

Uma proposta de aprimoramento na previsão de deslizamentos de terra com aprendizado de máquina

Renan Neves Rego¹, Thulio Aleixo Bezerra Brito¹,
Keisy Lizandra Silva Vasconcelos¹, Fernando Pontual de Souza Leão Júnior¹,
Jackson Raniel Florencio da Silva¹

¹Campus Caruaru - Universidade de Pernambuco (UPE)
Rod. BR 104, KM 62 - 55.002-971 - Caruaru – PE – Brasil

renan.nevesr, thulio.aleixo, keisy.lizandra
fernando.pontual, jackson.florencio@upe.br

Abstract. *In Brazil, millions of people live in areas at risk of floods and landslides, especially in the Metropolitan Region of Recife. The increasing frequency of natural disasters, such as the heavy rains that caused deaths and damage in 2021 and 2022, highlights the need for effective alert systems. This study proposes the implementation of dynamic triggers in a hydrometeorological monitoring and landslide warning system, using machine learning to improve prediction accuracy and strengthen urban resilience. The results show that combining algorithms such as SVM with oversampling techniques like ADASYN achieved a recall of 0.91 for the landslide class, correctly identifying 10 out of 11 critical events, demonstrating the model's potential to enhance event detection and mitigate risks through more precise monitoring. The main methodology used in this project is Design Science Research (DSR). In conclusion, the implementation of dynamic triggers, combined with the collection and standardization of relevant data, such as precipitation and landslide information, strengthens urban resilience and significantly contributes to the protection of lives and property. However, the study also emphasizes the need to expand the monitoring network, collect additional data, and improve the algorithms to ensure the system's long-term accuracy and effectiveness.*

Resumo. *No Brasil, milhões de pessoas vivem em áreas de risco de enchentes e deslizamentos de terra, especialmente na Região Metropolitana do Recife. A crescente frequência de desastres naturais, como as chuvas intensas que causaram mortes e danos em 2021 e 2022, destaca a necessidade de sistemas de alerta eficazes. Este trabalho propõe a implementação de gatilhos dinâmicos em um sistema de monitoramento hidrometeorológico e de alerta para deslizamentos de terra, utilizando aprendizado de máquina para melhorar a precisão das previsões e fortalecer a resiliência urbana. Os resultados demonstram que a combinação de algoritmos como SVM com técnicas de oversampling, como ADASYN, alcançou um recall de 0,91 para a classe de deslizamentos, identificando corretamente 10 de 11 eventos críticos, evidenciando o potencial do modelo em aprimorar a detecção desses eventos e mitigar riscos por meio de um monitoramento mais preciso. A metodologia utilizada neste projeto é a Pesquisa em Ciência do Design (DSR). Em conclusão, a implementação de gatilhos*

dinâmicos, combinada com a coleta e padronização de dados relevantes, como informações sobre precipitação e deslizamentos de terra, fortalece a resiliência urbana e contribui significativamente para a proteção de vidas e propriedades. No entanto, o trabalho também enfatiza a necessidade de expandir a rede de monitoramento, coletar dados adicionais e melhorar os algoritmos para garantir a precisão e a eficácia do sistema a longo prazo.

1. Introdução

No Brasil, aproximadamente 9 milhões de pessoas vivem em áreas de risco, sendo 75% suscetíveis a deslizamentos de terra [de Assis Dias et al. 2018]. Eventos extremos como esses causam significativas perdas humanas e econômicas, como observado em 2022 na Região Metropolitana do Recife, onde chuvas resultaram em cerca de 130 mortes e demandaram R\$ 142 milhões para assistência [Moraes 2023]. Nos anos de 2023, 2024 e 2025, houve apenas registros de deslizamentos, porém sem vítimas fatais.

Para mitigar tais impactos, a Conferência Mundial sobre Redução de Desastres propôs o Quadro de Ação de Hyogo, incentivando investimentos em sistemas de alerta [Marchezini 2017]. No Brasil, o Centro Nacional de Monitoramento e Alertas de Desastres Naturais (Cemaden) e o Centro Nacional de Gerenciamento de Riscos e Desastres (CENAD) monitoram variáveis ambientais e emitem alertas, mas a comunicação dos riscos à população ainda apresenta falhas.

Atualmente, o sistema de alerta utilizado dispõe de nove estações pluviométricas do Cemaden e duas da ANA. No entanto, apenas os dados de duas estações do Cemaden são efetivamente utilizadas. Essa limitação pode comprometer a precisão das previsões e a eficácia dos alertas emitidos.

Diante disso, este estudo busca aprimorar um sistema avançado de alerta para previsão de deslizamentos, integrando modelos matemáticos, sensores e aprendizado de máquina. Atualmente, a prefeitura de Jaboatão dos Guararapes utiliza o sistema OLHACHUVA, que se baseia em gatilhos estáticos fundamentados em dados históricos (2008-2021). No entanto, esse modelo apresenta limitações para lidar com as variações climáticas em curso, tornando necessária sua evolução para uma abordagem mais dinâmica e adaptável [Costa et al. 2023].

A proposta é implementar gatilhos dinâmicos baseados em dados em tempo real e algoritmos de aprendizado de máquina, com o objetivo de otimizar os alertas para deslizamentos. Essa abordagem visa aumentar a detecção de casos reais, reduzindo os alarmes falsos e aprimorando a capacidade de resposta das autoridades diante de eventos críticos. Para tal, utilizam-se séries temporais de deslizamentos, construídas a partir dos registros da Defesa Civil de Jaboatão dos Guararapes, seguindo metodologias de [Brunetti et al. 2010] e [Kanungo and Sharma 2014], e séries temporais de chuva acumulada baseadas em dados das estações do Cemaden.

O aprendizado de máquina permite identificar padrões e prever deslizamentos com maior precisão [Jordan and Mitchell 2015]. Entre os métodos empregados, destacam-se árvores de decisão e *Random Forest* para segmentação dos dados [Breiman 2001], técnicas de *Boosting* como AdaBoost e XGBoost para otimização preditiva [Friedman 2001], e SVM para classificação eficiente em espaços de alta dimensão.

nalidade [Cortes 1995]. A seleção de hiperparâmetros ocorre via *Grid Search*, garantindo um ajuste eficiente do modelo [Hutter et al. 2019].

A avaliação do desempenho dos modelos foi realizada por meio de métricas como matriz de confusão, acurácia, precisão, *recall* e *F1-score* [Han et al. 2012]. O objetivo principal é fortalecer a resiliência urbana frente a eventos climáticos extremos, otimizando a previsão e resposta a deslizamentos por meio da integração de dados hidrometeorológicos e técnicas avançadas de aprendizado de máquina.

Nas seções a seguir, a Seção 2 apresenta os principais trabalhos relacionados, destacando abordagens anteriores para previsão de deslizamentos de terra. A Seção 3 descreve a metodologia adotada neste estudo, incluindo a estruturação dos dados, os algoritmos utilizados e os critérios de avaliação. Em seguida, a Seção 4 apresenta e discute os resultados obtidos, comparando o desempenho dos modelos e analisando a eficácia das técnicas empregadas. A Seção 5 compara o desempenho do melhor modelo identificado com o critério estabelecido no projeto OLHACHUVA. Por fim, a Seção 6 traz as considerações finais, discutindo as contribuições do trabalho, suas limitações e possíveis direções para pesquisas futuras.

2. Trabalhos relacionados

Souza e Ebecken (2012) desenvolveram uma metodologia para prever deslizamentos no Rio de Janeiro utilizando Redes Neurais Artificiais, Regras de Classificação e o algoritmo Apriori. Trabalharam com um inventário de 1.033 eventos categorizados por bairro e 233 amostras de não deslizamento. A qualidade dos dados foi um desafio, devido à ausência de coordenadas geográficas precisas e ao desbalanceamento de classes.

Farahmand e Aghakouchak (2013) empregaram SVM para previsão de deslizamentos com dados de precipitação por satélite, mapas de uso do solo e informações topográficas. Utilizaram um inventário de 581 eventos e geraram 5.810 amostras negativas aleatórias. Aplicaram subamostragem aleatória 100 vezes e obtiveram um erro médio de 7%, com desafios na precisão da localização dos eventos.

Tehrani et al. (2019) aplicaram Regressão Logística e Árvores de Decisão em um conjunto de dados global da NASA, contendo 4.542 deslizamentos. Geraram uma quantidade equivalente de amostras negativas e analisaram atributos como precipitação, altitude e tipo de solo. Obtiveram uma AUC superior a 0,88, mas relataram grandes incertezas na geolocalização dos eventos.

Laedon et al. (2022) integraram múltiplas fontes de dados e compararam quatro algoritmos de aprendizado de máquina: RNA, Árvore de Decisão, Random Forest e LightGBM. Aplicaram validação cruzada estratificada e otimização por grid search. RF e LGBM apresentaram os melhores desempenhos, com F1 superior a 0,865. Destacaram desafios na definição de janelas temporais para precipitação e na avaliação de métodos para geração de amostras negativas.

O presente trabalho se diferencia por abordar diversas técnicas de balanceamento de dados, como SMOTE e ADASYN, aliadas a um conjunto diversificado de algoritmos, incluindo SVM, Árvore de Decisão, Gradient Boosting e XGBoost. Além disso, a otimização dos modelos é realizada por validação cruzada e holdout, permitindo uma análise robusta do desbalanceamento de dados e sua mitigação. A adaptação de métodos

para melhorar a qualidade dos dados e lidar com a imprecisão da precipitação reforça a eficácia e aplicabilidade do estudo na mitigação dos impactos dos deslizamentos de terra.

3. Metodologia

Esta pesquisa utilizou a abordagem Design Science Research, que foca na criação e avaliação de artefatos para solucionar problemas práticos [Hevner et al. 2004]. O estudo seguiu as seguintes etapas, iniciando com a identificação do problema, que neste caso, envolve a melhoria da eficiência de sistemas de monitoramento e alerta de deslizamentos por meio de aprendizado de máquina. A definição dos objetivos considerou o desenvolvimento de modelos preditivos [Peffers et al. 2008]. O *design* e desenvolvimento dos artefatos incluiu simulações para validar os gatilhos dinâmicos propostos [Gregor 2013], enquanto a avaliação comparou modelos de aprendizado de máquina com gatilhos estáticos [March 1995].

Para alimentar os modelos preditivos, foram utilizados dados pluviométricos do Centro Nacional de Monitoramento e Alertas de Desastres Naturais (CEMADEN), abrangendo o período de 2015 a 2022, e registros de deslizamentos da Defesa Civil de Jaboatão dos Guararapes.

O sistema de alerta dispõe de nove estações pluviométricas do CEMADEN e duas da Agência Nacional de Águas e Saneamento Básico (ANA). No entanto, apenas duas estações do CEMADEN foram utilizadas na modelagem, pois as demais apresentavam gaps significativos nos registros, comprometendo a consistência e continuidade das séries temporais.

Os registros de deslizamentos passaram por um processo de limpeza e normalização, incluindo a remoção de registros inconsistentes e a padronização das unidades de medida. A fusão dos dados pluviométricos e de deslizamentos garantiu um conjunto consistente para modelagem preditiva. A variável alvo foi definida como binária, classificando os dias com dez ou mais deslizamentos como eventos críticos (1) e os demais como não críticos (0).

O estudo identificou um forte desbalanceamento de classes, com predominância de registros sem deslizamentos (2072 casos) em relação aos eventos críticos (55 casos), o que pode comprometer a performance dos modelos [Blagus 2010]. Para mitigar esse problema, aplicamos as técnicas SMOTE [Nitesh et al. 2002] e ADASYN [He et al. 2008], que aumentam a representatividade da classe minoritária ao gerar exemplos sintéticos.

A divisão entre treino e teste seguiu a abordagem holdout (80/20), garantindo uma distribuição adequada dos dados [Kohavi 2001]. Como entrada, os modelos receberam o acumulado pluviométrico em diferentes períodos (24h, 3, 6, 7 e 15 dias), e a saída indicava a ocorrência ou não de deslizamentos. Foram testados os algoritmos SVM, Decision Tree, Random Forest, AdaBoost, Gradient Boosting e XGBoost, avaliados em três cenários distintos: sem tratamento, com SMOTE e com ADASYN. O uso do GridSearch garantiu a consistência na escolha dos hiperparâmetros, permitindo uma comparação justa entre os métodos.

4. Resultados

Tabela 1. Resultados dos Algoritmos com e sem Técnicas de Oversampling

Algoritmo	Técnica	Acurácia	Recall		Precisão		F1-score	
			Sem Deslizamento	Com Deslizamento	Sem Deslizamento	Com Deslizamento	Sem Deslizamento	Com Deslizamento
DT	Holdout	0.96	0.97	0.45	0.99	0.28	0.98	0.34
DT	SMOTE	0.90	0.91	0.55	0.99	0.14	0.95	0.23
DT	ADASYN	0.91	0.92	0.64	0.99	0.17	0.95	0.26
RF	Holdout	0.97	0.99	0.36	0.98	0.40	0.98	0.38
RF	SMOTE	0.91	0.92	0.55	0.99	0.16	0.95	0.24
RF	ADASYN	0.91	0.91	0.64	0.99	0.16	0.95	0.26
AdaBoost	Holdout	0.62	0.61	0.91	1.00	0.06	0.75	0.11
AdaBoost	SMOTE	0.62	0.61	0.91	1.00	0.06	0.75	0.11
AdaBoost	ADASYN	0.62	0.61	0.91	1.00	0.06	0.75	0.11
G.Boost	Holdout	0.95	0.96	0.55	0.99	0.27	0.97	0.36
G.Boost	SMOTE	0.93	0.94	0.64	0.99	0.22	0.96	0.33
G.Boost	ADASYN	0.91	0.91	0.73	0.99	0.18	0.95	0.29
XGBoost	Holdout	0.96	0.98	0.36	0.98	0.29	0.98	0.32
XGBoost	SMOTE	0.92	0.93	0.55	0.99	0.16	0.96	0.25
XGBoost	ADASYN	0.88	0.89	0.64	0.99	0.13	0.94	0.22
SVM	Holdout	0.96	0.98	0.27	0.98	0.27	0.98	0.27
SVM	SMOTE	0.81	0.82	0.73	0.99	0.10	0.90	0.17
SVM	ADASYN	0.79	0.79	0.91	1.00	0.10	0.88	0.18

Tabela 2. Verdadeiros Positivos, Falsos Positivos, Falsos Negativos e Verdadeiros Negativos

Algoritmo	Técnica	TP	FP	FN	TN
Decision Tree	Holdout	5	10	6	405
Decision Tree	SMOTE	6	5	5	379
Decision Tree	ADASYN	7	4	4	380
Random Forest	Holdout	4	7	7	409
Random Forest	SMOTE	6	5	5	383
Random Forest	ADASYN	7	4	4	379
AdaBoost	Holdout	10	1	1	252
AdaBoost	SMOTE	10	1	1	252
AdaBoost	ADASYN	10	1	1	252
GradientBoost	Holdout	6	5	5	399
GradientBoost	SMOTE	7	4	4	390
GradientBoost	ADASYN	8	3	3	379
XGBoost	Holdout	4	7	7	405
XGBoost	SMOTE	6	5	5	384
XGBoost	ADASYN	7	4	4	368
SVM	Holdout	3	8	8	407
SVM	SMOTE	8	3	3	339
SVM	ADASYN	10	1	1	326

As técnicas de oversampling foram aplicadas para lidar com o desbalanceamento da base de dados, especialmente para melhorar a predição da classe minoritária (deslizamento). Observou-se que o uso do ADASYN proporcionou um melhor equilíbrio entre as métricas avaliadas, aumentando o recall da classe de interesse. O SMOTE também melhorou o desempenho, mas apresentou limitações em alguns modelos. O SVM + ADASYN, como podemos analisar na Tabela 1, por exemplo, conseguiu aumentar o recall da classe de deslizamento de 0,27 (sem oversampling) para 0,91, enquanto o Random Forest + ADASYN melhorou de 0,36 para 0,64.

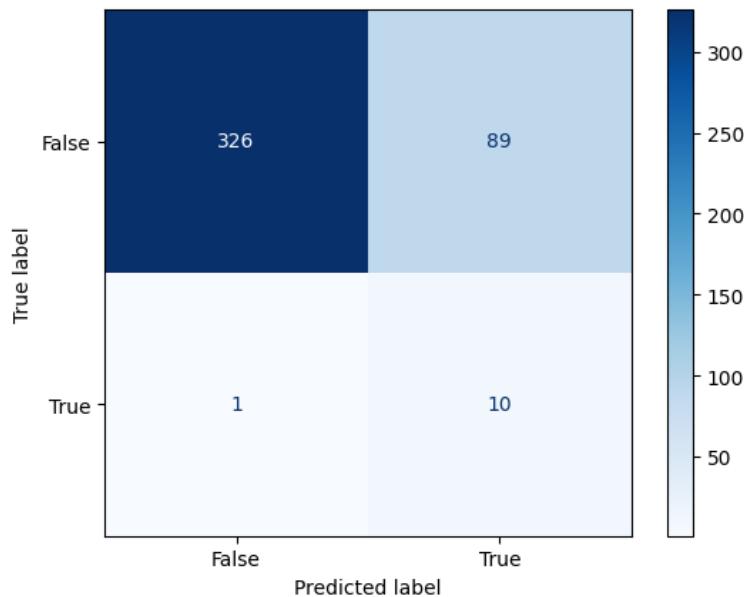
Os resultados indicam variações significativas no desempenho dos modelos testados. O SVM, quando combinado com o ADASYN, alcançou o melhor recall para a classe deslizamento (0,91), identificando corretamente 10 de 11 casos positivos, vide Tabela 2. O Gradient Boosting + ADASYN também apresentou um bom desempenho, com recall de 0,73 e precisão de 0,18 para a classe de deslizamento. O Random Forest, ao utilizar ADASYN, alcançou recall de 0,64 e precisão de 0,16, demonstrando uma melhora significativa em relação à versão sem oversampling (recall de 0,36).

Por outro lado, algoritmos como Decision Tree apresentaram dificuldades na predição da classe minoritária. Sem técnicas de oversampling, a árvore de decisão alcançou apenas 0,45 de recall para deslizamentos, melhorando para 0,64 com ADASYN. Já o Adaboost apresentou um recall relativamente alto (0,91) para a classe de deslizamento sem oversampling, mas com baixa precisão (0,06), resultando em um F1-score de apenas 0,11.

No contexto de detecção de deslizamentos, o recall é a métrica mais crítica, pois mede a capacidade do modelo de identificar corretamente os casos positivos. A baixa quantidade de amostras positivas torna essencial maximizar a identificação de verdadeiros positivos, reduzindo o número de falsos negativos. Um recall baixo pode resultar em falhas na identificação de deslizamentos, aumentando os riscos associados. Embora a precisão seja uma métrica relevante, neste estudo priorizou-se modelos que maximizam o recall da classe deslizamento.

A análise entre precisão e recall revelou que os modelos com alto recall frequentemente apresentavam baixa precisão para a classe deslizamento. O trade-off entre essas métricas deve ser considerado de acordo com a aplicação prática do modelo. Como a detecção de deslizamentos é crucial para a mitigação de desastres, priorizou-se o recall em detrimento da precisão. O Gradient Boosting, por exemplo, teve recall de 0,73 com ADASYN, mas precisão de apenas 0,18. Já o Adaboost apresentou recall de 0,91, mas sua precisão extremamente baixa de 0,06 compromete a confiabilidade das previsões.

Figura 1. matriz de confusão: SVM + ADASYN



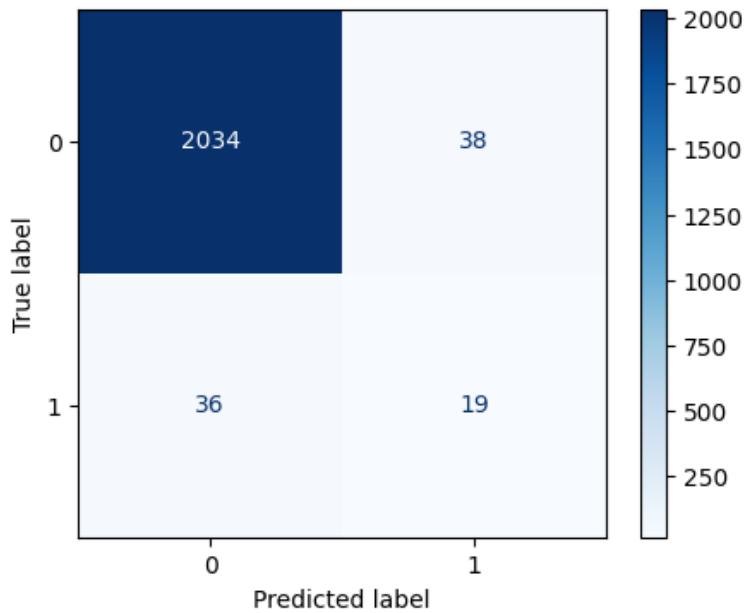
Com base nos resultados obtidos e como podemos observar na matriz de confusão, conclui-se que a melhor abordagem para a predição de deslizamentos foi a combinação do SVM com a técnica de oversampling ADASYN. Essa configuração proporcionou o melhor equilíbrio entre recall e precisão, garantindo a identificação da maioria dos eventos de deslizamento sem comprometer excessivamente a performance geral do modelo. A utilização de técnicas de oversampling mostrou-se essencial para melhorar a predição da classe minoritária, sendo o ADASYN a alternativa mais eficaz para este estudo.

Dessa forma, recomenda-se a aplicação do SVM + ADASYN em cenários onde a detecção de deslizamentos de terra é uma prioridade, garantindo maior segurança na tomada de decisões baseadas nos resultados do modelo.

5. Comparação dos resultados obtidos pelo algoritmo de melhor desempenho com o critério estabelecido no projeto original

O projeto original do OlhaChuva adotava uma abordagem simplificada para a predição de deslizamentos, baseando-se exclusivamente em um modelo matemático com dados estáticos. Sua previsão de ocorrência de chuvas considerava apenas se o volume acumulado de água atingia um determinado limite para a tomada de decisões. Apesar da simplicidade do método, ele obteve um desempenho razoável ao prever o acumulado de chuva de 100mm em três dias, alcançando 19 verdadeiros positivos e 2034 verdadeiros negativos, enquanto registrou 38 falsos negativos e 36 falsos positivos.

Figura 2. Matriz confusão do OLHACHUVA



Com a nova proposta baseada em técnicas de *machine learning*, o modelo que apresentou o melhor desempenho foi a combinação de SVM, ADASYN e *GridSearch*. Esse modelo obteve 10 verdadeiros positivos e 326 verdadeiros negativos, reduzindo significativamente o número de falsos positivos para apenas 1, embora tenha aumentado os falsos negativos para 89.

Ao comparar os falsos negativos entre os dois modelos, observamos que, no projeto original, a taxa de falsos negativos (36 casos) indicava uma limitação considerável na capacidade de prever eventos críticos de deslizamento. No entanto, a nova abordagem, apesar de reduzir drasticamente os falsos positivos, apresentou um aumento nos falsos negativos, evidenciando que o modelo ainda precisa de refinamentos para melhorar sua capacidade preditiva.

A seguir, comparamos os relatórios de classificação dos dois modelos:

Tabela 3. Comparação de Relatórios de Classificação: OLHACHUVA x SVM

Classe	Precisão	Recall	F1-score	Support	Precisão	Recall	F1-score	Support
Não desliz.	0.98	0.98	0.98	2072	1.00	0.79	0.88	415
Deslizamento	0.33	0.35	0.34	55	0.10	0.91	0.18	11
Accuracy			0.97	2127			0.79	426
Média Macro	0.66	0.66	0.66	2127	0.55	0.85	0.53	426
Média Pond.	0.97	0.97	0.97	2127	0.97	0.79	0.86	426

A análise da tabela revela que, embora o modelo baseado em SVM + ADASYN tenha uma acurácia geral inferior ao modelo original, ele apresenta um desempenho muito superior na detecção de deslizamentos, conforme demonstrado pelo alto valor de *recall* para a classe de deslizamentos. Isso significa que ele consegue identificar um número maior de ocorrências reais, mesmo que, para isso, aceite um aumento na taxa de falsos

negativos. Essa característica pode ser crucial em cenários onde a falha na previsão de um evento crítico pode ter consequências severas.

Outro ponto importante a ser considerado é a limitação dos dados utilizados para o treinamento dos modelos. Os registros de precipitação estão disponíveis apenas a partir de 2015, enquanto os dados de deslizamentos abrangem o período de 2005 a 2022. Dessa forma, apenas o intervalo entre 2015 e 2022 pode ser explorado para análise, reduzindo a quantidade de dados históricos disponíveis. Essa discrepância temporal pode impactar a robustez e a generalização dos modelos preditivos, tornando necessário o uso de técnicas que lidem com conjuntos de dados desbalanceados e limitados.

Diante dos resultados obtidos, fica evidente que a introdução de técnicas de *machine learning* trouxe melhorias significativas na identificação de deslizamentos, apesar dos desafios ainda existentes. Trabalhos futuros podem explorar o uso de dados meteorológicos complementares, aprimoramento no balanceamento de dados e ajuste fino dos hiperparâmetros para melhorar a performance do modelo sem comprometer sua precisão.

6. CONSIDERAÇÕES FINAIS

A implementação de um sistema de monitoramento hidrometeorológico e alerta de deslizamentos baseado em aprendizado de máquina representa um avanço significativo na previsão e mitigação de desastres naturais. A substituição de gatilhos estáticos por dinâmicos melhora a precisão das previsões e fortalece a resiliência urbana, contribuindo para a proteção de comunidades vulneráveis.

O desenvolvimento e teste de algoritmos que ajustam dinamicamente os parâmetros de alerta com base em dados em tempo real, como níveis de precipitação e previsões meteorológicas, aumentam a eficácia do sistema, permitindo respostas mais rápidas e precisas a eventos críticos.

Portanto, a integração de gatilhos dinâmicos ao monitoramento hidrometeorológico potencializa a gestão de riscos, aprimora a precisão das previsões e fortalece a capacidade de resposta a desastres, promovendo maior segurança para a população e o meio ambiente.

6.1. Limitações do trabalho

Apesar dos avanços alcançados, este estudo apresenta algumas limitações que impactam a precisão e generalização do sistema de alerta de deslizamentos. Uma das principais restrições é o período de dados disponíveis. Os registros de precipitação abrangem apenas 2015 até o presente, o que pode não ser suficiente para capturar padrões climáticos de longo prazo. Além disso, a análise conjunta com os dados de deslizamentos se restringe a 2015-2022, reduzindo a amostra para correlação entre eventos.

A cobertura espacial também é limitada, pois, embora existam outras estações meteorológicas na região, muitas apresentam lacunas nos dados (*gaps*), dificultando uma análise mais abrangente e representativa. A dependência da estação de Piedade compromete a precisão das previsões para outras áreas. Além disso, a qualidade dos dados representa um desafio. A presença de falhas e inconsistências nos registros pode introduzir incertezas nos modelos, afetando a confiabilidade das previsões. A ausência de padronização completa nos dados históricos de deslizamentos também dificulta análises mais robustas.

Outro fator crítico é a falta de variáveis complementares. O modelo se baseia apenas em dados de precipitação, sem considerar geologia, topografia e umidade do solo, elementos essenciais para uma previsão mais precisa. Cabe salientar que esse modelo foi desenvolvido para o atendimento de uma necessidade local, para a qual não havia a coleta nessas outras variáveis citadas.

Por fim, a capacidade de generalização dos modelos é limitada, pois foram testados com dados de uma única estação. Isso restringe a aplicabilidade do sistema em regiões com diferentes características climáticas e geográficas. Superar essas limitações, por meio da expansão da rede de monitoramento, melhoria na qualidade dos dados e inclusão de variáveis ambientais, será essencial para aumentar a precisão e a eficácia do sistema de alerta.

6.2. Trabalhos futuros

Diante dos avanços alcançados neste projeto, futuras pesquisas podem focar em aprimoramentos que tornem o sistema de monitoramento hidrometeorológico e alerta de deslizamentos ainda mais eficiente.

Embora existam outras estações além da de Piedade, muitas apresentam lacunas nos registros. Estratégias para minimizar esses gaps possibilitariam uma cobertura espacial mais confiável, aumentando a precisão das previsões. Além disso, a ampliação do período de análise, incluindo registros históricos e contínuos, permitirá identificar padrões mais complexos.

A inclusão de novas variáveis ambientais, como umidade do solo, índice de vegetação e dados geotécnicos, pode aumentar a precisão das previsões. Sensores específicos para essas medições podem fornecer informações complementares ao modelo.

No campo do aprendizado de máquina, a adoção de modelos híbridos e adaptativos, bem como a aplicação de técnicas de otimização como busca em grade e algoritmos genéticos, pode melhorar o desempenho preditivo. Essas estratégias podem revelar padrões ainda não explorados, tornando o sistema mais eficiente e dinâmico.

Dessa forma, avanços nessas áreas contribuirão para um sistema mais robusto e preciso, fortalecendo a capacidade de previsão e mitigação de desastres naturais.

Referências

- Blagus, Rok, L. L. (2010). Class prediction for high-dimensional class-imbalanced data. *BMC Bioinformatics*, 11:523–523.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1):5–32.
- Brunetti, M. T., Peruccacci, S., Rossi, M., Luciani, S., Valigi, D., and Guzzetti, F. (2010). Rainfall thresholds for the possible occurrence of landslides in italy. *Natural Hazards and Earth System Sciences*, 10(3):447–458.
- Cortes, Corinna, V. V. (1995). Support-vector networks. *Machine Learning*, 20:273–297.
- Costa, A., Cipriano, J., Rego, R., Oliveira, C., Orengo, J., Brito, C., Guedes, R., Silva, T., and Jr, F. (2023). DESENVOLVIMENTO DE UM SISTEMA DE ALERTA DE DESLIZAMENTOS DE TERRA EM ESCALA MUNICIPAL: MONITORAMENTO, GATILHOS E PREVISÃO METEOROLÓGICA.

- de Assis Dias, M. C., Saito, S. M., dos Santos Alvalá, R. C., Stenner, C., Pinho, G., Nobre, C. A., de Souza Fonseca, M. R., Santos, C., Amadeu, P., Silva, D., et al. (2018). Estimation of exposed population to landslides and floods risk areas in brazil, on an intra-urban scale. *International Journal of Disaster Risk Reduction*, 31:449–459.
- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29(5):1189–1232.
- Gregor, Shirley, H. A. R. (2013). Positioning and presenting design science research for maximum impact. *MIS Quarterly*, 37(2):337–355.
- Han, Jiawei, Kamber, Micheline, Pei, and Jian (2012). *Data Mining: Concepts and Techniques*. Morgan Kaufmann, 3 edition.
- He, H., Bai, Y., Garcia, E. A., and Li, S. (2008). Adasyn: Adaptive synthetic sampling approach for imbalanced learning. In *2008 IEEE international joint conference on neural networks (IEEE world congress on computational intelligence)*, pages 1322–1328. Ieee.
- Hevner, A. R., March, S. T., Park, J., and Ram, S. (2004). Design science in information systems research. *MIS quarterly*, pages 75–105.
- Hutter, F., Kotthoff, L., and Vanschoren, J. (2019). *Automated machine learning: methods, systems, challenges*. Springer Nature.
- Jordan, M. I. and Mitchell, T. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349:255–260.
- Kanungo, D. and Sharma, S. (2014). Rainfall thresholds for prediction of shallow landslides around chamoli-joshimath region, garhwal himalayas, india. *Landslides*, 11(4):629–638.
- Kohavi, R. (2001). A study of cross-validation and bootstrap for accuracy estimation and model selection. 14.
- March, Salvatore, S. G. (1995). Design and natural science research on information technology. *Decision Support Systems*, 15:251–266.
- Marchezini, V., o., editor (2017). *Reduction of vulnerability to disasters: from knowledge to action*. RiMa Editora São Carlos - SP.
- Moraes, Katarina e Ferreira, C. (2023). Desastre das chuvas em pernambuco completa um ano com 134 mortes e nenhuma responsabilização. <https://jc.ne10.uol.com.br/pernambuco/2023/05/15468725-desastre-das-chuvas-em-pernambuco-completa-um-ano-com-134-mortes-e-nenhuma-responsabilizacao.html>.
- Nitesh, Chawla, V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). Smote: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16:321–357.
- Peffers, K., Tuunanen, T., Rothenberger, M. A., and Chatterjee, S. (2008). A design science research methodology for information systems research. *Journal of Management Information Systems*, 24:45–77.