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

Cemaden
Centro Nacional de Monitoramento e Alertas de Desastres Naturais

UNIDADE DE PESQUISA DO
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA, INOVAÇÕES E COMUNICAÇÕES

PÁTRIA AMADA BRASIL
GOVERNO FEDERAL

FELICIA FRANÇA PEREIRA

COMPARAÇÃO DE DADOS DERIVADOS DE LIDAR E VANT PARA O MAPEAMENTO DE SUSCETIBILIDADE DE DESLIZAMENTO USANDO O ALGORITMO RANDOM FOREST

2022

FELICIA FRANÇA PEREIRA

COMPARAÇÃO DE DADOS DERIVADOS DE LIDAR E VANT PARA O MAPEAMENTO DE SUSCETIBILIDADE DE DESLIZAMENTO USANDO O ALGORITMO RANDOM FOREST

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

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

Orientador: Profa. Dra. Tatiana Sussel Gonçalves Mendes

Coorientador: Prof. Dr. Silvio Jorge Coelho Simões

São José dos Campos

2022

Instituto de Ciência e Tecnologia [internet]. Normalização de tese e dissertação [acesso em 2022]. Disponível em http://www.ict.unesp.br/biblioteca/normalizacao

Apresentação gráfica e normalização de acordo com as normas estabelecidas pelo Serviço de Normalização de Documentos da Seção Técnica de Referência e Atendimento ao Usuário e Documentação (STRAUD).

Pereira, Felicia França
Comparação de dados derivados de LiDAR e VANT para o mapeamento de suscetibilidade de deslizamento usando o algoritmo Random Forest / Felicia França Pereira. - São José dos Campos : [s.n.], 2022.
56 f. : il.

Dissertação (Mestrado em Desastres Naturais) - Pós-graduação em Desastres Naturais - Universidade Estadual Paulista (Unesp), Instituto de Ciência e Tecnologia; Centro Nacional de Monitoramento e Alertas de Desastres Naturais (Cemaden), São José dos Campos, 2022.
Orientadora: Tatiana Sussel Gonçalves Mendes
Coorientador: Silvio Jorge Coelho Simões

1. Floresta Aleatória. 2. Modelo de Suscetibilidade a Deslizamentos. 3. MDT. 4. LiDAR. 5. VANT. I. Mendes, Tatiana Sussel Gonçalves , orient. II. Simões, Silvio Jorge Coelho, coorient. III. Universidade Estadual Paulista (Unesp), Instituto de Ciência e Tecnologia, São José dos Campos. IV. Universidade Estadual Paulista 'Júlio de Mesquita Filho' - Unesp. V. Universidade Estadual Paulista (Unesp). VI. Centro Nacional de Monitoramento e Alertas de Desastres Naturais (Cemaden). VII. Título.

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi e Seção Técnica de Informática, ICMC/USP
com adaptações - STATI, STRAUD e DTI do ICT/UNESP.
Renata Aparecida Couto Martins CRB-8/8376

BANCA EXAMINADORA

Profa. Dra. Tatiana Sussel Gonçalves Mendes (Orientador)
Universidade Estadual Paulista (Unesp)
Instituto de Ciência e Tecnologia
Campus de São José dos Campos

Prof. Dr. Rogério Galante Negri
Universidade Estadual Paulista (Unesp)
Instituto de Ciência e Tecnologia
Campus de São José dos Campos

Dra. Alessandra Cristina Corsi
Instituto de Pesquisas Tecnológicas - IPT
Seção de Investigações, Riscos e Gerenciamento Ambiental – SIRGA

São José dos Campos, 02 de agosto de 2022.

AGRADECIMENTOS

Agradeço a Deus, que me iluminou e me deu forças para nunca desistir dos meus objetivos.

Aos meu pais Maria França e Alberto Carvalho, que me deram o total apoio ao longo da minha vida e nunca me deixaram faltar nada, sempre torcendo por mim. À minha irmã Alicia França que sempre esteve me apoiando, e a minha família pelo suporte. Ao meu querido Arian, que me aconselhou e me apoiou nos momentos difíceis, obrigada pela paciência em meus momentos de estresse e desanimo e por ser meu amigo e amor.

Em especial à minha orientadora Dra. Tatiana Sussel, que contribuiu de forma significativa em minha carreira acadêmica e profissional, e que sempre esteve ao meu lado. Agradeço pela oportunidade, toda a paciência e conselhos durante esse período. Ao meu coorientador Dr. Silvio Simões por todo o ensinamento, discussões e contribuições durante o período do mestrado.

À Jennifer e a Tatiany, pelas noites de estudo, inúmeras reuniões e todo o aprendizado compartilhado ao longo dessa pesquisa.

Meu muito obrigado aos amigos Fábio e Ana, com quem dividi moradia na república Cab. Da Leila, vocês foram especiais nessa caminhada.

Á todas as outras pessoas não citadas, mas que contribuíram direta ou indiretamente os meus sinceros agradecimentos, essa conquista não é só minha.

SUMÁRIO

LISTA DE ABREVIATURAS E SIGLAS **3**
RESUMO **4**
ABSTRACT **5**
1 INTRODUÇÃO **8**
2 ARTIGO **13**
2.1 Artigo "Comparison of LiDAR and UAV derived data for landslide susceptibility mapping using Random Forest algorithm" **13**
3 CONSIDERAÇÕES FINAIS **51**
REFERÊNCIAS **52**

LISTA DE ABREVIATURAS E SIGLAS

AUC	*Area Under the Curve*
ANN	*Artificial Neural Networks*
CRED	Centro de Pesquisas de Epidemiologia em Desastres
DT	*Decision Tree*
LiDAR	*Light Detection And Ranging*
MDE	Modelo Digital de Elevação
MDT	Modelo Digital de Terreno
PMRR	Plano Municipal de Redução de Riscos
PPDC	Plano Preventivo de Defesa Civil
RF	*Random Forest*
ROC	*Receiver Operating Characteristic Curve*
RP	Reconhecimento de Padrões
SR	Sensoriamento Remoto
SVM	*Support Vector Machine*
UNISDR	Estratégia Internacional de Nações Unidas para a Redução de Desastres
VANT	Veículo Aéreo Não-Tripulado

PEREIRA, F.F. **Comparação de dados derivados de LiDAR e VANT para o mapeamento de suscetibilidade de deslizamento usando o algoritmo Random Forest**. Dissertação. São José dos Campos: Universidade Estadual Paulista (Unesp), Instituto de Ciência e Tecnologia; Centro Nacional de Monitoramento e Alertas de Desastres Naturais (Cemaden), 2022.

RESUMO

Impulsionados pela tectônica, clima e/ou atividade humana, os deslizamentos de terra são onipresentes. Este fenômeno é uma das principais ameaças ao desenvolvimento econômico e social das zonas montanhosas, muitas vezes resultando em vítimas e consequências econômicas. O mapeamento de suscetibilidade a deslizamentos é considerado uma importante ferramenta para os gestores locais localizarem as áreas que possuem maior probabilidade de ocorrência de eventos e pensarem em estratégias de redução de risco. O Modelo Digital de Terreno (MDT) é um dos dados fundamentais para os modelos de predição e é usado para derivar importantes fatores condicionantes para análises de suscetibilidade de deslizamento em escala detalhada. Desse modo, a utilização de técnicas de Sensoriamento Remoto e Fotogrametria, usando as tecnologias de varredura a LASER (Light Detection and Range - LiDAR) e de imageamento por Veículo Aéreo Não-Tripulado (VANT), podem impulsionar esse levantamento, produzindo dados de altíssima resolução. Nesse sentido, este estudo teve como objetivo comparar mapas de suscetibilidade a deslizamentos gerados pelo algoritmo de aprendizado de máquina floresta aleatória (Random Forest - RF) com dados de LiDAR e VANT. Para tanto, o desempenho alcançado na previsão foi avaliado usando medidas de avaliação estatística baseadas em conjuntos de dados de treinamento e validação. Os resultados obtidos mostraram que a precisão de ambos os modelos é maior que 0,70, a área sob a curva (AUC) é maior que 0,80, e o modelo gerado a partir dos dados LiDAR é mais preciso. Os resultados também mostraram que os dados do VANT têm potencial para uso no mapeamento de suscetibilidade a deslizamentos em escala intraurbana, contribuindo para estudos em áreas de risco sem dados disponíveis.

Palavras-chave: floresta aleatória; modelo de suscetibilidade a deslizamentos; MDT; LiDAR; VANT.

PEREIRA, F.F. ***Comparison of LiDAR and UAV derived data for landslide susceptibility using Random Forest algorithm****. Dissertation. São José dos Campos: São Paulo State University (Unesp), Institute of Science and Technology; National Center for Monitoring and Early Warning of Natural Disasters (Cemaden), 2022.*

ABSTRACT

Driven by tectonics, weather and/or human activity, landslides are ubiquitous. This phenomenon is one of the main threats to the economic and social development of mountainous areas, often resulting in victims and economic consequences. The landslide susceptibility mapping is considered an important tool for local managers to locate the most vulnerable areas and think about risk reduction strategies. The Digital Terrain Model (DTM) is one of the fundamental modeling data and is used to derive important conditional factors for detailed scale landslide susceptibility analyses. Thus, the use of Remote Sensing and Photogrammetry techniques, using LASER (Light Detection and Range - LiDAR) and Unmanned Aerial Vehicle (UAV) imaging technologies, can boost this survey, producing data of very high resolution. In this sense, this study aimed to compare landslide susceptibility maps generated by the Random Forest (RF) machine learning algorithm with LiDAR and UAV data. To this end, the performance achieved in the prediction was evaluated using statistical evaluation measures based on training and validation data sets. The results obtained showed that the accuracy of both models is greater than 0.70, the area under the curve (AUC) is greater than 0.80, and the model generated from LiDAR data is more accurate. The results also showed that UAV data have potential for use in mapping landslide susceptibility on an intra-urban scale, contributing to studies in risk areas without available data.

Keywords: random forest; landslide susceptibility model; DTM; LiDAR; UAV

1 INTRODUÇÃO

Um deslizamento de terra é um fenômeno natural em que o material em uma encosta é movido para baixo devido à gravidade e outros fatores externos. Impulsionados pela tectônica, clima e/ou atividade humana, os deslizamentos de terra são onipresentes em qualquer ambiente terrestre inclinado (FROUDE; PETLEY, 2018). São fenômenos naturais extremamente complexos capazes de causar danos significativos às pessoas, propriedades e redes de transporte (ABDULWAHID; PRADHAN, 2017).

Em um contexto global, o fenômeno é uma das principais ameaças ao desenvolvimento econômico e social das zonas montanhosas. Segundo o Centro de Pesquisas de Epidemiologia em Desastres (CRED) e a Estratégia Internacional de Nações Unidas para a Redução de Desastres (UNISDR) (2018), entre 1900 e 2017 cerca de 130.000 pessoas perderam a vida por causa de deslizamentos de terra e inundações. Ademais, conforme os resultados estatísticos do Banco Mundial (2005) a perda econômica causada por deslizamentos de terra é de cerca de US$ 20 bilhões por ano e 300 milhões de pessoas vivem em áreas propensas a tais eventos (KAO; CHIU; TSAI, 2017).

No Brasil, os desastres provocados por deslizamentos geralmente são associados a períodos de precipitação intensa e, principalmente em regiões montanhosas e serranas, onde predomina clima úmido, com relevos colinosos e/ou tabulares, além de um expansivo processo de urbanização (PICHLER, 1957). Os municípios brasileiros com situações consideradas de alto e muito alto risco perante enchentes, enxurradas e deslizamentos, estão situados no Rio de Janeiro, Bahia, São Paulo e Pernambuco (SANTOS, 2012).

Na Serra do Mar, que é uma das feições orográficas mais proeminentes do litoral sul e sudeste do país medindo 1.500 km de extensão, foi palco de grandes eventos de deslizamentos como o de Santos em 1928, Caraguatatuba em 1967 e Cubatão em 1985; (PICHLER, 1957; LISTO; VIEIRA, 2012; LISTO; GOMES; FERREIRA, 2021). Nota-se que entre 1990 e 2012, os deslizamentos tiveram uma grande variação, aumentando em 21,7 vezes, e a maioria dos óbitos situaram-se na região sudeste (CEPED, 2013).

O município de Santos, situado no litoral do estado de São Paulo, possui áreas de morro residuais que apresentam intensa ação do intemperismo químico e físico. Nessas áreas ocorreram vários deslizamentos catastróficos, como o ocorrido em 1928 e em 1956 (PICHLER, 1957). Esses desastres motivaram a realização de um amplo trabalho de contenção das encostas e monitoramento da situação geomórfica nos morros santistas, porém a expansão da urbanização nessas áreas intensifica problemas estruturais em relação aos desastres de deslizamentos na região.

No intuito de reduzir o risco de desastres, em 1989 teve início a operação do Plano Preventivo de Defesa Civil (PPDC) do município, o qual se mostrou muito eficiente visto a queda expressiva nos números de óbitos do município (FERREIRA, 2016). Em 2005 o primeiro Plano Municipal de Redução de Riscos (PMRR) de Santos foi elaborado (CANIL, 2010). Sendo

atualizado em 2012 com medidas de prevenção e correção para eliminar ou reduzir os riscos geológico que incluem deslizamentos (FERREIRA; ABIKO, 2019). Dessa forma, o mapeamento de suscetibilidade a deslizamentos desempenha um papel importante no desenvolvimento de medidas preventivas, pois visa indicar a probabilidade de ocorrência de processos naturais e induzidos (CARVALHO; GALVÃO, 2016).

Para o mapeamento da suscetibilidade a deslizamento diversas metodologias já foram aplicadas, as quais podem ser dividas em qualitativas e quantitativas (XIE et al., 2018). Análises qualitativas são baseadas em julgamentos de especialistas sobre fatores condicionantes a deslizamentos. Enquanto que, em abordagens quantitativas, são utilizadas as relações matemáticas entre locais de deslizamentos e fatores casuais (NEUHÄUSER; DAMM; TERHORST, 2012). As abordagens quantitativas tem se destacado nos últimos anos, com a utilizacao de dados de inventário e dados morfométricos de cicatrizes de deslizamentos já ocorridos devido a perfomance obtida (LIU et al., 2021; YOUSSEF; POURGHASEMI, 2021).

Os modelos estatísticos usados para a análise de suscetibilidade à deslizamentos tornaram-se, dentre os modelos quantitativos, um dos métodos mais utilizados para este tipo de análise, pois baseiam-se na relação funcional entre os fatores condicionantes aos deslizamentos e os deslizamentos pré-existentes. Tais modelos possibilitam determinar o peso das variáveis envolvidas e, também, a validação dos resultados (ZÊZERE et al., 2017). Este modelo utiliza diferentes técnicas estatísticas clássicas como, técnicas probabilísticas, Regressão Logística, Índice Estatístico, Razão de Frequência e Fator de Certeza (LEE; MIN, 2001; WU et al., 2016; ZHAO et al., 2015; PIRASTEH; LI, 2017; JULIEV et al., 2019). Porém, os métodos estatísticos exigem a coleta de grandes quantidades de dados para produzir resultados confiáveis (YILMAZ, 2009).

Os métodos baseados em apredizado de máquina, conceitualmente, aprendem com os dados e não dependem de funções baseadas em regras. Se tratando de relações não lineares os dados podem ser modelados de forma eficaz de acordo com o método adotado. Estes métodos mostram-se superiores aos métodos estatísticos clássicos em detrimento da maior capacidade de precisão de previsão. O uso de métodos de aprendizado de máquina para aprimorar o mapeamento tem sido cada vez mais utilizado nos últimos anos a partir de técnicas como Redes Neurais Artificiais (do inglês, Artificial Neural Networks - ANN) (HUANG et al., 2020a; HUANG et al., 2020b; LUO et al., 2019), máquina de vetor de suporte (do inglês, Support Vector Machine - SVM) (WANG; DUAN; HONG, 2019; HUANG; ZHAO, 2018), floresta aleatória (do inglês, Random Forest - RF) (ARABAMERI et al., 2020; CHEN et al., 2018; FANG et al., 2021), árvore de decisão (do inglês, Decision Tree - DT) (PHAM et al., 2019; HUANG et al., 2020b).

Dentre os algoritmos de aprendizado de máquina, o algoritmo RF destaca-se por sua capacidade de processar dados massivos, que podem ser categóricos ou numéricos, possibilitando a obtenção de altos valores de precisão na modelagem de suscetibilidade a deslizamentos (SHIRVANI, 2020). Este método usa estimativa imparcial na construção do modelo, o que significa que o treinamento é rápido e simples de implementar (RODRIGUEZ-GALIANO et al., 2012). O

algoritmo mostrou-se uma técnica de alto desempenho com aplicações para o mapeamento de suscetibilidade a deslizamentos (CHEN et al., 2017; TAALAB; CHENG; ZHANG, 2018; PARK; KIM, 2019; DOU et al., 2019). Dou et al. (2019) expuseram que, em geral, o RF demostrou melhor desempenho em ambas as amostras com Area Under the Curve (AUC) > 0,9. em relação o DT para mapear a suscetibilidade ao deslizamento de terra induzida pela chuva em escala regional no Japão. Youssef e Pourghasemi (2021) realizaram uma análise com sete métodos de aprendizado de máquina para a modelagem de suscetibilidade a deslizamento de terra e a comparação de seus desempenhos e resultados mostraram que o RF foi um dos métodos que produziu melhor desempenho em comparação aos outros, obtendo precisão igual a 95,1%.

Diante da complexidade dos eventos de deslizamento de terra, além das implementações de técnicas avançadas de aprendizado de máquina, a qualidade dos dados de entrada é um fator determinante para o desempenho dos modelos (LIMA; STEGER; GLADE, 2021). Nos modelos de predição, os dados de entrada são fatores condicionantes que, em sua maioria, são derivados de um Modelo Digital de Elevação (MDE), os quais podem ser obtidos a partir de diversas fontes como: a digitalização de mapas topográficos, técnicas de Sensoriamento Remoto (SR) (SALEEM et al., 2019). Porém para uma escala detalhada regional é adequado o uso de modelos digitais de terreno (MDT) para que artefatos não sejam considerados na modelagem da superfície.

Atualmente, vários MDEs estão disponíveis gratuitamente (por exemplo, Shuttle Radar Topographic Mission (SRTM); Advanced Spaceborne Thermal Emission and Reflection (ASTER) (RABBY; ISHTIAQUE; RAHMAN, 2020), esses possuem resolução espacial que podem ser consideradas apropriadas para a análise em escalas pequenas e médias, mas é demasiadamente grosseira para trabalhos em escala de detalhe. Além disso esses modelos foram obtidos há mais de duas décadas, podendo comprometer a capacidade de previsão do modelo (HAWKER et al., 2018). Em visto disso faz-se necessário o estudo com MDTs atuais e com alta resolução espacial (menor que 1 metro).

Considerando o desenvolvimento e o avanço da tecnologia, as técnicas de SR, como a fotogrametria e técnicas de varredura a LASER[1], denominado aqui como Light Detection And Ranging (LiDAR) podem produzir dados com resolução espacial sub-métrica. Essas técnicas podem ser usadas para dar suporte a modelos de previsão de deslizamento em escala detalhada, aplicados a áreas específicas de declividade, para refinar mapas de suscetibilidade a deslizamentos de mesoescala produzidos usando bancos de dados gratuitos ou MDEs derivados de mapas topográficos como base para a área.

O uso da tecnologia LiDAR está se tornando cada vez mais popular, estudos recentes têm mostrado a utilidade de dados topográficos derivados de LiDAR para mapeamento de áreas suscetíveis a deslizamentos e inventários de deslizamentos. Gaidzik et al. (2017), ao usarem MDTs derivados de LiDAR, alcançaram alta precisão, conforme evidenciado pela precisão geral das previsões com valores de AUC entre 0,77 e 0,87. Gorsevski (2021) ao implementar uma

[1]Palavra inglesa, do acrônimo Light Amplification by Stimulated Emission Radiation que significa amplificação de luz por emissão estimulada de radiação (raio LASER).

abordagem evolutiva usando classificação simbólica com programação genética, o potencial de previsão espacial da suscetibilidade ao deslizamento foi examinado e obteve-se uma precisão de 85,0%, classificando um total de 13,4% como regiões de alta suscetibilidade, com uma precisão quantitativa geral exponencial correspondente para 0,9082.

Pesquisas recentes utilizando-se de produtos cartográficos produzidos a partir da tecnologia de Veículo Aéreo Não-Tripulado (VANTs) evidenciaram a produção de dados altamente detalhados, pois a tecnologia é capaz de fornecer imagens com alta resolução espacial e uma densa nuvem de pontos com coordenadas tridimensionais. Ademais, ela proporciona imageamento de eventos temporais, flexibilidade de trajetória de voo e o acoplamento de diferentes sensores para a captação de uma grande variedade de dados de interesse, especialmente em ambientes perigosos e de difícil acesso (COLOMINA; MOLINA, 2014; BERNARDO; PALAMARA; BOIMA, 2021). Os dados obtidos por VANT estão cada vez mais presentes no monitoramento e gestão de desastres naturais (GIORDAN et al., 2017; GOMEZ; PURDIE, 2016), suscetibilidade a desastres geológicos (WANG et al., 2020), suscetibilidade a deslizamentos (TEMPA et al., 2021; GANTIMUROVA; PARSHIN; EROFEEV, 2021), monitoramento e mapeamento de deslizamentos (GIORDAN et al., 2018; PEPPA, 2018), e avaliação de deslizamento (SESTRAS et al., 2021; EKER; AYDIN; HÜBL, 2018; EKER; AYDIN, 2021).

Neste contexto, observa-se que os dados de VANT têm sido amplamente utilizados para fins de estudos de deslizamentos (PEPPA, 2018; EKER; AYDIN; HÜBL, 2018; SESTRAS et al., 2021). No entanto, devido a natureza do processo fotogramétrico, existem limitações no que tangem a modelagem de regiões com cobertura vegetal densa, o que pode impactar a geração de modelos nestas regiões (HAWKER et al., 2018). A ocupação extensiva das encostas agrava ainda mais o problema pois dificulta o imageamento e a modelagem do terreno. Esta condição é comum nas regiões montanhosas de centros urbanos do Brasil, levando a problemas geológicos e hidrológicos com consequências catastróficas (MODENESI-GAUTTIERI; HIRUMA, 2004).

Portanto, embora existam alguns trabalhos utilizando dados de LiDAR e VANTs para mapeamento de suscetibilidade a deslizamentos, até o momento poucos estudos aplicaram um método para comparar mapas de suscetibilidade a deslizamentos obtidos a partir de dados derivados dessas duas técnicas de levantamento. Diante disso, surge a hipótese de que, em modelos de suscetibilidade a deslizamentos, os dados de VANT podem render bons resultados comparados aos dados de LiDAR.

Nesse sentido, este trabalho tem como objetivo a comparação dos dois modelos de mapeamento de suscetibilidade a deslizamentos usando algoritmo de RF gerado a partir da tecnologia LiDAR e VANT para verificar se os dados derivados de VANT podem atingir precisão compatível com dados derivados de LiDAR.

Assim, essa pesquisa contribui para avaliar a capacidade dos dados de VANT para realizar estudos de suscetibilidade a deslizamentos na ausência de dados de escala mais detalhada. Este documento encontra-se organizado da seguinte forma: o Capítulo 2 apresenta o artigo "*Comparison of LiDAR and UAV derived data for landslide susceptibility mapping using Random*

Forest algorithm" e o Capítulo 3 aborda as considerações finais do presente trabalho.

2 ARTIGO

2.1 Artigo "**Comparison of LiDAR and UAV derived data for landslide susceptibility mapping using Random Forest algorithm**"

O artigo a seguir foi submetido para apreciação do corpo editorial do periódico LANDSLIDES em 16 de junho de 2022.

Comparison of LiDAR and UAV derived data for landslide susceptibility mapping using Random Forest algorithm

Felicia França Pereira[1*], Tatiana Sussel Gonçalves Mendes[2†], Silvio Jorge Coelho Simões[2,5†], Márcio Roberto Magalhães de Andrade[3†], Mário Luiz Lopes Reiss[4†], Jennifer Fortes Cavalcante Renk[3†] and Tatiany Correia da Silva Santos[1†]

[1*]Graduate Program in Natural Disasters, UNESP/CEMADEN, Estrada Doutor Altino Bondesan, São José dos Campos, São Paulo, Brazil.
[2]Department of Environmental Engineering, Institute of Science and Technology, São Paulo State University – Unesp, Estrada Doutor Altino Bondesan, São José dos Campos, São Paulo, Brazil.
[3]National Center for Monitoring and Early Warning of Natural Disasters - CEMADEN, MCTI, Estrada Doutor Altino Bondesan, São José dos Campos, São Paulo, Brazil.
[4]LAFOTO – Laboratory of Photogrammetry Research, Department of Geodesy, UFRGS – Federal University of Rio Grande do Sul, Av. Bento Gonçalves, Porto Alegre, Rio Grande do Sul, Brazil.
[5]Center for Marine and Environmental Research – CIMA, University of Algarve, Estr. da Penha, Faro, Portugal.

*Corresponding author(s). E-mail(s): felicia.franca@unesp.br;
Contributing authors: tatiana.mendes@unesp.br;
silvio.simoes@unesp.br ; marcio.andrade@cemaden.gov.br;
mario.reiss@ufrgs.br; jennyfgeo@gmail.com;
tatiany.corriea@unesp.br;
†These authors contributed equally to this work.

Abstract

Earthquakes, extreme rainfall, or human activity can all cause landslides. Several landslides occur each year around the world, often resulting in casualties and economic consequences. Landslide susceptibility mapping is considered to be the main technique for predicting the likelihood of an event based on the characteristics of the physical environment. Digital Terrain Model (DTM) is one of the fundamental data of modeling and is used to derive important conditional factors for detailed-scale landslide susceptibility analyses. With this in mind, this study aimed to compare landslide susceptibility maps generated by Random Forest (RF) machine learning algorithm with data from Light Detection and Range (LiDAR) and Unmanned Aerial Vehicle (UAV). To this end, the performance achieved in prediction was evaluated using statistical evaluation measures based on training and validation datasets. The obtained results showed that the accuracy of both models is greater than 0.70, the area under the curve (AUC) is greater than 0.80, and the model generated from the LiDAR data is more accurate. The results also showed that the data from UAV have potential to use in landslide susceptibility mapping on an intra-urban scale, contributing to studies in risk areas without available data.

Keywords: Random Forest, Landslide Susceptibility Model, DTM, LiDAR, UAV

1 Introduction

Landslides can be triggered by earthquakes, extreme precipitation, or human activities and are highly complex natural phenomena that can cause significant damage to people, property, and transportation networks (Saro et al, 2016). Annually, several landslide events occur worldwide, often resulting in the loss of human life and economic consequences (Froude and Petley, 2018; Chen et al, 2019).

The mapping of areas susceptible to landslides and knowledge of the mechanisms of the event are essential tools to assist in urban planning (Abdulwahid and Pradhan, 2017) and to support decision-making in situations of critical events (Tien Bui et al, 2017). In this sense, researchers focus on predicting locations of future landslide events to mitigate disasters (Dou et al, 2019) and identify priority areas for the installation of early warning systems and propose the reorganization of urban space (Aditian et al, 2018).

The landslide susceptibility mapping is considered a primary procedure to predict the probability of occurrence of landslides based on the spatial distribution of conditioning factors related to the characteristics of the physical environment (topographic and geomorphological factors), triggering factors, and human activities existing in the region (Chen et al, 2018; Wang et al, 2015). However, for susceptibility modeling, the landslide inventory and the

conditioning factors of a given area are essential data (Gaidzik and Ramírez-Herrera, 2021).

Susceptibility models are increasingly employing morphometric data from landslide scars or inventories (Tien Bui et al, 2017; Liu et al, 2021; Youssef and Pourghasemi, 2021). The development of quantitative methods makes statistical models (bivariate and multivariate) the most used method for this type of analysis, as they allow determining the weight of each conditioning factor and verifying the results. One example is Logistic Regression (RL), which is widely used successfully in landslide susceptibility mapping work (Saro et al, 2016; Tsangaratos and Ilia, 2016; Du et al, 2017). With fewer statistical constraints, machine learning techniques have been applied in landslide susceptibility studies, such as Support Vector Machines (SVMs) (Pourghasemi et al, 2013; Peng et al, 2014; Dou et al, 2019; Zhao and Zhao, 2021), Artificial Neural Networks (ANN) (Park et al, 2013; Saro et al, 2016; Aditian et al, 2018; Shahri et al, 2019), Decision Trees (DT) (Althuwaynee et al, 2014; Shirzadi et al, 2018; Dou et al, 2019) and Random Forest (RF) (Youssef et al, 2016; Park and Kim, 2019; Hong et al, 2019; Abu El-Magd et al, 2021).

Due to the complex nature of landslide events, it is difficult to predict them. One of the problems is that the quality of the input data directly influences susceptibility predictive models. Most of the time, the prediction model uses conditioning factors derived from a Digital Elevation Model (DEM), which may not be suitable for detailed scale mapping (Pradhan and Sameen, 2017). In this case, the Digital Terrain Model (DTM) is an essential attribute to deriving several significant conditioning factors used in susceptibility analysis (Brock et al, 2020). DTM can provide all geomorphic features such as slope, aspect, curvature, and hydrological resources, such as the Topographic Wetness Index (TWI) and the Stream Power Index (SPI). Due to DTM factors, the prediction accuracy of the Landslide Susceptibility Model (LSM) is highly dependent on the resolution of the DTM (Chen et al, 2020; Brock et al, 2020; Abraham et al, 2021; Gaidzik and Ramírez-Herrera, 2021).

There are different techniques for generating DEM, such positioning on the earth's surface, digitizing of topographic maps, or Remote Sensing (SR) techniques (Saleem et al, 2019). The freely available DEMs are derived from SR techniques, having a horizontal resolution of 30 meters. The fact that these models were obtained more than two decades ago can also impair the prediction capacity (Hawker et al, 2018). The use of SR techniques and Photogrammetry, using Light Detection and Range (LiDAR) and imaging technologies using Unmanned Aerial Vehicle (UAV), together with image processing and computer vision techniques, become essential to obtaining a high-resolution DTM (< 1 meters). Centimeter-scale LiDAR-derived DTM improves landslide susceptibility mapping, increasing accuracy in identifying high and very high susceptibility zones (Gaidzik et al, 2017; Reichenbach et al, 2018; Gaidzik and Ramírez-Herrera, 2021).

The use of LiDAR technology to study landslides is becoming increasingly popular as LiDAR pulse can penetrate tree cover to obtain ground elevation,

thus providing more detail on the microtopography of the study area (Miner et al, 2010; Abdulwahid and Pradhan, 2017; Lin et al, 2013). LiDAR technology has been successfully used not only in landslide detection and mapping (Trigila et al, 2015; Mora et al, 2015; Niculiţă et al, 2016; Niculiţă, 2016; Gaidzik and Ramírez-Herrera, 2021) but also in sliding dynamics and hazard assessment (Haas et al, 2012; Abdulwahid and Pradhan, 2017; Liu et al, 2019; Jena et al, 2020). Centimeter-scale LiDAR-derived DTM improves landslide susceptibility mapping, increasing accuracy in identifying high and very high susceptibility zones (Reichenbach et al, 2018; Gaidzik et al, 2017; Gaidzik and Ramírez-Herrera, 2021).

In the same way, due to the technological development of autonomous driving systems, high-quality digital cameras, micro GNSS trackers, and advances in lightweight carbon fiber airframes, the use of UAVs for research purposes has become increasingly common (Nebiker et al, 2016; Sestras et al, 2021), allowing to collect of data for large-scale land assessment in a cost-effective and relatively inexpensive way at large scales, which was not previously available (Colomina and Molina, 2014; Hosseinalizadeh et al, 2019; Bernardo et al, 2021). Applications of the data obtained from UAVs range from archaeological (Fernández-Hernandez et al, 2015; Campana, 2017) and geological (Saroglou et al, 2018) research, monitoring and modeling of urban areas (Huang et al, 2021), monitoring and management of natural hazards (Giordan et al, 2017; Gomez and Purdie, 2016), monitoring and landslide mapping (Giordan et al, 2018; Peppa, 2018), and landslide assessment (Ramírez-Herrera and Gaidzik, 2017; Sestras et al, 2021; Eker et al, 2018; Eker and Aydın, 2021).

Considering the ease of use of UAV technology and its potential, since it allows mapping, at low cost, areas with a high spatial and temporal resolution (Colomina and Molina, 2014), the data derived from this technology have been widely used for landslide studies (Sestras et al, 2021; Eker et al, 2018; Eker and Aydın, 2021). The UAV data can be used to support landslide prediction models at a detailed scale, applied in specific regions of slope to refine the maps of susceptibility to mass movement available on a medium scale, which were made using DEM base from accessible databases or regional topographic maps. However, its use to map the terrain in dense vegetation cover is hampered due to the lack of exposed terrain surface (Hawker et al, 2018). This problem can be further aggravated due to the extensive occupation of the slopes of the hills. This situation is common in mountainous regions of Brazil, giving rise to geo-hydrological problems with catastrophic consequences (Modenesi-Gauttieri and Hiruma, 2004).

Thus, although there are works related to landslide susceptibility mapping using LiDAR and UAV-derived data, up to this time, few studies have applied an approach that compares the landslide susceptibility maps obtained using data derived from these two technologies. Considering this comparison, a question was raised: Among the products generated from LiDAR and UAV

technologies, in susceptibility models, UAV technology data manages to obtain good results due to its limitation?

In this regard, aiming to contribute to the abovementioned issue, this paper compares landslide susceptibility maps obtained by the Random Forest machine learning algorithm from LiDAR and UAV-derived data. For that, the performances obtained in the predictions were evaluated using statistical evaluation measures, such as receiver operating characteristic curve (ROC), the AUC, and F1-Score, based on training data sets and validation.

This investigation is fundamental to evaluating the capacity of UAV data for landslide susceptibility studies in hillside areas without more detailed scale data, considering the low cost and time for data acquisition.

2 Study area and Data Preparation

2.1 Study area and landslide inventory

The study area for the experiment is located in the municipality of Santos (Figures 1c), which has an important history of landslides, including recent events, as seen in Table 1.

Table 1 Mass movement events of great magnitude in the history of the Municipality of Santos. Source (Sao Paulo, 1988).

Year	Socioeconomic Consequences	Processes
1928	80 deaths 8 buried houses Destruction of part of the Santa Casa de Santos hospital	Landslides mobilizing around 130,000 m² on the slopes of Monte Serrat
1950	3 deaths and several injuries	Landslides on the slope of Monte Serrat
1956	21 deaths and dozens injured 50 houses destroyed	Several landslides on the Santa Terezinha and Monte Serrat hills
1956	43 deaths 100 houses partially or totally destroyed	Landslides in Santos and São Vicente (neighboring municipality)
1959	5 deaths	Landslides in Santos
1978	100 people hit, 4 deaths (January)	Landslides in Monte Serrat and Vila Progresso
1978	11 deaths and several injured (March)	Landslide in Morro Jabaquara and Nova Cintra
1979	13 deaths	Landslides in Santos
1988	1 death 166 homeless 21 homes banned	Landslides in 34 points in the cities of Santos and São Vicente

The municipality of Santos, situated in the Baixada Santista region in the state of São Paulo, Brasil (Figure 1a and 1b), is located at latitude 23°56'27" S and longitude 45°19'48" W. Its area comprises 280,674 km^2 and an estimated population of 430,000 inhabitants, of which 419,000 refer to the population residing in the municipality, with a population density of 1494.26 (hab./km^2) (IBGE, 2017). The city comprises an island with an area of 39.4 km^2 and a mainland area of 231.6 km^2. Santos has a tropical climate and has significant rainfall throughout the year. In the background of the Köppen

climate classification, the climate categorization is Af. The annual average rainfall is 2.143 mm. The rainiest month is January, with an average of 284 mm (CEPAGRI, 2020).

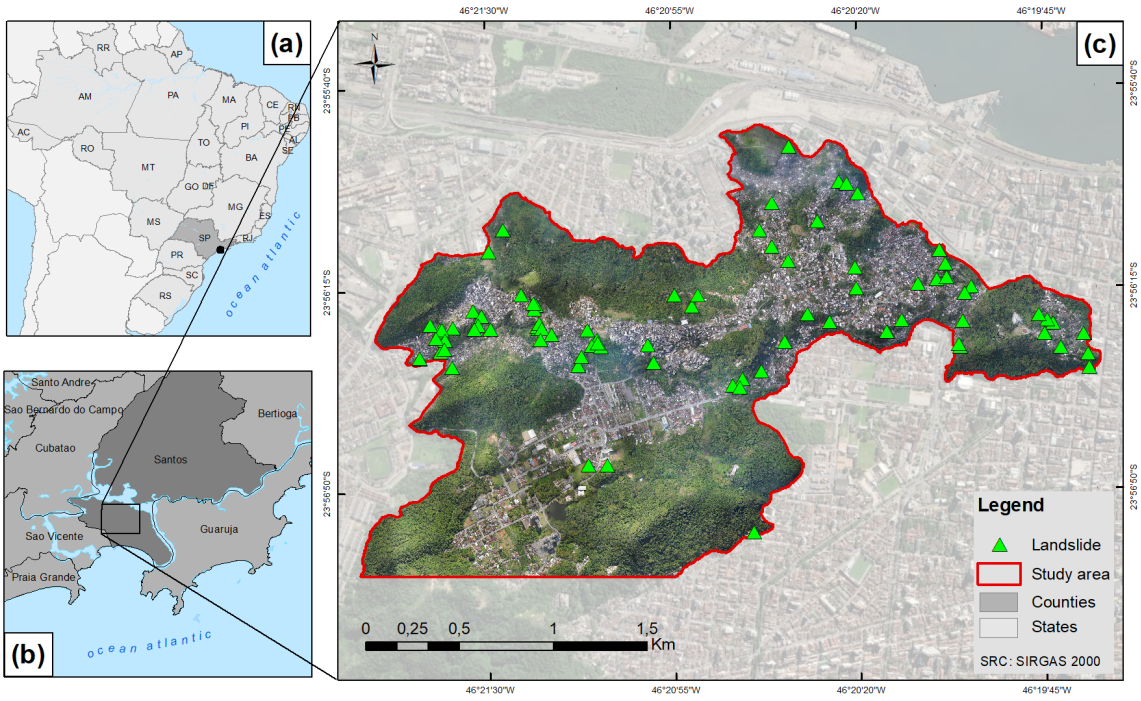

Fig. 1 Landslide location map with: (a) Brasil, (b) state of São Paulo, and (c) Study area: hills of Santos.

This region occupies an area of about 4650 km^2 and has a rugged topography with altitudes that can reach 210 meters above mean sea level (Pichler, 1957). The hills have a steep slope. For slopes no higher than 45°, these are covered by soil, but when these slopes are greater than 45°, rocky outcrops usually appear. The remaining soil layer reaches more than 20 meters at the top of the mountain, while on the hillside, the layer thickness is reduced to a few meters (Pichler, 1957). This area is located in a narrow strip of the plain between the sea and an orographic barrier known as Serra do Mar, where several landslide events have occurred over the years, due to the presence of many elements that favorable these processes, such as high rainfall and high slopes. The location demonstrates the municipality's great potential for landslide events (Moura, 2014; Gobbi et al, 2017).

The inventory of landslides in the study area was collected from reports made available by the Civil Defense of the municipality of Santos and by the database of the Geological Institute of the state of São Paulo. From these, 215 georeferenced occurrences were collected between the years 2014 to 2021, and 379 occurrences between 1993 and 2013.

2.2 LiDAR data

The LiDAR data were processed and made available by the Municipality of Santos from a survey carried out in 2013. LiDAR points were acquired with a Trimble Harrier 68i sensor aboard an AS 350 Ecureuil Helicopter. During data acquisition, the platform was operated at an altitude of 700 meters with an average speed of 101 km/h. The LiDAR sidelap was 35% and the LiDAR endlap was 65%, resulting in a point cloud with an approximate density of 2.38/points per m^2. LiDAR points were divided into ground and surface classes; only ground points were used in this study to obtain the DTM.

2.3 UAV images

The survey for UAV photos captures was carried out by the REMADEN/REDEGEO Project of the National Center for Monitoring and Early Warning of Natural Disasters (Cemaden, in the Portuguese acronym), in partnership with Sao Paulo State University (UNESP, in the Portuguese acronym), on August 23, 2019. Support and geodetic control points were also collected for the quality control of the DTM. The processing and analysis of the UAV photo captures were published by Reiss et al (2022). A SenseFly eBee-Plus UAV was used with a senseFly SODA (Sensor Optimized for Drone Applications) digital camera. The camera produces images with an RGB resolution of 5,472 × 3,648 pixels (20Mpx).

The UAV was operated at the height of 500 m above the ground, with a bow overlap rate of 75% and a lateral overlap rate of 70%. Totaling 417 photographs taken from the study area. GPS UAV device allows to get the geolocation information from optical photos. Agisoft PhotoScan Metashape Professional (v. 1.8.2) was the software operated for processing the images. It is a popular SfM software to process the resulting product, converting aerial photos into dense SfM points. Upon recognizing the photo's metadata (such as focal length and pixel size), the software automatically corrects the camera's intrinsic parameters, calculates the camera's position, and generates a point cloud from the match points of the images (Iglhaut et al, 2019).

2.4 Landslide conditioning factors

Choosing the appropriate conditioning factors is an essential step in mapping landslide susceptibility, as the selection of these factors influences the predictive capacity of the models. The conditioning factors considered in the analysis of landslide susceptibility were chosen based on the literature review and data availability (Pradhan and Sameen, 2017; Sangchini et al, 2016; Zhang et al, 2019). Topographic and hydrological factors include aspect, elevation, plan curvature, profile curvature, slope, Terrain Ruggedness Index (TRI), TWI, SPI, and Sediment Transport Index (STI). In addition, geological factors such as geomorphology and lithology will be used in the analysis.

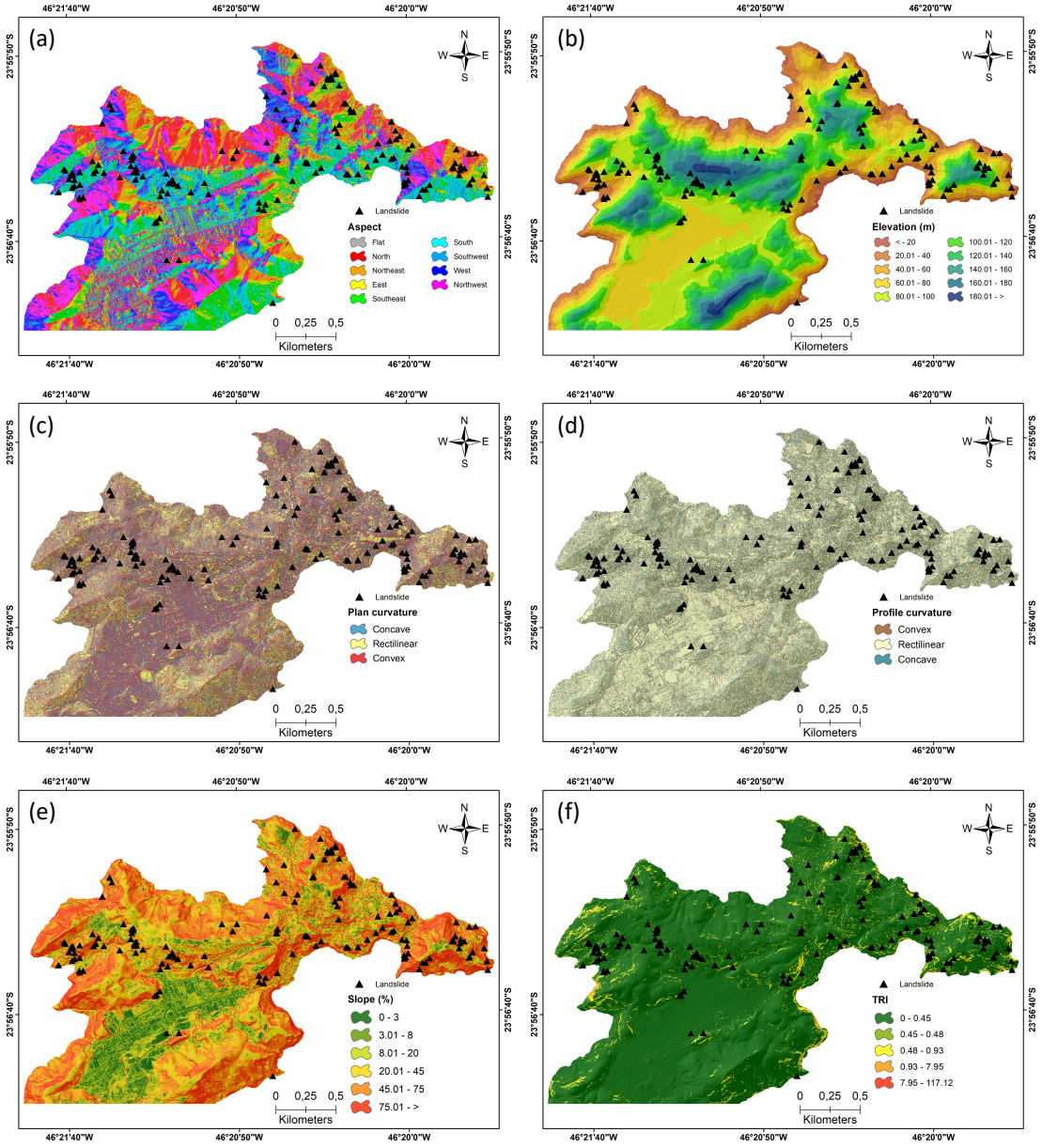

Fig. 2 Topographic landslide conditioning maps derived from LiDAR: (a) aspect, (b) elevation, (c) plan curvature, (d) profile curvature, (e) slope, and (f) TRI.

Topographic factors to be used are aspect, elevation, plan curvature, profile curvature, slope, and TRI, shown in Figures 2 and 3 from LiDAR and UAV-derived data, respectively. The aspect is defined as the azimuth corresponding to the steepest slope of the terrain, which determines the degree of solar radiation on the hillside and significantly impact on the physical environment and vegetation around the slope (Xiao et al, 2018). The elevation is affected by various geomorphological and geological processes, with relief classes denoting the range of elevation from the lowest to the highest point in the study area (Youssef and Pourghasemi, 2021; Raja et al, 2017).

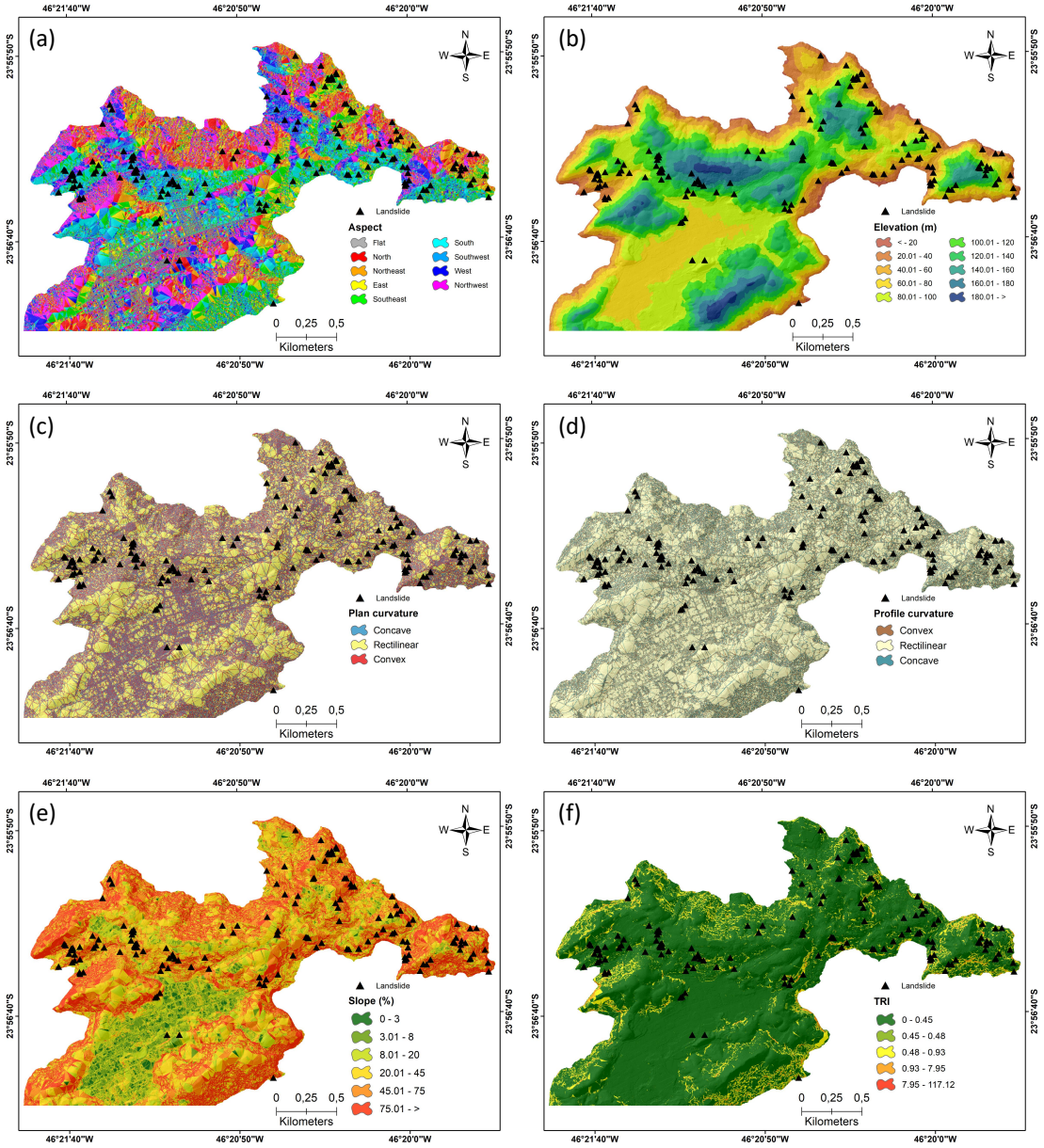

Fig. 3 Topographic landslide conditioning maps derived from the UAV: (a) aspect, (b) elevation, (c) plan curvature, (d) profile curvature, (e) slope, and (f) TRI.

The plan curvature represents the flow acceleration, and the erosion/deposition rate affects the convergence and dispersion of flows (Youssef et al, 2016). Positive values correspond to divergent topography, negative values to convergent topography, and null values to flat topography (Chen and Zhang, 2021). The profile curvature affects the change in slope flow velocity and the associated process of sediment transport influences the acceleration and deceleration of the flow. Positive values represent convex terrain, negative values, concave terrain, and null values represent straight terrain (Xiao et al, 2019).

The slope angle determines the terrain surface's steepness and landslide's stress degree the topographical surface concerning the horizontal slope (He et al, 2019). The shear stress in the soil or other unconsolidated materials generally increases with the slope, so that the slope directly affects the occurrence of landslides, as the increase in the inclination angle implies a decrease in the safety factor, increasing the occurrence of landslides (Tien Bui et al, 2017). For Liu et al (2021) the inclination angle proved to be the most significant conditioning factor.

The TRI is considered a morphometric measure that describes the heterogeneous condition of a land surface (Althuwaynee et al, 2014), is related to the texture of the terrain, and is used to represent the difference in elevation between the surface and the surrounding area, integrating the concave and convex regions (Riley et al, 1999; Smith, 2014). In Equation 1 for calculating the TRI, *max* denotes the largest pixel value in a mobile 3x3 pixel kernel and *min* denotes the minimum value (Pawluszek and Borkowski, 2017).

$$TRI = \sqrt{(|x|(max^2 - min^2))} \tag{1}$$

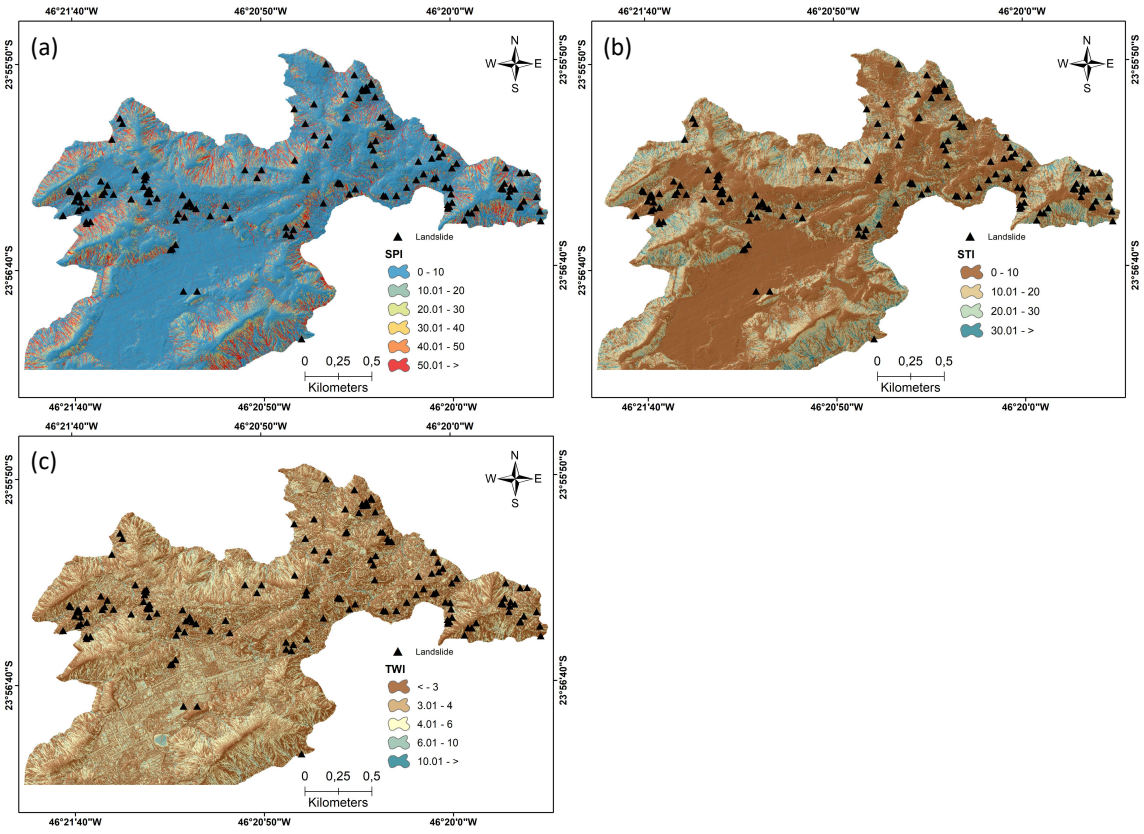

Fig. 4 LiDAR-derived hydrological landslide conditioning maps: (a) SPI, (b) STI, and (c) TWI.

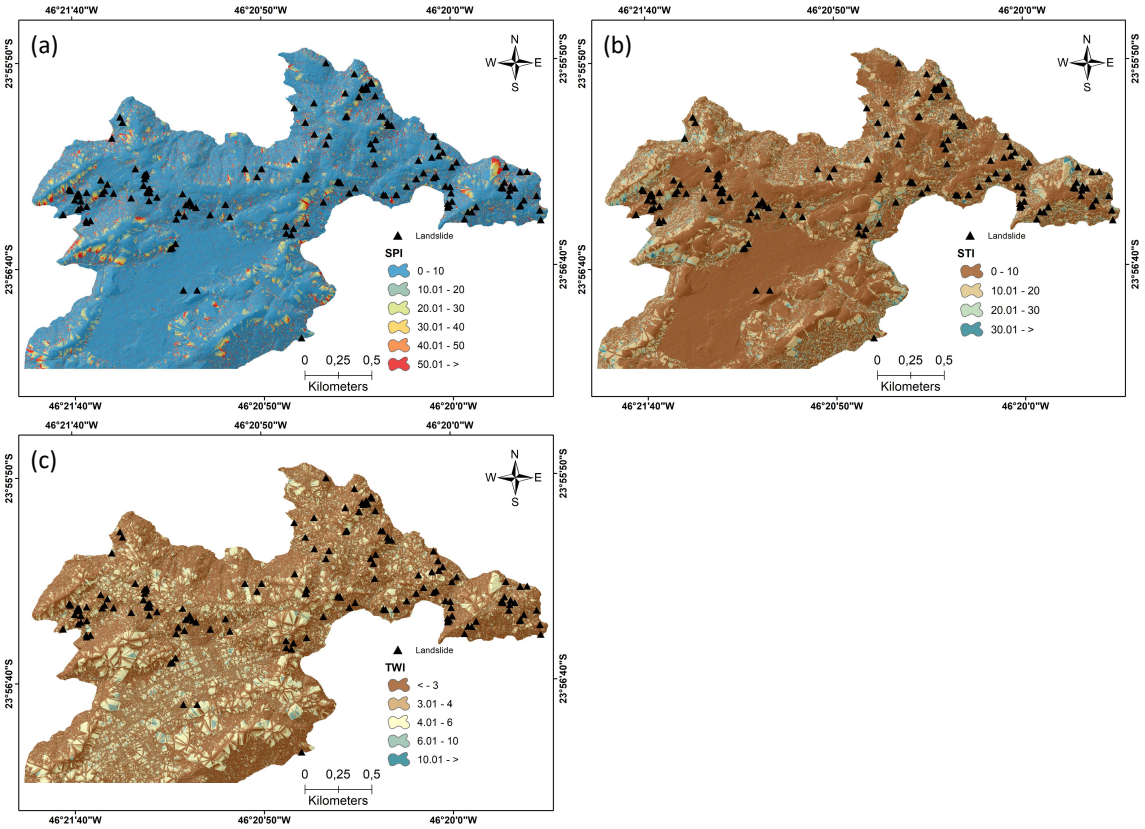

Fig. 5 UAV-derived hydrological landslide conditioning maps: (a) SPI, (b) STI, and (c) TWI.

Hydrological factors have a significant influence on the occurrence of landslides (Liu et al, 2019). As in studies carried out by Lee et al (2018); He et al (2019); Wang et al (2020), the SPI, STI, and TWI factors were used in this study, as presented in Figure 4 for LiDAR and in Figure 5 for UAV. Based on the DTM, flow accumulation is performed to simulate the flow or potential flow of streams and rivers. From the flow accumulation, the shape of the valley in the hydrological basin is derived. The SPI is characterized by the erosion/deposition power of a water flow, and with the presumption that it relates the discharge to a defined catchment area, it was proposed by Moore et al (1991). In Equation 2, A_s is the local ascending area that drains through a certain point per unit boundary length and b is the local slope gradient measured in degrees.

$$SPI = A_s * \tan b \tag{2}$$

The STI (Equation 3) is directly related to the delivery of sediments from the terrain to the channel during landslide events (Saleem et al, 2019). This represents slope failure and deposition procedures (Moore et al, 1991). In Equation 3, the elements A_s and b were defined in Equation 2, and m and n are respectively the exponent of the catchment area and and the slope.

$$STI = (m+1) * (A_s/22.13)^m * (\sin b/0.896)^n \tag{3}$$

TWI characterizes surface water saturation zones affecting slope material, resulting in decreased soil stability (Chen and Zhang, 2021). This conditioning factor can be calculated according to Equation 4, where a is the local ascending area drains through a certain point per unit boundary length, and tan b is the local slope gradient measured in degrees.

$$TWI = \ln\left(\frac{a}{\tan b}\right) \tag{4}$$

The geological factor applied in this work was the lithology and the geotechnical factor represented by the geotechnical map, as seen in Figure 6. These were reconstructed from studies carried out by the Technological Research Institute (IPT, in the Portuguese acronym) in 1979, the first one entitled Lithological Map of the Hills of Santos and São Vicente and the second geotechnical map of the mountains of Santos and São Vicente (municipality neighbor), both with a scale of 1:10,000. The lithology was characterized by the following classes: Migmatites, Granite Santos, Granite with Megacrystals, Gravitational Sediments, Alveolar Sediments Gravitacional Sediments, and Alveolar Sediments. Furthermore, the geological conditions represent one of the most important factors that influence the stability of the slopes. In this regard, the geotechnical map is divided into 11 classes, according to Table 2.

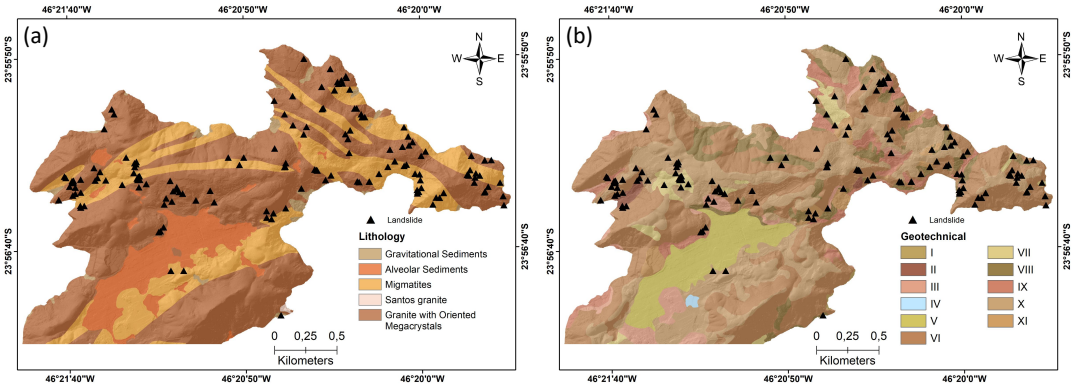

Fig. 6 The geologic and geotechnical factors used in the analysis: (a) Lithology, and (b) Geotechnical.

2.5 Frequency ratio

The Frequency Ratio (FR) model is commonly used for landslide susceptibility analysis (Ozdemir and Altural, 2013; Hong et al, 2016). FR is a geospatial assessment tool and one of the most essential bivariate statistical methods (Wang and Li, 2017), which involves bivariate statistical analysis based on spatial distribution, that is, on the probability of occurrence and non-occurrence of landslides in a specific area (Park et al, 2013). The bivariate probability of each independent landslide trigger can be determined by its relationship to the occurrence of inventory landslides (Oh et al, 2017). If the

Table 2 Geotechnical classes.

Legend	Geotechnical classes
I	Areas explored or under exploration for material removal
II	Talus bodies with strong evidence of movement
III	Bands situated below unstable zones
IV	Lake
V	Alluvial plain embedded in the top of the hills
VI	Segment of predominantly straight slopes with slope >40°
VII	Segment of predominantly straight slopes with slope 30° - 40°
VIII	Segment of predominantly straight slopes with slope 20° - 30°
IX	Segment of straight or convex slopes, slightly inclined <20°
X	Straight or convex hill tops and slopes, slightly inclined
XI	Deposition zone (talus) apparently stable and large blocks may occur on the surface

bivariate probability is greater than 1, the stronger the correlation between the incidence of landslides and the triggering factors, and if the probability is less than 1, the weaker the correlation (Aghdam et al, 2016). FR can be described by Equation 5, which was used to calculate all subclasses of conditioning factors based on their relationship with the scar inventory (Oh et al, 2017).

$$FR = \frac{(A/B)}{(C/D)} \tag{5}$$

Where A is the number of pixels with landslides for each subclass conditioning factor, B is the total number of landslides in the study area, C is the number of pixels in the subclass area of each factor and D is the total number of pixels in the study area. Then, the FR was normalized, generating the Normalized Frequency Ratio (NFR) on a scale of probability values from 0 to 1 using Equation 6, where the Factor class FR and FR by class (Yan et al, 2019).

$$NRF = \frac{Factor\ class\ FR}{\sum Factor\ class\ FR} \tag{6}$$

3 Methods

The methodology proposed by this work comprises the steps described in Figure 7. In the first part of the methodology, information on LiDAR and UAV data was obtained. Then, the conditioning factors were produced based on the DTM generated for each data with a resolution of 0.5 meters. All conditioning factors were made using the Whitebox Geospatial Analysis Tools (GAT) platform (Lindsay, 2016). Those with numerical data were classified into class intervals so that their classes are defined factors with nominal data (lithological and geotechnical) that have not been changed. In the event frequency distribution analysis stage, based on the inventory data, the conditioning factors were reclassified according to the probability of occurrence and non-occurrence of landslides. The reclassified conditioning factors were used as input to the RF

algorithm, as were the training samples, to generate the landslide susceptibility maps. From the validation samples, the susceptibility model evaluation metrics were calculated.

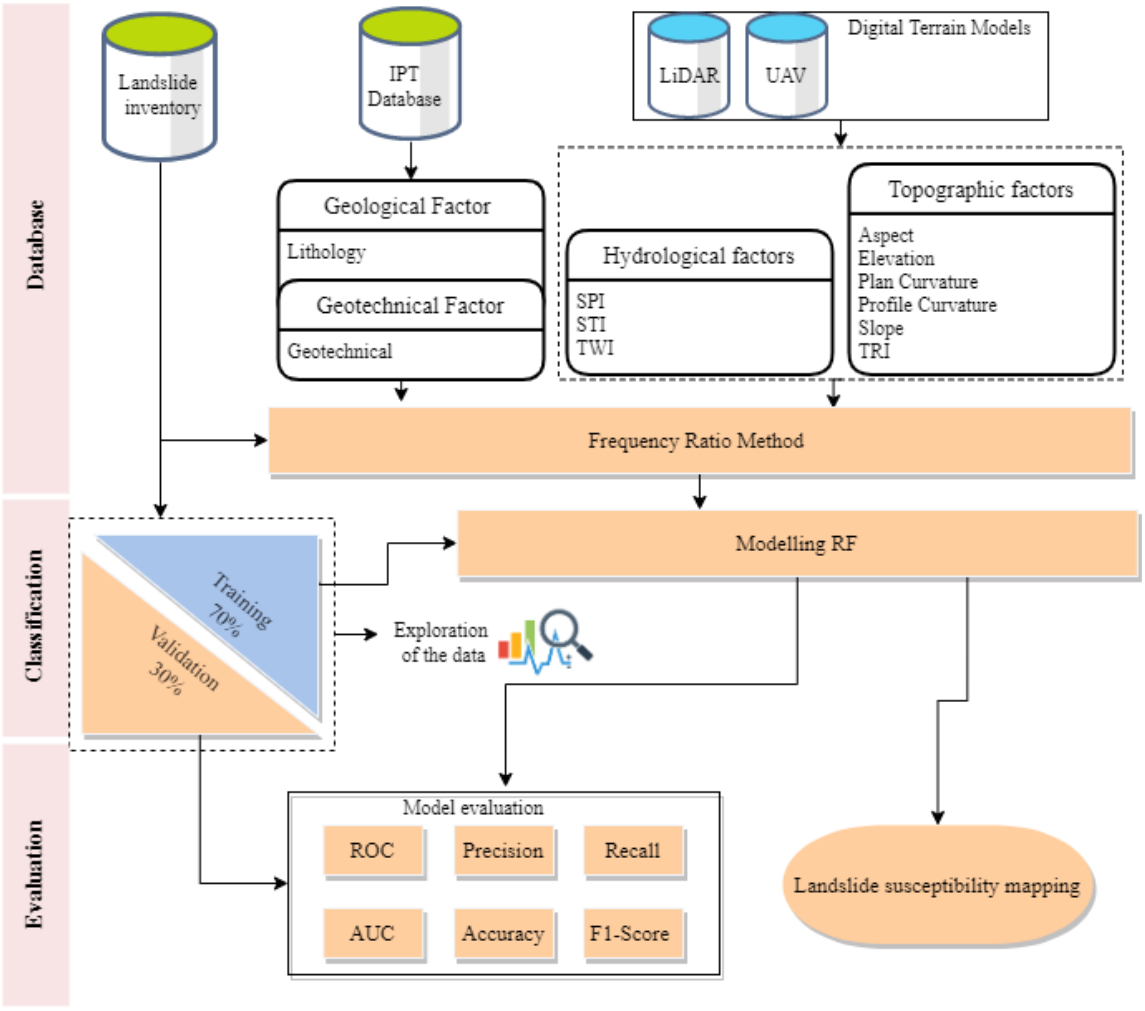

Fig. 7 Flowchart of the proposed methodology to landslide susceptibility mapping using RF algorithm.

3.1 Mapping units and dataset division

To produce susceptibility maps, ML techniques require both occurrence and non-occurrence samples for landslides. However, these samples must be divided into training and validation sets. The first is used to build the prediction model and the second is to analyze the model's predictive capacity. As proposed by Sahin et al (2020) and Dang et al (2020), data separation is due at 70% for training and 30% for validation. Thus, from the union of landslide occurrences inventories made available by the Civil Defense and IG (Instituto Geológico, in Portuguese), a previous selection of the landslide points was carried out considering that the landslide occurrences are related to occupied areas in dense urban settlements. This fact induces the localization of these

events in the study area. For the present study was used 215 samples of the landslide. The exact number of samples (215 points) was randomly generated within the study area to generate a set of non-landslide points, totaling 430 samples. Considering the 70:30 ratio for the division into training and validation samples, for the training dataset, 301 randomly divided points were used, and for the validation dataset, 129 points were used.

3.2 Multicollinearity diagnosis

To build a LSM is vital to select suitable variables for landslide modeling because conditioning factors within the training dataset can be correlated with each other, which can introduce noise that confuses the model (Sahin et al, 2020). In this study, the Spearman's correlation coefficient was used to quantify the multicollinearity of the conditioning factors.

Spearman's correlation analysis between the conditioning factors is a non-parametric statistical indicator of correlation that uses the classifications for calculation and measurement, reflecting the proximity of the correlation between two variables (Wang et al, 2020), this varies between -1 and 1, representing the negative and positive linear correlation, respectively. The value zero means no linear correlation. The Spearman's correlation coefficient is obtained using Equation 7.

$$r_s = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \tag{7}$$

Where: d_i is the difference between the ranks, and n is the number of members for each parameter. Given a pair of random variables, the linear correlation (r) can be summarized as follows:$|r| \leq 0.3$ there is no linear correlation; $0.3 \leq |r| \leq 0.5$ low linear correlation; $0.5 \leq |r| \leq 0.7$ significant linear correlation; and $|r| \geq 0.7$ high linear correlation (Wang et al, 2020). An advantage of the Spearman's rank coefficient is that if one or more data are much larger than others, their correlation is not affected because only their ranks are evaluated (Rezaei et al, 2018).

The importance analysis of the conditioning factors in the prediction model was performed using the implicit importance technique in the RF algorithm, which is implemented in the scikit-learn library (available in https://scikit-learn.org/stable/). This tool measures the importance of attributes or, in this case, the conditioning factors by analyzing the number of nodes in the tree that use a particular attribute to reduce the global impurities of the forest. It automatically calculates this value for each attribute after training. It normalizes the variables based on the standard deviation of differences in Out of Bag (OOB) error before and after the variables are shuffled and checked within the training data. This metric captures the model's sensitivity to each predictor variable. The more sensitive, the more important the variable for the performance of the model (Sahin et al, 2020; Tanyu et al, 2021).

3.3 Random Forest algorithm

The RF method, as an ensemble machine learning classification algorithm, aims to improve the flexibility and accuracy of classification trees (Zhang et al, 2019). This method classifies the data using a combination of decision trees. Each tree depends on the values of a random sample vector with independent configuration and is uniformly distributed over all the trees (Breiman, 2001). RF increases the diversity between classification trees, alters the data, and arbitrarily modifies a set of explanatory factors for different inductive processes in trees (Chen et al, 2017).

The final decision of the class members and the construction of the model is determined by majority votes in the tree (Dang et al, 2020; Sahin et al, 2020; Zhou et al, 2021). The model is built from several parameters, the most important being: the number of trees in the forest, the maximum number of features that RF can experience in a single tree, the number of variables or conditional factors tested at each node, and the number of nodes controlling the structure in each tree to minimize the OOB error and achieve good performance of the model (Taalab et al, 2018; Sahin et al, 2020).

The RF model was used to perform the classification of the predictive model. The Python language was used to build algorithms implemented with the scikit-learn library. To choose the number of trees to be used, the search was implemented by iteratively calculating the RF OOB errors with different numbers of random attributes. The smaller the OOB errors, the greater the prediction accuracy of the corresponding model. After the analysis, the number of predictors used to split the nodes was 3, and 200 was used for the number of decision trees in RF. Then the trained RF model can be used to calculate the prediction models.

3.4 Evaluation and comparison of results

The ROC has been widely used to validate the performance of LSMs (Raja et al, 2017; Sun et al, 2020; Chen and Zhang, 2021). It is a valuable method to represent two-dimensionally the performance of prediction models, analyzing the relationship between sensitivity and specificity of a binary classifier, in this case, the occurrence or not of landslides. The ROC curve shows the value of specificity and sensitivity, respectively, on the y and x-axes. The true positive (TP) is a pixel correctly called the landslide area, and the false positive (FP) is a pixel incorrectly named landslide area (Liu et al, 2021). Thus, it is necessary to calculate the specificity and sensitivity given by Equations 8 and 9, respectively:

$$Sensitivity = \frac{TP}{TP + FN} \tag{8}$$

$$Specificity = \frac{TN}{TN + FP} \tag{9}$$

Where: TP and true negative (TN) is the number of correctly sorted pixels, FP and false negative (FN) is the number of wrongly sorted pixels

The AUC is as it indicates the accuracy of the models on a scale of 0 to 1, it is calculated from the area under the ROC curve. A maximum AUC or close to 1 indicates that the model produced an excellent result. On the other hand, an AUC value close to 0.5 means a bad result for the prediction. In general, it is considered that if the AUC of the model is more significant than 0.7, the model has high accuracy (Raja et al, 2017). ROC and AUC curves are among the most used metrics for evaluating a machine learning model in landslide susceptibility analyses.

In addition, overall accuracy (OA) in Equation 10, is widely used to evaluate the predictive ability of landslide models. OA is the proportion of landslide and non-landslide pixels correctly classified by the model (Chen et al, 2017; Lv et al, 2022). A higher OA value implies better classification precision (Wang et al, 2019).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{10}$$

On the other hand, the F1-Score (Equation 11), according to Azarafza et al (2021) represents the harmonic mean of Precision (Equation 12) and Recall (Equation 13), considering the performance of the positive class to evaluate a model, that is, how many samples were correctly classified as positive, as well as how robust is the classification to measure the ability of the model not to classify a negative sample as positive (Dang et al, 2020). Thus, according to Sahin et al (2020), the results vary between 0 and 1. The closer to 1, the more correctly the instance is classified, so if any of the sensitivity and precision values decrease, the F1-Score decreases and vice versa.

$$F1 - Score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{11}$$

$$Precision = \frac{TP}{TP + FP} \tag{12}$$

$$Recall = \frac{TP}{TP + FN} \tag{13}$$

4 Results

The frequency ratio model is built based on the spatial distribution, that is, based on the probability of landslides. This step allows analyzing for each factor which class intervals are most relevant for the triggering of landslides, and what are the possible thresholds for each variable. Nevertheless, for some conditioning factors, the distributions of these processes are associated with different intervals, and it is not possible to clearly define which class is more likely.

Concerning the topographic parameters, elevation presents high values in both models derived from LiDAR and UAV, in the range of 20m - 60m and 100m - 120m, representing approximately 34% of the study area. The South (S) and East (E) classes of the aspect factor presented the highest ratio values. These were the most expressive and exceeded all estimates. In addition, the

Southeast (SE) model class using UAV-derived data has a value of 17%. Considering the TRI, the class that stood out the most for the UAV data model was 0.45 - 0.48, and for the LiDAR data, it was 0.48 - 0.93.

For the slope in the model using UAV data, the more expressive classes (with their percentage of landslides in each class) were 20.01 - 45 (18.60%), 45.01 - 75 (34.88%), > 75.01 (42.79%). However, for the model using LiDAR data the occurrences of landslides were concentrated in the following classes (with their percentage of landslides in each class): 20.01 - 45 (10.23%), 45.01 - 75 (18.14%), > 75.01 (65.12%). Thus, totaling 75.43% and 73.46% of the study area, respectively, for models using UAV and LiDAR data. The difference was more evident in the class 45.01 - 75, verifying the most significant presence of vegetation in this class, which makes modeling for the UAV difficult.

Regarding the hydrological parameters, the TWI of the study area was classified into five classes: (< 3), (3.01 - 4), (4.01 - 6), (6.01 - 10), and (10.01>). The most representative class was (< 3). For the SPI the representative classes were 40.01 - 50 for UAV data-based model and 10.01 - 20 for LiDAR data-based model. The STI presented high ratio values between the classes 10-30 for a model using LiDAR data, and for a model using UAV data the class that stood out the most was the 20 - 30.

For the geotechnical condition, the classes of predominantly straight slopes with slopes ranging from 30 to 40, followed by the class of predominantly straight slope segments with slope >40, were the highest. These classes are directly related to the slope of the relief. In lithology, landslides were present in migmatites, Granite with Oriented Megacrystals, and deposition zone (talus) in addition to Gravitational Sediments.

4.1 Multicollinearity of factors

Spearman's correlation analysis was performed for the conditioning factors to assess the correlations between variables. Figure 8 shows the correlations in color scale, where it can be seen that the factors that have the highest correlation for the model using LiDAR data (Fig. 8a) are the Slope with TRI, and STI and the SPI with STI maximum correlation value of (Slope - TRI 0.5), (Slope - STI 0.59), (SPI - STI 0.65) and for the model using UAV data (Fig. 8b) (Slope - TRI 0.57), (Slope - STI 0.45), (Slope - TWI 0.41). A high correlation between Slope - TRI also was obtained by Zhao and Zhao (2021) and in the study by Wang et al (2020) obtained high correlations between SPI - STI, and Slope - TWI.

The correlations between the TWI and the slope are explained as evidenced by the fact that one is inversely proportional to the other, while the TWI describes the effects of land change on surface runoff, so when the slope is relatively flat, the impact of terrain changes on the flow rate is higher (Xiao et al, 2019). For SPI and STI, high values represent the water contribution of slope and high flow velocities, as well as the effect of topography on erosion (Pradhan and Kim, 2014). The result demonstrates no strong correlations between

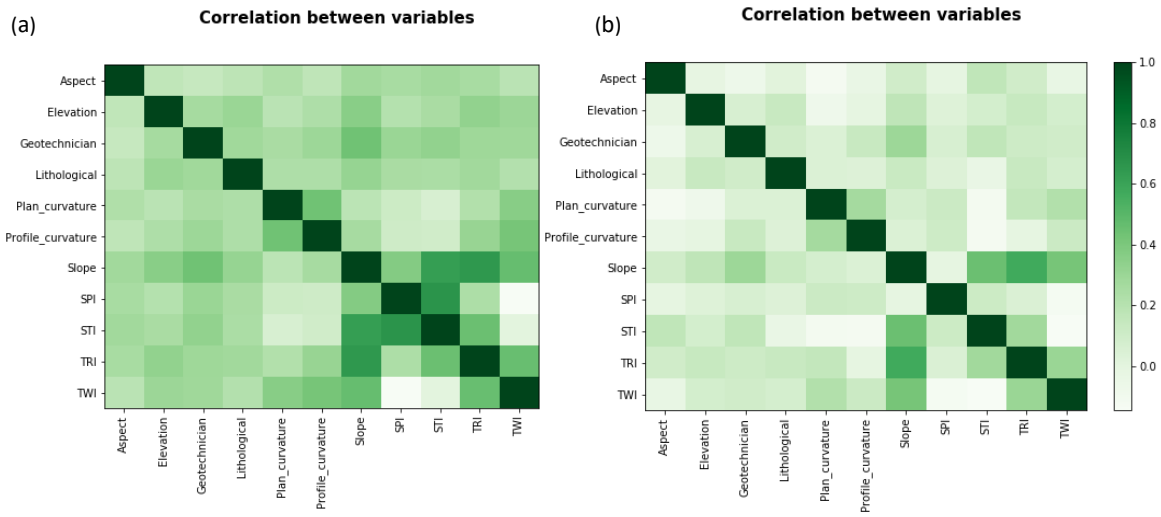

Fig. 8 Spearman's correlation for models using (a) LiDAR data, and (b) UAV data.

the conditioning factors selected to be used in the landslide prediction model as they do not present values above 0.7, indicating high collinearity.

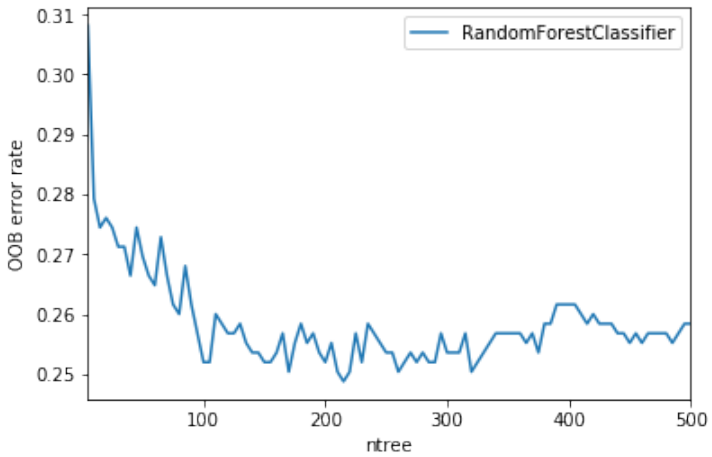

Fig. 9 RF model OOB error graph.

To perform the landslide prediction model using RF, an investigation was performed between the OOB error against the number of trees employed in the model. According to Sahin et al (2020), in this investigation, we sought to know, from the training samples, the number of trees that present the smallest OOB error. In this process, it was found that the model with 200 trees would be ideal, as shown in the graph in Figure 9.

Thus, the classification of the LSM was performed using the parameters described in the methodology, and the importance of each conditioning factor

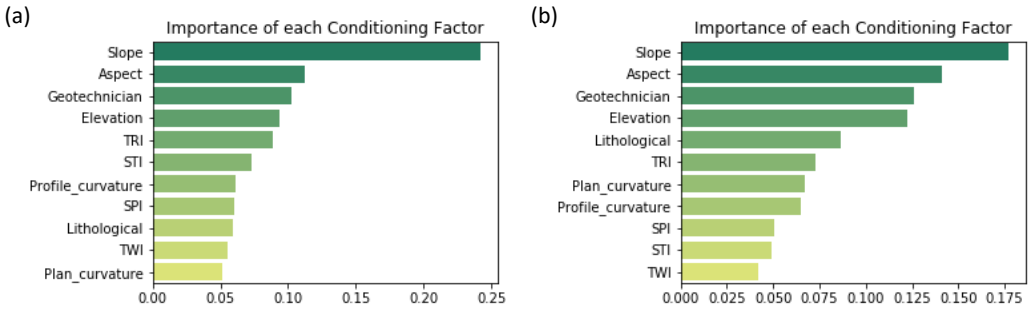

Fig. 10 Importance RF for models using (a) LiDAR data, and (b) UAV data.

was verified. As a result, Figure 10 presents the values of the importance of the variables given by the RF algorithm.

The importance of each factor is shown as an average of the impurity decrease that appears on the x-axis, as shown in Figure 10. The Slope presented the most expressive values among the variables for LiDAR data (0.24, Figure 10a) and UAV data (0.18, Figure 10b) , as it is a geomorphological unit of hills with a morphodynamic that tends to a high dissection of slopes and width relevant. This variable was also of great importance in the studies presented by Abu El-Magd et al (2021); Sahin (2020); Saleem et al (2019); Youssef and Pourghasemi (2021). The result of the other factors for LiDAR was Aspect (0.11), Geotechnician (0 .10), Elevation (0.09), TRI (0.09), STI (0.07), Profile curvature (0.06), SPI (0.06), Lithological (0.06), TWI (0 .05), and Plan curvature (0.05). For the UAV we have Aspect (0.14), Geotechnician (0.13), Elevation (0.12), Lithological (0.09), TRI (0.07), Plan curvature (0.07), Profile curvature (0.07), SPI (0.05), STI (0.05), and TWI (0.04).

Landslides are typically binary problems, and the confusion matrix is often used to analyze the model's prediction accuracy. As in studies by Taalab et al (2018); Sun et al (2020); Zhou et al (2021), the value of 0.5 was used as a limit to judge the predicted value. When the probability is >0.5, it is considered a landslide (assigned 1); otherwise, it is considered a non-landslide (assigned 0). Therefore, the analysis of Precision, Accuracy, Recall, and F1-Score as indicators of model evaluation.

Table 3 Performance of the four models.

Statistical measures	LiDAR	UAV
Precision	0.800	0.797
Recall	0.862	0.712
Accuracy	0.862	0.760
F1-Score	0.830	0.752

The results obtained in this study applied different performance metrics for the models using LiDAR and UAV-derived data. The performance results in terms of all metrics were given in Table 3. According to the model using LiDAR data, the estimated values for Precision, Accuracy, Recall, and F1-Score are 0.800, 0.862, 0.862, and 0.830, respectively, and for the model using UAV data, the Precision, Accuracy, Recall, and F1-Score 0.797, 0.712, 0.760 0.752, respectively.

High Precision values indicate that every pixel predicted to be a scar is a landslide scar, no matter how many landslide scars are not detected. However, a high Recall value means that all landslide scars are detected, no matter how many objects are misclassified as landslide scars. Now, analyzing the Accuracy, it is noted that the two technologies were approximated when the predicted values of landslides are in fact, landslides. However, the recall had a differential, emphasizing that the LiDAR technology could better classify all the scars concerning the UAV. As a result, when evaluating the confusion matrix considering the adopted threshold of 0.5 and measuring Accuracy between known class values and predicted class values, it can be said that the model using LiDAR data outperformed the model using UAV data in terms of metric analysis.

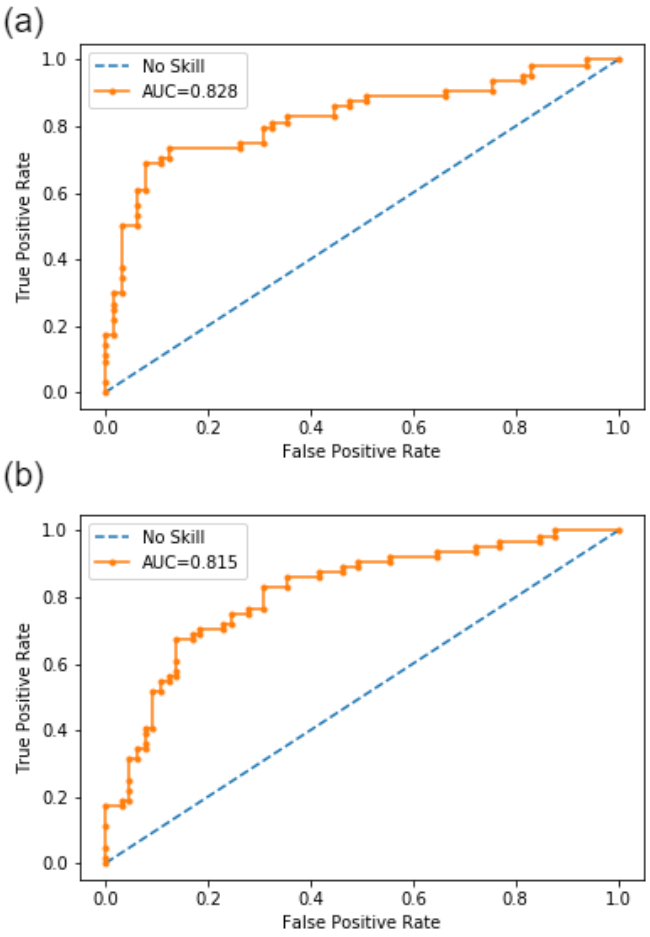

Fig. 11 The prediction ROC and AUC values for models using: (a) LiDAR data; and (b) UAV data.

According to the analysis of the ROC curve and prediction capabilities with the AUC value (Figure 11), the values of the models were above 0.8,(0.828 and 0.815, for LiDAR and UAV, respectively), which indicates that the models were effective for landslide susceptibility analysis. The landslide density (LD) can quantitatively reflect the spatial distribution of the landslide susceptibility maps, defined as the percentage of landslide pixels divided by the percentage of susceptible class pixels. As a result, in Table 4 it is observed that the value of LD increases progressively as the level of landslide susceptibility raises from very low to very high. Similar results were also found by Yan et al (2019); Wang et al (2020) showing similar LD distribution trends in susceptibility models.

Table 4 Landslide distribution in predicted landslide susceptible zones.

Landslide-susceptible zones	LiDAR Area of zones (%)	UAV Area of zones (%)	No. of landslide LIDAR	No. of landslide UAV	Landslide LIDAR (%)	Landslide UAV (%)	LIDAR Density (LD)	UAV Density (LD)
Very low (I)	18.91	24.79	3.00	2.00	1.40	0.93	0.07	0.04
Low (II)	42.99	25.98	8.00	9.00	3.72	4.19	0.09	0.16
Moderate (III)	21.42	26.68	56.00	31.00	26.05	14.42	1.22	0.54
High (IV)	10.97	18.32	54.00	111.00	25.12	51.63	2.29	2.82
Very high (V)	5.70	4.24	94.00	62.00	43.72	28.84	7.67	6.80

The maps of susceptibility to landslides resulting from the models were reclassified into five categories (Figure 12), considering the method used by (Chen and Zhang, 2021), which defines the classes by equal intervals: very low (0.005 - 0.2), low (0.2 - 0.4), moderate (0.4 - 0.6), high (0.6 - 0.8) and very high (0.8 - 1), because, to compare the susceptibility models produced by different technologies, it is necessary to use the same standard to identify levels of susceptibility to landslides.

Figure 12 shows the landslide susceptibility maps. Both results have a similar spatial distribution, i.e., most historical landslide sites are in areas of high susceptibility. The areas of very low susceptibility are mainly distributed in the central part of the study area with low altitude and low slope, with the alluvial plain embedded in the top of the hills.

5 Discussion

The number of studies using different methods in LSM and comparing their performance to obtain the best models is increasing. However, landslides are a natural phenomenon and a very complex process, affected by several geo-environmental factors; thus, it is not easy to be modeled accurately (Pham et al, 2020). Therefore, it is crucial to improve the performance of forecasting methods so that managers and decision-makers obtain more accurate maps of susceptibility to landslides. Recently, machine learning techniques have been successfully adopted and applied to evaluate landslide susceptibility (Huang et al, 2020). The performance of these models mainly depends on the input data (Gaidzik and Ramírez-Herrera, 2021). So, it is essential to have quality data to provide them as input to learning models, either by the landslide inventories or concerning the DTMs, which are bases for the conditioning factors.

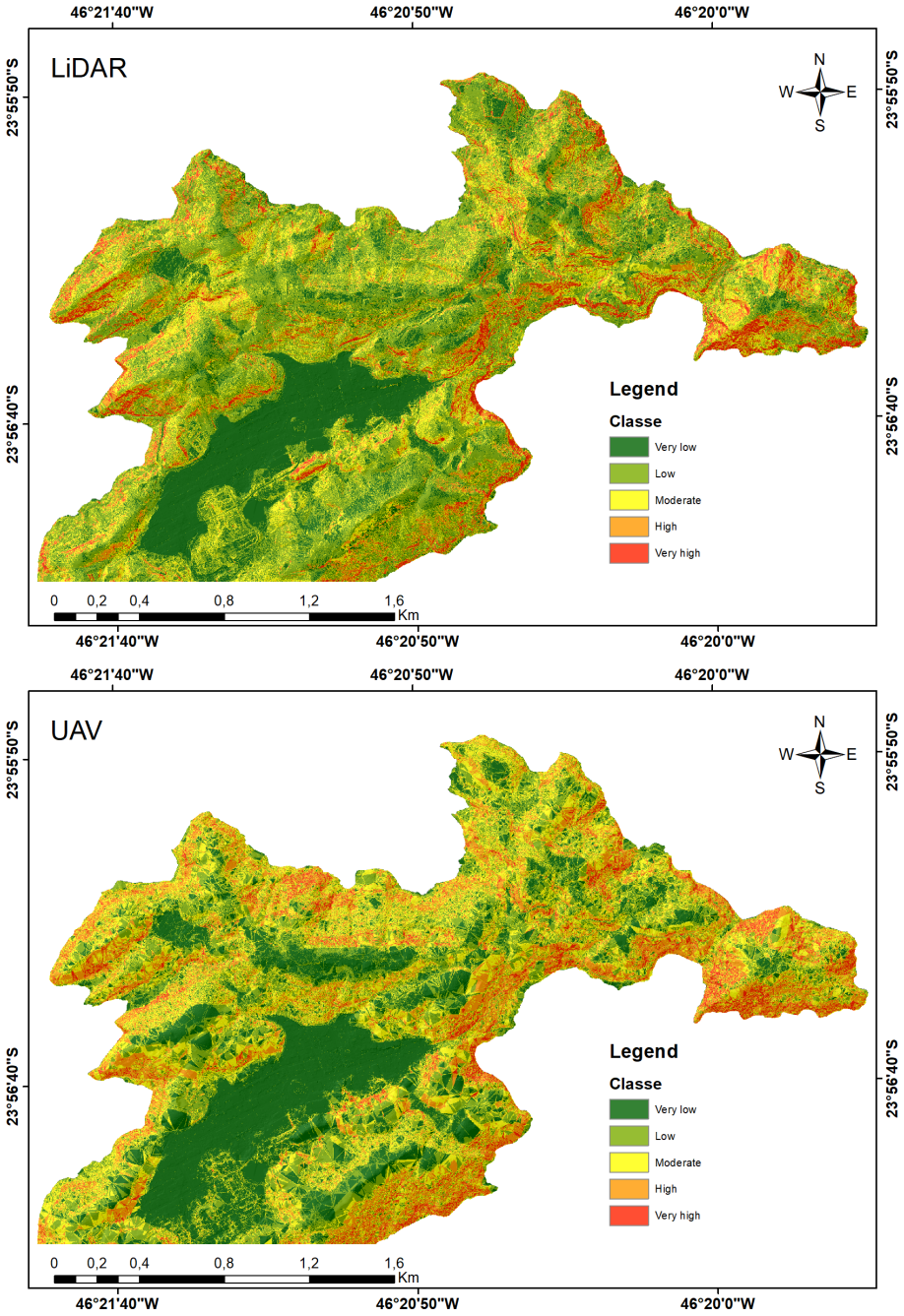

Fig. 12 Landslide susceptibility maps produced by proposed method using: (a) LiDAR data, and (b) UAV data.

The importance of inventories helps to understand which pattern of terrain attributes is related to occurrences (Liu et al, 2021; Youssef and Pourghasemi, 2021). Furthermore, the DTM is the data source for most of the input factors in susceptibility modeling; as a result, the quality of the DTM data and the data acquisition form is a significant input for producing an accurate LSM output. This fact was verified by Gaidzik and Ramírez-Herrera (2021); Mahalingam et al (2016), which showed that landslide susceptibility mapping using high-resolution DTM LiDAR-derived had its accuracy improved when compared with a lower resolution DTM. In this context, the comparison between two

acquisition technologies for DTM was evidenced in this study, considering the UAV technology as a form of rapid acquisition and a positive cost-benefit ratio.

The susceptibility models showed great performance results, proven by the evaluation metrics. However, differences were observed in the classification of the conditioning factors values. These values were more divergent in densely forested areas, a relationship that the nature of the sensor can easily explain to obtain the topographic data. With the passive nature of the optical sensor of the UAV technology, which data produced is impenetrable to the vegetation. Even with robust filtering algorithms to generate the DTM, removing vegetation and inferring the relief in densely forested areas, is a complex task. Regarding LiDAR data, the density of points in forest areas is higher due to the active nature of the sensor.

However, the limitation with LiDAR technology data concerns its high acquisition cost and, consequently, low availability. In Brazil, few municipalities have a survey using LiDAR technology and make such data available for research.

This study allowed shows that the data from UAV technology have potential to contribute to more risk areas being studied at the detail level. This potential also was verified by Borrelli et al (2019), who used LiDAR and UAV data from different dates (2012 and 2019) to evaluate the topographical changes that occurred within a small catchment. Moreover, the use of the UAV data can subside data for other studies that depend on the quality of susceptibility or risk maps, especially regarding to the accuracy of risk areas delimitation, as proposed by Assis Dias et al (2018), who estimates the population in an area at risk on an intra-urban scale.

The landslide susceptibility mapping helps identify areas that researchers can pay more attention to for more detailed investigations (Cao et al, 2021). It is in this way that susceptibility mapping using high resolution DTM like those evaluated in this article has the potential to complement studies and applications in disaster risk areas, such as the definition of suitable locations for the installation of collection platforms of data and collection of subsurface data, besides providing subsidies for the elaboration of contingency plans.

Thus, considering a large country like Brazil, with 6,227,740 vulnerable people in landslide risk areas (Assis Dias et al, 2020), and studies indicating Brazil's increased vulnerability to landslides and flash floods in the context of climate change (Debortoli et al, 2017), the landslide susceptibility mapping with more easily accessible and cheaper technologies as UAV is a need and urgent.

6 Conclusion

This article compares the two models of landslide susceptibility mapping using RF algorithm generated from the LiDAR and UAV technology to verify if the UAV-derived data can achieve accuracy commensurate with LiDAR-derived data. The study area is characterized by steep slopes with frequent landslides

in the rainy seasons and susceptibility to landslides. LiDAR and UAV surveys produced high-resolution DTMs (spatial resolution 0.5 meters). Topographic and hydrological (generated from each DTM) and geological conditioning factors were used for the prediction models. The results showed that both models achieved accuracy above 0.70 and AUC greater than 0.80 and that, as expected, the model generated from LiDAR data achieved greater accuracy. However, despite the limitation of UAV-derived data regarding areas of dense vegetation for the generation of DTM, the landslide susceptibility map from UAV data was considered consistent with the landslide susceptibility map obtained by LiDAR data. Therefore, this study allows concluded that the UAV technology can be an alternative adequate to generate landslide susceptibility maps in areas without a DTM with high-resolution data.

Acknowledgments. All authors thank the FINEP (REMADEN/REDEGEO Project - Invitation Letter MCTI/FINEP/FNDCT 01/2016) for the financial support and to FUNCATE for the administrative support of the resources; to the Civil Defense of Santos for logistical help and field acquisitions support; Eduardo Kimoto Hosokawa, currently Chief of the Climate Change Section of the SEMAM Environment Department and, previously, engineer of SEDURB Urban Development Department for providing the LiDAR data; and the LAFOTO (Laboratory of Photogrammetry Research, Department of Geodesy, Institute of Geoscience, UFRGS) through the Coordinator Mário Luiz Lopes Reiss for equipment to support the data processing.

Declarations

- Funding FINEP (REMADEN/REDEGEO Project - Invitation Letter MCTI/FINEP/FNDCT 01/2016)
- Conflict of interest The authors declare no competing interests
- Ethics approval Not applicable.
- Availability of data and materials Not applicable.
- Code availability Not applicable.

Appendix A Relative frequency

References

Abdulwahid WM, Pradhan B (2017) Landslide vulnerability and risk assessment for multi-hazard scenarios using airborne laser scanning data (LiDAR). Landslides 14(3):1057–1076. https://doi.org/10.1007/s10346-016-0744-0

Abraham MT, Satyam N, Jain P, et al (2021) Effect of spatial resolution and data splitting on landslide susceptibility mapping using different machine learning algorithms. Geomatics, Natural Hazards and Risk 12(1):3381–3408. https://doi.org/10.1080/19475705.2021.2011791

Table A1 Relative frequency values for LiDAR and UAV data.

		uav				lidar			
Conditioning factor	Classes	Pixels number in domain (in %)	Number of landslide (in %)	FR value	NFR value	Pixels number in domain (in %)	Number of landslide (in %)	FR value	NFR value
Elevation (m)	0 - 20	1.17	0.47	0.40	0.05	2.49	1.40	0.56	0.07
	20.01 - 40	7.72	12.09	1.57	0.19	9.13	12.56	1.38	0.16
	40.01 - 60	9.93	17.21	1.73	0.21	9.83	17.67	1.80	0.21
	60.01 - 80	17.95	15.35	0.86	0.10	18.19	13.49	0.74	0.09
	80.01 - 100	20.69	17.67	0.85	0.10	20.74	18.60	0.90	0.11
	100.01 - 120	15.20	18.14	1.19	0.14	14.92	18.60	1.25	0.15
	120.01 - 140	12.73	11.63	0.91	0.11	11.97	11.63	0.97	0.11
	140.01 - 160	9.66	6.51	0.67	0.08	8.43	5.12	0.61	0.07
	160.01 - 180	3.89	0.93	0.24	0.03	3.37	0.93	0.28	0.03
	180.01 - 200	1.06	0.00	0.00	0.00	0.94	0.00	0.00	0.00
Geotechnical	Areas explored or under exploration for material removal	0.10	0.00	0.00	0.00	0.10	0.00	0.00	0.00
	Talus bodies with strong evidence of movement	0.63	0.00	0.00	0.00	0.63	0.00	0.00	0.00
	Bands situated below unstable zones	3.72	3.26	0.88	0.09	3.72	3.26	0.88	0.09
	Lake	0.19	0.00	0.00	0.00	0.19	0.00	0.00	0.00
	Alluvial plain embedded in the top of the hills	8.65	0.00	0.00	0.00	8.65	0.00	0.00	0.00
	Segment of predominantly straight slopes with slope (>40°)	41.29	56.74	1.37	0.14	41.29	56.74	1.37	0.14
	Segment of predominantly straight slopes with slope (30°- 40°)	2.81	8.84	3.14	0.32	2.81	8.84	3.14	0.32
	Segment of predominantly straight slopes with slope (20°- 30°)	7.33	6.98	0.95	0.10	7.33	6.98	0.95	0.10
	Segment of straight or convex slopes. slightly inclined (<20°)	5.34	2.79	0.52	0.05	5.34	2.79	0.52	0.05
	Straight or convex hill tops and slopes. slightly inclined	27.63	15.81	0.57	0.06	27.63	15.81	0.57	0.06
	Deposition zone (talus) apparently stable and large blocks may occur on the surface	2.32	5.58	2.41	0.24	2.32	5.58	2.41	0.24
Lithology	Gravitational Sediments	1.59	0.93	0.58	0.19	1.59	0.93	0.58	0.19
	Migmatites	27.31	42.33	1.55	0.50	27.31	42.33	1.55	0.50
	Alveolar Sediments	10.61	0.00	0.00	0.00	10.61	0.00	0.00	0.00
	Granite with Oriented Megacrystals	60.41	56.74	0.94	0.31	60.41	56.74	0.94	0.31
	Santos granite	0.08	0.00	0.00	0.00	0.08	0.00	0.00	0.00
Slope %	0 - 3	2.00	0.47	0.23	0.05	5.39	0.00	0.00	0.00
	3.01 - 8	6.42	0.00	0.00	0.00	8.71	1.40	0.16	0.03
	8.01 - 20	14.79	4.19	0.28	0.07	12.45	5.12	0.41	0.09
	20.01 - 45	28.02	19.53	0.70	0.16	22.49	10.23	0.45	0.10
	45.01 - 75	24.66	29.30	1.19	0.27	28.69	18.14	0.63	0.14
	75.01 ->	24.10	46.51	1.93	0.45	22.28	65.12	2.92	0.64
Aspect	Flat (-1)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	North (0-22.5 ; 337.5-360)	14.47	10.23	0.71	0.09	15.68	6.98	0.44	0.05
	Northeast (22.5-67.5)	12.16	6.51	0.54	0.07	12.04	13.02	1.08	0.13
	East (67.5-112.5)	10.47	14.42	1.38	0.17	9.19	10.70	1.16	0.14
	Southeast (112.5-157.5)	12.23	20.93	1.71	0.21	12.73	11.16	0.88	0.11
	South (157.5-202.5)	13.60	14.42	1.06	0.13	13.70	26.51	1.93	0.24
	Southwest (202.5-247.5)	10.31	12.56	1.22	0.15	9.50	8.84	0.93	0.11
	West (247.5-292.5)	12.01	12.09	1.01	0.12	11.14	9.30	0.84	0.10
	Northwest (292.5-337.5)	14.74	8.84	0.60	0.07	16.01	13.49	0.84	0.10
Plan curvature	Concave	27.03	32.56	1.20	0.38	34.07	31.63	0.93	0.31
	Rectilinear	45.10	32.56	0.72	0.23	30.46	29.77	0.98	0.33
	Convex	27.87	34.88	1.25	0.39	35.47	38.60	1.09	0.36
Profile curvature	Convex	19.41	27.44	1.41	0.43	20.98	25.58	1.22	0.39
	Rectilinear	60.50	51.16	0.85	0.25	57.32	52.09	0.91	0.29
	Concave	20.09	21.40	1.06	0.32	21.70	22.33	1.03	0.33
TWI	<3	57.28	70.70	1.23	0.35	45.89	61.86	1.35	0.34
	3.01 - 4	18.46	12.09	0.65	0.18	23.64	15.81	0.67	0.17
	4.01 - 6	17.78	11.16	0.63	0.18	22.19	16.74	0.75	0.19
	6.01 - 10	5.73	6.05	1.06	0.30	7.34	5.12	0.70	0.18
	10.01 >	0.75	0.00	0.00	0.00	0.93	0.47	0.50	0.13
SPI	0 - 10	77.79	70.70	0.91	0.10	67.54	50.70	0.75	0.10
	10.01 - 20	9.55	7.91	0.83	0.09	11.81	21.86	1.85	0.24
	20.01 - 30	3.43	5.58	1.63	0.18	5.48	6.98	1.27	0.17
	30.01 - 40	1.76	2.79	1.59	0.18	3.14	3.72	1.18	0.15
	40.01 - 50	1.06	2.33	2.19	0.25	2.03	2.33	1.15	0.15
	50.01 >	6.42	10.70	1.67	0.19	10.00	14.42	1.44	0.19
STI	0 - 10	71.37	50.70	0.71	0.12	61.38	31.16	0.51	0.09
	10 - 20	20.56	35.35	1.72	0.29	25.06	44.65	1.78	0.30
	20 - 30	4.22	8.84	2.09	0.36	7.97	13.95	1.75	0.30
	30 >	3.86	5.12	1.33	0.23	5.59	10.23	1.83	0.31
TRI	0 - 0.45214161	86.99	65.12	0.75	0.09	91.87	70.23	0.76	0.07
	0.45214161 - 0.481239951	1.65	5.12	3.11	0.37	1.65	5.58	3.38	0.33
	0.481239951 - 0.933381561	9.67	26.98	2.79	0.34	5.87	22.79	3.88	0.38
	0.933381561 >	1.70	2.79	1.64	0.20	0.61	1.40	2.30	0.22

Abu El-Magd SA, Ali SA, Pham QB (2021) Spatial modeling and susceptibility zonation of landslides using random forest, naïve bayes and K-nearest neighbor in a complicated terrain. Earth Science Informatics 14(3):1227–1243. https://doi.org/10.1007/s12145-021-00653-y

Aditian A, Kubota T, Shinohara Y (2018) Comparison of GIS-based landslide susceptibility models using frequency ratio, logistic regression, and artificial neural network in a tertiary region of Ambon, Indonesia. Geomorphology 318:101–111. https://doi.org/10.1016/j.geomorph.2018.06.006

Aghdam IN, Varzandeh MHM, Pradhan B (2016) Landslide susceptibility mapping using an ensemble statistical index (Wi) and adaptive neuro-fuzzy inference system (ANFIS) model at Alborz mountains (Iran). Environmental Earth Sciences 75(7):1–20. https://doi.org/10.1007/s12665-015-5233-6

Althuwaynee OF, Pradhan B, Park HJ, et al (2014) A novel ensemble decision tree-based CHi-squared automatic interaction detection (CHAID) and multivariate logistic regression models in landslide susceptibility mapping. Landslides 11(6):1063–1078. https://doi.org/10.1007/s10346-014-0466-0

Assis Dias MCd, Saito SM, dos Santos Alvalá RC, 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. https://doi.org/10.1016/j.ijdrr.2018.06.002

Assis Dias MCd, Saito SM, dos Santos Alvalá RC, et al (2020) Vulnerability index related to populations at-risk for landslides in the Brazilian Early Warning System (BEWS). International journal of disaster risk reduction 49:101,742. https://doi.org/10.1016/j.ijdrr.2020.101742

Azarafza M, Azarafza M, Akgün H, et al (2021) Deep learning-based landslide susceptibility mapping. Scientific Reports 11(1):1–16. https://doi.org/10.1038/s41598-021-03585-1

Bernardo E, Palamara R, Boima R (2021) UAV and soft computing methodology for monitoring landslide areas (susceptibility to landslides and early warning). WSEAS Transactions on Environment and Development 17:490–501. https://doi.org/10.37394/232015.2021.17.47

Borrelli L, Conforti M, Mercuri M (2019) Lidar and uav system data to analyse recent morphological changes of a small drainage basin. ISPRS International Journal of Geo-Information 8(12):536. https://doi.org/10.3390/ijgi8120536

Breiman L (2001) Random forests. Machine learning 45(1):5–32. https://doi.org/10.1023/a:1010933404324

Brock J, Schratz P, Petschko H, et al (2020) The performance of landslide susceptibility models critically depends on the quality of digital elevation models. Geomatics, Natural Hazards and Risk 11(1):1075–1092. https://doi.org/10.1080/19475705.2020.1776403

Campana S (2017) Drones in archaeology. state-of-the-art and future perspectives. Archaeological Prospection 24(4):275–296. https://doi.org/10.1002/arp.1569

Cao Y, Wei X, Fan W, et al (2021) Landslide susceptibility assessment using the Weight of Evidence method: A case study in Xunyang area, China. PLoS

one 16(1):e0245,668. https://doi.org/10.1371/journal.pone.0245668

CEPAGRI (2020) Centro de Pesquisas Meteorológicas e Climáticas Aplicadas à Agricultura - CEPAGRI/UNICAMP. https://www.cpa.unicamp.br/, (Accessed on 05/31/2022)

Chen W, Zhang S (2021) GIS-based comparative study of Bayes network, Hoeffding tree and logistic model tree for landslide susceptibility modeling. Catena 203:105,344. https://doi.org/10.1016/j.catena.2021.105344

Chen W, Xie X, Wang J, et al (2017) A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility. Catena 151:147–160. https://doi.org/10.1016/j.catena.2016.11.032

Chen W, Peng J, Hong H, et al (2018) Landslide susceptibility modelling using GIS-based machine learning techniques for Chongren County, Jiangxi Province, China. Science of the Total Environment 626:1121–1135. https://doi.org/10.1016/j.scitotenv.2018.01.124

Chen W, Sun Z, Han J (2019) Landslide susceptibility modeling using integrated ensemble weights of evidence with logistic regression and random forest models. Applied Sciences 9(1):171. https://doi.org/10.3390/app9010171

Chen Z, Ye F, Fu W, et al (2020) The influence of DEM spatial resolution on landslide susceptibility mapping in the Baxie River basin, NW China. Natural Hazards 101(3):853–877. https://doi.org/10.1007/s11069-020-03899-9

Colomina I, Molina P (2014) Unmanned aerial systems for photogrammetry and remote sensing: A review. ISPRS Journal of Photogrammetry and Remote Sensing 92:79–97. https://doi.org/10.1016/j.isprsjprs.2014.02.013

Dang VH, Hoang ND, Nguyen LMD, et al (2020) A novel GIS-based random forest machine algorithm for the spatial prediction of shallow landslide susceptibility. Forests 11(1):118. https://doi.org/10.3390/f11010118

Debortoli NS, Camarinha PIM, Marengo JA, et al (2017) An index of Brazil's vulnerability to expected increases in natural flash flooding and landslide disasters in the context of climate change. Natural Hazards 86(2):557–582. https://doi.org/10.1007/s11069-016-2705-2

Dou J, Yunus AP, Bui DT, et al (2019) Assessment of advanced random forest and decision tree algorithms for modeling rainfall-induced landslide susceptibility in the Izu-Oshima Volcanic Island, Japan. Science of the Total Environment 662:332–346. https://doi.org/10.1016/j.scitotenv.2019.01.221

Du Gl, Zhang Ys, Iqbal J, et al (2017) Landslide susceptibility mapping using an integrated model of information value method and logistic regression in the Bailongjiang watershed, Gansu Province, China. Journal of Mountain Science 14(2):249–268. https://doi.org/10.1007/s11629-016-4126-9

Eker R, Aydın A (2021) Long-term retrospective investigation of a large, deep-seated, and slow-moving landslide using InSAR time series, historical aerial photographs, and UAV data: The case of Devrek landslide (NW Turkey). Catena 196:104,895. https://doi.org/10.1016/j.catena.2020.104895

Eker R, Aydın A, Hübl J (2018) Unmanned aerial vehicle (UAV)-based monitoring of a landslide: Gallenzerkogel landslide (Ybbs-Lower Austria) case study. Environmental Monitoring and Assessment 190(1):1–14. https://doi.org/10.1007/s10661-017-6402-8

Fernández-Hernandez J, González-Aguilera D, Rodríguez-Gonzálvez P, et al (2015) Image-based modelling from unmanned aerial vehicle (UAV) photogrammetry: an effective, low-cost tool for archaeological applications. Archaeometry 57(1):128–145. https://doi.org/10.1111/arcm.12078

Froude MJ, Petley DN (2018) Global fatal landslide occurrence from 2004 to 2016. Natural Hazards and Earth System Sciences 18(8):2161–2181. https://doi.org/10.5194/nhess-18-2161-2018

Gaidzik K, Ramírez-Herrera MT (2021) The importance of input data on landslide susceptibility mapping. Scientific Reports 11(1):1–14. https://doi.org/10.1038/s41598-021-98830-y

Gaidzik K, Ramírez-Herrera MT, Bunn M, et al (2017) Landslide manual and automated inventories, and susceptibility mapping using LiDAR in the forested mountains of Guerrero, Mexico. Geomatics, Natural Hazards and Risk 8(2):1054–1079. https://doi.org/10.1080/19475705.2017.1292560

Giordan D, Manconi A, Remondino F, et al (2017) Use of unmanned aerial vehicles in monitoring application and management of natural hazards. Geomatics, Natural Hazards and Risk 8(1):1–4. https://doi.org/10.1080/19475705.2017.1315619

Giordan D, Hayakawa Y, Nex F, et al (2018) The use of remotely piloted aircraft systems (RPASs) for natural hazards monitoring and management. Natural hazards and Earth System Sciences 18(4):1079–1096. https://doi.org/10.5194/nhess-18-1079-2018

Gobbi ES, et al (2017) Variação do manto intempérico nas encostas dos morros residuais em Santos-SP. PhD thesis, Universidade Estadual de Campinas, Instituto de Geociências, https://doi.org/https://doi.org/10.47749/T/UNICAMP.2017.1056631, (Accessed on 03/16/2022)

Gomez C, Purdie H (2016) UAV-based photogrammetry and geocomputing for hazards and disaster risk monitoring–a review. Geoenvironmental Disasters 3(1):1–11. https://doi.org/10.1186/s40677-016-0060-y

Haas F, Heckmann T, Hilger L, et al (2012) Quantification and modelling of debris flows in the Proglacial area of the Gepatschferner/Austria using ground-based LiDAR. In: Collins, Adrian L. ; Golosov, Valentin ; Horowitz, Arthur J. ; Lu, Xixi ; Stone, Mike ; Walling, Des E. ; Zhang, Xinbao (Hrsg.): Erosion and Sediment Yields in the Changing Environment : proceedings of an IAHS International Commission on Continental Erosion Symposium, held at the Institute of Mountain Hazards and Environment, CAS-Chengdu, China, 11-15 October 2012, IAHS publication, vol 356. Wallingford, p 293–302, URL https://edoc.ku.de/id/eprint/9362/

Hawker L, Bates P, Neal J, et al (2018) Perspectives on digital elevation model (DEM) simulation for flood modeling in the absence of a high-accuracy open access global DEM. Frontiers in Earth Science 6:233. https://doi.org/10.3389/feart.2018.00233

He Q, Shahabi H, Shirzadi A, et al (2019) Landslide spatial modelling using novel bivariate statistical based naïve bayes, RBF classifier, and RBF network machine learning algorithms. Science of the Total Environment 663:1–15. https://doi.org/10.1016/j.scitotenv.2019.01.329

Hong H, Naghibi SA, Pourghasemi HR, et al (2016) GIS-based landslide spatial modeling in Ganzhou City, China. Arabian Journal of Geosciences 9(2):1–26. https://doi.org/10.1007/s12517-015-2094-y

Hong H, Miao Y, Liu J, et al (2019) Exploring the effects of the design and quantity of absence data on the performance of random forest-based landslide susceptibility mapping. Catena 176:45–64. https://doi.org/10.1016/j.catena.2018.12.035

Hosseinalizadeh M, Kariminejad N, Chen W, et al (2019) Gully headcut susceptibility modeling using functional trees, naïve bayes tree, and random forest models. Geoderma 342:1–11. https://doi.org/10.1016/j.geoderma.2019.01.050

Huang F, Cao Z, Guo J, et al (2020) Comparisons of heuristic, general statistical and machine learning models for landslide susceptibility prediction and mapping. Catena 191:104,580. https://doi.org/10.1016/j.catena.2020.104580

Huang H, Savkin AV, Huang C (2021) Decentralized autonomous navigation of a UAV network for road traffic monitoring. IEEE Transactions on Aerospace and Electronic Systems 57(4):2558–2564. https://doi.org/10.1109/TAES.2021.3053115.

IBGE (2017) Instituto Brasileiro de Geografia e Estatística. Estimates of the resident population with reference date july 1st, 2017 (estimativas da população residente com data de referência 1º de julho de 2017). https://cidades.ibge.gov.br/brasil/sp/santos/panorama, (Accessed on 03/16/2022)

Iglhaut J, Cabo C, Puliti S, et al (2019) Structure from motion photogrammetry in forestry: A review. Current Forestry Reports 5(3):155–168. https://doi.org/10.1007/s40725-019-00094-3

Jena R, Pradhan B, Beydoun G, et al (2020) Integrated model for earthquake risk assessment using neural network and analytic hierarchy process: Aceh province, Indonesia. Geoscience Frontiers 11(2):613–634. https://doi.org/10.1016/j.gsf.2019.07.006

Lee JH, Sameen MI, Pradhan B, et al (2018) Modeling landslide susceptibility in data-scarce environments using optimized data mining and statistical methods. Geomorphology 303:284–298. https://doi.org/10.1016/j.geomorph.2017.12.007

Lin ML, Chen TW, Lin CW, et al (2013) Detecting large-scale landslides using LiDAR data and aerial photos in the Namasha-Liuoguey area, Taiwan. Remote Sensing 6(1):42–63. https://doi.org/10.3390/rs6010042

Lindsay JB (2016) Whitebox GAT: A case study in geomorphometric analysis. Computers & Geosciences 95:75–84. https://doi.org/10.1016/j.cageo.2016.07.003

Liu C, Shao X, Wu H, et al (2019) Giant landslide displacement analysis using a point cloud set conflict technique: a case in Xishancun landslide, Sichuan, China. International Journal of Remote Sensing 40(8):3247–3266. https://doi.org/10.1080/01431161.2018.1541331

Liu Z, Gilbert G, Cepeda JM, et al (2021) Modelling of shallow landslides with machine learning algorithms. Geoscience Frontiers 12(1):385–393. https://doi.org/10.1016/j.gsf.2020.04.014

Lv L, Chen T, Dou J, et al (2022) A hybrid ensemble-based deep-learning framework for landslide susceptibility mapping. International Journal of Applied Earth Observation and Geoinformation 108:102,713. https://doi.org/10.1016/j.jag.2022.102713

Mahalingam R, Olsen MJ, O'Banion MS (2016) Evaluation of landslide susceptibility mapping techniques using lidar-derived conditioning factors (oregon case study). Geomatics, Natural Hazards and Risk 7(6):1884–1907. https://doi.org/10.1080/19475705.2016.1172520

Miner A, Flentje P, Mazengarb C, et al (2010) Landslide Recognition using LiDAR derived Digital Elevation Models-Lessons learnt from selected Australian examples. In A. L. Williams, G. M. Pinches, C. Y. Chin, T. J. McMorran and C. I. Massey (Eds.), Geologically Active: Proceedings of the 11th IAEG Congress of the International Association of Engineering Geology and the Environment, Auckland, New Zealand, 2010, (pp. 1-9). Boca Raton, FL, USA: CRC Press. URL https://ro.uow.edu.au/cgi/viewcontent.cgi?article=1590&context=engpapers

Modenesi-Gauttieri MC, Hiruma ST (2004) A expansão urbana no planalto de Campos do Jordão: diagnóstico geomorfológico para fins de planejamento. Revista do Instituto Geológico 25(1-2):1–28. https://doi.org/10.5935/0100-929X.20040001

Moore ID, Grayson R, Ladson A (1991) Digital terrain modelling: a review of hydrological, geomorphological, and biological applications. Hydrological processes 5(1):3–30. https://doi.org/10.1002/hyp.3360050103

Mora OE, Liu Jk, Gabriela Lenzano M, et al (2015) Small landslide susceptibility and hazard assessment based on airborne LiDAR data. Photogrammetric Engineering & Remote Sensing 81(3):239–247. https://doi.org/10.14358/PERS.81.3.239

Moura CAd (2014) Análise da suscetibilidade do meio físico a movimentos gravitacionais de massa por meio de sistematização de zoneamento geoambiental. PhD thesis, Universidade Estadual Paulista, Instituto de Geociências e Ciências Exatas, URL http://hdl.handle.net/11449/127586, (Accessed on 09/31/2021)

Nebiker S, Lack N, Abächerli M, et al (2016) Light-weight multispectral UAV sensors and their capabilities for predicting grain yield and detecting plant diseases. International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences 41. https://doi.org/10.5194/isprsarchives-XLI-B1-963-2016

Niculiţă M, Mărgărint MC, Santangelo M (2016) Archaeological evidence for Holocene landslide activity in the Eastern Carpathian lowland. Quaternary International 415:175–189. https://doi.org/10.1016/j.quaint.2015.12.048

Niculiţă M (2016) Automatic landslide length and width estimation based on the geometric processing of the bounding box and the geomorphometric analysis of DEMs. Natural Hazards and Earth System Sciences 16(8):2021–2030. https://doi.org/10.5194/nhess-16-2021-2016, URL https://nhess.copernicus.org/articles/16/2021/2016/

Oh HJ, Lee S, Hong SM (2017) Landslide susceptibility assessment using frequency ratio technique with iterative random sampling. Journal of Sensors

2017. https://doi.org/10.1155/2017/3730913

Ozdemir A, Altural T (2013) A comparative study of frequency ratio, weights of evidence and logistic regression methods for landslide susceptibility mapping: Sultan Mountains, SW Turkey. Journal of Asian Earth Sciences 64:180–197. https://doi.org/10.1016/j.jseaes.2012.12.014

Park S, Kim J (2019) Landslide susceptibility mapping based on random forest and boosted regression tree models, and a comparison of their performance. Applied Sciences 9(5):942. https://doi.org/10.3390/app9050942

Park S, Choi C, Kim B, et al (2013) Landslide susceptibility mapping using frequency ratio, analytic hierarchy process, logistic regression, and artificial neural network methods at the Inje area, Korea. Environmental earth sciences 68(5):1443–1464. https://doi.org/10.1007/s12665-012-1842-5

Pawluszek K, Borkowski A (2017) Impact of DEM-derived factors and analytical hierarchy process on landslide susceptibility mapping in the region of Rożnów Lake, Poland. Natural Hazards 86(2):919–952. https://doi.org/10.1007/s11069-016-2725-y

Peng L, Niu R, Huang B, et al (2014) Landslide susceptibility mapping based on rough set theory and support vector machines: A case of the Three Gorges area, China. Geomorphology 204:287–301. https://doi.org/10.1016/j.geomorph.2013.08.013

Peppa MV (2018) Morphology-based landslide monitoring with an unmanned aerial vehicle. PhD thesis, Newcastle University, URL http://theses.ncl.ac.uk/jspui/handle/10443/4115, (Accessed on 11/16/2021)

Pham BT, Prakash I, Dou J, et al (2020) A novel hybrid approach of landslide susceptibility modelling using rotation forest ensemble and different base classifiers. Geocarto International 35(12):1267–1292. https://doi.org/10.1080/10106049.2018.1559885

Pichler E (1957) Aspectos geológicos dos escorregamentos de santos. Boletim da Sociedade Brasileira de Geologia 6(2):68–77. URL https://ppegeo.igc.usp.br/index.php/BSBG/article/view/12614

Pourghasemi HR, Jirandeh AG, Pradhan B, et al (2013) Landslide susceptibility mapping using support vector machine and GIS at the Golestan Province, Iran. Journal of Earth System Science 122(2):349–369. https://doi.org/10.1007/s12040-013-0282-2

Pradhan AMS, Kim YT (2014) Relative effect method of landslide susceptibility zonation in weathered granite soil: a case study in Deokjeok-ri Creek, South Korea. Natural Hazards 72(2):1189–1217. https://doi.org/10.1007/

s11069-014-1065-z

Pradhan B, Sameen MI (2017) Effects of the Spatial Resolution of Digital Elevation Models and Their Products on Landslide Susceptibility Mapping, Springer International Publishing, Cham, pp 133–150. https://doi.org/10.1007/978-3-319-55342-9_7

Raja NB, Çiçek I, Türkoğlu N, et al (2017) Landslide susceptibility mapping of the sera river basin using logistic regression model. Natural Hazards 85(3):1323–1346. https://doi.org/10.1007/s11069-016-2591-7

Ramírez-Herrera MT, Gaidzik K (2017) La pintada landslide—a complex double-staged extreme event, guerrero, mexico. Cogent Geoscience 3(1):1356,012. https://doi.org/10.1080/23312041.2017.1356012

Reichenbach P, Rossi M, Malamud BD, et al (2018) A review of statistically-based landslide susceptibility models. Earth-Science Reviews 180:60–91. https://doi.org/10.1016/j.earscirev.2018.03.001

Reiss M, Mendes T, Pereira F, et al (2022) Evaluation of an unmanned aerial vehicle (uav) for measuring and monitoring natural disaster risk areas. The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 43:1077–1083. https://doi.org/10.5194/isprs-archives-XLIII-B2-2022-1077-2022

Rezaei S, Shooshpasha I, Rezaei H (2018) Empirical correlation between geotechnical and geophysical parameters in a landslide zone (case study: Naigeschal landslide). Earth Sciences Research Journal 22(3):195–204. https://doi.org/10.15446/esrj.v22n3.69491

Riley SJ, DeGloria SD, Elliot R (1999) Index that quantifies topographic heterogeneity. Intermountain Journal of Sciences 5(1-4):23–27

Sahin EK (2020) Assessing the predictive capability of ensemble tree methods for landslide susceptibility mapping using XGBoost, gradient boosting machine, and random forest. SN Applied Sciences 2(7):1–17. https://doi.org/10.1007/s42452-020-3060-1

Sahin EK, Colkesen I, Acmali SS, et al (2020) Developing comprehensive geocomputation tools for landslide susceptibility mapping: LSM tool pack. Computers & Geosciences 144:104,592. https://doi.org/10.1016/j.cageo.2020.104592

Saleem N, Huq M, Twumasi NYD, et al (2019) Parameters derived from and/or used with digital elevation models (DEMs) for landslide susceptibility mapping and landslide risk assessment: a review. ISPRS International Journal of Geo-Information 8(12):545. https://doi.org/10.3390/ijgi8120545

Sangchini EK, Emami SN, Tahmasebipour N, et al (2016) Assessment and comparison of combined bivariate and ahp models with logistic regression for landslide susceptibility mapping in the Chaharmahal-e-Bakhtiari Province, Iran. Arabian Journal of Geosciences 9(3):1–15. https://doi.org/10.1007/s12517-015-2258-9

Sao Paulo (1988) Instabilidade da serra do mar no estado de são paulo – situações de risco. Tech. Rep. 4v, Secretaria de Estado da Ciência e Tecnologia (SCT): Secretaria de Estado do Meio Ambiente (SMA) IPT – Instituto de Pesquisas Tecnológicas, IG - Instituto Geológico, IF – instituto Florestal, IBt – Instituto de Botânica., São Paulo

Saro L, Woo JS, Kwan-Young O, et al (2016) The spatial prediction of landslide susceptibility applying artificial neural network and logistic regression models: A case study of Inje, Korea. Open Geosciences 8(1):117–132. https://doi.org/10.1515/geo-2016-0010

Saroglou C, Asteriou P, Zekkos D, et al (2018) UAV-based mapping, back analysis and trajectory modeling of a coseismic rockfall in Lefkada island, Greece. Natural Hazards and Earth System Sciences 18(1):321–333. https://doi.org/10.5194/nhess-18-321-2018

Sestras P, Bilașco S, Roșca S, et al (2021) Geodetic and UAV monitoring in the sustainable management of shallow landslides and erosion of a susceptible urban environment. Remote Sensing 13(3):385. https://doi.org/10.3390/rs13030385

Shahri AA, Spross J, Johansson F, et al (2019) Landslide susceptibility hazard map in southwest sweden using artificial neural network. Catena 183:104,225. https://doi.org/10.1016/j.catena.2019.104225

Shirzadi A, Soliamani K, Habibnejhad M, et al (2018) Novel GIS based machine learning algorithms for shallow landslide susceptibility mapping. Sensors 18(11):3777. https://doi.org/10.3390/s18113777

Smith MW (2014) Roughness in the earth sciences. Earth-Science Reviews 136:202–225. https://doi.org/10.1016/j.earscirev.2014.05.016

Sun D, Wen H, Wang D, et al (2020) A random forest model of landslide susceptibility mapping based on hyperparameter optimization using bayes algorithm. Geomorphology 362:107,201. https://doi.org/10.1016/j.geomorph.2020.107201

Taalab K, Cheng T, Zhang Y (2018) Mapping landslide susceptibility and types using random forest. Big Earth Data 2(2):159–178. https://doi.org/10.1080/20964471.2018.1472392

Tanyu BF, Abbaspour A, Alimohammadlou Y, et al (2021) Landslide susceptibility analyses using random forest, C4.5, and C5.0 with balanced and unbalanced datasets. Catena 203:105,355. https://doi.org/10.1016/j.catena.2021.105355

Tien Bui D, Tuan TA, Hoang ND, et al (2017) Spatial prediction of rainfall-induced landslides for the Lao Cai area (Vietnam) using a hybrid intelligent approach of least squares support vector machines inference model and artificial bee colony optimization. Landslides 14(2):447–458. https://doi.org/10.1007/s10346-016-0711-9

Trigila A, Iadanza C, Esposito C, et al (2015) Comparison of Logistic Regression and Random Forests techniques for shallow landslide susceptibility assessment in giampilieri (NE Sicily, Italy). Geomorphology 249:119–136. https://doi.org/10.1016/j.geomorph.2015.06.001

Tsangaratos P, Ilia I (2016) Comparison of a logistic regression and naïve bayes classifier in landslide susceptibility assessments: The influence of models complexity and training dataset size. Catena 145:164–179. https://doi.org/10.1016/j.catena.2016.06.004

Wang LJ, Guo M, Sawada K, et al (2015) Landslide susceptibility mapping in Mizunami City, Japan: A comparison between logistic regression, bivariate statistical analysis and multivariate adaptive regression spline models. Catena 135:271–282. https://doi.org/10.1016/j.catena.2015.08.007

Wang Q, Li W (2017) A GIS-based comparative evaluation of analytical hierarchy process and frequency ratio models for landslide susceptibility mapping. Physical Geography 38(4):318–337. https://doi.org/10.1080/02723646.2017.1294522

Wang Y, Fang Z, Hong H (2019) Comparison of convolutional neural networks for landslide susceptibility mapping in Yanshan County, China. Science of the Total Environment 666:975–993. https://doi.org/10.1016/j.scitotenv.2019.02.263

Wang Y, Fang Z, Wang M, et al (2020) Comparative study of landslide susceptibility mapping with different recurrent neural networks. Computers & Geosciences 138:104,445. https://doi.org/10.1016/j.cageo.2020.104445

Xiao L, Zhang Y, Peng G (2018) Landslide susceptibility assessment using integrated deep learning algorithm along the China-Nepal highway. Sensors 18(12):4436. https://doi.org/10.3390/s18124436

Xiao T, Yin K, Yao T, et al (2019) Spatial prediction of landslide susceptibility using GIS-based statistical and machine learning models in Wanzhou County, Three Gorges Reservoir, China. Acta Geochimica 38(5):654–669.

https://doi.org/10.1007/s11631-019-00341-1

Yan F, Zhang Q, Ye S, et al (2019) A novel hybrid approach for landslide susceptibility mapping integrating analytical hierarchy process and normalized frequency ratio methods with the cloud model. Geomorphology 327:170–187. https://doi.org/10.1016/j.geomorph.2018.10.024

Youssef AM, Pourghasemi HR (2021) Landslide susceptibility mapping using machine learning algorithms and comparison of their performance at Abha basin, Asir Region, Saudi Arabia. Geoscience Frontiers 12(2):639–655. https://doi.org/10.1016/j.gsf.2020.05.010

Youssef AM, Pourghasemi HR, Pourtaghi ZS, et al (2016) Landslide susceptibility mapping using random forest, boosted regression tree, classification and regression tree, and general linear models and comparison of their performance at Wadi Tayyah Basin, Asir Region, Saudi Arabia. Landslides 13(5):839–856. https://doi.org/10.1007/s10346-015-0614-1

Zhang Y, Ge T, Tian W, et al (2019) Debris flow susceptibility mapping using machine-learning techniques in Shigatse area, China. Remote Sensing 11(23):2801. https://doi.org/10.3390/rs11232801

Zhao S, Zhao Z (2021) A comparative study of landslide susceptibility mapping using SVM and PSO-SVM models based on grid and slope units. Mathematical Problems in Engineering 2021. https://doi.org/10.1155/2021/8854606

Zhou X, Wen H, Zhang Y, et al (2021) Landslide susceptibility mapping using hybrid random forest with GeoDetector and RFE for factor optimization. Geoscience Frontiers 12(5):101,211. https://doi.org/10.1016/j.gsf.2021.101211

3 CONSIDERAÇÕES FINAIS

O mapeamento da suscetibilidade a deslizamentos de terra é um dos métodos mais importantes usados como uma medida não estrutural fundamental para prevenir vítimas humanas e perdas socioeconômicas em eventos de desastres. Portanto, o modelo de previsão de deslizamento de terra é uma ferramenta útil para agências governamentais desenvolverem estratégias para prevenir e mitigar riscos de deslizamento. Os modelos baseados em métodos de aprendizado de máquina, como o RF, podem ser utilizados por tomadores de decisão, planejadores, pesquisadores acadêmicos, engenheiros e agências governamentais para mitigar a perda de vidas humanas e propriedades

Nesse sentido este trabalho tem como objetivo a comparação dos dois modelos de mapeamento de suscetibilidade a deslizamentos usando algoritmo de RF gerado a partir da tecnologia LiDAR e VANT para verificar se os dados derivados de VANT podem atingir precisão compatível com dados derivados de LiDAR. O resultado do desempenho do modelo de predição foi avaliado considerando as métricas estatísticas. O modelo usando os dados LiDAR atingiu 0,800, 0,862, 0,862 e 0,830, respectivamente, para Precisão, Acurácia, Recall e F1-Score. Para o modelo usando os dados de VANT, os valores atingidos foram 0,797, 0,712, 0,760 e 0,752, respectivamente, para Precisão, Acurácia, Recall e F1-Score.

A área de estudo é caracterizada por encostas íngremes com deslizamentos frequentes nas estações chuvosas e suscetibilidade a deslizamentos. Os levantamentos LiDAR e VANT produziram MDTs de alta resolução (resolução espacial de 0,5 metros). Fatores de condicionantes topográficos e hidrológicos (gerados a partir de cada MDT) e geológicos foram usados para os modelos de previsão. Os resultados mostraram que ambos os modelos alcançaram acurácia acima de 0,70 e AUC maior que 0,80 e que, como esperado, o modelo gerado a partir dos dados LiDAR obteve maior acurácia.

Os mapas de suscetibilidade a deslizamentos obtidos foram reclassificados em cinco categorias, definindo as classes por intervalos iguais para comparar os modelos de suscetibilidade produzidos por diferentes tecnologias, utilizando assim o mesmo padrão para identificar níveis de suscetibilidade a deslizamentos nos dois mapas. Os intervalos usados foram: muito baixo (0,005 - 0,2), baixo (0,2 - 0,4), moderado (0,4 - 0,6), alto (0,6 - 0,8) e muito alto (0,8 - 1).

Os resultados obtidos com os dados de VANTs foram considerados consistentes com aqueles obtidos por dados de LiDAR, apesar das limitações dos dados de VANTs para áreas densamente vegetadas. Portanto, este estudo nos leva a concluir que a tecnologia VANT pode ser uma alternativa adequada para gerar mapas de suscetibilidade a deslizamentos em áreas urbanas em nível de detalhe com dados de alta resolução em áreas sem dados de alta resolução disponíveis.

Como perspectivas futuras, espera-se contribuir para implementações de modelos mais robustos para uma plataforma de alertas para deslizamentos, incluindo a pluviosidade em tempo real com dados de estações meteorológicas, além da extrapolação dos dados para a região.

REFERÊNCIAS

ABDULWAHID, W. M.; PRADHAN, B. Landslide vulnerability and risk assessment for multi-hazard scenarios using airborne laser scanning data (LiDAR). **Landslides**, v. 14, n. 3, p. 1057–1076, 2017.

ARABAMERI, A. et al. Landslide susceptibility evaluation and management using different machine learning methods in the Gallicash River Watershed, Iran. **Remote Sensing**, v. 12, n. 3, p. 475, 2020.

BERNARDO, E.; PALAMARA, R.; BOIMA, R. UAV and soft computing methodology for monitoring landslide areas (susceptibility to landslides and early warning). **WSEAS Transactions on Environment and Development**, v. 17, p. 490–501, 2021.

CANIL, K. Áreas de risco aos processos de escorregamentos no município de Santos, SP: Análise e indicadores de vulnerabilidade. In: ENCONTRO NACIONAL DOS GEÓGRAFOS, XVI, Porto Alegre. **Anais [...]**. Porto Alegre: AGB, 2010.

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

CEPED, U. **Atlas brasileiro de desastres naturais 1991 a 2012**: volume brasil. Florianópolis: Universidade Federal de Santa Catarina, 2013.

CHEN, W. et al. GIS-based landslide susceptibility evaluation using a novel hybrid integration approach of bivariate statistical based random forest method. **Catena**, v. 164, p. 135–149, 2018.

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

COLOMINA, I.; MOLINA, P. Unmanned aerial systems for photogrammetry and remote sensing: a review. **ISPRS Journal of photogrammetry and remote sensing**, v. 92, p. 79–97, 2014.

DOU, J. et al. Assessment of advanced random forest and decision tree algorithms for modeling rainfall-induced landslide susceptibility in the Izu-Oshima Volcanic Island, Japan. **Science of the Total Environment**, v. 662, p. 332–346, 2019.

EKER, R.; AYDIN, A. Long-term retrospective investigation of a large, deep-seated, and slow-moving landslide using InSAR time series, historical aerial photographs, and UAV data: The case of Devrek landslide (NW Turkey). **Catena**, v. 196, p. 104895, 2021.

EKER, R.; AYDIN, A.; HÜBL, J. Unmanned aerial vehicle (UAV)-based monitoring of a landslide: Gallenzerkogel landslide (Ybbs-Lower Austria) case study. **Environmental Monitoring and Assessment**, v. 190, n. 1, p. 1–14, 2018.

FANG, Z. et al. Landslide susceptibility mapping using rotation forest ensemble technique with different decision trees in the Three Gorges Reservoir area, China. **Remote Sensing**, v. 13, n. 2, p. 238, 2021.

FERREIRA, K.; ABIKO, A. Urban resilience and landslide risk management: The case of santos (brazil). In: BRUNETTA, G.; CALDARICE, O.; TOLLIN, N.; ROSAS-CASALS, M.; MORATÓ, J. **Urban resilience for risk and adaptation governance**: theory and practice. Cham: Springer International Publishing, 2019. p. 207–228.

FERREIRA, K. A. **Resiliência urbana e a gestão de riscos de escorregamentos**: uma avaliação da defesa civil do município de santos-sp. Tese (Doutorado) — Universidade de São Paulo, 2016.

FROUDE, M. J.; PETLEY, D. N. Global fatal landslide occurrence from 2004 to 2016. **Natural Hazards and Earth System Sciences**, v. 18, n. 8, p. 2161–2181, 2018.

GAIDZIK, K. et al. Landslide manual and automated inventories, and susceptibility mapping using LiDAR in the forested mountains of Guerrero, Mexico. **Geomatics, Natural Hazards and Risk**, v. 8, n. 2, p. 1054–1079, 2017.

GANTIMUROVA, S.; PARSHIN, A.; EROFEEV, V. GIS-based landslide susceptibility mapping of the Circum-Baikal Railway in Russia using uav data. **Remote Sensing**, v. 13, n. 18, p. 3629, 2021.

GIORDAN, D. et al. The use of remotely piloted aircraft systems (RPASs) for natural hazards monitoring and management. **Natural Hazards and Earth System Sciences**, v. 18, n. 4, p. 1079–1096, 2018.

GIORDAN, D. et al. Use of unmanned aerial vehicles in monitoring application and management of natural hazards. **Geomatics, Natural Hazards and Risk**, v. 8, n. 1, p. 1–4, 2017.

GOMEZ, C.; PURDIE, H. UAV-based photogrammetry and geocomputing for hazards and disaster risk monitoring–a review. **Geoenvironmental Disasters**, v. 3, n. 1, p. 1–11, 2016.

GORSEVSKI, P. V. An evolutionary approach for spatial prediction of landslide susceptibility using LiDAR and symbolic classification with genetic programming. **Natural Hazards**, v. 108, n. 2, p. 2283–2307, 2021.

HAWKER, L. et al. Perspectives on digital elevation model (DEM) simulation for flood modeling in the absence of a high-accuracy open access global DEM. **Frontiers in Earth Science**, v. 6, p. 233, 2018.

HUANG, F. et al. Comparisons of heuristic, general statistical and machine learning models for landslide susceptibility prediction and mapping. **Catena**, v. 191, p. 104580, 2020.

HUANG, F. et al. Landslide susceptibility prediction considering regional soil erosion based on machine-learning models. **ISPRS International Journal of Geo-Information**, v. 9, n. 6, p. 377, 2020.

HUANG, Y.; ZHAO, L. Review on landslide susceptibility mapping using support vector machines. **Catena**, v. 165, p. 520–529, 2018.

JULIEV, M. et al. Comparative analysis of statistical methods for landslide susceptibility mapping in the Bostanlik District, Uzbekistan. **Science of the Total Environment**, v. 653, p. 801–814, 2019.

KAO, L.-S.; CHIU, Y.-H.; TSAI, C.-Y. An evaluation study of urban development strategy based on of extreme climate conditions. **Sustainability**, v. 9, n. 2, p. 284, 2017.

LEE, S.; MIN, K. Statistical analysis of landslide susceptibility at Yongin, Korea. **Environmental Geology**, v. 40, n. 9, p. 1095–1113, 2001.

LIMA, P.; STEGER, S.; GLADE, T. Counteracting flawed landslide data in statistically based landslide susceptibility modelling for very large areas: a national-scale assessment for Austria. **Landslides**, v. 18, n. 11, p. 3531–3546, 2021.

LISTO, F. d. L. R.; GOMES, M. C. V.; FERREIRA, F. S. Evaluation of shallow landslide susceptibility and factor of safety variation using the TRIGRS model, Serra do Mar Mountain Range, Brazil. **Journal of South American Earth Sciences**, v. 107, p. 103011, 2021.

LISTO, F. d. L. R.; VIEIRA, B. C. Mapping of risk and susceptibility of shallow-landslide in the city of São Paulo, Brazil. **Geomorphology**, v. 169, p. 30–44, 2012.

LIU, Z. et al. Modelling of shallow landslides with machine learning algorithms. **Geoscience Frontiers**, v. 12, n. 1, p. 385–393, 2021.

LUO, X. et al. Mine landslide susceptibility assessment using ivm, ann and SVM models considering the contribution of affecting factors. **PLoS One**, v. 14, n. 4, p. e0215134, 2019.

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

NEUHÄUSER, B.; DAMM, B.; TERHORST, B. Gis-based assessment of landslide susceptibility on the base of the weights-of-evidence model. **Landslides**, v. 9, n. 4, p. 511–528, 2012.

PARK, S.; KIM, J. Landslide susceptibility mapping based on random forest and boosted regression tree models, and a comparison of their performance. **Applied Sciences**, v. 9, n. 5, p. 942, 2019.

PEPPA, M. V. **Morphology-based landslide monitoring with an unmanned aerial vehicle**. Tese (Doutorado) — Newcstle University, 2018.

PHAM, B. T. et al. Landslide susceptibility assessment by novel hybrid machine learning algorithms. **Sustainability**, v. 11, n. 16, p. 4386, 2019.

PICHLER, E. Aspectos geológicos dos escorregamentos de Santos. **Boletim da Sociedade Brasileira de Geologia**, v. 6, n. 2, p. 68–77, 1957.

PIRASTEH, S.; LI, J. Probabilistic frequency ratio (PFR) model for quality improvement of landslide susceptibility mapping from LiDAR-derived DEMs. **Geoenvironmental Disasters**, v. 4, n. 1, p. 1–17, 2017.

RABBY, Y. W.; ISHTIAQUE, A.; RAHMAN, M. S. Evaluating the effects of digital elevation models in landslide susceptibility mapping in Rangamati district, Bangladesh. **Remote Sensing**, v. 12, n. 17, p. 2718, 2020.

RODRIGUEZ-GALIANO, V. F. et al. An assessment of the effectiveness of a random forest classifier for land-cover classification. **ISPRS journal of photogrammetry and remote sensing**, v. 67, p. 93–104, 2012.

SALEEM, N. et al. Parameters derived from and/or used with digital elevation models (DEMs) for landslide susceptibility mapping and landslide risk assessment: a review. **ISPRS International Journal of Geo-Information**, v. 8, n. 12, p. 545, 2019.

SANTOS, A. d. **Enchentes e deslizamentos**: causas e soluções: áreas de risco no brasil. São Paulo: Pini, 2012.

SESTRAS, P. et al. Geodetic and UAV monitoring in the sustainable management of shallow landslides and erosion of a susceptible urban environment. **Remote Sensing**, v. 13, n. 3, p. 385, 2021.

SHIRVANI, Z. A holistic analysis for landslide susceptibility mapping applying geographic object-based random forest: a comparison between protected and non-protected forests. **Remote Sensing**, v. 12, n. 3, p. 434, 2020.

TAALAB, K.; CHENG, T.; ZHANG, Y. Mapping landslide susceptibility and types using random forest. **Big Earth Data**, v. 2, n. 2, p. 159–178, 2018.

TEMPA, K. et al. UAV technique to localize landslide susceptibility and mitigation proposal: a case of Rinchending Goenpa landslide in Bhutan. **Natural Hazards Research**, v. 1, n. 4, p. 171–186, 2021.

WANG, J. et al. Refined micro-scale geological disaster susceptibility evaluation based on UAV tilt photography data and weighted certainty factor method in Mountainous Area. **Ecotoxicology and Environmental Safety**, v. 189, p. 110005, 2020.

WANG, Y.; DUAN, H.; HONG, H. A comparative study of composite kernels for landslide susceptibility mapping: a case study in Yongxin County, China. **Catena**, v. 183, p. 104217, 2019.

WU, Y. et al. Landslide susceptibility assessment using frequency ratio, statistical index and certainty factor models for the Gangu County, China. **Arabian Journal of Geosciences**, v. 9, n. 2, p. 1–16, 2016.

XIE, P. et al. Application and comparison of logistic regression model and neural network model in earthquake-induced landslides susceptibility mapping at mountainous region, China. **Geomatics, Natural Hazards and Risk**, v. 9, n. 1, p. 501–523, 2018.

YILMAZ, I. Landslide susceptibility mapping using frequency ratio, logistic regression, artificial neural networks and their comparison: a case study from Kat landslides (Tokat—Turkey). **Computers & Geosciences**, v. 35, n. 6, p. 1125–1138, 2009.

YOUSSEF, A. M.; POURGHASEMI, H. R. Landslide susceptibility mapping using machine learning algorithms and comparison of their performance at Abha Basin, Asir Region, Saudi Arabia. **Geoscience Frontiers**, v. 12, n. 2, p. 639–655, 2021.

ZÊZERE, J. et al. Mapping landslide susceptibility using data-driven methods. **Science of the Total Environment**, v. 589, p. 250–267, 2017.

ZHAO, C. et al. A comparative study of statistical index and certainty factor models in landslide susceptibility mapping: a case study for the Shangzhou District, Shaanxi Province, China. **Arabian Journal of Geosciences**, v. 8, n. 11, p. 9079–9088, 2015.