modelos de Redes Neurais Artificiais. Essa contribuição foi significativa para alcançar a alta precisão nos resultados obtidos. O modelo foi treinado com o objetivo de minimizar os falsos negativos e apresentou métricas de desempenho, como Acurácia, Precisão, Revocação e F1-Score, todas superiores a 90% tanto no conjunto de teste quanto no conjunto de treino. Esse desempenho é considerado satisfatório, especialmente considerando que os dados utilizados se limitaram às características geomorfológicas do local de estudo.

A comparação entre o mapa de susceptibilidade a deslizamentos e o mapa de Coutinho *et al.* (2023) demonstra a complementaridade entre abordagens baseadas em geomorfologia e aquelas que incluem fatores socioeconômicos e de ocupação do solo. Enquanto o mapa de susceptibilidade oferece uma visão detalhada da propensão natural do terreno a deslizamentos, o mapa de Coutinho *et al.* (2023) proporciona uma avaliação mais ampla, incorporando fatores antropogênicos que influenciam o risco.

Como pesquisa futura, visando obter resultados mais assertivos e precisos, é recomendável incluir variáveis adicionais como uso e ocupação do solo, vegetação, espessura e formação do solo. Com uma maior riqueza de características, a rede neural pode aprimorar a classificação e corrigir parte dos erros identificados anteriormente. A utilização de um Modelo Digital de Terreno (MDT) de alta qualidade e um banco de dados de deslizamentos

com pontos de instabilidade precisamente localizados nos locais exatos onde os deslizamentos ocorreram provavelmente resultará em melhorias significativas nos resultados.

A aplicação de Redes Neurais Artificiais demonstrou ser uma ferramenta eficaz na identificação de áreas de risco, superando limitações das metodologias tradicionais em termos de velocidade e abrangência de área, pela sua capacidade de aprendizado e generalização de resultados. A combinação de Inteligência Artificial e SIG representa um avanço significativo no campo da geotecnologia aplicada à gestão de riscos naturais.

7 REFERÊNCIAS BIBLIOGRÁFICAS

AALST, Wil MP et al. Process mining: a two-step approach to balance between underfitting and overfitting. *Software & Systems Modeling*, v. 9, p. 87-111, 2010.

ADITIAN, A.; KUBOTA, T.; SHINOHARA, Y. Comparison of GIS-based landslide susceptibility models using frequency ratio, logistic regression, and artificial neural network in a tertiary region of Ambon, Indonesia. *Geomorphology*, v. 318, p. 101–111, 2018.

AGGARWAL, Charu C. et al. *Neural networks and deep learning*. Cham: springer, 2018.

ALMEIDA, H. de M. et al. Relação de Fatores Ambientais e Antrópicos na Ocorrência de Deslizamentos no Município de São Luís–MA, BRASIL. In: *Proceedings of Safety, Health and Environment World Congress*. Cubatão, Brazil, COPEC-Science and Education Research Council. 2014. p. 126-129.

ALMEIDA, Lutiane Queiroz de. Riscos ambientais e vulnerabilidades nas cidades brasileiras: conceitos, metodologias e aplicações. *Coleção PROPG Digital (UNESP)*, 2012.

ALVES, Camila; COUTINHO, Katherine. Tragédia provocada por chuvas e deslizamento de barreiras é o maior desastre de Pernambuco do século 21, dizem especialistas. *Portal G1, Pernambuco*, 30 de maio de 2022. Disponível em: <<https://g1.globo.com/pe/pernambuco/noticia/2022/05/30/tragedia-provocada-por-chuvas-e-deslizamento-de-barreiras-e-o-maior-desastre-de-pernambuco-do-seculo-21-dizem-especialistas.ghtml>>. Acesso em: 24 de out. de 2023.

AMORIM, Samuel. Estudio comparativo de métodos para la evaluación de la susceptibilidad del terreno a la formacion de deslizamientos superficiales: Aplicación al Pirineo Oriental. Tese de Doutorado em Engenharia Civil, Universitat Politècnica de Catalunya, Barcelona, 2012.

APURV, Tushar et al. Impact of climate change on floods in the Brahmaputra basin using CMIP5 decadal predictions. *Journal of Hydrology*, v. 527, p. 281-291, 2015.

AVEN, T. *Risk Analysis*. 2. ed. Chichester: John Wiley & Sons, 2015.

AZEVEDO, Caio da Silva. Uso de redes neurais artificiais na elaboração de mapas de suscetibilidade para escorregamentos planares rasos. 2022. Tese de Doutorado. Universidade de São Paulo.

BAI, Yuhan. RELU-function and derived function review. SHS Web of Conferences. EDP Sciences, 2022. p. 02006.

BISHOP, Christopher M.; NASRABADI, Nasser M. Pattern recognition and machine learning. New York: springer, 2006.

BRADLEY, Andrew P. The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern recognition, v. 30, n. 7, p. 1145-1159, 1997.

BRAGAGNOLO, Lucimara; DA SILVA, Roberto Valmir; GRZYBOWSKI, José Mario Vicensi. Mapeamento da suscetibilidade a deslizamentos com Redes Neurais Artificiais a partir do módulo gratuito e de código aberto r. landslide. In: XII Simpósio de Práticas de Engenharia Geotécnica da Região Sul. 2019.

BRASIL. Agência Gov. Ministério das Cidades vai financiar 20 planos municipais de redução de riscos. Brasília, 2023. Disponível em:<<https://agenciagov.ebc.com.br/noticias/202310/ministerio-das-cidades-vai-financiar-20-planos-municipais-de-reducao-de-riscos>>. Acesso em: 01 de nov. de 2023.

BRAUN, Anika et al. Landslide susceptibility mapping in Tegucigalpa, Honduras, using data mining methods. In: IAEG/AEG Annual Meeting Proceedings, San Francisco, California, 2018-Volume 1: Slope Stability: Case Histories, Landslide Mapping, Emerging Technologies. Springer International Publishing, 2019. p. 207-215.

CARVALHO, Celso Santos; MACEDO, Eduardo Soares de; OGURA, Agostinho Tadashi. Mapeamento de riscos em encostas e margem de rios. Brasília: Ministério das Cidades, 2007.

CAVALCANTI, Rogério; ANDRADE, Pollyana; FERREIRA, Rodrigo; SILVA, Ricardo; CARVALHO, João. LAGOA ENCANTADA: CAUSAS E CONSEQUÊNCIAS DOS MOVIMENTOS DE MASSA, UM EXEMPLO NA PERIFERIA DO RECIFE/PE. III Simpósio Nacional de Geomorfologia, v. 3, 2010.

CEMADEN - Centro Nacional de Monitoramento e Alerta de Desastres Naturais. Municípios monitorados, 2023. Disponível em: <<http://www2.cemaden.gov.br/municipios-monitorados-2/>>. Acesso em: 25 de out. de 2023.

CEMADEN - Centro Nacional de Monitoramento e Alerta de Desastres Naturais. Parceiros, 2023. Disponível em: <<http://www2.cemaden.gov.br/parceiros/>>. Acesso em: 25 de out. de 2023.

CEPED UFSC - Centro de Estudos e Pesquisas em Engenharia e Defesa Civil. Atlas brasileiro de desastres naturais: 1991 a 2012. Florianópolis: CEPED UFSC, 2013. Disponível em: <https://www.ceped.ufsc.br/wp-content/uploads/2012/01/AMAZONAS_mioloWEB.pdf>; Acesso em: 24 de out. de 2023.

CODA, Felipe A. et al. Modelagem e análise de uma arquitetura do sistema de aquisição de big data no contexto da Indústria 4.0. In: Congresso Brasileiro de Automática-CBA. 2020.

COUTINHO, Roberto Quental et al. Mapeamento de áreas de risco de deslizamento na comunidade lagoa encantada, Recife-PE. CONTRIBUCIONES A LAS CIENCIAS SOCIALES, v. 16, n. 10, p. 23106-23124, 2023.

CRUDEN, David Milne. Cruden, DM, Varnes, DJ, 1996, Landslide Types and Processes, Transportation Research Board, US National Academy of Sciences, Special Report, 247: 36-75. Landslides Eng. Pract, v. 24, p. 20-47, 1993.

DA SILVA, Adriana Cassiano et al. Mudanças do uso e ocupação da terra a partir da expansão urbana e as influências climáticas sobre a morfodinâmica no município do Recife/PE entre 1975 e 2015. Revista Cerrados (Unimontes), v. 13, n. 1, p. 33-49, 2015.

DA SILVA, Aline Pimentel; BARROSO, Emílio Velloso; POLIVANOV, Helena. Índices pluviométricos críticos para prevenção de desastres por deslizamentos na cidade de Niterói, RJ. Geologia USP. Série Científica, v. 22, n. 3, p. 47-60, 2022.

DA SILVA, Eduardo Almeida et al. ANÁLISE DA VARIABILIDADE TEMPORAL DA PRECIPITAÇÃO EM RECIFE-PE. RECIMA21-Revista Científica Multidisciplinar-ISSN 2675-6218, v. 2, n. 7, p. e27512-e27512, 2021.

DE ASSIS MOREIRA, Richard Eustáquio; BARBIERI, Alisson Flávio. Populações em Áreas de Risco de Deslizamentos em Belo Horizonte, Brasil, 2020.

DE LIMA SANTOS, David et al. Limiar de precipitação com potencial de gerar deslizamentos nos complexos de risco em uma zona urbana. Revista de Gestão de Água da América Latina, v. 16, n. 2019, 2019.

DE SANTANA, John Kennedy Ribeiro et al. ANÁLISE DA EVOLUÇÃO DAS ÁREAS DE PERIGO A ESCORREGAMENTOS NO BAIRRO DA MACAXEIRA, RECIFE (PE). REVISTA EQUADOR, v. 8, n. 2, p. 240-256, 2019.

DE SOUZA, Werônica Meira et al. Áreas de risco mais vulneráveis aos desastres decorrentes das chuvas em Recife-PE. Brazilian Journal of Environmental Sciences (RBCIAMB), n. 34, p. 79-94, 2014.

DIAS, Fernando Peres; HERRMANN, ML de P. Susceptibilidade a deslizamentos: estudo de caso no bairro Saco Grande, Florianópolis-SC. Caminhos de Geografia, v. 3, n. 6, p. 57-73, 2002.

DIAZ, Gonzalo I. et al. An effective algorithm for hyperparameter optimization of neural networks. IBM Journal of Research and Development, v. 61, n. 4/5, p. 9: 1-9: 11, 2017.

EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA – EMBRAPA. Clima: Recursos naturais da Mata Sul Pernambucana, 2021. Disponível em: <https://www.embrapa.br/agencia-de-informacao-tecnologica/territorios/territorio-mata-sul-pernambucana/caracteristicas-do-territorio/recursos-naturais/clima>. Acesso em: 11 julho 2024.

FAWCETT, Tom. An introduction to ROC analysis. Pattern recognition letters, v. 27, n. 8, p. 861-874, 2006.

FREITAS, Carlos Machado de et al. Desastres naturais e saúde: uma análise da situação do Brasil. Ciência & Saúde Coletiva, v. 19, p. 3645-3656, 2014.

FRIGERIO, Ivan et al. A GIS-based approach to identify the spatial variability of social vulnerability to seismic hazard in Italy. Applied geography, v. 74, p. 12-22, 2016.

GAMA, Joao. A survey on learning from data streams: current and future trends. *Progress in Artificial Intelligence*, v. 1, p. 45-55, 2012.

GAMEIRO, Samuel. Capacidade de generalização e extrapolação espacial de redes neurais artificiais no mapeamento a suscetibilidade de deslizamentos. 2020.

GÉRON, Aurélien. Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow. O'Reilly Media Inc., ed. 2, 2022.

GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. Deep learning. MIT press, 2016.

GOSWAMI, Saptarsi et al. A review on application of data mining techniques to combat natural disasters. *Ain Shams Engineering Journal*, v. 9, n. 3, p. 365-378, 2018.

HAHNLOSER, Richard; SARPESHKAR, Rahul; MAHOWALD, Misha; DOUGLAS, Rodney; SEUNG, Sebastian. Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit. *NATURE*, v. 405, p. 947 – 951, 2000.

HARMENDANI, Pedro; ZANETTI, Márcia. Aprendizado de máquina por reforço aplicado no Jogo de Cartas Uno. *Revista de Sistemas e Computação*, Salvador, v. 9, n. 2, p. 245-251, 2019.

HASTIE, Trevor et al. The elements of statistical learning: data mining, inference, and prediction. New York: springer, 2008.

HAYKIN, Simon. Neural networks: a comprehensive foundation. Prentice Hall PTR, 1998.

HAYKIN, Simon. Neural networks: A guided tour. *Nonlinear Biomedical Signal Processing*, v. 1, p. 53-68, 2000.

HEINE, Reuben A.; LANT, Christopher L.; SENGUPTA, Raja R. Development and comparison of approaches for automated mapping of stream channel networks. *Annals of the Association of American Geographers*, v. 94, n. 3, p. 477-490, 2004.

HIGHLAND, Lynn M.; BOBROWSKY, Peter. O Manual de Deslizamento—Um Guia para a Compreensão de Deslizamentos. Reston, Virginia, US Geological Survey Circular, v. 1325, 2008.

HIGHLAND, Lynn. Landslide types and processes. 2004.

HOLANDA, Maria Júlia de Oliveira. Solos colapsíveis e expansivos no Brasil: classificação da suscetibilidade de ocorrência aplicando redes neurais artificiais. 2022.

HUNGR, Oldrich; LEROUEIL, Serge; PICARELLI, Luciano. The Varnes classification of landslide types, an update. Landslides, v. 11, p. 167-194, 2014.

IBGE – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. Susceptibilidade a deslizamentos no Brasil: Primeira aproximação. Rio de Janeiro: IBGE, 2019. Disponível em: <<https://biblioteca.ibge.gov.br/index.php/biblioteca-catalogo>>; Acesso em: 24 de out. de 2023.

INPE - Instituto Nacional de Pesquisas Espaciais. I ESCUELA DE PRIMAVERA SOBRE SOLUCIONES ESPACIALES PARA EL MANEJO DE DESASTRES NATURALES Y RESPUESTAS DE EMERGENCIAS INUNDACIONES. Rio Grande do Sul: INPE, 2008. Disponível em: <<http://www3.inpe.br/crs/crectalc/escuela.php>>; Acesso em: 24 de out. de 2023.

IPCC - Painel Intergovernamental sobre Mudanças Climáticas. Managing the risks of extreme events and disasters to advance climate change adaptation. Estados Unidos da América: IPCC, 2012. Disponível em: <<https://www.ipcc.ch/report/managing-the-risks-of-extreme-events-and-disasters-to-advance-climate-change-adaptation/>>. Acesso em: 25 de out. de 2023.

IZBICKI, Rafael; SANTOS, Tiago. APRENDIZADO DE MÁQUINA: UMA ABORDAGEM ESTATÍSTICA. ed. 1, São Carlos, SP, 2020.

JORDAN, Michael I.; MITCHELL, Tom M. Machine learning: Trends, perspectives, and prospects. Science, v. 349, n. 6245, p. 255-260, 2015.

KAUFMAN, Dora. Inteligência Artificial e os desafios éticos: a restrita aplicabilidade dos princípios gerais para nortear o ecossistema de IA. PAULUS: Revista de Comunicação da FAPCOM, v. 5, n. 9, 2021.

KOVÁCS, Zsolt László. Redes neurais artificiais. Editora Livraria da Física, 2002.

KUHN, Max et al. Discriminant analysis and other linear classification models. Applied predictive modeling, p. 275-328, 2013

KULATHUNGA, Nalinda et al. Effects of the nonlinearity in activation functions on the performance of deep learning models. arXiv preprint arXiv:2010.07359, 2020.

LANTZ, Brett. Machine learning with R: expert techniques for predictive modeling. Packt publishing ltd, 2019.

LUCCHESI, Luisa Vieira. Modelagem de suscetibilidade e de limiares de precipitação para deslizamentos de terra utilizando métodos de aprendizagem de máquina. 2022.

LOPES, Kelvin S. et al. Behavioral Biometrics for Continuous Authentication on Mobile Devices: Anomaly Detection through Keystroke Dynamics with Machine Learning. In: Anais do XX Encontro Nacional de Inteligência Artificial e Computacional. SBC. p. 1129-1141. 2023.

MACEDO, ES de; SANDRE, L. H. Mortes por deslizamentos no Brasil: 1988 a 2022. Revista Brasileira de Geologia de Engenharia e Ambiental, v. 12, n. 1, p. 110-117, 2022.

MAGALHAES, Eduardo Nunes; FERREIRA, Gilmar Gonçalves; SOARES, Julia Aparecida. APLICAÇÃO DE REDES NEURAS ARTIFICIAS DO TIPO MULTI LAYER PERCEPTRONS (MLP) NA PROPOSIÇÃO DE UM NOVO MODELO PARA ESTUDO DE RESISTÊNCIA DE TALUDES: Application of artificial neural networks of the Multi Layer Perceptrons Type (MLP) in the proposal of a new model for study of slope resistance. Geosciences= Geociências, v. 41, n. 2, p. 347-358, 2022.

MARANDOLA JR, Eduardo; HOGAN, Daniel Joseph. Vulnerabilidades e riscos: entre geografia e demografia. Revista Brasileira de Estudos de População, v. 22, n. 1, p. 29-53, 2005.

MELO, Cristiane; CIRILO, José; AMORIM, Samuel; ALVES, Fellipe; GEDES, Paulo. AVALIAÇÃO DOS ÍNDICES DE PRECIPITAÇÃO NOS DESLIZAMENTOS DE TERRA NA COMUNIDADE DE LAGOA ENCANTADA (IBURA – RECIFE) ENTRE OS ANOS DE 2013 E 2017. XV Simpósio de Recursos Hídricos do Nordeste, v. 15, 2020.

MELO, Cristiane Ribeiro. UTILIZAÇÃO DE INFORMAÇÕES ESPACIAIS DE ALTA RESOLUÇÃO PARA ANÁLISE DE SUSCETIBILIDADE DE DESLIZAMENTO SOB SITUAÇÃO DE PRECIPITAÇÕES INTENSAS. Dissertação de Mestrado em Engenharia Civil – Universidade Federal de Pernambuco, Recife, 2021.

MODESTO, Akilas Girão et al. Mapeamento Tecnológico da Aplicação de Redes Neurais para Eficiência Energética de Sistemas de Bombeamento. REVISTA DE TECNOLOGIA APLICADA, v. 11, n. 3, p. 82-95, 2023.

MOLINA, Ellen AN; CARDOSO, Andréa O.; NOGUEIRA, Fernando R. Relação precipitação-deslizamento no Município de São Bernardo do Campo-SP. Ciência e Natura, v. 37, n. 1, p. 46-54, 2015.

MONARD, Maria Carolina; BARANAUSKAS, José Augusto. Conceitos sobre aprendizado de máquina. Sistemas inteligentes-Fundamentos e aplicações, v. 1, n. 1, p. 32, 2003.

NASCIMENTO, Kayo Renato Da Silva; ALENCAR, Marcelo Hazin. Management of risks in natural disasters: A systematic review of the literature on NATECH events. Journal of Loss Prevention in the Process Industries, v. 44, p. 347-359, 2016.

NG, Andrew Yan-Tak. Machine Learning Specialization. 2023. Curso disponível no YouTube. Imagem localizada em Course 1, Week 1, Lesson 4. Disponível em: https://www.youtube.com/watch?v=WtlvKq_zxPI. Acesso em: 10 jul. 2024.

NG, Andrew Yan-Tak. Machine Learning Specialization. 2023. Curso disponível no YouTube. Imagem localizada em Course 1, Week 3, Lesson 4. Disponível em: <https://www.youtube.com/watch?v=8upNQI-40Q8>. Acesso em: 10 jul. 2024.

NIELSEN, Michael A. Neural networks and deep learning. San Francisco, CA, USA: Determination press, 2019.

OLIVEIRA, G.G.; GUASSELLI, L.A.; QUEVEDO, R.P.; RUIZ, L.F.C.; BRESSANI, L.A.; RIFFEL, E.S. Identificação e análise de áreas suscetíveis a fluxos de detritos na bacia hidrográfica do Rio Taquari-Antas, RS. *Pesquisas em Geociências*, v. 45, n. 2, p. 1–25, 2018.

OLIVEIRA, Mateus Nicoladelli. UNIVERSIDADE DO ESTADO DE SANTA CATARINA–UDESC, CENTRO DE CIÊNCIAS TECNOLÓGICAS–CCT PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA–PPGEEL. Joinville, 2020.

MENDES, André Silva; TRIGO, Paulo M.; PAULO, Joel Preto. OTIMIZAÇÃO DE HÍPER-PARÂMETROS EM CLASSIFICADORES DE SOM. *Tecniacústica*. 2023.

PACHECO, M.; COUTINHO, R. Q. Avaliação de risco de deslizamentos nas encostas de petrópolis (RJ) e da região metropolitana de recife (PE). *Solos Rochas*, v. 29, p. 223-235, 2006.

PEREIRA, Raphael Crespo; AMORIM, Samuel França; CAVALCANTI, Eduardo Jorge Nunes. Aplicação de Redes Neurais Artificiais e Geoprocessamento para Previsão da Susceptibilidade de Deslizamentos Superficiais com o Uso de Programas Gratuitos, Recife/PE, 2020.

SANTOS, Hellen Geremias dos et al. Machine learning para análises preditivas em saúde: exemplo de aplicação para predizer óbito em idosos de São Paulo, Brasil. *Cadernos De saúde pública*, v. 35, p. e00050818, 2019.

ROCHA, Brenda Oliveira; KÖRTING, Thales Sehn. BOAS PRÁTICAS DE PROCESSAMENTO DE IMAGENS PARA MAPEAMENTO RÁPIDO DE DESLIZAMENTOS. In: XX Simpósio Brasileiro de Sensoriamento Remoto. Florianópolis, SC, 2023.

SANTOS, AR dos. Enchentes e deslizamentos: causas e soluções. Áreas de risco no Brasil. São Paulo: Pini, 2012.

SHALEV-SHWARTZ, Shai; BEN-DAVID, Shai. Understanding machine learning: From theory to algorithms. Cambridge university press, 2014.

SOUZA, Celia Regina de Gouveia et al. Desastres naturais: conhecer para prevenir. 1. Ed. São Paulo: Instituto Geológico, 2009. 196 p.

SOUSA, Willian Lino; DE MEDEIROS, Marcio Avelino; FREIRE, Lívio Antônio Melo. Modelo em Aprendizagem de Máquina para Determinação da Capacidade de Carga em Fundações Diretas. 2023.

SULAIMAN, Samia Nascimento; ALEDO, Antonio. Desastres naturais: convivência com o risco. Estudos Avançados, v. 30, p. 11-23, 2016.

TOMINAGA, L.K.; SANTORO, J.; AMARAL, R. (organização). 2009. Desastres Naturais: Conhecer para Prevenir. São Paulo: Instituto Geológico, 2009. 160 p. : il. color. ; 24 cm. ISBN 978-85-87235-09-1

TRALLI, David M. et al. Satellite remote sensing of earthquake, volcano, flood, landslide and coastal inundation hazards. ISPRS Journal of Photogrammetry and Remote Sensing, v. 59, n. 4, p. 185-198, 2005.

VARNES, David J. Slope movement types and processes. Special report, v. 176, p. 11-33, 1978.

XU, Jiuping et al. Natural disasters and social conflict: A systematic literature review. International journal of disaster risk reduction, v. 17, p. 38-48, 2016.

ZARE, Mohammad et al. Landslide susceptibility mapping at Vaz Watershed (Iran) using an artificial neural network model: a comparison between multilayer perceptron (MLP) and radial basic function (RBF) algorithms. Arabian Journal of Geosciences, v. 6, p. 2873-2888, 2013.

ZHANG, Haotian; ZHANG, Lin; JIANG, Yuan. Overfitting and underfitting analysis for deep learning based end-to-end communication systems. 2019 11th international conference on wireless communications and signal processing (WCSP). IEEE, 2019. p. 1-6.