Mapeamento Sistemático de Sistemas de Detecção de Intrusão Baseados em Aprendizado de Máquina para Redes Wi-Fi

Gabriele Pereira da Silva [1], Julia Milioranza Gomes[1], Luiz Eduardo Reis Martins[1], Karen da Silva Figueiredo Medeiros Ribeiro[1]

[1]Instituto de Computação – Universidade Federal do Mato Grosso (UFMT)
Cuiabá – MT – Brasil

julia.gomes1,gabriele.silva1,luiz.martins@sou.ufmt.br,

karen@ic.ufmt.br

Abstract. *This paper presents the results of a systematic mapping of literature about applied research on machine learning techniques for intrusion detection in Wi-Fi networks. Our goal is to identify the main algorithms and datasets used and which performance metrics were adopted. Studies published since 2020 were analyzed, focusing on anomaly detection methods and machine learning-based intrusion detection systems (IDS). The results of this paper could assist researchers and information security professionals in developing more effective solutions and selecting appropriate approaches to enhance Wi-Fi network protection.*

Resumo. *Este artigo apresenta os resultados de um mapeamento sistemático de pesquisas que aplicam técnicas de aprendizado de máquina para a detecção de intrusão em redes Wi-Fi. O mapeamento tem como objetivo identificar os principais algoritmos utilizados, as bases de dados mais empregadas para treinar os modelos e as métricas de desempenho adotadas. Foram analisados estudos publicados a partir de 2020, com foco em métodos de detecção de anomalias e sistemas de detecção de intrusão (IDS) baseados em aprendizado de máquina. Os resultados deste estudo podem auxiliar pessoas pesquisadoras e profissionais da área de segurança da informação no desenvolvimento de soluções mais eficazes e na escolha de abordagens adequadas para aprimorar a proteção de redes Wi-Fi.*

1. Introdução

Os Sistemas de Detecção de Intrusão (IDS) são de grande importância em uma sociedade cada vez mais conectada, uma vez que servem para monitorar e identificar atividades anômalas em rede, procurando manter a integridade, confidencialidade e disponibilidade dos dados [Liu and Lang 2019]. Com isso, inúmeros modelos e métodos de IDS surgiram para ajudar na segurança cibernética.

Este artigo, desenvolvido como trabalho da disciplina de Metodologia Científica do curso de Ciência da Computação, realiza um mapeamento sistemático a fim de identificar os modelos de Aprendizado de Máquina mais frequentes em IDS's aplicados em redes Wi-Fi, seguindo a metodologia PICOC [Kitchenham 2007]. Essa metodologia foi escolhida por ser uma metodologia de revisão sistemática da literatura para a área da

ciência da computação. O objetivo com a realização deste mapeamento é fornecer uma visão abrangente dos diferentes algoritmos, banco de dados e métricas de desempenho utilizados nesta área, destacando suas principais características e desafios reportados nos artigos.

O trabalho está organizado da seguinte forma: na Seção 2 descreve-se a metodologia aplicada na pesquisa. A seguir, na Seção 3 apresentam-se os resultados encontrados, com uma discussão sobre as principais tendências observadas na literatura. Finalmente, na Seção 4 conclui-se com as limitações e sugestões para trabalhos futuros.

2. Metodologia

O mapeamento sistemático conduzido neste estudo seguiu as recomendações da metodologia PICOC [Kitchenham 2007] e foi dividido em três fases principais: planejamento, condução e extração de resultados.

Para garantir a organização e eficiência do processo, foi utilizada a ferramenta **Parsifal**[1], que auxiliou na gestão das referências, na aplicação dos critérios de seleção e análise dos estudos. O mapeamento realizado teve como objetivo responder às seguintes questões de pesquisa (QP):

- QP01: Quais são os principais algoritmos de aprendizado de máquina utilizados na detecção de intrusão em redes Wi-Fi?
- QP02: Quais conjuntos de dados são mais frequentemente empregados para treinar e avaliar modelos de detecção de intrusão em redes Wi-Fi?
- QP03: Quais métricas de desempenho são mais utilizadas para avaliar a eficácia dos modelos?

2.1. Planejamento

Na etapa de planejamento, foi estabelecido o protocolo de revisão sistemática para definir as diretrizes a serem seguidas ao longo do estudo. O protocolo incluiu as questões de pesquisa mencionadas anteriormente, a string de busca `("Artificial Intelligence"OR "Machine Learning") AND ("Anomaly Detection"OR "Intrusion Detection System"OR "IDS") AND ("NSL-KDD"OR "KDD99")`.

A partir da string de busca, foram definidas as bases de dados científicas que seriam utilizadas no mapeamento: IEEE Xplore e Science Direct. A escolha dessas bases foi baseada em abrangência e relevância na área de Ciência da Computação, além de serem bases de acesso aberto dentro da instituição executora da pesquisa.

Os critérios de inclusão (CI) definidos foram:

- CI01: Artigos publicados a partir do ano de 2021, para garantir a relevância dos métodos atuais.
- CI02: Estudos primários que aplicam técnicas de aprendizado de máquina para detecção de intrusão em redes Wi-Fi
- CI03: Estudos que apresentam avaliação quantitativa dos modelos (ex.: precisão, recall, F1-score)

[1]Parsifal `https://parsif.al/`

- CI04: Trabalhos que utilizam conjuntos de dados públicos, tais como NSL-KDD, AWID, KDD'99, entre outros.

E os critérios de exclusão (CE) também foram estabelecidos, a saber:

- CE01: Estudos indisponíveis em inglês ou português.
- CE02: Estudos que não utilizam aprendizado de máquina para detecção de intrusão.
- CE03: Estudos secundários, como revisões sistemáticas, surveys ou meta-análises
- CE04: Trabalhos com acesso restrito ou sem texto completo disponível.
- CE05: Trabalhos que analisam apenas redes cabeadas (wired networks), sem foco em Wi-Fi.
- CE06: Estudos não quantitativos.

2.2. Classificação e Seleção dos Artigos

Conforme ilustrado na Figura 1, após a aplicação da string de busca, foram encontrados 360 artigos publicados entre 2021 e 2024, provenientes nas bases IEEE Digital Library e Science@Direct. Em seguida, identificaram-se e removeram-se os artigos duplicados, resultando em 334 artigos.

Após essa etapa, realizou-se a leitura dos títulos e resumos dos artigos, e com base nos critérios de inclusão (CI) e exclusão (CE), foram eliminados aqueles que não atendiam aos requisitos, restando 68 artigos.

Além disso, para avaliar a aderência dos artigos aos critérios de inclusão, foi empregada uma planilha contendo perguntas específicas, elaboradas com base nos critérios estabelecidos. Essas perguntas, embora não sejam as mesmas que as questões de pesquisa (QP), estão diretamente relacionadas aos critérios de inclusão (CI). Cada artigo foi classificado de acordo com sua correspondência a essas perguntas, utilizando a seguinte escala:

- 0: Quando o artigo não atendia ao critério.
- 0,5: Quando o artigo era neutro ou apresentava informações parciais relacionadas ao critério.
- 1: Quando o artigo atendia plenamente ao critério.

Após a classificação por pares, foram selecionados para análise apenas os artigos que obtiveram uma pontuação total superior a 0,8 em relação aos critérios de inclusão e que não zeraram a primeira pergunta: *"O estudo apresenta um objetivo claro relacionado à detecção de intrusão em redes Wi-Fi usando aprendizado de máquina?"*. Essa abordagem permitiu filtrar os estudos mais alinhados aos objetivos do mapeamento. No total, 9 artigos atenderam a esse critério e foram incluídos na análise final.

3. Discussão e Resultados

Nesta seção, serão apresentados e discutidos os principais resultados obtidos a partir da análise dos dados coletados no mapeamento. O objetivo é compreender as tendências e os padrões identificados, bem como suas implicações para o campo de estudo em questão.

Em relação à primeira pergunta de pesquisa,*"Quais são os principais algoritmos de aprendizado de máquina utilizados na detecção de intrusão em redes Wi-Fi"*, os dados revelaram os seguintes resultados:

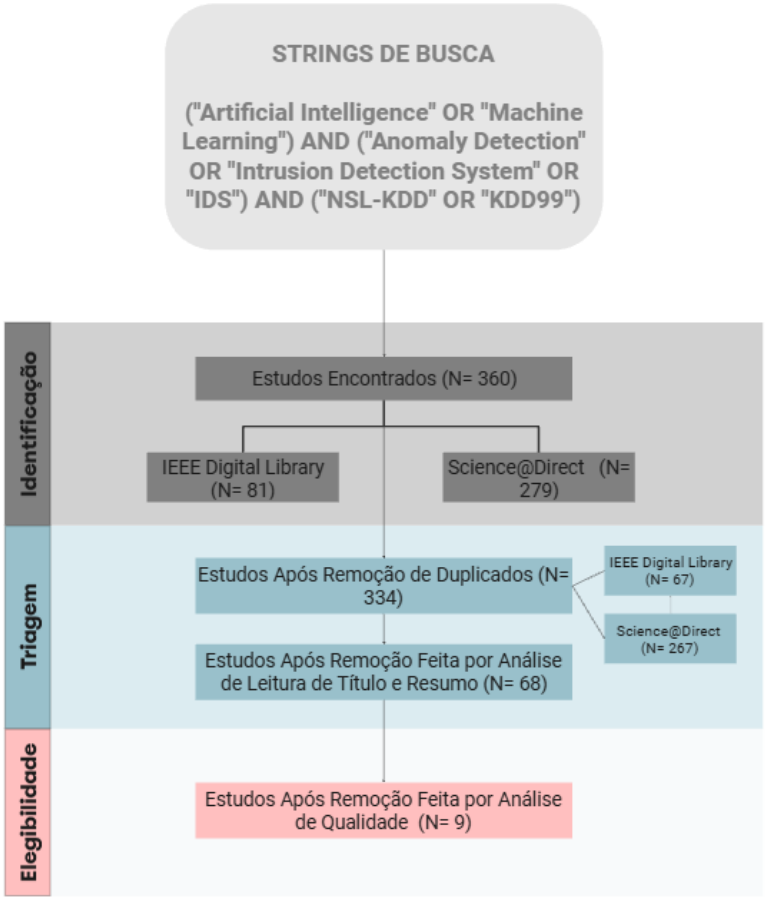

Figura 1. Fluxograma de seleção dos artigos.

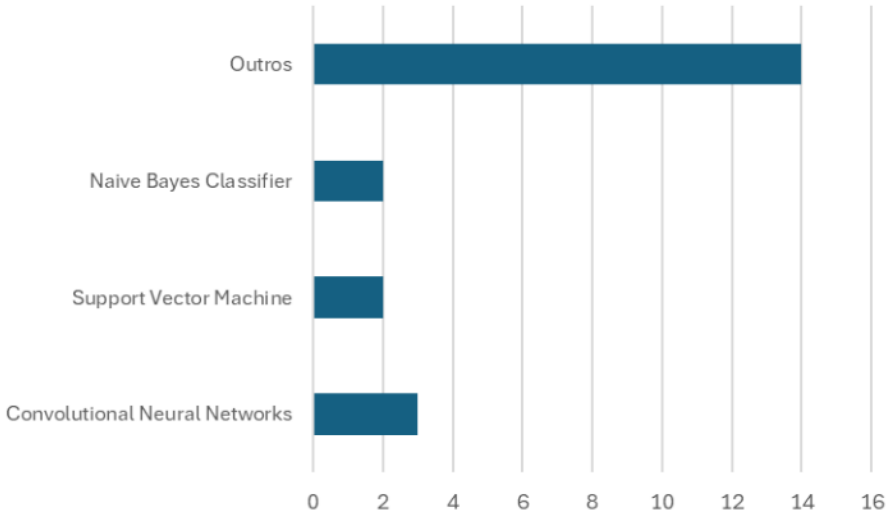

Figura 2. Frequência dos algoritmos utilizados nas pesquisas.

Conforme a Figura 2, o levantamento identificou 21 algoritmos utilizados em modelos de detecção de intrusões. O algoritmo mais utilizado foi a CNN (Convolutional Neural Network), presente em 3 (14,29%) modelos do total. A CNN é por sua vez a

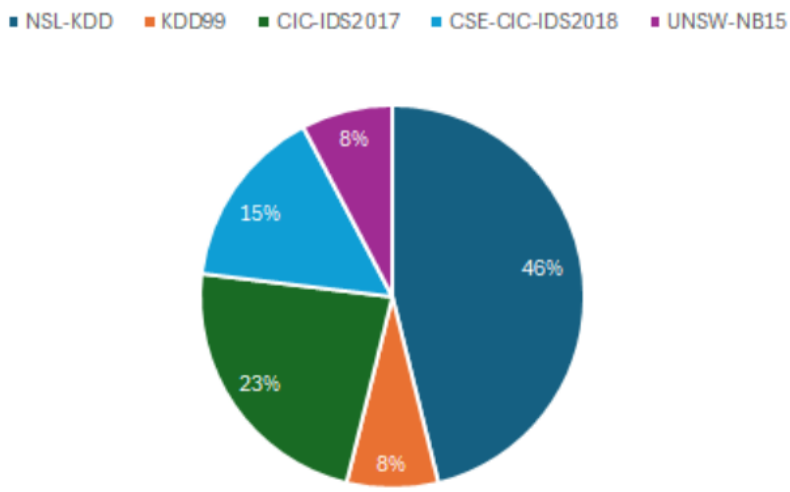

Figura 3. Utilização das bases de dados nas pesquisas.

mais escolhida como demonstrado na Tabela 1, por sua capacidade de identificar padrões complexos em dados, como tráfego de rede, de forma eficiente. Ela usa camadas convolucionais para extrair características importantes e pode ser combinada com mecanismos de atenção, como em [ALR].

Em seguida, os algoritmos Naive Bayes e SVM (Support Vector Machine) foram os segundos mais utilizados, cada um aparecendo em 2 (9,52%) trabalhos. Segundo [Wisanwanichthan and Thammawichai 2021], o Naive Bayes foi aplicado para detectar ataques comuns, enquanto o SVM foi usado para identificar ataques complexos, alcançando uma acurácia de 88,97%.

Os algoritmos CNN, Naive Bayes e SVM são frequentemente usados em IDS para redes Wi-Fi por algumas razões relacionadas às características dos ataques, à natureza dos dados de tráfego de rede e às capacidades desses modelos, a saber:

- **CNN:** As CNNs são excelentes para identificar padrões complexos e características espaciais nos dados. Além disso, ataques em redes Wi-Fi podem ter assinaturas sutis que exigem um modelo capaz de capturar relações não triviais entre os dados, algo que CNNs fazem bem. Quando treinadas com grandes conjuntos de dados, CNNs podem aprender a diferenciar tráfego normal de tráfego malicioso sem precisar de extração manual de características [Alrayes et al. 2024].
- **Naive Bayes:** O modelo Naive Bayes é leve computacionalmente e eficiente para classificações rápidas. Este modelo possui boa generalização para tráfego categorizado, uma vez que muitos ataques podem ser descritos por padrões estatísticos distintos, a abordagem probabilística do Naive Bayes pode ser útil na detecção de anomalias baseadas em distribuições de frequência. O Naive Bayes também pode oferecer uma boa interpretabilidade, o que é útil para análise forense de intrusões [Yang 2018].
- **SVM:** O SVM é útil quando o conjunto de dados de treinamento não é muito grande, algo comum em detecção de intrusão quando não há muitos exemplos rotulados de ataques. Este modelo possui capacidade de lidar com dados de alta dimensionalidade e o SVM pode separar classes eficazmente em

espaços de alta dimensão. Por fim, o SVM usa margens máximas para encontrar a melhor fronteira de decisão entre tráfego normal e intrusivo, sendo particularmente eficaz quando há distinções bem definidas entre essas classes [Wisanwanichthan and Thammawichai 2021].

Os 14 algoritmos restantes foram encontrados em apenas 1 modelo cada, representando 66,67% do total. Esses algoritmos incluem: Biogeography-Based Optimization Algorithm, Transformer, Regressão Logística (LR), Árvore de Decisão, Random Forest, Perceptron Multicamadas, Algoritmo Genético, Temporal Convolutional Network, Long Short-Term Memory, ANU-Net, U-Net, Black Hole Algorithm e Firefly Algorithm. A Tabela 1 apresenta a listagem dos algoritmos utilizados em cada um dos trabalhos analisados.

Como demonstrado na Figura 3, o conjunto de dados NSL-KDD foi o mais utilizado, representando 46,2% do total (6 ocorrências) [Jiang et al. 2024], [ALR], [Assy et al. 2023], [Babu and Rao 2023], [Yong and Gao 2023] . Em seguida, os conjuntos CIC-IDS2017 e CSE-CIC-IDS2018 apresentaram 23,1% de utilização (3 ocorrências) [Babu and Rao 2023], [Jiang et al. 2024], [Çetin 2022] e 15,4% (2 ocorrências) [Babu and Rao 2023], [Ashiku and Dagli 2021] de uso, respectivamente. Por fim, tanto o conjunto KDD99 [Ashiku and Dagli 2021] quanto o UNSW-NB15 [Mezina et al. 2021] apresentaram 7,7% (1 ocorrência) de uso.

Essa distribuição indica uma preferência significativa pelo conjunto NSL-KDD, possivelmente por seus avanços em comparação ao KDD99, que, embora amplamente adotado, apresentava deficiências relevantes, como um excesso de dados redundantes e repetidos. O NSL-KDD superou essas limitações ao eliminar registros duplicados, resultando em um conjunto de dados mais equilibrado e representativo. Além disso, promoveu uma distribuição mais adequada entre tráfego normal e ataques, assegurando que os diferentes níveis de complexidade das amostras maliciosas fossem distribuídos de forma mais homogênea entre os conjuntos de treinamento e teste. Esses aprimoramentos tornam o NSL-KDD uma base mais confiável para a avaliação de métodos de detecção de anomalias, possibilitando comparações mais precisas entre diversas abordagens de machine learning.conferindo maior relevância no contexto das pesquisas analisadas. [Wisanwanichthan and Thammawichai 2021]

As bases de dados NSL-KDD, CIC-IDS2017 e CSE-CIC-IDS2018 são amplamente utilizadas para treinar e avaliar modelos de Aprendizado de Máquina aplicados à detecção de intrusões em redes Wi-Fi porque possuem características, tais como: cenários realistas de ataques e conjunto balanceado de tráfego normal e tráfego malicioso, que as tornam adequadas para esse cenário:

- **NSL-KDD:** Contém um conjunto balanceado de ataques conhecidos e tráfego legítimo, permitindo que os modelos aprendam a distinguir os padrões com maior precisão. Classifica ataques em quatro categorias principais: DoS (Denial of Service), Probe, U2R (User to Root) e R2L (Remote to Local), cobrindo uma ampla gama de ameaças reais. É relativamente pequena e fácil de processar, sendo ideal para experimentos rápidos e comparação de algoritmos [Alrayes et al. 2024].
- **CIC-IDS2017:** Inclui dados de fluxo de rede extraídos de pacotes, fornecendo um contexto mais rico para a detecção de anomalias. Contém registros completos de tráfego em protocolos essenciais de redes Wi-Fi, como HTTP, HTTPS, FTP,

SSH e DNS. Permite a criação de modelos mais robustos, pois cobre variações de ataques modernos, algo essencial para sistemas Wi-Fi que estão sujeitos a ataques avançados [Jiang et al. 2024].

- **CSE-CIC-IDS2018:** Representa tráfego de rede mais recente, refletindo ameaças emergentes que podem comprometer a segurança de redes sem fio. Possui um grande volume de dados, permitindo o treinamento de modelos mais complexos, como Redes Neurais Profundas (DNN) e CNN. Inclui tanto tráfego normal quanto tráfego malicioso etiquetado, facilitando a avaliação de desempenho dos modelos [Babu and Rao 2023].

Na terceira parte da pesquisa, foram identificadas 11 métricas de desempenho utilizadas para avaliar a eficácia dos modelos como é mostrado na Tabela 2. A acurácia foi a mais comum, presente em todos os artigos, por medir a proporção de previsões corretas. O F1-score apareceu em 8 artigos [Assy et al. 2023], [Babu and Rao 2023], [Jiang et al. 2024], [Mezina et al. 2021], [Yong and Gao 2023], [Wisanwanichthan and Thammawichai 2021], [Çetin 2022], [ALR], combinando precisão e recall. Outras métricas também citadas foram: Matriz de confusão, Precisão e Recall onde apareceram em [Assy et al. 2023], [Babu and Rao 2023], [Jiang et al. 2024], [Mezina et al. 2021], [ALR], [Wisanwanichthan and Thammawichai 2021], [Çetin 2022]. FAR em [Wisanwanichthan and Thammawichai 2021]. G-mean, BAS e Especificidade em [Çetin 2022]. Matthews Correlation Coefficient e Kappa Score em [Babu and Rao 2023].

Tabela 1. Modelos de IDS e seus respectivos autores

Modelo	Acurácia	Autor
CNN	99,72%	[ALR]
CNN	92,20%	[Assy et al. 2023]
ANU-Net	98,00%	[Babu and Rao 2023]
BBO + Transformer	99,10%	[Jiang et al. 2024]
Black Hole Algorithm (BHA) + Firefly Algorithm (FA)	86,10%	[Yong and Gao 2023]
Naive Bayes + SVM	88,97%	[Wisanwanichthan and Thammawichai 2021]
Temporal Convolutional Network (TCN)	97,00%	[Mezina et al. 2021]
CNN	94,40%	[Ashiku and Dagli 2021]
SVM + Árvore de Decisão Algoritmo Genético + Regressão Logística Random Forest + Perceptron Multicamadas	91,00%	[Çetin 2022]

Tabela 2. Frequência das métricas utilizadas nos estudos

Métrica	Frequência
Acurácia	9
F1-Score	8
Recall	7
Precisão	7
Matriz Confusão	7
Kappa Score	1
Taxa de Alarme Falso	1

As métricas acurácia, F1-score, matriz de confusão, precisão e recall são amplamente utilizadas para avaliar modelos de aprendizado de máquina para detecção de in-

trusão em redes Wi-Fi porque fornecem uma visão abrangente do desempenho do sistema em um problema de classificação desbalanceada e sensível a falsos positivos e falsos negativos. Quando essas métricas são usadas juntas, podem oferecer uma visão equilibrada do desempenho do modelo para este cenário, sendo ideal otimizar F1-score, precisão e recall para equilibrar a detecção de ameaças sem gerar muitos alarmes falsos.

4. Conclusão

Este trabalho apresentou um mapeamento sistemático sobre Sistemas de Detecção de Intrusão Baseados em Aprendizado de Máquina para Redes Wi-Fi, com auxílio da metodologia PICOC [Kitchenham 2007].

Foram avaliados 9 artigos, quantidade limitada devido ao prazo curto para desenvolvimento do trabalho, destacando-se o uso do modelo CNN como o mais frequente e eficaz, com acurácia superior a 90%. Modelos clássicos, como Naive Bayes Classifier e Support Vector Machine, também foram utilizados para obter resultados satisfatórios. Esses achados sugerem que futuras pesquisas aprofundem o uso de CNNs ou explorem abordagens híbridas, enquanto usuários podem considerar esses modelos para aumentar a segurança em redes Wi-Fi.

Quanto às bases de dados, o conjunto NSL-KDD foi o mais utilizado, aparecendo em 46,2% dos estudos, seguido pelo CIC-IDS2017 e CSE-CIC-IDS2018. A preferência pelo NSL-KDD deve-se às suas vantagens em relação ao KDD99, como a eliminação de redundâncias e um melhor equilíbrio entre tráfego normal e malicioso, características que o tornam uma referência para testes em detecção de intrusões.

Sobre as métricas de avaliação, a acurácia foi a mais frequente, presente em todos os trabalhos analisados. Outras métricas importantes foram F1-score, matriz de confusão, precisão e recall, fundamentais para avaliar modelos em contextos de dados desbalanceados, como é comum em detecção de intrusões em redes Wi-Fi.

Apesar dos avanços, a detecção de intrusões em redes Wi-Fi ainda enfrenta desafios, especialmente quanto à alta dimensionalidade dos dados, exigindo um pré-processamento eficiente. Estudos futuros podem desenvolver novas arquiteturas baseadas em aprendizado de máquina para otimizar esse processo, preservando as características mais relevantes, como observado nos datasets KDD99 e NSL-KDD.

Referências

Alrayes, F. S., Zakariah, M., Amin, S. U., Khan, Z. I., and Alqurni, J. S. (2024). Cnn channel attention intrusion detection system using nsl-kdd dataset. *Computers, Materials & Continua*, 79(3).

Ashiku, L. and Dagli, C. (2021). Network intrusion detection system using deep learning. *Procedia Computer Science*, 185:239–247. Big Data, IoT, and AI for a Smarter Future.

Assy, A. T., Mostafa, Y., El-khaleq, A. A., and Mashaly, M. (2023). Anomaly-based intrusion detection system using one-dimensional convolutional neural network. *Procedia Computer Science*, 220:78–85. The 14th International Conference on Ambient Systems, Networks and Technologies Networks (ANT) and The 6th International Conference on Emerging Data and Industry 4.0 (EDI40).

Babu, K. S. and Rao, Y. N. (2023). Improved monarchy butterfly optimization algorithm (imbo): Intrusion detection using mapreduce framework based optimized anunet. *Computers, Materials and Continua*, 75(3):5887–5909.

Jiang, T., Fu, X., and Wang, M. (2024). Bbo-cfat: Network intrusion detection model based on bbo algorithm and hierarchical transformer. *IEEE Access*, 12:54191–54201.

Kitchenham, B. (2007). Guidelines for performing systematic literature reviews in software engineering. EBSE Technical Report EBSE-2007-01, Keele University and University of Durham.

Liu, H. and Lang, B. (2019). Machine learning and deep learning methods for intrusion detection systems: A survey. *Applied Sciences*, 9(20).

Mezina, A., Burget, R., and Travieso-González, C. M. (2021). Network anomaly detection with temporal convolutional network and u-net model. *IEEE Access*, 9:143608–143622.

Wisanwanichthan, T. and Thammawichai, M. (2021). A double-layered hybrid approach for network intrusion detection system using combined naive bayes and svm. *IEEE Access*, 9:138432–138450.

Yang, F.-J. (2018). An implementation of naive bayes classifier. In *2018 International Conference on Computational Science and Computational Intelligence (CSCI)*, pages 301–306.

Yong, X. and Gao, Y. (2023). Hybrid firefly and black hole algorithm designed for xgboost tuning problem: An application for intrusion detection. *IEEE Access*, 11:28551–28564.

Çetin, G. (2022). An effective classifier model for imbalanced network attack data. *Computers, Materials and Continua*, 73(3):4519–4539.