Detecção de Ataques de GPS em Veículos Aéreos Não Tripulados com Classificação Multiclasse

Gustavo Gualberto Rocha de Lemos[1], **Rodrigo Augusto Cardoso da Silva**[1]

[1]Centro de Matemática, Computação e Cognição
Universidade Federal do ABC (UFABC)
Av. dos Estados, 5001 – 09210-580 – Santo André – SP – Brasil

gustavo.gualberto@aluno.ufabc.edu.br,cardoso.rodrigo@ufabc.edu.br

Abstract. *Unmanned aerial vehicles (UAVs) are increasingly being employed across various domains. These vehicles usually rely on the Global Positioning System (GPS), which makes them vulnerable to attacks based on fake GPS signals. Hence, this paper proposes an Intrusion Detection System (IDS) that utilizes machine learning techniques to detect and identify GPS Jamming and three types of GPS Spoofing attacks. The proposed multiclass classifier enables the identification of the attack type – an essential factor in determining the most effective protective measures. The achieved accuracy rate was 98.08%, with 2.6% false negatives, lowering the likelihood of overlooking attacks, which is crucial in real UAV deployments.*

Resumo. *Veículos aéreos não tripulados (VANTs) têm sido cada vez mais utilizados em diversos domínios. Esses veículos geralmente dependem do Sistema de Posicionamento Global (GPS), o que os torna vulneráveis a ataques baseados em sinais de GPS falsos. Assim, este artigo propõe um Sistema de Detecção de Intrusão (IDS) que utiliza técnicas de aprendizado de máquina para detectar e identificar GPS Jamming e três tipos de ataques de GPS Spoofing. O classificador multiclasse proposto permite a identificação do tipo de ataque – algo essencial para determinar as medidas de proteção mais eficazes. A acurácia alcançada foi de 98,08%, com 2,6% de falsos negativos, diminuindo a probabilidade de ignorar ataques, algo essencial em infraestruturas com VANTs reais.*

1. Introdução

Os avanços tecnológicos dos veículos aéreos não tripulados (VANTs) têm facilitado sua utilização em diversos setores, como agricultura, vigilância, cidades inteligentes, e redes de computadores [Khan et al. 2022, da Silva et al. 2021]. A versatilidade dos VANTs, que deixaram de ser apenas dispositivos mecânicos para também serem nós de processamento e rede, os torna alvos potenciais para ataques cibernéticos [Ranyal and Jain 2021].

Os VANTs podem ser operados de forma autônoma. Nesse caso, eles geralmente dependem de sinais fornecidos por um Sistema Global de Navegação por Satélite (GNSS) para determinar sua posição. O *Global Positioning System* (GPS) é o GNSS mais conhecido, mantido pelos EUA. Com isso, a sigla GPS é comumente usada para se referir a GNSSs como um todo; utilizaremos essa convenção neste artigo. Existem dois tipos de sinais GPS: um projetado para uso civil, que não possui de mecanismos para garantir a

confidencialidade e a integridade dos dados, e outro sinal criptografado reservado exclusivamente para fins militares. A maioria dos dispositivos, incluindo VANTs não militares, depende do sinal civil, tornando-os mais suscetíveis a ataques de GPS.

Dois ataques comuns envolvendo sinais de GPS são os de GPS *Jamming* e de GPS *Spoofing* [Derhab et al. 2023]. No GPS *Jamming*, o atacante emite energia de radiofrequência com potência suficiente para interferir nos sinais recebidos pelos dispositivos na área-alvo. No GPS *Spoofing*, o atacante utiliza antenas terrestres para gerar sinais contendo informações de localização falsas. Um ataque de GPS *Spoofing* bem-sucedido, por exemplo, pode direcionar o VANT para a localização do atacante, permitindo que ele sequestre a aeronave, roube sua carga e potencialmente acesse informações sensíveis. Portanto, a detecção de ataques de GPS é crítica em missões autônomas de VANTs.

Uma maneira de mitigar ataques de GPS é empregar software para realizar a análise de sinais – um Sistema de Detecção de Intrusão (IDS) capaz de identificar autonomamente sinais falsos de GPS. Ao detectar um sinal falso, o IDS pode, por exemplo, alertar o proprietário da aeronave, ou o VANT poderia acionar instrumentos alternativos para navegação. Uma abordagem comum na literatura foi de usar um IDS para determinar se a aeronave está sujeita a um ataque de GPS ou não, independente do tipo de ataque. No entanto, diferentes ataques têm implicações diversas para os VANTs. Em um ataque de GPS *Jamming*, por exemplo, o objetivo do atacante pode ser apenas desorientar a aeronave, enquanto um ataque de GPS *Spoofing* visa obter o controle da trajetória do VANT. Portanto, este artigo examina três classes distintas de ataques de GPS *Spoofing* e o ataque de GPS *Jamming*. Em vez de apenas indicar que o VANT está sob ataque (IDS de classe binária), o IDS identifica o tipo específico de ataque em progresso (IDS de multiclasse). Esta é uma informação chave para decidir a resposta mais apropriada pelo VANT e permite melhores investigações forenses do ataque.

Diversos autores [Aissou et al. 2021, Talaei Khoei et al. 2022, Khoei et al. 2023, Gasimova et al. 2022, Khoei et al. 2022, Aissou et al. 2022, Talaei Khoei et al. 2023] propuseram IDSs baseados em aprendizado de máquina para detectar ataques de GPS *Jamming* e GPS *Spoofing*. No entanto, esses IDSs foram baseados em classificadores binários. A contribuição deste artigo é um IDS multiclasse para detectar ataques de GPS *Jamming* e GPS *Spoofing* em três níveis de sofisticação. Foram usados classificadores de conjunto (*ensemble*), que utilizam múltiplos classificadores base, para melhorar o desempenho do IDS. A solução final foi baseada em um classificador *Stack* multiclasse que alcançou uma acurácia de 98,08% com 2,6% de falsos negativos nas avaliações. Foram avaliadas tanto soluções de classe binária quanto multiclasse. Os resultados obtidos indicam que o IDS multiclasse alcançou uma alta precisão, que, embora ligeiramente menor da obtida pela versão binária, reduziu efetivamente os falsos negativos comparado ao estado da arte. Um falso negativo implica em um ataque não detectado, potencialmente resultando em consequências graves, como destruição ou sequestro do VANT, demonstrando assim os benefícios do IDS baseado em um classificador multiclasse aqui proposto.

O restante deste artigo está organizado da seguinte forma. A Seção 2 revisa a literatura relacionada a ataques de GPS direcionados a VANTs. A Seção 3 introduz o cenário considerado e IDS proposto. A Seção 4 descreve a metodologia aplicada nos experimentos, enquanto a Seção 5 apresenta os resultados numéricos obtidos. Finalmente, a Seção 6 conclui este artigo.

2. Trabalhos relacionados

Esta seção revisa artigos relacionados a ataques de GPS *Spoofing* e GPS *Jamming* em VANTs. Os autores de [Derhab et al. 2023] avaliaram o impacto e o risco associados a vários tipos de ataques em VANTs, classificando os ataques de GPS *Spoofing* e GPS *Jamming* como extremamente arriscados devido à facilidade de acesso dos atacantes ao sistema GPS. Por exemplo, os autores de [da Silva 2017] realizaram testes em VANTs reais e confirmaram que esses ataques são facilmente executados com equipamentos baratos e software gratuito. Consequentemente, esses ataques requerem atenção e contramedidas mais robustas em comparação com outros tipos de ataques.

A detecção de ataques de GPS Spoofing e GPS Jamming em VANTs é um grande tema de pesquisa na literatura [Davidovich et al. 2022, Whelan et al. 2022, Wei et al. 2022, Aissou et al. 2021, Talaei Khoei et al. 2022]. As soluções empregaram vários métodos envolvendo dados de sinal de GPS e outros sensores. A solução em [Davidovich et al. 2022] avaliou imagens capturadas pela câmera do VANT para detectar a presença de um ataque de GPS *Spoofing*. Essa detecção é alcançada analisando a correlação entre os quadros de vídeo e a localização do VANT usando algoritmos de visão computacional como BFMatcher e SURF. Este método oferece a vantagem de não exigir hardware adicional. No entanto, ele tem limitações, como depender de condições de iluminação ideais e ser menos eficaz em áreas com terrenos uniformes.

A aprendizagem de máquina é uma técnica comum para detectar ataques de GPS *Spoofing* e GPS *Jamming* [Whelan et al. 2022, Aissou et al. 2021, Talaei Khoei et al. 2022, Khoei et al. 2023, Gasimova et al. 2022, Khoei et al. 2022, Aissou et al. 2022, Talaei Khoei et al. 2023]. Por exemplo, os autores de [Whelan et al. 2022] introduziram o IDS chamado MAVIDS, que emprega um método de detecção baseado em anomalias para identificar ataques de GPS *Spoofing* e GPS *Jamming*. Os autores treinaram o classificador usando dados coletados de sensores do VANT que utilizaria o IDS, oferecendo uma solução viável diante da disponibilidade limitada desse tipo de dado. Os autores empregaram Máquina de Vetores de Suporte de Classe Única (OC-SVM), Fator Local de Anomalia (LOF) e Autoencoder, resultando em F1-score de 90,57% para GPS *Spoofing* e 94,3% para GPS *Jamming*.

O artigo [Aissou et al. 2021] apresentou um IDS que utiliza classificadores baseados em árvores para detectar ataques de GPS *Spoofing*. O treinamento utiliza um conjunto de dados composto por dados de características de sinal de GPS obtidos de um receptor de GPS real, juntamente com dados modificados para representar as assinaturas dos três tipos de ataques de GPS *Spoofing*. Os classificadores testados foram XGBoost, Random Forests, Gradient Boost e Light Gradient Boost Machine, com o XGBoost alcançando a maior precisão (95,52%).

Os dados coletados por [Aissou et al. 2021] também foram utilizados em outros estudos [Talaei Khoei et al. 2022, Khoei et al. 2023, Gasimova et al. 2022, Khoei et al. 2022, Aissou et al. 2022, Talaei Khoei et al. 2023]. Em [Talaei Khoei et al. 2022], dois métodos de seleção dinâmica de classificadores, Seleção Dinâmica Otimizada por Métrica (MOD) e Seleção Dinâmica Otimizada por Métrica Ponderada (WMOD), foram empregados para otimizar o desempenho de dez classificadores tradicionais, levando a uma acurácia de 99,6% para ambos os métodos.

Os autores de [Khoei et al. 2023] propuseram três técnicas de *Deep Learning* para detecção de ataques de GPS *Spoofing*: Rede Neural Profunda, Rede Neural U e Memória de Longo Prazo com Rede Neural. A técnica de Rede Neural U obteve o melhor resultado, alcançando uma precisão de 98,80%. Em [Gasimova et al. 2022], classificadores de conjunto *Stacking*, *Boosting* e *Bagging* foram aplicados para detectar ataques de GPS *Spoofing*, com o método de *Stacking* alcançando a maior precisão de 95,43%.

O estudo de [Khoei et al. 2022] comparou modelos de aprendizado de máquina supervisionado e não supervisionado para detecção de ataques de GPS, sendo que o modelo de Árvore de Decisão demonstrou a maior precisão de 99,87%. Os autores de [Aissou et al. 2022] avaliaram classificadores baseados em instâncias, sendo que o modelo Nu-SVM alcançou a melhor precisão de 92,78%. Além disso, [Talaei Khoei et al. 2023] investigou o impacto de mudanças nos parâmetros do classificador e nas características do conjunto de dados no desempenho do classificador, sendo que o modelo de Árvore de Decisão alcançou uma precisão notável de 99,99% em um *data set* balanceado com parâmetros otimizados.

Trabalhos anteriores propuseram soluções para detectar ataques de GPS *Jamming* e *Spoofing*; no entanto, eles apenas empregaram classificação binária, isto é, apenas detectaram a ocorrência de um ataque, sem detectar seu tipo. O presente artigo considera várias classes de ataques de GPS (*Jamming* e três tipos de ataques de *Spoofing*), corroborando para a adoção de contramedidas mais eficazes. Este artigo também aborda ataques de GPS *Jamming*, que foram parcialmente negligenciados em trabalhos anteriores na literatura. Por fim, os resultados dos classificadores do presente artigo trazem avanços comparados às soluções existentes na literatura.

3. Cenário considerado e método de classificação proposto

Um VANT em voo autônomo deve determinar sua localização com precisão para cumprir sua missão. Por exemplo, em aplicações de vigilância aérea, a aeronave necessita determinar sua localização precisa durante o voo. Se um atacante executar um ataque de GPS *Jamming* contra este VANT, este último seria impedido de seguir sua rota de vigilância. No caso de um ataque de GPS *Spoofing*, a rota da aeronave poderia ser alterada para redirecioná-la ao atacante ou mesmo para causar uma colisão.

Para mitigar os problemas causados pelos ataques de GPS em VANTs, assumimos que o computador de bordo executará um software IDS para identificar possíveis ataques de GPS *Spoofing* ou *Jamming*. Este sistema analisará dados dos sinais de GPS recebidos, indicando se um ataque está ocorrendo e, em caso positivo, especificando o tipo de ataque. O IDS passará por uma fase de treinamento, utilizando dados representando sinais normais e sinais de GPS gerados por um atacante.

O restante desta seção classifica os tipos de ataques de GPS em VANTs (Subseção 3.1) e apresenta o IDS proposto neste artigo (Subseção 3.2).

3.1. Classificação dos ataques

Esta subseção detalha ataques de GPS *Spoofing* e GPS *Jamming*. Os ataques de GPS *Spoofing* podem ser classificados em três tipos [Haider and Khalid 2016]: *Spoofing* simples, *Spoofing* intermediário e *Spoofing* sofisticado. Esses ataques serão descritos no restante desta subseção. Para caracterização e detecção desses ataques, treze características

do sinal GPS são consideradas: *Pseudo Random Number* (PRN), *Doppler Shift Measurement* (DO), *Pseudo Range* (PR), *Receiver Time* (RX), *Time of Week* (TOW), *Carrier Phase Cycles* (CP), *Early Correlators* (EC), *Late Correlators* (LC), *Prompt Correlator* (PC), *Prompt In-Phase Component* (PIP), *Prompt Quadrature Component* (PQP), *Tracking Carrier Doppler* (TCD), e *Carrier to Noise Ratio* (CN0). Essas características são detalhadas em [Talaei Khoei et al. 2022].

No ataque de *Spoofing* simples, o atacante utiliza uma única antena de GPS e não conhece a posição do VANT. Ele gera um sinal de GPS falso dessincronizado, fazendo com que as medições do efeito Doppler se desviem além da faixa normal de cerca de 20 Hz. Neste tipo de ataque, os sinais de GPS falsos chegam com maior potência em comparação com os autênticos, uma vez que são gerados em uma antena próxima, resultando em um maior CN0. Isso torna este ataque fácil de detectar [Aissou et al. 2021].

Em um ataque de *Spoofing* intermediário, o atacante ainda utiliza uma única antena, mas sabe a posição real do VANT, manipulando o sinal de GPS gerado para manter os valores normais do DO e PR. Detectar tal ataque requer monitoramento meticuloso das seguintes características do sinal: TOW, CP, e deslocamento informações de amplitude do correlator [Aissou et al. 2021].

Em um ataque de *Spoofing* sofisticado, o atacante utiliza várias antenas de GPS sincronizadas para gerar sinais falsos em vários canais, imitando a funcionalidade de uma constelação de satélites GPS legítima, ganhando controle completo sobre o sistema de localização. Sincronizar múltiplas antenas é desafiador, mas pode ser feito usando tecnologias avançadas de Software Definido por Rádio. As características do sinal mais afetadas por este tipo de ataque são os correlatores [Aissou et al. 2021].

Por fim, durante um ataque de *Jamming*, um sinal de alta potência é enviado para o receptor de GPS do VANT. Dada a potência tipicamente baixa do sinal de GPS, aproximadamente -160dB [Misra and Enge 2006], esta interferência impede o acesso do VANT ao sinal de GPS genuíno. Isso impacta diretamente no CN0, calculado a partir da potência do sinal de GPS e da potência do ruído recebido.

3.2. IDS proposto

A solução proposta é um IDS para detectar ataques em andamento a fim de alertar o VANT. O veículo aéreo, por sua vez, pode tomar diversas ações, como notificar seu dono de um possível ataque em andamento, pousar a aeronave, ou mesmo usar instrumentos de navegação alternativos para o voo autônomo. O IDS é baseado em um modelo de aprendizado de máquina multiclasse pré-treinado a partir de um *data set* contendo dados de sinais de GPS representando operações normais de VANTs, bem como dados representando ataques de GPS *Spoofing* e *Jamming*. Após o treinamento, o IDS é utilizado para reconhecer potenciais ataques com base nas características dos sinais recebidos.

Devido à sua natureza multiclasse, o IDS não apenas identifica a ocorrência de um ataque, mas também categoriza seu tipo. Esta informação é crucial nos processos de tomada de decisão para elaborar estratégias para mitigar a ameaça. Além disso, ela contribui significativamente para futuras investigações forenses, auxiliando na adoção de contramedidas personalizadas para cada tipo de ataque.

Ao detectar um ataque de GPS *Jamming*, por exemplo, o VANT poderia suspender sua missão e escanear áreas onde o parâmetro CN0 se restaure a valores normais. Se isso não acontecer dentro de um limite de distância predefinido, ele poderia acionar temporariamente sensores de navegação secundários como câmera, giroscópio ou barômetro enquanto aguarda novas instruções. Diferentes respostas podem ser pré-programadas com base no nível de sofisticação do ataque de GPS *Spoofing*. Como o IDS pode produzir falsos positivos, o sistema pode incorporar limiares de tolerância, ativando medidas de proteção apenas após um certo número de detecções positivas durante o monitoramento. Este limiar mínimo pode diferir dependendo do nível de sofisticação, por exemplo, sendo inversamente proporcional ao nível de sofisticação do ataque, já que ataques mais sofisticados precisam ser mitigados mais rapidamente. O conjunto de estratégias para enfrentar ataques de GPS pode ser extenso e está além do escopo do presente trabalho. Não obstante, o IDS multiclasse aqui proposto fornece o necessário para que essas estratégias sejam possíveis, oferecendo uma solução mais eficiente em comparação com IDSs baseados em classificação binária.

4. Metodologia

Esta seção apresenta a metodologia empregada nos experimentos para avaliar o IDS proposto. A Subseção 4.1 detalha os dados utilizados, os tipos de ataque cobertos e sua aplicação no treinamento dos classificadores. A Seção 4.2 descreve os classificadores utilizados e está dividida em duas subseções: a Subseção 4.2.1 descreve os classificadores de forma isolada, enquanto a Subseção 4.2.2 detalha os classificadores de conjunto (*ensemble*) empregados. Por fim, a Subseção 4.3 apresenta as métricas utilizadas na avaliação.

4.1. Dados

Os dados utilizados na fase de treinamento e nos testes subsequentes foram derivados do *data set* introduzido em [Aissou et al. 2021]. O conjunto de dados original compreende dados em tempo real extraídos de um receptor GPS de 8 canais, resultando nas 13 características distintas já listadas na Subseção 3.1. Além disso, esse *data set* possui dados simulando os diferentes tipos de ataques de GPS *Spoofing* (Subseção 3.1), obtidos a partir da manipulação dos dados originais para representar a assinatura dos ataques de *Spoofing*.

O *data set* original [Aissou et al. 2021] empregado neste trabalho contém amostras para operação normal e as três categorias de ataques de GPS *Spoofing*. Esse *data set* foi ampliado para o presente estudo a fim de incluir amostras que simulam a assinatura do ataque de *Jamming* (Subseção 3.1). A simulação desse ataque baseou-se em [da Silva 2017], que analisou o impacto do GPS *Jamming*, revelando que níveis elevados de ruído levam a uma redução nos valores de CN0. Quando esses valores ficam abaixo de 25 dB-Hz, o ruído é suficiente para interromper a função de localização do VANT. Adicionalmente, os autores de [Misra and Enge 2006] demonstraram que obstáculos ambientais, como espaços fechados e edifícios, também degradam esse parâmetro. No entanto, considerando que VANTs operam predominantemente em espaços abertos e a altitudes significativas, é razoável supor que níveis altos de ruído sejam devido a ataques de *Jamming* e não a obstruções ambientais. Portanto, as amostras que simulam este ataque apresentam valores do parâmetro CN0 variando entre 21 e 24 dB-Hz. Esta faixa foi selecionada pois ataques de *Jamming* que reduzem o $C/N0$ para esses níveis já são suficientes para prejudicar a operação do receptor GPS [Bauernfeind et al. 2011]. Consequentemente, incluir valores abaixo de 21 dB/Hz é desnecessário.

Tabela 1. Ajuste de Parâmetros

Classificador	Configuração dos parâmetros	Melhores valores
GNB	var_smoothing = range(1e-9, 1e+9)	var_smoothing = 410103958,85331386
KNN	n_neighbors = range(5, 100), p = range(1, 3)	n_neighbors = 85, p = 1
DT	criterion = ['gini', 'entropy'] max_depth = range(1, 35)	criterion = 'entropy' max_depth = 19
NN	activation = ['identity', 'logistic', 'tanh', 'relu']	activation = 'tanh'
LDA	solver = ['svd', 'lsqr', 'eigen']	solver = 'svd'
LR	solver = ['lbfgs', 'liblinear', 'newton-cg', 'newton-cholesky', 'sag', 'saga']	solver = 'liblinear'
AB	n_estimators = range(10, 1000)	n_estimators = 1000
RF	criterion = ['gini', 'entropy', 'log_loss']	criterion = 'entropy'
GB	learning_rate = range(0,1, 1)	learning_rate = 0,1

Uma forma para reduzir o tempo necessário para o treinamento e a predição dos classificadores é eliminar características redundantes. Através do cálculo da Correlação de Spearman, estudos anteriores [Aissou et al. 2021, Talaei Khoei et al. 2022, Khoei et al. 2023, Gasimova et al. 2022, Khoei et al. 2022, Aissou et al. 2022, Talaei Khoei et al. 2023] identificaram dois pares de características altamente correlacionadas no conjunto de dados utilizado neste trabalho: DO e TCD (95% de correlação) e TOW e RX (94% de correlação). Essa correlação é esperada, pois DO e TCD são ambos baseados em medições de Efeito Doppler, enquanto TOW e RX são baseados em medições de tempo. Assim, removemos as características com o menor Ganho de Informação de cada par (TCD e RX), sem comprometer os resultados dos classificadores.

O conjunto de dados final consiste então em 397.646 amostras (73,46%) representando a operação normal, 36.438 amostras (6,73%) de ataques de *Spoofing simples*, 44.212 amostras (8,17%) de *Spoofing intermediário*, 31.995 amostras (5,91%) de *Spoofing sofisticado* e 31.022 amostras (5,73%) de *Jamming*. Este *data set* usado no presente artigo pode ser acessado em [Lemos 2023].

4.2. Classificadores

O *data set* foi particionado aleatoriamente como entrada para os classificadores, alocando 75% das amostras para treinamento e 25% para fins de teste. Todos os classificadores empregados são do scikit-learn, uma biblioteca Python de código aberto projetada para aprendizado de máquina [Pedregosa et al. 2011]. Inicialmente, foram testados os classificadores base, seguidos pelos classificadores de conjunto, que são métodos que utilizam múltiplos classificadores individuais para obter resultados aprimorados.

Para determinar os melhores parâmetros para cada classificador, foi utilizado o algoritmo de Otimização Bayesiana da biblioteca scikit-optimize. A Otimização Bayesiana é uma técnica eficiente para ajustar os parâmetros dos modelos de aprendizado de máquina, pois converge rapidamente para valores adequados. A Tabela 1 exibe os valores dos parâmetros usados com cada classificador e seus valores ótimos determinados pela Otimização Bayesiana.

4.2.1. Classificadores Base

Esta subseção introduz os classificadores base usados neste trabalho, selecionados com base em sua adaptabilidade a cenários de problemas de classificação binária e multiclasse. Os classificadores *Naive Bayes* são baseados no teorema de Bayes e assumem que todas as características são independentes. Este classificador assume independência entre as características. A versão *Gaussian Naive Bayes* (GNB) foi empregada, versão que assume uma distribuição Gaussiana das características dos dados.

O classificador *K-Nearest Neighbors* (KNN) é baseado na proximidade de uma nova amostra (amostra de teste) a exemplos conhecidos dentro do conjunto de treinamento. O valor de K determina o número de vizinhos considerados para a previsão, e várias medidas de distância podem ser utilizadas. Para determinar a classe de uma nova amostra, o KNN realiza uma votação majoritária entre os k-vizinhos mais próximos. Já o classificador *Decision Tree* (DT) constrói uma árvore de decisão dividindo os dados em subconjuntos menores com base nas características. Ele começa com todos os dados como um conjunto e os divide em dois subconjuntos usando uma característica específica escolhida. A característica selecionada é aquela que proporciona o maior ganho de informação ou a diminuição mais significativa na impureza do subconjunto. O processo de divisão continua recursivamente até que todos os subconjuntos sejam puros ou quase puros em relação a uma classe ou valor de saída.

A rede neural *Multilayer Perceptron* (MLP) usa o algoritmo de retropropagação para seu processo de treinamento. Ela compreende várias camadas conectadas de neurônios, com uma camada de saída que transforma valores da última camada oculta em um valor de saída. A Análise Discriminante Linear (LDA) gera uma matriz de dispersão entre classes e uma matriz de dispersão dentro da classe. Essas matrizes ajudam a identificar direções discriminantes que efetivamente separam as classes, e os dados são projetados ao longo dessas direções, resultando em um novo conjunto de características que distinguem otimamente entre as classes.

Por fim, a Regressão Logística (LR) modela a probabilidade de uma instância pertencer a uma determinada classe usando a função logística sigmoide, transformando a saída linear em valores de probabilidade que variam de 0 a 1. Subsequentemente, essa probabilidade é atribuída a uma classe particular com base em um limiar de decisão.

4.2.2. Classificadores de conjunto

Quatro classificadores (ou métodos) de conjunto foram empregados: *Boosting*, *Bagging*, *Voting* e *Stacking*. O método *Boosting* utiliza seus próprios classificadores internos, enquanto os restantes exigem a especificação dos classificadores base a serem utilizados. O restante desta subseção apresenta os classificadores de conjunto empregados.

Os classificadores do tipo *Boosting* combinam múltiplos classificadores base fracos para criar um classificador capaz de fazer previsões mais precisas. Por exemplo, o *Adaptive Boosting* (AdaBoost) combina classificadores fracos de forma ponderada. Ele começa ajustando um classificador aos dados originais e, em seguida, ajusta cópias adicionais do classificador ao mesmo conjunto de dados, dando mais peso às amostras classifi-

cadas incorretamente em cada iteração. Assim, o AdaBoost foca nas amostras mais desafiadoras de classificar, melhorando seu desempenho geral. Diferentemente, o classificador *Random Forests* cria múltiplas árvores de decisão independentes e combina suas previsões, determinando a decisão final por votação majoritária entre as árvores. Similar ao *Random Forests*, o *Gradient Boosting* também combina diversas árvores de decisão, mas difere na construção e metodologia de combinação. Ele treina sequencialmente múltiplas árvores, ajustando cada árvore para corrigir erros cometidos por suas predecessoras. Esta técnica progressiva visa melhorar o desempenho do classificador, aproveitando os pontos fortes de múltiplas árvores de decisão fracas.

Para os métodos *Bagging*, *Voting* e *Stacking*, foram selecionados os 5 classificadores que tiveram melhor desempenho nos testes anteriores: *Decision Tree*, KNN, *Neural Network*, *Random Forests* e *Gradient Boosting*. O método *Bagging* (*Bootstrap Aggregating*) treina múltiplas cópias do mesmo classificador em subconjuntos aleatórios e suas previsões são combinadas por votação majoritária. O método *Voting* combina classificadores conceitualmente diferentes e utiliza votação majoritária ou médias das probabilidades previstas para prever os rótulos dos resultados. O método *Stacking* treina um conjunto de classificadores em um conjunto de dados, e as previsões desses classificadores base são usadas para treinar um classificador final, também conhecido como meta-classificador. O *Gradient Boost* foi escolhido como meta-classificador para o método Stacking devido ao seu desempenho.

4.3. Métricas

As métricas empregadas para avaliar os classificadores neste trabalho foram Accuracy, Matriz de Confusão, Precision, Recall e F1-Score. A Accuracy é a razão entre as previsões corretas e o número total de amostras, sendo particularmente útil quando os erros na predição de todos os rótulos possuem o mesmo peso [Burkov 2019]. Essa métrica serviu como base para comparar o desempenho dos classificadores durante a avaliação inicial, uma vez que o objetivo é identificar um classificador com os melhores resultados gerais. A Equação (1) ilustra como essa pontuação é calculada, onde $\hat{y}_i$ representa o valor previsto da i-ésima amostra, y_i é o valor verdadeiro correspondente, e $1(x)$ é a Função Indicadora.

$$\texttt{Accuracy}(y, \hat{y}) = \frac{1}{n_{\text{amostras}}} \sum_{i=0}^{n_{\text{amostras}}-1} 1(\hat{y}_i = y_i) \tag{1}$$

A matriz de confusão é uma tabela que demonstra a distribuição das previsões do classificador entre os rótulos do conjunto de dados. Nessa matriz, o eixo horizontal denota os rótulos previstos pelo classificador, enquanto o eixo vertical corresponde aos rótulos reais [Burkov 2019]. Essa métrica foi empregada para avaliar o comportamento do classificador com a maior Accuracy em termos da distribuição dos erros de previsão entre os rótulos.

Depois da avaliação inicial, as métricas de Precision, Recall e F1-Score também foram empregadas para fornecer uma avaliação mais profunda do classificador que apresentou a maior Accuracy. A Precision é a razão entre as amostras positivas previstas corretamente em relação ao total de amostras previstas como positivas. Um valor de Precision mais alto implica uma capacidade mais forte do classificador de reduzir o número

Tabela 2. Valores de Accuracy (em %) obtidos pelos classificadores base.

Classificador	Multiclasse	Binário	Classificador	Multiclasse	Binário
GNB	73,38	73,38	LR	73,05	72,74
KNN	86,44	88,09	AB	63,25	94,75
DT	94,25	95,17	RF	90,82	92,38
NN	93,06	93,91	GB	93,51	95,30
LDA	79,22	78,81	—	—	—

Tabela 3. Valores de Accuracy (em %) obtidos pelos classificadores de conjunto.

Classificador de conjunto	Classificadores base	Multiclasse	Binário
Bagging	KNN	86,62	88,18
	DT	93,19	94,06
	NN	73,39	76,81
	RF	90,71	92,36
	GB	93,13	95,30
Hard Voting	KNN, DT, NN, RF, GB	92,97	94,40
Soft Voting	KNN, DT, NN, RF, GB	92,51	93,73
Stacking	KNN, DT, NN, RF, GB	98,08	98,48

de amostras falsamente positivas. O Recall representa a razão entre as instâncias positivas previstas corretamente em relação ao total de instâncias que são genuinamente positivas. Valores de Recall mais altos implicam uma melhor capacidade do classificador de identificar todas as instâncias positivas. O F1-Score é a média harmônica entre Precision e Recall.

5. Resultados numéricos

Esta seção apresenta a avaliação numérica do IDS proposto neste artigo. A Tabela 2 exibe os resultados de Accuracy obtidos a partir dos Classificadores Base em suas versões de classe binária e multiclasse. Os classificadores com melhor desempenho foram KNN, *Decision Tree*, *Neural Network*, *Random Forests* e *Gradient Boost*. Posteriormente, esses cinco classificadores foram empregados em métodos de conjunto, como *Bagging*, *Hard Voting*, *Soft Voting* e *Stacking*, na tentativa de melhorar os resultados, cujos resultados são exibidos na Tabela 3. Conforme descrito na Subseção 4.2.2, o método *Bagging* é aplicado individualmente a cada classificador, enquanto os outros métodos de conjunto utilizam os classificadores coletivamente. O método Stacking obteve a maior Accuracy, alcançando 98,08% de Accuracy com a versão multiclasse e 98,48% com a versão de classe binária.

A Figura 1 compara os maiores valores de Accuracy obtidos por todos os classificadores em ambas as versões multiclasse e binária. Para os classificadores *Neural Network*, *Decision Tree*, *Random Forests* e *Gradient Boost*, a versão binária obteve melhores resultados. Isso ocorre para esses classificares pois, apesar do suporte nativo à classificação multiclasse, prever múltiplas classes é uma tarefa mais complexa do que prever apenas duas, gerando maior confusão entre elas.

O KNN apresenta melhor desempenho na versão de classe binária devido ao número reduzido de opções na votação dos vizinhos, minimizando assim a confusão de rótulos em comparação com a versão multiclasse. Entre os classificadores, apenas a LDA e a LR mostraram menor Accuracy na versão de classe binária, enquanto o GNB teve desempenho semelhante em ambas as versões. Isso mostra que, para esses classificadores,

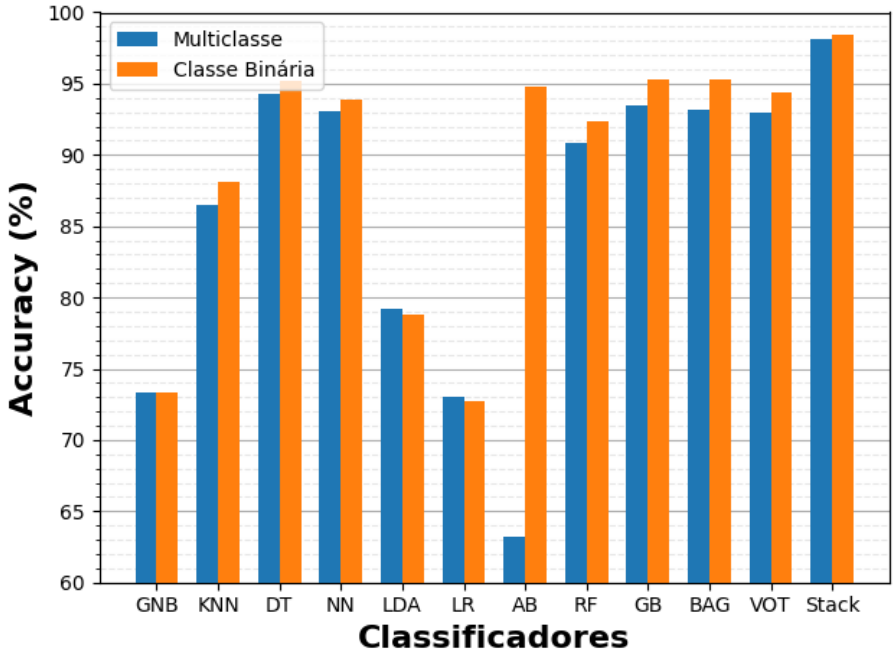

Figura 1. Accuracy obtida pelos classificadores avaliados.

a complexidade maior dos dados com múltiplas classes não foi um problema.

O AdaBoost mostrou uma diferença de desempenho mais significativa entre as versões em comparação com todos os outros classificadores, o que provavelmente se deve à dificuldade de seus aprendizes fracos em prever resultados corretos ao lidar com inúmeras classes. No geral, as discrepâncias nos resultados destacam que certos classificadores se saem melhor do que outros para os mesmos dados. O Teorema de Sem Almoço Grátis (*No Free Lunch Theorem*) [Wolpert 1996] afirma que é impossível determinar o melhor classificador para um conjunto de dados antecipadamente, enfatizando a importância de testar vários classificadores a fim de obter resultados ótimos [Géron 2023].

Como o método de Stacking apresentou a maior Accuracy, ele passou por uma avaliação usando métricas adicionais para detalhar o IDS proposto. A Figura 2 exibe a matriz de confusão com os rótulos previstos em relação aos rótulos reais no conjunto de dados de teste multiclasse. O rótulo 0 denota operação normal, enquanto os rótulos 1, 2 e 3 representam os ataques de *Spoofing* simples, intermediário e sofisticado, respectivamente. O rótulo 4 indica um ataque de *Jamming*. O *Stacking* resultou em 948 instâncias de falsos negativos, onde um ataque estava presente, mas foi classificado erroneamente como operação normal; esse número é a soma de três tipos de classificação incorreta, relativo a 359 instâncias com o rótulo 1, 496 com o rótulo 2 e 93 com o rótulo 3. Isso constitui 2,6% das 36.017 amostras de ataque totais no teste, e 1.000 instâncias de falsos positivos, onde nenhum ataque estava presente, mas o classificador indicou o contrário, representando 1% das 99.310 amostras de operação normal totais no teste.

Além disso, houve 338 amostras de ataques de *Spoofing* simples classificadas erroneamente como *Spoofing* intermediário, 314 amostras de ataques de *Spoofing* intermediário classificadas erroneamente como *Spoofing* simples e uma única amostra de *Spoofing* simples classificada erroneamente como *Spoofing* sofisticado. O ataque de *Jamming*

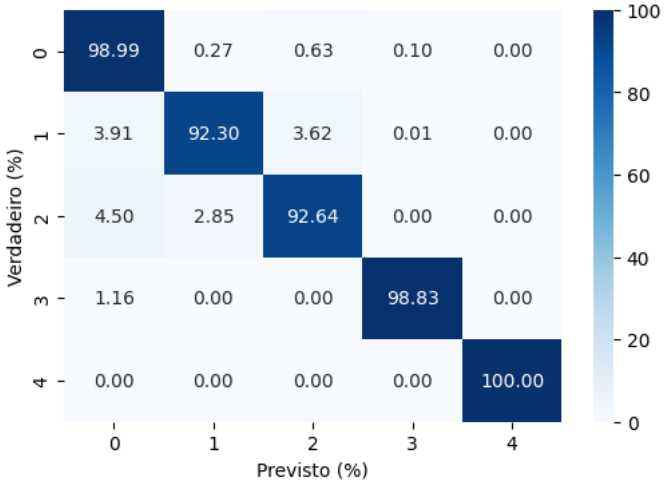

Figura 2. Matriz de confusão para o *Stacking* multiclasse.

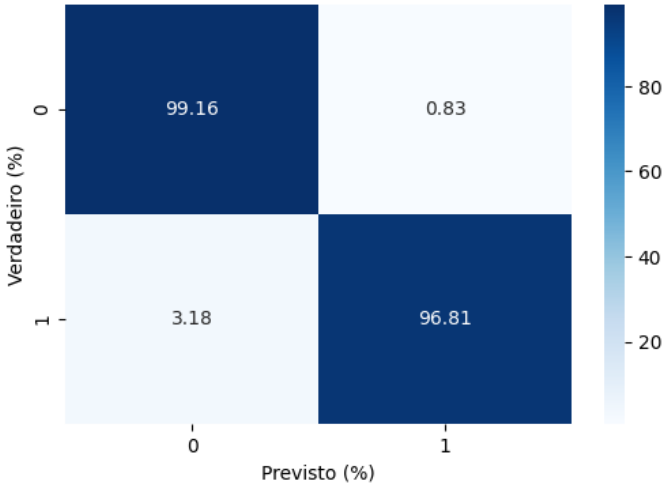

Figura 3. Matriz de confusão para o *Stacking* binário.

permaneceu distinto e não foi classificado erroneamente. Essa distinção decorre dos valores consistentes de CN0 abaixo de 25 dB-Hz, uma característica que os classificadores podem aprender facilmente.

A Figura 3 ilustra a matriz de confusão gerada pelo método de *Stacking* quando aplicado à classe binária. Ocorreram 1148 falsos negativos, correspondendo a 3,1% do total de 36.018 entradas de ataque, e 833 falsos positivos, representando 0,8% do total de 99.310 amostras de operação normal no conjunto de testes.

O Stacking multiclasse exibe uma Accuracy ligeiramente menor em comparação com sua versão de classe binária. No entanto, uma inspeção mais detalhada de suas respectivas matrizes de confusão revela que a versão multiclasse produz menos falsos negativos, que representam os erros mais críticos em um IDS, pois indicam ataques não detectados. A versão de classe binária teve 1148 casos de falsos positivos, enquanto a multiclasse teve 948, implicando que, em um cenário real, o IDS multiclasse deixaria passar mais 200 ataques. A Accuracy reduzida da versão multiclasse surge principalmente

Tabela 4. Comparação do resultados do Stacking para classificação multiclasse e binária.

Versão	Accuracy (%)	Falsos Negativo (%)
Multiclasse	98,08	2,6
Binária	98,48	3,1

Tabela 5. Precision, Recall, e F1-score para *Stacking* multiclasse e binário.

Multiclasse			
Classe	**Precision (%)**	**Recall (%)**	**F1-Score (%)**
Operação normal	99	99	99
Spoofing simples	94	92	93
Spoofing intermediário	91	93	92
Spoofing sofisticado	99	99	99
Jamming	100	100	100
Binário			
Classe	**Precision (%)**	**Recall (%)**	**F1-Score (%)**
Operação normal	99	99	99
Ataque	98	97	97

dos desafios em distinguir entre tipos de ataques, o que, dentro do contexto de um IDS, é muito menos grave do que um falso negativo. A Tabela 4 detalha essa comparação.

Quando o IDS identifica com precisão o tipo de ataque em andamento, o VANT pode tomar medidas necessárias para mitigar o ataque específico sendo sofrido. Um falso negativo indica que a aeronave está vulnerável a um ataque; por exemplo, em um ataque de *Jamming*, o VANT perde suas informações de localização, potencialmente resultando em colisões ou navegação em áreas restritas, o que levaria à perda da aeronave e prejuízos financeiros. Ataques de *Spoofing* tem um risco adicional de levar o VANT fisicamente aos atacantes, o que causaria danos mais significantes.

A Tabela 5 apresenta as métricas de Precision, Recall e F1-score para cada classe do classificador *Stacking* multiclasse. O melhor o desempenho foi obtido para as classes Operação normal, *Spoofing* sofisticado e *Jamming*, enquanto os resultados menos favoráveis foram obtidos para *Spoofing* simples e intermediário. Uma análise das métricas sugere que o classificador enfrenta desafios para identificar e distinguir com precisão entre os ataques simples e intermediários, o que deteriora o desempenho do classificador para ataques intermediários em comparação com os sofisticados.

A Tabela 5 também apresenta os resultados das métricas para a versão de classe binária do classificador *Stacking*. Os resultados permaneceram consistentes para a classe Operação normal, mantendo-se estáveis em 99% para todas as métricas. No entanto, para a classe de Ataque, a Precision atingiu 98%, enquanto o Recall e o F1-Score alcançaram 97%. Observa-se que o valor de Recall é menor que o de Accuracy na classe de Ataque, indicando que este classificador se destaca mais em minimizar falsos positivos do que em identificar com precisão todas as amostras positivas de ataque. Essencialmente, ele prioriza a redução de falsos positivos em relação a falsos negativos. Como observado anteriormente, para um IDS, evitar falsos negativos é primordial; portanto, um valor de Recall mais alto seria mais benéfico do que uma Accuracy mais alta.

Em resumo, este avaliação numérica visou identificar o classificador mais eficaz para múltiplas classes de ataques GPS, visando utilizá-lo em um IDS confiável para sinais GPS recebidos por VANTs. Os cinco principais classificadores base em termos de alta precisão foram empregados dentro de um método de conjunto *Stacking* para alcançar o melhor resultado, um classificador multiclasse com 98,08% de Accuracy e apenas 2,6% de falsos negativos. Este método não apenas detecta ataques, mas também diagnostica com precisão os tipos específicos de ataques, facilitando a implementação de contramedidas após a identificação dos ataques, fornecendo registros para futuras investigações forenses. Além disso, a versão multiclasse demonstrou notavelmente uma quantidade menor de falsos negativos em comparação com sua versão de classe binária. Isso é uma vantagem significativa pois, para um IDS, o erro de falso negativo é o mais grave, visto que isso implica em um ataque não detectado.

6. Conclusões

Este artigo propôs um IDS para detectar ataques de GPS *Jamming* e três variações de ataques de GPS *Spoofing* direcionados a VANTs. O IDS foi criado a partir de testes com classificadores a fim de se obter o melhor desempenho. Trabalhos relacionados da literatura se apoiaram em classificadores binários e, diferentemente, a presente pesquisa empregou testes com classificadores binários e multiclasse. A versão multiclasse do método *Stack* alcançou uma Accuracy de 98,08% com apenas 2,6% de falsos negativos. Os resultados obtidos mostraram que a abordagem multiclasse tem a vantagem de reduzir a ocorrência de falsos negativos, que é o erro mais crítico de um IDS em um cenário real. Além disso, a abordagem multiclasse permite ao IDS identificar o tipo de ataque em andamento, permitindo que o VANT implemente contramedidas mais eficientes contra o ataque sofrido, dando mais suporte para potenciais investigações forenses.

Ataques de GPS são viáveis, mas montar um ambiente real para replicá-los é uma tarefa que exige diversos equipamentos e também o devido cuidado para não danificá-los. Sendo assim, como trabalhos futuros, sugere-se avaliar o IDS com VANTs reais. Além disso, continuar a investigação com modelos de aprendizado de máquina baseados em técnicas alternativas pode potencialmente melhorar o desempenho em comparação com os resultados obtidos neste estudo.

Agradecimentos

Este estudo foi parcialmente financiado pela Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), processo nº 2015/24494-8.

Referências

Aissou, G., Benouadah, S., El Alami, H., and Kaabouch, N. (2022). Instance-based supervised machine learning models for detecting GPS spoofing attacks on UAS. In *2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC)*, pages 0208–0214.

Aissou, G., Slimane, H. O., Benouadah, S., and Kaabouch, N. (2021). Tree-based supervised machine learning models for detecting GPS spoofing attacks on UAS. In *2021 IEEE 12th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*, pages 0649–0653.

Bauernfeind, R., Kraus, T., Dötterböck, D., Eissfeller, B., Loehnert, E., and Wittmann, E. (2011). Car jammers: Interference analysis. *GPS World*, 22:28–35.

Burkov, A. (2019). *The Hundred-Page Machine Learning Book*. Andriy Burkov, Quebec.

da Silva, D. A. M. (2017). Gps jamming and spoofing using software defined radio. Master's thesis, University Institute of Lisbon.

da Silva, R. A. C., da Fonseca, N. L. S., and Boutaba, R. (2021). Evaluation of the employment of UAVs as fog nodes. *IEEE Wireless Communications*, 28(5):20–27.

Davidovich, B., Nassi, B., and Elovici, Y. (2022). Towards the detection of GPS spoofing attacks against drones by analyzing camera's video stream. *Sensors*, 22(7).

Derhab, A., Cheikhrouhou, O., Allouch, A., Koubaa, A., Qureshi, B., Ferrag, M. A., Maglaras, L., and Khan, F. A. (2023). Internet of drones security: Taxonomies, open issues, and future directions. *Vehicular Communications*, 39:100552.

Gasimova, A., Khoei, T. T., and Kaabouch, N. (2022). A comparative analysis of the ensemble models for detecting GPS spoofing attacks on UAVs. In *2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC)*, pages 0310–0315.

Géron, A. (2023). *Hands-On Machine Learning with ScikitLearn, Keras, and TensorFlow*. O'Reilly Media, Sebastopol.

Haider, Z. and Khalid, S. (2016). Survey on effective GPS spoofing countermeasures. In *2016 Sixth International Conference on Innovative Computing Technology (INTECH)*, pages 573–577.

Khan, A., Gupta, S., and Gupta, S. K. (2022). Emerging UAV technology for disaster detection, mitigation, response, and preparedness. *Journal of Field Robotics*, 39(6):905–955.

Khoei, T. T., Aissou, G., Al Shamaileh, K., Devabhaktuni, V. K., and Kaabouch, N. (2023). Supervised deep learning models for detecting GPS spoofing attacks on unmanned aerial vehicles. In *2023 IEEE International Conference on Electro Information Technology (eIT)*, pages 340–346.

Khoei, T. T., Gasimova, A., Ahajjam, M. A., Shamaileh, K. A., Devabhaktuni, V., and Kaabouch, N. (2022). A comparative analysis of supervised and unsupervised models for detecting GPS spoofing attack on UAVs. In *2022 IEEE International Conference on Electro Information Technology (eIT)*, pages 279–284.

Lemos, G. (2023). Original dataset. `https://docs.google.com/spreadsheets/d/1srN7w4dO2NU8XKeyeLlwjNZbaz9Sx4Wj/edit?usp=sharing&ouid=107994669384648426370&rtpof=true&sd=true`.

Misra, P. and Enge, P. (2006). *Global Position System Signals, Measurement and Performance*. Ganga Jamuna Press, Lincoln.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

Ranyal, E. and Jain, K. (2021). Unmanned aerial vehicle's vulnerability to GPS spoofing a review. *Indian Society of Remote Sensing*, 49:585–591.

Talaei Khoei, T., Ismail, S., and Kaabouch, N. (2022). Dynamic selection techniques for detecting GPS spoofing attacks on UAVs. *Sensors*, 22(2).

Talaei Khoei, T., Ismail, S., Shamaileh, K. A., Devabhaktuni, V. K., and Kaabouch, N. (2023). Impact of dataset and model parameters on machine learning performance for the detection of GPS spoofing attacks on unmanned aerial vehicles. *Applied Sciences*, 13(1).

Wei, X., Wang, Y., and Sun, C. (2022). Perdet: Machine-learning-based UAV GPS spoofing detection using perception data. *Remote Sensing*, 14(19).

Whelan, J., Almehmadi, A., and El-Khatib, K. (2022). Artificial intelligence for intrusion detection systems in unmanned aerial vehicles. *Computers and Electrical Engineering*, 99:107784.

Wolpert, D. H. (1996). The Lack of A Priori Distinctions Between Learning Algorithms. *Neural Computation*, 8(7):1341–1390.