

Predição de Descargas Atmosféricas em Ambiente de Mineração utilizando Métodos de Detecção de Anomalias

RESUMO

Descargas atmosféricas são eventos raros, belos e igualmente perigosos. Por isso, prever a ocorrência antecipada de descargas permite pessoas e equipamentos tomar ações de segurança. Neste contexto, o objetivo dessa pesquisa foi propor um modelo de predição de descargas atmosféricas capaz de classificar, 15 minutos antes, a possibilidade de descarga(s) atingir uma região alvo, que corresponde a uma área de mineração. Dados de descargas atmosféricas provenientes de múltiplas fontes foram utilizados. Ocorrências na proximidade da área investigada foram utilizadas para avaliar os modelos de detecção de anomalias SVM One Class e Florestas Isoladas. O modelo gerado com Florestas Isoladas conseguiu prever, com antecedência, mais de 61% das descargas, considerando a métrica F1-score. Os resultados alcançados por esta pesquisa são promissores e fornecem tanto uma metodologia que considera padrões espaciais, quanto uma modelagem que busca por padrões espúrios nos dados.

PALAVRAS CHAVE. Detecção de anomalias, Aprendizado de máquina, Detecção de outliers, Raios, Descargas atmosféricas.

IC – Inteligência Computacional; POI – PO na Indústria

ABSTRACT

Lightning discharges are rare, mesmerizing and equally dangerous events. Due to this, early prediction of lightning allows people and companies to take safety actions. In this context, this paper aims to propose a lightning discharge model capable of classifying, 15 minutes before, the possibility of lightning striking a target place, which corresponds to a mining area. Lightning data from multiple sources were used. Occurrences in grids around the area were used to evaluate the SVM One Class and Isolation Forest outliers detection models. The model generated with Isolation Forest was able to predict, in advance, more than 61% of the discharges, considering the F1-score metric. The results achieved in this research are promising and provide both a methodology that considers spatial patterns and a modelling that looks for spurious patterns in the data.

KEYWORDS. Anomaly detection. Machine learning. Outliers detection. Lightning. Atmospheric discharges.

IC – Computational Intelligence; POI – OR in Industry

1. Introdução

Descargas atmosféricas têm gerado inúmeros incidentes e acidentes ao longo da história. Os riscos sobre pessoas e as operações das empresas têm forçado governos e empresas a buscarem soluções de mitigação. Porém, há desafios tanto na detecção do fenômeno em si [Dwyer e Uman, 2014], quanto nos modelos capazes de gerar alertas de advertência [Oliveira et al., 2022].

Toda a rede de detecção de descargas possui limitações na mensuração do fenômeno, de tal modo que apenas um percentual delas, em geral entre 40% e 80%, são detectadas [Pinto Jr e Pinto, 2018]. Consequentemente, muitas soluções existentes tornam-se ineficientes. Para contornar o problema sem haver perda de informação, redes integradas como a BrasilDAT Dataset [Pinto Jr e Pinto, 2018], Rede Integrada Nacional de Detecção de Descargas atmosféricas (RINDAT) [RINDAT, 2022] e ERA5 atmospheric reanalysis [Hersbach et al., 2020] tem sido amplamente utilizadas. Os autores Pinto Jr e Pinto [2018], por exemplo, argumentam que a BrasilDAT Dataset é capaz de captar mais de 90% das descargas e que em algumas regiões ela detecta mais que o dobro de qualquer rede individual.

A literatura também reporta, além do efeito caótico no espaço-tempo [Dwyer e Uman, 2014], que os dados de descargas são altamente desbalanceados [Mostajabi et al., 2019; Alves et al., 2022]. No campo da ciência de dados, isto quer dizer que na construção de modelos de aprendizado de máquina que tendem a classificar se um evento (futuro) é ou não a ocorrência de uma descarga atmosférica, há muito mais amostras de uma classe que de outra. Por exemplo, no Sudeste, onde o período de chuvas está mais concentrado entre os meses de setembro a março, as amostras de dados para o período de abril a agosto será majoritariamente negativa (i.e., sem descargas). Mesmo no período chuvoso, se forem consideradas a probabilidade de raios em uma área específica, por exemplo em uma área de mineração, haverá (possivelmente) muito mais amostras sem descargas do que com descargas.

Dentre as técnicas existentes no campo da inteligência computacional para lidar com este tipo de problema, os métodos de detecção de anomalias têm ganhado destaque, vide Agrawal e Agrawal [2015]. Inclui-se, ainda, recentes aplicações desta abordagem na prevenção de problemas causados por descargas atmosféricas em turbinas eólicas, como reportado por Matsui et al. [2020, 2022a,b].

No contexto de aprendizado de máquina, especialmente em mineração de dados, uma anomalia, ou outlier, refere-se a uma amostra que é consideravelmente diferente do restante dos dados. Detecção de anomalias é definida, em termos básicos, como a tarefa de identificar ocorrências raras em um conjunto de dados, seja por uma ou mais alterações dentro de um padrão de dados, um valor atípico ou um evento que esteja fora de uma tendência padrão.

Na literatura já existem pesquisas que empregaram métodos de detecção de anomalias para prever a ocorrência de descargas atmosféricas. Matsui et al. [2020, 2022a,b] utilizaram tais abordagens para detectar danos causados por raios em pás de turbinas eólicas. Todavia, estes trabalhos se diferem da presente pesquisa tanto em relação à base de dados utilizada, na localização onde a solução deve ser utilizada, e também no objetivo proposto.

O objetivo deste artigo é prever a ocorrência antecipada de descargas atmosféricas sobre uma área de mineração com raio de 10km no Sudeste do Brasil. Dados históricos da BrasilDAT Dataset foram utilizados. Verifica-se, então, a partir da ocorrência de descargas em sub-regiões próximas da área de proteção e se houve descarga (sim/não) na região a ser protegida 15 minutos à frente. Dois métodos de aprendizado de máquina de detecção de anomalias foram utilizados e os resultados se mostraram promissores.

O restante do artigo é dividido da seguinte forma: A Seção 2 apresenta os principais con-

ceitos e trabalhos relacionados a detecção de descargas atmosféricas utilizando aprendizado de máquina. A Seção 3 detalha a fonte de dados e os procedimentos metodológicos. A Seção 4 descreve os resultados encontrados e o desempenho dos modelos, incluindo, ainda, uma discussão sobre o tema de descargas atmosféricas e trabalhos futuros. Por fim, a Seção 5 apresenta a conclusão dos resultados.

2. Estado da Arte

A descarga atmosférica é um fenômeno da natureza muito atrativo para os seres humanos devido o brilho e beleza que ela proporciona no céu. Contudo, se trata de um evento altamente perigoso, causando mortes e danos em equipamentos industriais em todo o mundo Santos et al. [2018]; Holle [2014]. O número de acidentes fatais com raios ainda é elevado em muitos países [Holle, 2014], o que indica a necessidade de mais entendimento do fenômeno e mais medidas mitigadoras de riscos [Dwyer e Uman, 2014]. Há de 30-100 descargas do tipo nuvem-nuvem e nuvem-solo por segundo no mundo, ou 9 milhões por dia [Dwyer e Uman, 2014], sendo o Brasil o país com o maior número de incidências [Pinto Jr e Pinto, 2021].

Além disso, empresas que possuem equipamentos e pessoas em campo precisam seguir protocolos de proteção contra descargas, como aqueles descritos na Norma NBR 5419 – Proteção de estruturas contra descargas atmosféricas [ABNT, 2001].

Os sistemas de detecção de descargas variam em relação à capacidade de identificação do fenômeno no espaço e tempo [Oliveira et al., 2022]. As descargas podem ser do tipo intra-nuvem, nuvem-nuvem, solo-ar e nuvem-solo, sendo esta última aquela que mais danos traz à sociedade [Santos et al., 2018, 2016; Oliveira et al., 2022]. Elas podem ter cargas positivas ou negativas (maioria) [Pinto Jr e Pinto, 2021].

Já os modelos para predição de descargas são variados. Oliveira et al. [2022] investigaram os sensores de campo elétrico como ferramentas de geração de alertas antecipados de descargas. Embora tais equipamentos sejam utilizados por muitas empresas, os autores reportaram diferentes casos onde eles não captaram variação do campo elétrico quando ocorriam descargas – i.e., falhas na predição. O contrário também se aplicava, pois houveram casos em que havia variação no campo elétrico e não havia descargas – i.e., falsa predição.

Mostajabi et al. [2019] propuseram uma abordagem baseada em aprendizado de máquina que utilizava dados climáticos (pressão do ar local, temperatura do ar, umidade relativa e velocidade do vento) para prever os riscos de raios em 12 estações meteorológicas na Suíça com antecedência de 0-10, 10-20 e 20-30 minutos. Bons resultados foram reportados e foram superiores tanto ao modelo de Persistência, quanto o método baseado no campo eletrostático. Alves et al. [2017], utilizando dados de sondagem atmosférica por satélite, geraram modelos baseados em redes neurais artificiais para previsão de raios em diferentes janelas de previsões (1-5 horas).

Além disso, métodos de detecção de anomalias têm sido utilizados com sucesso na tarefa de predição de descargas atmosféricas. Apenas por definição, embora haja discussões entre os significados de *outliers* e anomalias, neste artigo eles serão considerados como sinônimos. Embasado também na definição de Aggarwal [2017] que refere-se a *outlier* como como anormalidades, discordantes, desvios ou anomalias na literatura da mineração de dados e estatística.

Modelos de aprendizado de máquina para detecção de anomalias possuem aplicação em problemas onde os dados são fortemente desbalanceados. A abordagem é utilizada na construção de um modelo focado no reconhecimento da classe majoritária, considerando como anomalia tudo que foge dos padrões identificados. Aplicações sólidas podem ser encontradas, por exemplo, em detecção de intrusos, fraudes, diagnóstico médico, aplicações da lei e ciências da terra, vide revisão

proposta por Aggarwal [2017]. Mais recentemente, Matsui et al. [2020, 2022a,b] desenvolveram diferentes propostas de controle supervisorio para pás de turbinas eólicas via modelos de detecção de anomalias. Danos causados nos equipamentos e propostas de verificação/inspeção foram reportadas.

Neste trabalho, buscou-se avaliar a aplicação de métodos de detecção de anomalias na predição de descargas atmosféricas em uma região de mineração situada em Minas Gerais. Como os dados de descargas são altamente desbalanceados, acredita-se que esta abordagem pode ser útil na identificação antecipada de tais fenômenos. Para isto, utilizando uma base de dados proveniente de múltiplas fontes, modelos de aprendizado de máquina foram treinados e os resultados foram positivos.

3. Procedimentos Metodológicos

3.1. Base de Dados

A base de dados utilizada é composta por ocorrências de descargas nuvem-solo e nuvem-nuvem com o horário e localização, no formato: yyyy-MM-dd HH:mm:ss. Foram utilizados dois anos de dados, de 2019/4 a 2021/3. A área de cobertura é de 50km ao redor do ponto a ser investigado.

Os dados foram fornecidos pelo Grupo de Eletricidade Atmosférica do Instituto Nacional de Pesquisas Espaciais (ELAT/INPE). Eles são provenientes de múltiplas fontes: três redes de superfície formadas, no total, por cerca de 110 sensores espalhados pelo Brasil que detectam radiação nas faixas de frequência de VLF e LF e, também, pelo sensor *Geostationary Lightning Mapper* (GLM) a bordo do satélite geoestacionário GOES-16, que detecta radiação na faixa do visível. Como antes mencionado, fontes integradas permitem uma maior eficiência de detecção das descargas, visto que nenhum sistema individual detecta todas as descargas que ocorrem [Pinto Jr e Pinto, 2021, 2018].

3.2. Preparação dos Dados

A quantidade de amostras analisadas foi 10727 amostras em 2019 (31%), 18063 em 2020 (52%) e 5784 em 2021 (17%). Dados de 2019 e 2020 foram utilizados para treino e 2021 para teste. Utilizou-se esta estratégia porque acredita-se que os padrões ocorridos em 2019 e 2020 podem ser úteis para o aprendizado do método e posterior predição para 2021.

Para simular uma aplicação capaz de fazer predição de descargas na área-alvo em tempo real, estes dados foram discretizados em intervalos de 5 minutos. Inspirado no trabalho de Santos et al. [2016], cuja proposta de agrupar regiões homogêneas no estado de São Paulo, a área monitorada nesta presente pesquisa de 100km de diâmetro foi dividida em partes menores, grades de 5km x 5km. Para cada intervalo de tempo, contabilizava-se a quantidade de descargas que atingia cada grade. A resposta, saída para o método de aprendizado de máquina, era se durante 15 minutos, 15 minutos à frente, alguma descarga atingiu a área alvo. Se sim, a resposta era (classe) 1. Caso contrário, (classe) 0. Cada amostra era, portanto, um vetor de 101 posições, sendo (as grades de) $x_1 \dots x_{100}$ a entrada e a posição 101 a saída desejada. Posteriormente, varreu-se os vetores de características e os valores maiores ou iguais a 1 (descarga) era substituído por 1 – ou seja, o vetor de características era também um vetor de valores binários. Para ilustrar, a Figura 1 representa estes passos.

Após este processamento a quantidade de dados de treino e teste, por casse, obtida é representada na Tabela 1. É possível observar que a classe 0 representa o rótulo de uma amostra que não teve uma ou mais descargas atmosféricas dentro dos 10km da região monitorada nos 15

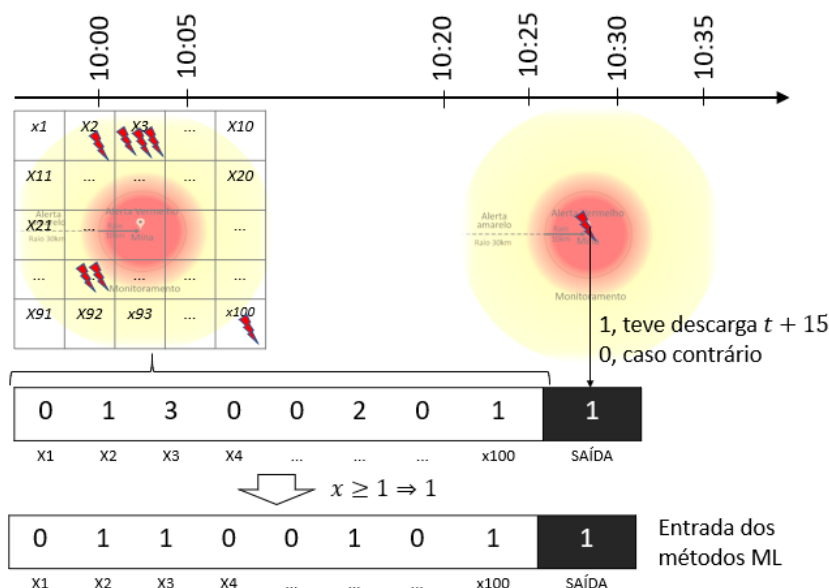


Figura 1: Esquema ilustrativo dos procedimentos utilizados para geração dos dados de entrada dos métodos de detecção de anomalias. Em cada intervalo de 5", verificava-se a quantidade de descargas em cada grade. A resposta era se durante 15", 15" à frente, alguma descarga atingiu a região a ser protegida.

minutos à frente. Já a classe 1 representa o rótulo de uma amostra que teve uma ou mais descargas atmosféricas dentro dos 10km da região monitorada 15 minutos à frente.

Tabela 1: Quantidade de descargas atmosféricas por ano do dataset.

Conjunto	Classe	Quantidade
Treino	0	26733
	1	0
Teste	0	4184
	1	2410

Ainda na Tabela 1, pode-se observar que não existe nenhuma amostra com o rótulo 1 no conjunto de treinamento. A justificativa para isso é pelo fato de a classe 1 ser considerada como anomalia, uma vez que existe poucas observações de ocorrências de descargas atmosféricas nas regiões alvo.

3.3. Métodos de Detecção de Anomalias

Dado esse problema claro de desbalanceamento entre as classes, entende-se que faz sentido explorar métodos de detecção de anomalias. Por isso, duas técnicas foram abordadas nessa pesquisa investigativa para a modelagem de um sistema para predição de descargas atmosféricas. São eles: SVM One Class e Florestas Isoladas [Liu et al., 2012], ambos implementados no pacote Scikit-learn [Pedregosa et al., 2011].

Sabe-se que um algoritmo de aprendizado de máquina tem vários parâmetros que não são ajustados pelo conjunto de treinamento. Esses parâmetros controlam a precisão do modelo. Por isto,

os hiperparâmetros são particularmente importantes na construção de um modelo de aprendizado. Eles são configurados antecipadamente e utilizados pelo método de aprendizado antes que o modelo seja treinado.

Dessa forma, existem algumas abordagens de otimização de hiperparâmetros, sendo que uma delas é a pesquisa em grade, que é uma técnica de ajuste que tenta calcular os valores ótimos dos hiperparâmetros. A técnica usa a busca exaustiva que é realizada nos valores dos parâmetros específicos de um modelo de aprendizado. O resultado do uso dessa abordagem é o melhor conjunto de hiperparâmetros possíveis para o modelo.

A Tabela 2 apresenta os valores de hiperparâmetros testados na construção do modelo SVM One Class, usando a técnica de pesquisa em grade.

Tabela 2: Hiperparâmetros avaliados na pesquisa em grade para o método SVM One Class.

Abreviação	Hiperparâmetro	Valores testados
(1)	kernel	linear, poly, rbf, sigmoid
(2)	gamma	scale, auto
(3)	nu	0.01, 0.1, 0.5, 0.8, 0.99

A Tabela 3 apresenta os valores de hiperparâmetros testados na construção do modelo Florestas Isoladas, usando a técnica de pesquisa em grade.

Tabela 3: Hiperparâmetros avaliados na pesquisa em grade para o método de Florestas Isoladas.

Abreviação	Hiperparâmetro	Valores testados
(1)	n_estimators	100 a 800, com salto de 25
(2)	max_samples	100 a 500, com salto de 25
(3)	contamination	0.001, 0.01, 0.1, 0.2, 0.3, 0.4, 0.5
(4)	max_features	1, 3, 7, 10, 12
(5)	bootstrap	True, False

Pré-processados os dados e definidos os métodos de aprendizado bem como a abordagem de otimização de hiperparâmetros, deu-se início a etapa de treinamento. Após cada geração de um novo modelo com uma combinação de hiperparâmetros, foi feita a avaliação deles com as amostras do conjunto de teste.

No problema atual que se trata de uma classificação estatística, emprega-se uma matriz de confusão para a avaliação de modelos de aprendizado de máquina. Ela é uma tabela que permite a visualização do desempenho de um modelo. Cada linha dessa tabela representa as instâncias de uma classe prevista, enquanto cada coluna ilustra as instâncias de uma classe real.

Uma matriz de confusão é composta por quatro valores: verdadeiro positivo, falso negativo, falso positivo e verdadeiro negativo. Essa matriz é muito útil para a avaliação do modelo, uma vez que seus dados descrevem o resultado da classificação de cada registro, além de que é através dela que se torna possível encontrar as demais métricas, como acurácia, precisão, revocação e F1-score. Essas métricas foram as utilizadas na avaliação dos modelos e cada uma delas é obtida sendo:

- **Acurácia:** é a divisão entre todos os acertos pelo total de exemplos.
- **Precisão:** também conhecida como especificidade, indica a relação entre os valores previstos positivos realizados corretamente e todas as previsões positivas, sendo as verdadeiras e as falsas.
- **Revocação:** também conhecida como sensibilidade, é utilizada para indicar a relação entre as previsões positivas realizadas corretamente e todas as previsões que realmente são positivas (verdadeiro positivo e falso negativo).
- **F1-score:** é uma maneira de se visualizar as métricas de precisão e de revocação juntas. O cálculo dessa métrica se dá através da média harmônica, uma vez que os valores calculados são iguais, em que se obtém resultados muito próximos da média aritmética simples. No entanto, sempre que os valores são diferentes, essa média se aproxima mais dos valores menores.

Através dessas métricas foi possível avaliar e comparar cada um dos modelos que foram construídos por meio de uma combinação diferente de hiperparâmetros. Assim sendo, foi possível concluir as diferenças de desempenho entre o SVM One Class e Florestas Isoladas, especificamente para o problema de predição de descargas atmosféricas para os 15 minutos à frente.

3.4. Ambiente de Desenvolvimento

Para o desenvolvimento do estudo foi utilizada a linguagem de programação Python 3.6, no ambiente de desenvolvimento Google Colaboratory [Bisong, 2019].

O ambiente de desenvolvimento foi preparado para uso da Scikit-learn, que é uma biblioteca de aprendizado de máquina de código livre para a linguagem de programação Python. O pacote possibilita o uso de algoritmos de classificação, regressão e agrupamento, incluindo o SVM One Class e as Florestas Isoladas, empregados nessa pesquisa. Além disso, a biblioteca foi projetada para interoperar com outros pacotes numéricos e científicos da linguagem de programação Python, tais como o NumPy e o SciPy [Pedregosa et al., 2011].

4. Resultados

Os resultados descritos nesta seção referem-se à predição antecipada de descargas atmosféricas em uma região de mineração em Minas Gerais. A proposta faz parte de uma aplicação, onde busca-se proteger pessoas e equipamentos contra os riscos trazidos por raios.

A Tabela 4 apresenta os resultados dos cinco melhores modelos utilizando o método SVM One Class. A métrica escolhida para ordenar os resultados foi a F1-score, uma vez que ela é uma maneira de visualizar as métricas de precisão e de revocação juntas. Por mais que a acurácia tenha sido obtida, tem que se levar em conta que o conjunto de teste está desbalanceado e, por isso, seus resultados devem ser analisados com cautela.

A Tabela 5 apresenta os cinco melhores resultados avaliados no conjunto de teste, considerando o método Florestas Isoladas.

Considerando os resultados obtidos por cada um dos métodos, percebe-se que o modelo gerado utilizando o algoritmo de Florestas Isoladas conseguiu um desempenho melhor em termos de F1-score.

A Tabela 6 apresenta a matriz de confusão calculada com as amostras do conjunto de teste, gerada com o melhor modelo de Florestas Isoladas. Percebe-se que das 4184 amostras da classe 0

Tabela 4: Resultados dos cinco melhores modelos gerados com o método SVM One Class ordenados por F1-score.

Hiperparâmetros			Métricas			
(1)	(2)	(3)	Acurácia	Precisão	Revocação	F1-score
rbf	scale	0.5	0.53	0.42	0.80	0.56
rbf	auto	0.5	0.53	0.42	0.80	0.55
rbf	scale	0.8	0.45	0.39	0.95	0.55
rbf	auto	0.8	0.45	0.39	0.95	0.55
rbf	auto	0.99	0.36	0.36	0.99	0.53

Tabela 5: Resultados dos cinco melhores modelos gerados com o método de Florestas isoladas ordenado por F1-score.

Hiperparâmetros					Métricas			
(1)	(2)	(3)	(4)	(5)	Acurácia	Precisão	Revocação	F1-score
250	250	0.3	1	0	0.67	0.53	0.70	0.61
300	175	0.3	1	0	0.66	0.53	0.70	0.60
350	300	0.3	1	0	0.66	0.52	0.70	0.60
125	100	0.3	3	1	0.66	0.53	0.69	0.60
200	100	0.3	3	0	0.65	0.52	0.71	0.60

contidas no conjunto de teste, o modelo conseguiu classificar corretamente 2610 (62%), enquanto 1574 (38%) foram rotuladas como sendo da classe 1, ou anomalia. Em relação as 2410 amostras da classe 1 (anomalias), 740 (31%) delas foram incorretamente rotuladas como classe 0, enquanto 1670 (69%) foram classificadas corretamente como 1.

Tabela 6: Matriz de confusão considerando o melhor conjunto de hiperparâmetros para o método Florestas Isoladas.

		Previsto pelo modelo	
		Sem descarga (0)	Com descarga (1)
Real	Sem descarga (0)	2610	1574
	Com descarga (1)	740	1670

Quando o modelo classifica uma amostra como 0 (normal), significa que nos próximos 15 minutos não irá ocorrer uma ou mais descargas atmosféricas no alvo (10km em torno do ponto do interesse). Em contrapartida, quando o modelo rotula como 1 (anomalia) uma determinada amostra, quer dizer que o modelo de aprendizado está considerando que nos próximo 15 minutos uma ou mais descargas atmosféricas podem, possivelmente, atingir a região onde os trabalhadores desempenham atividades em campo aberto.

4.1. Discussão

Os resultados encontrados foram promissores e confirmam que os métodos de detecção de anomalias pode ser utilizados como modelos de identificação antecipada de descargas atmosféricas. O modelo, quando implementado, deve ser entendido como uma ferramenta de auxílio à equipe de segurança do trabalho e utilizado juntamente com outras ferramentas para geração de alertas ao pessoal que trabalha em campo aberto e precisam buscar por locais seguros dadas as instabilidades climáticas.

Todavia, algumas discussões ainda são pertinentes quanto aos sistemas de identificação de descargas e/ou geração de alertas e devem ser levados em consideração na implementação de tais modelos:

- Sistemas para detecção de descargas devem ser aperfeiçoados. Toda rede de detecção de descargas atmosféricas possui limitações, de tal modo que apenas um percentual das descargas, em geral entre 40% e 80%, são detectadas. De modo a detectar-se mais descargas, a integração de diversas redes individuais tem sido propostas, como a BrasilDAT Dataset [Pinto Jr e Pinto, 2018];
- Redes integradas são capazes de identificar descargas com mais precisão [Pinto Jr e Pinto, 2018]. O modelo de identificação de descargas fica, com isso, totalmente dependente da acurácia das redes de detecção, sendo que o oposto não se aplica. Este argumento justifica novos estudos na área;
- Embora neste estudo foram utilizados dados espaciais (e implicitamente temporais) de descargas, diversos outros estudos foram propostos na literatura utilizando outros parâmetros, por exemplo meteorológicos [Mostajabi et al., 2019], campo elétrico [Oliveira et al., 2022; Srivastava et al., 2015] e imagens de satélite [Sakaino, 2012];
- Em caso de implementação de um sistema de alertas a partir da identificação das descargas, é necessário implementar lógicas de envio de alertas e de renovação.
- Sobre o último tópico, os sistemas de comunicação dos alertas aos trabalhadores deve ser eficaz e também tratado com importância pelas empresas. Podem ser utilizados, por exemplo, mensagens de texto como SMS, Telegram, WhatsApp e avisos sonoros, como bipes, sirenes e auto-falantes;
- A ocorrência de uma descarga tem que ser entendida como um real perigo à vida. Holle [2014] apontou que educação e conscientização são necessários para haver diminuição no número de vítimas. Este aviso é válido para todos, inclusive serve como argumento para que a equipe de segurança no trabalho possa oferecer treinamento aos trabalhadores com atividades em campo aberto.

A literatura ainda fornece recomendações de proteção contra descargas, como aquelas descritas na Norma NBR 5419 – Proteção de estruturas contra descargas atmosféricas (técnicas) e outras gerais como aquelas na cartilha elaborada pelo ELAT/INPE [ELAT/INPE, 2022]. Na lista abaixo é feita uma compilação delas que podem ser úteis para empresas, pessoas e trabalhadores em campo.

- Buscar abrigo dentro de edificações;

- Permanecer distante de árvores;
- Veículo fechado, com os vidros fechados, também é uma boa opção para se proteger;
- Evitar utilizar veículos sem capota, motos e bicicletas
- Permanecer distante de postes e fiação elétrica, cercas de arame, varais metálicos e linhas férreas;
- Evitar se abrigar em tendas;
- Evitar ficar em alto de morros, topos de prédios e telhados;
- Evitar permanecer em estruturas altas como torres de linhas telefônicas e de energia elétrica;
- Não permanecer em áreas abertas, como praias, campos de futebol, quadras esportivas e estacionamentos;
- Na ausência de abrigo, optar por locais mais baixos do terreno. Permanecer agachado, com os pés juntos, cabeça abaixada e as mãos sob o joelho. Não deitar.

4.2. Aplicação em ambiente de mineração

No caso de ambiente de mineração, entende-se que abrigo seguro é qualquer edificação de alvenaria ou edificação que possua SPDA (Sistema de Proteção de Descargas Atmosférica) com aterramento adequado conforme preconiza a NBR 5419. Qualquer veículo, caminhão ou equipamento fora de estrada, empilhadeira, perfuratriz, máquina de pátio, por exemplo, pode ser considerado um abrigo. Todo ambiente fechado metálico funciona como uma “Gaiola de Faraday” e é considerado um ambiente de risco reduzido frente uma descarga atmosférica.

Neste ambiente são feitos projetos e estudo eletrogeométrico onde se utiliza o princípio da esfera rolante para verificação do raio de proteção ou cobertura de um SPDA/para-raios tipo Franklin. Estes estudos são úteis e necessários para representar ambientes de risco reduzido frente a uma descarga atmosférica – em caso de alerta vermelho.

Nas trocas de turno, há locais próprios e dotados de um sistema de SPDA e aterramento onde os caminhões do tipo fora de estrada estacionam e os empregados descem do veículo e entram nas caminhonetes ou veículos de pequeno porte. Estes ambientes são concebidos e fundamentados em normas técnicas nacionais, internacionais e melhores práticas de engenharia.

5. Conclusão

Este artigo apresentou uma metodologia para detecção automática e antecipada de descargas atmosféricas em uma área específica. A região investigada possui um raio de 10km e corresponde a uma área de mineração, onde diversos trabalhadores desempenham atividades a céu aberto. Os métodos SVM One Class e Florestas Isoladas foram testados considerando a presença de descargas em regiões próximas e pode ser utilizado em uma aplicação que funciona em tempo real, sendo carregado para fazer predições a cada 5 minutos. O modelo gerado com o método Florestas Isoladas teve um melhor desempenho, acima de 60% em falhas e falsas identificações.

A abordagem proposta é facilmente replicável para outras localidades, rurais ou urbanas, como agricultura e centros urbanos. Contudo, precisa-se analisar cuidadosamente o desbalanceamento dos dados, uma vez que haverá menos amostras positivas. Precisa validar também se os resultados vão atender o time envolvido no projeto (cliente). Além disso, vale o argumento de que

um bom modelo de predição de descargas sempre vai depender de uma boa base de dados, capaz de identificar corretamente as descargas no espaço e tempo.

Como discutido amplamente na literatura, a metodologia proposta e os resultados apresentados podem servir como modelo de apoio à equipe de segurança do trabalho. É válido, ainda, mencionar que se faz necessária a validação da proposta com o time gerencial para buscar uma melhor adequação às atividades desenvolvidas em campo aberto.

É necessário entender, portanto, que a solução de prevenção contra descargas é, na verdade, um conjunto de soluções. Elas vão desde a conscientização de que raios realmente são perigosos, definição prática e treinamento contínuo de ações a serem tomadas após um alerta iminente de descarga, sistemas efetivos de notificação de alertas e possíveis retreinos nos modelos de aprendizado de máquina.

Referências

- ABNT (2001). Associação Brasileira de Normas Técnicas. NBR 5419 - Proteção de estruturas contra descargas atmosféricas. <https://gsea.com.br/normasabnt/ABNT%205419%20-%20PARA%20RAIOS.pdf>. Online; acessado em 01/07/22.
- Aggarwal, C. C. (2017). An introduction to outlier analysis. In *Outlier analysis*, p. 1–34. Springer.
- Agrawal, S. e Agrawal, J. (2015). Survey on anomaly detection using data mining techniques. *Procedia Computer Science*, 60:708–713.
- Alves, E. R., Tavares da Costa Jr, C., Lopes, M. N. G., da Rocha, B. R. P., e de Sá, J. A. S. (2017). Lightning prediction using satellite atmospheric sounding data and feed-forward artificial neural network. *Journal of Intelligent & Fuzzy Systems*, 33(1):79–92.
- Alves, M. A., Oliveira, B. A. S., Silvestrow, F. P., Rodrigues, L. F. M., Daher, E. L., Maia, W., Soares, W. S., Santos, A. P. P., Ferreira, D. B. S., e Pinto Junior, O. (2022). Predição de descargas atmosféricas utilizando machine learning para prevenção de acidentes. In *IX Simpósio Brasileiro de Sistemas Elétricos (IX SBSE)*. SBA.
- Bisong, E. (2019). Google colabatory. In *Building Machine Learning and Deep Learning Models on Google Cloud Platform*, p. 59–64. Springer.
- Dwyer, J. R. e Uman, M. A. (2014). The physics of lightning. *Physics Reports*, 534(4):147–241.
- ELAT/INPE (2022). Proteção contra Raios. http://www.inpe.br/webelat/docs/Cartilha_Protecao_Contra_Raios_Brasil_2020.pdf. Online; acessado em 30/06/22.
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., et al. (2020). The era5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730):1999–2049.
- Holle, R. L. (2014). Some aspects of global lightning impacts. In *2014 International Conference on Lightning Protection (ICLP)*, p. 1390–1395. IEEE.
- Liu, F. T., Ting, K. M., e Zhou, Z.-H. (2012). Isolation-based anomaly detection. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 6(1):1–39.

- Matsui, T., Yamamoto, K., e Ogata, J. (2022a). Anomaly detection for wind turbine damaged due to lightning strike. *Electric Power Systems Research*, 209:107918.
- Matsui, T., Yamamoto, K., e Ogata, J. (2022b). Anomaly detection using a scada feature extractor and machine learning to detect lightning damage on wind turbine blades. *IEEE Transactions on Electrical and Electronic Engineering*.
- Matsui, T., Yamamoto, K., Sumi, S., e Triruttanapiruk, N. (2020). Detection of lightning damage on wind turbine blades using the scada system. *IEEE Transactions on Power Delivery*, 36(2): 777–784.
- Mostajabi, A., Finney, D. L., Rubinstein, M., e Rachidi, F. (2019). Nowcasting lightning occurrence from commonly available meteorological parameters using machine learning techniques. *Npj Climate and Atmospheric Science*, 2(1):1–15.
- Oliveira, B. A. S., Alves, M. A., Maia, W., Silvestrow, F. P., Soares, W. S., Rodrigues, L. F. M., Santos, A. P. P., Daher, E. L., Ferreira, D. B. S., e Pinto Junior, O. (2022). Uso do campo elétrico na geração de alarmes para risco de descargas atmosféricas. In *IX Simpósio Brasileiro de Sistemas Elétricos (IX SBSE)*. SBA.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, 12:2825–2830.
- Pinto Jr, O. e Pinto, I. (2018). Brasil dataset: Combining data from different lightning locating systems to obtain more precise lightning information. In *25th International Conference on Lightning Detection*.
- Pinto Jr, O. e Pinto, I. R. C. A. (2021). *Brasil campeão mundial de raios*. Artliber.
- RINDAT (2022). Rede Integrada Nacional de Detecção de Descargas atmosféricas. <http://www.rindat.com.br/>. Online; acessado em 30/06/22.
- Sakaino, H. (2012). Spatio-temporal image pattern prediction method based on a physical model with time-varying optical flow. *IEEE Transactions on Geoscience and Remote Sensing*, 51(5): 3023–3036.
- Santos, A. P. P., Coelho, C. A., Pinto Junior, O., dos Santos, S. R. Q., de Lima, F. J. L., e de Souza, E. B. (2018). Climatic diagnostics associated with anomalous lightning incidence during the summer 2012/2013 in southeast brazil. *International Journal of Climatology*, 38(2):996–1009.
- Santos, A. d., Lima, F. d., Souza, E. d., Pinto Júnior, O., e Santos, S. d. (2016). Aplicação da análise cluster na identificação de regiões homogêneas para incidência de descargas atmosféricas do tipo nuvem-solo no estado de são paulo, sudeste do brasil. *Revista Brasileira de Geografia Física*, 9 (07):2188–2195.
- Srivastava, A., Mishra, M., e Kumar, M. (2015). Lightning alarm system using stochastic modelling. *Natural Hazards*, 75(1):1–11.