UNIVERSIDADE FEDERAL DE CAMPINA GRANDE
CENTRO DE ENGENHARIA ELÉTRICA E INFORMÁTICA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

IGOR MONTEIRO ABREU DOS SANTOS

DETECÇÃO DE ATAQUES DE INJEÇÃO DE DADOS FALSOS EM SISTEMAS DE POTÊNCIA UTILIZANDO *AUTOENCODERS* COM MECANISMO DE ATENÇÃO

CAMPINA GRANDE - PB
2025

IGOR MONTEIRO ABREU DOS SANTOS

DETECÇÃO DE ATAQUES DE INJEÇÃO DE DADOS FALSOS EM SISTEMAS DE POTÊNCIA UTILIZANDO *AUTOENCODERS* COM MECANISMO DE ATENÇÃO

Dissertação apresentada ao Programa de Pós-graduação em Engenharia Elétrica da Universidade Federal de Campina Grande, como requisito parcial para a obtenção do título de Mestre em Ciências em Engenharia Elétrica.

Orientador(es): George Rossany Soares de Lira, D.Sc.
Pablo Bezerra Vilar, D.Sc.

Campina Grande - PB

2025

S237d Santos, Igor Monteiro Abreu dos.

Detecção de ataques de injeção de dados falsos em sistemas de potência utilizando *autoencoders* com mecanismo de atenção / Igor Monteiro Abreu dos Santos. – Campina Grande, 2025.

82 f. : il. color.

Dissertação (Mestrado em Engenharia Elétrica) – Universidade Federal de Campina Grande, Centro de Engenharia Elétrica e Informática, 2025.

"Orientação: Prof. Dr. George Rossany Soares de Lira, Prof. Dr. Pablo Bezerra Vilar".

Referências.

1. Ataques de Injeção de Dados Falsos. 2. Sistemas de Potência. 3. *Autoencoders*. 4. Mecanismo de Atenção *Multi-head*. 5. Detecção de Anomalias. 6. Segurança Cibernética I. Lira, George Rossany Soares de. II. Vilar, Pablo Bezerra. III. Título.

CDU 621.3:004.056(043.3)

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECÁRIA MEIRE EMANUELA DA SILVA MELO CRB-15/568

IGOR MONTEIRO ABREU DOS SANTOS

Detecção de Ataques de Injeção de Dados Falsos em Sistemas de Potência Utilizando *Autoencoders* com Mecanismo de Atenção

Dissertação apresentada ao Programa de Pós-graduação em Engenharia Elétrica da Universidade Federal de Campina Grande, como requisito parcial para a obtenção do título de Mestre em Ciências em Engenharia Elétrica.

Orientador(es): George Rossany Soares de Lira e Pablo Bezerra Vilar.

Aprovada em 15 de julho de 2025, pela seguinte banca examinadora:

Prof. **George Rossany Soares de Lira** - D.Sc. da UFCG
Orientador

Prof. **Pablo Bezerra Vilar** - D.Sc. da UFCG
Orientador

Prof. **Washington Luiz Araújo Neves** - Ph.D. da UFCG
Examinador Interno

Prof. **Danilo Freire de Souza Santos** - D.Sc. da UFCG
Examinador Interno

Campina Grande - PB
2025

Processo: 23096.046089/2025-09 Documento: 5651579

MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DE CAMPINA GRANDE
POS-GRADUACAO EM ENGENHARIA ELETRICA
Rua Aprigio Veloso, 882, - Bairro Universitario, Campina Grande/PB, CEP 58429-900

REGISTRO DE PRESENÇA E ASSINATURAS

1 - ATA DA DEFESA PARA CONCESSÃO DO GRAU DE MESTRE EM ENGENHARIA ELÉTRICA, REALIZADA EM 15 DE JULHO DE 2025.

(Nº 783)

CANDIDATO(A): **IGOR MONTEIRO ABREU DOS SANTOS.** COMISSÃO EXAMINADORA: DANILO FREIRE DE SOUZA SANTOS, Dr., UFCG - Presidente da Comissão e Examinador Interno, GEORGE ROSSANY SOARES DE LIRA, D.Sc., UFCG – Orientador, PABLO BEZERRA VILAR, D.Sc., UFCG – Orientador e WASHINGTON LUIZ ARAÚJO NEVES, Dr., UFCG - Examinador Interno. TÍTULO DA DISSERTAÇÃO: DETECÇÃO DE ATAQUES DE INJEÇÃO DE DADOS FALSOS EM SISTEMAS DE POTÊNCIA UTILIZANDO AUTOENCODERS COM MECANISMO DE ATENÇÃO. ÁREA DE CONCENTRAÇÃO: Processamento da Energia. HORA DE INÍCIO: **10h** – LOCAL: **Auditório Ricardo Loureiro/LAT.** Em sessão pública, após exposição de cerca de 45 minutos, o(a) candidato(a) foi arguido(a) oralmente pelos membros da Comissão Examinadora, tendo demonstrado suficiência de conhecimento e capacidade de sistematização, no tema de sua dissertação, obtendo o conceito APROVADO. Face à aprovação, declara o(a) presidente da Comissão, achar-se o examinado, legalmente habilitado a receber o Grau de Mestre em Engenharia Elétrica, cabendo a Universidade Federal de Campina Grande, como de direito, providenciar a expedição do Diploma, a que o(a) mesmo(a) faz jus. Na forma regulamentar, foi lavrada a presente ata, que é assinada por mim, LEANDRO FERREIRA DE LIMA, e os membros da Comissão Examinadora. Campina Grande, 15 de Julho de 2025.

LEANDRO FERREIRA DE LIMA
Secretário

DANILO FREIRE DE SOUZA SANTOS, Dr., UFCG
Presidente da Comissão e Examinador Interno

GEORGE ROSSANY SOARES DE LIRA, D.Sc., UFCG
Orientador

PABLO BEZERRA VILAR, D.Sc., UFCG
Orientador

WASHINGTON LUIZ ARAÚJO NEVES, Dr., UFCG

Processo: 23096.046089/2025-09 Documento: 5651579

IGOR MONTEIRO ABREU DOS SANTOS
Candidato

2 - APROVAÇÃO

2.1. Segue a presente Ata de Defesa de Dissertação de Mestrado do candidato **IGOR MONTEIRO ABREU DOS SANTOS** assinada eletronicamente pela Comissão Examinadora acima identificada.

2.2. No caso de examinadores externos que não possuam credenciamento de usuário externo ativo no SEI, para igual assinatura eletrônica, os examinadores internos signatários **certificam** que os examinadores externos acima identificados participaram da defesa da tese e tomaram conhecimento do teor deste documento.

Documento assinado eletronicamente por **LEANDRO FERREIRA DE LIMA**, **SECRETÁRIO (A)**, em 15/07/2025, às 14:21, conforme horário oficial de Brasília, com fundamento no art. 8º, caput, da Portaria SEI nº 002, de 25 de outubro de 2018.

Documento assinado eletronicamente por **Igor Monteiro Abreu dos Santos**, **Usuário Externo**, em 15/07/2025, às 14:47, conforme horário oficial de Brasília, com fundamento no art. 8º, caput, da Portaria SEI nº 002, de 25 de outubro de 2018.

Documento assinado eletronicamente por **GEORGE ROSSANY SOARES DE LIRA**, **PROFESSOR(A) DO MAGISTERIO SUPERIOR**, em 15/07/2025, às 14:51, conforme horário oficial de Brasília, com fundamento no art. 8º, caput, da Portaria SEI nº 002, de 25 de outubro de 2018.

Documento assinado eletronicamente por **DANILO FREIRE DE SOUZA SANTOS**, **PROFESSOR(A) DO MAGISTERIO SUPERIOR**, em 15/07/2025, às 16:11, conforme horário oficial de Brasília, com fundamento no art. 8º, caput, da Portaria SEI nº 002, de 25 de outubro de 2018.

Documento assinado eletronicamente por **WASHINGTON LUIZ ARAUJO NEVES**, **PROFESSOR(A) DO MAGISTERIO SUPERIOR**, em 22/07/2025, às 16:59, conforme horário oficial de Brasília, com fundamento no art. 8º, caput, da Portaria SEI nº 002, de 25 de outubro de 2018.

Documento assinado eletronicamente por **PABLO BEZERRA VILAR**, **PROFESSOR(A) DO MAGISTERIO SUPERIOR**, em 25/07/2025, às 15:03, conforme horário oficial de Brasília, com fundamento no art. 8º, caput, da Portaria SEI nº 002, de 25 de outubro de 2018.

A autenticidade deste documento pode ser conferida no site https://sei.ufcg.edu.br/autenticidade, informando o código verificador **5651579** e o código CRC **B1B9B592**.

Referência: Processo nº 23096.046089/2025-09 SEI nº 5651579

A Deus, pela força e luz nos momentos de incerteza. Aos que acreditam que o conhecimento é uma forma de transcendência. E a todos que veem na educação um caminho para transformar o mundo.

AGRADECIMENTOS

Agradeço a Deus pelo dom da vida! Obrigado, meu Senhor, por me acompanhar, proteger, abençoar e conceder boa saúde e força. Ao Senhor todo louvor, honra e glória para sempre!

Expresso, com muito amor, minha eterna gratidão aos meus pais, Neuma e Martins. Obrigado por todo carinho, dedicação, presença, incentivo, bons conselhos, por serem minha fonte de inspiração e por nunca terem poupado esforços para me fazer todo tipo de bem.

Agradeço a todos os meus familiares, gratidão pela torcida, orações, palavras de apoio e gestos de companheirismo. Todo apoio de vocês tornam a caminhada mais leve e agradável.

Agradeço aos meus orientadores, professores George Lira e Pablo Vilar, pela constante disposição em me auxiliar e pelas contribuições essenciais para o desenvolvimento de todas as etapas deste trabalho de Mestrado, além de todas as oportunidades que me concederam e me confiaram durante o período. Estendo o meu agradecimento a todos os professores que tive ao longo da vida, pois sem dúvidas eles foram extremamente fundamentais no caminho que trilhei até chegar aqui.

Agradeço também a todos os amigos, professores e funcionários do Laboratório de Alta Tensão, pela boa convivência e partilha de aprendizados.

"Ainda que não exista céu, e a Bíblia se torne inútil,
viver em santidade se torne loucura,
foi um prazer ser cristão."
(Billy Graham)

RESUMO

Esta dissertação propõe um método para a detecção de ataques de injeção de dados falsos (FDI, do inglês, *False Data Injection*) em sistemas de potência, utilizando *autoencoders* com mecanismo de atenção *multi-head*. Os ataques FDI representam uma ameaça significativa à segurança e estabilidade das redes elétricas modernas, pois manipulam dados de medição para enganar sistemas de controle sem serem detectados. O modelo desenvolvido combina a capacidade dos *autoencoders* de aprender padrões normais de operação com a eficiência do mecanismo de atenção multi-head para identificar anomalias sutis causadas por ataques. A metodologia foi validada em sistemas de diferentes complexidades (IEEE-14, 30, 118 e 300 barras), utilizando dados gerados pelo MATPOWER. Os resultados demonstraram alta eficácia na detecção, com acurácia superior a 97% para variações de ataque de $\pm 50\%$ e $\pm 25\%$, e desempenho robusto mesmo em ataques mais sutis ($\pm 15\%$). A abordagem proposta supera métodos tradicionais ao eliminar a necessidade de dados rotulados e modelos complexos do sistema, destacando-se como uma solução escalável e adaptável para a segurança de infraestruturas críticas.

Palavras-chave: ataques de injeção de dados falsos; sistemas de potência; *autoencoders*; mecanismo de atenção *multi-head*; detecção de anomalias; segurança cibernética.

ABSTRACT

This dissertation proposes a method for detecting False Data Injection (FDI) attacks in power systems using autoencoders with a multi-head attention mechanism. FDI attacks pose a significant threat to the security and stability of modern electrical grids by manipulating measurement data to deceive control systems undetected. The developed model combines the ability of autoencoders to learn normal operation patterns with the efficiency of the multi-head attention mechanism to identify subtle anomalies caused by attacks. The methodology was validated on systems of varying complexity (IEEE-14, 30, 118, and 300 buses) using data generated by MATPOWER. The results demonstrated high detection efficacy, with accuracy exceeding 97% for attack variations of ±50% and ±25%, and robust performance even for more subtle attacks (±15%). The proposed approach outperforms traditional methods by eliminating the need for labeled data and complex system models, standing out as a scalable and adaptable solution for critical infrastructure security.

Keywords: false data injection attacks; power systems; autoencoders; multi-head attention mechanism; anomaly detection; cybersecurity.

PRODUÇÕES BIBLIOGRÁFICAS, PRODUÇÕES TÉCNICAS E PREMIAÇÕES

I. M. A. Santos, G. R. S. Lira and P. B. Vilar, "Detection of False Data Injection Attacks in Power Systems Using Deep Autoencoder with Attention Mechanism," 2024 Workshop on Communication Networks and Power Systems (WCNPS), Brasilia, Brazil, 2024, pp. 1-7, doi: 10.1109/WCNPS65035.2024.10814586.

I. M. A. Santos, G. R. S. Lira and P. B. Vilar. Monitoramento de Tráfego de Rede e Detecção de Anomalias em Sistemas de Potência Usando Python e Protocolos de Comunicação. In: Simpósio de Automação de Sistemas Elétricos, 2025, Foz do Iguaçu. Anais do XIV SIMPASE, 2025. (Trabalho aprovado)

I. M. A. Santos, G. R. S. Lira and P. B. Vilar. Avaliação de Inteligência Artificial com Mecanismo de Atenção na Detecção de Ataques de Injeção de Dados Falsos em Sistemas de Potência. In: Seminário Nacional de Produção e Transmissão de Energia Elétrica, 2025, Recife. Anais do XXVIII SNPTEE - Seminário Nacional de Produção e Transmissão de Energia Elétrica, 2025. (Trabalho aprovado)

I. M. A. Santos, G. R. S. Lira and P. B. Vilar, "Application of Deep Autoencoder with Multi-head Attention Mechanism in Detecting False Data Injection Attacks in Power Systems,". IEEE Transactions on Smart Grid. 2025 (Em fase de revisão para submissão à revista)

LISTA DE ILUSTRAÇÕES

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

SG	Rede Inteligente (*Smart Grid*)
SEP	Sistema Elétrico de Potência
DER	Recursos Energéticos Distribuídos (*Distributed Energy Resources*)
DS	Armazenamento Distribuído (*Distributed Storage*)
BESS	Sistemas de Armazenamento de Baterias (*Battery Energy Storage System*)
TESS	Sistema de Armazenamento de Energia Térmica (*Thermal Energy Storage System*)
TIC	Tecnologia da Informação e Comunicação
IoT	Internet das Coisas (*Internet of Things*)
FDI	Injeção de Dados Falsos (*False Data Injection*)
DoS	Negação de Serviço (*Denial of Service*)
DDoS	Negação de Serviço Distribuído (*Distributed Denial of Service*)
MiTM	Homem no Meio (*Man in-the-Middle*)
ICS	Sistema de Controle Industrial (*Industrial Control System*)
BDD	Detecção de Dados ruins (*Bad Data Detection*)
GD	Geração Distribuída

SUMÁRIO

1 INTRODUÇÃO

A evolução dos Sistemas Elétricos de Potência (SEP) ao longo das últimas décadas reflete um avanço tecnológico significativo, marcado pela transição dos modelos convencionais de geração, transmissão e distribuição de energia para as chamadas redes inteligentes, ou *Smart Grids* (SG). Inicialmente, os sistemas elétricos convencionais, estabelecidos ao longo do século XX, operavam de maneira linear e centralizada, com pouca ou nenhuma interação entre os componentes do sistema e os consumidores, além de uma capacidade limitada de resposta à eventos não planejados e uma automação e controle remoto restritos. No entanto, as crescentes demandas por eficiência energética, sustentabilidade e a integração dos recursos energéticos distribuídos impulsionaram a adoção de tecnologias mais avançadas, resultando em uma transformação na forma como a energia é gerenciada (ABDALLAH, 2016).

Este novo modelo é capaz de suportar tanto as formas tradicionais de geração de energia quanto os recursos energéticos distribuídos (DER, do inglês, *Distributed Energy Resources*), incluindo a geração distribuída (GD) como as fontes solar e eólica, fontes de armazenamento distribuído (DS, do inglês, *Distributed Storage*) como sistemas de armazenamento de energia de bateria (BESS, do inglês, *Battery Energy Storage System*) e sistemas de armazenamento de energia térmica (TESS, do inglês, *Thermal Energy Storage System*). Estas transformações sucitaram uma mudança do SEP convencional para um Sistema Elétrico Ciberfísico, viabilizadas principalmente pela introdução de Tecnologias de Informação e Comunicação (TIC), sistemas de controle automatizados e sensoriamento remoto.

Entretanto, essa modernização expôs o sistema elétrico a novas vulnerabilidades (SAKHNINI; KARIMIPOUR; DEHGHANTANHA, 2019). A interconexão e a digitalização dos sistemas de energia, embora fundamentais para o funcionamento das redes inteligentes, criaram superfícies de ataque cibernético que antes não existiam. Esses riscos se agravam à medida que os sistemas elétricos se tornam mais dependentes de tecnologia da informação, o que torna a segurança cibernética uma preocupação central para a estabilidade, disponibilidade e a resiliência das redes elétricas modernas. Desta forma, os sistemas de energia modernos estão sujeitos a vários ataques cibernéticos como ataques de injeção de dados falsos (FDI, do inglês, *False Data Injection*), ataques de negação de serviço (DoS, do inglês, *Denial of Service*), ataques de negação de serviço distribuídos (DDoS, do inglês, *Distributed Denial of Service*), ataques homem no meio (MiTM, do inglês, *Man-in-the-Middle*), ataques de análise de pacotes, ataques de injeção de pacotes, falsificação de dados, entre outros (HABIB et al., 2023). A Tabela 1 apresenta alguns incidentes mundiais relevantes ocorridos mediante ataques cibernéticos.

Tabela 1 – Incidentes relevantes ocorridos por ataques cibernéticos (HABIB et al., 2023).

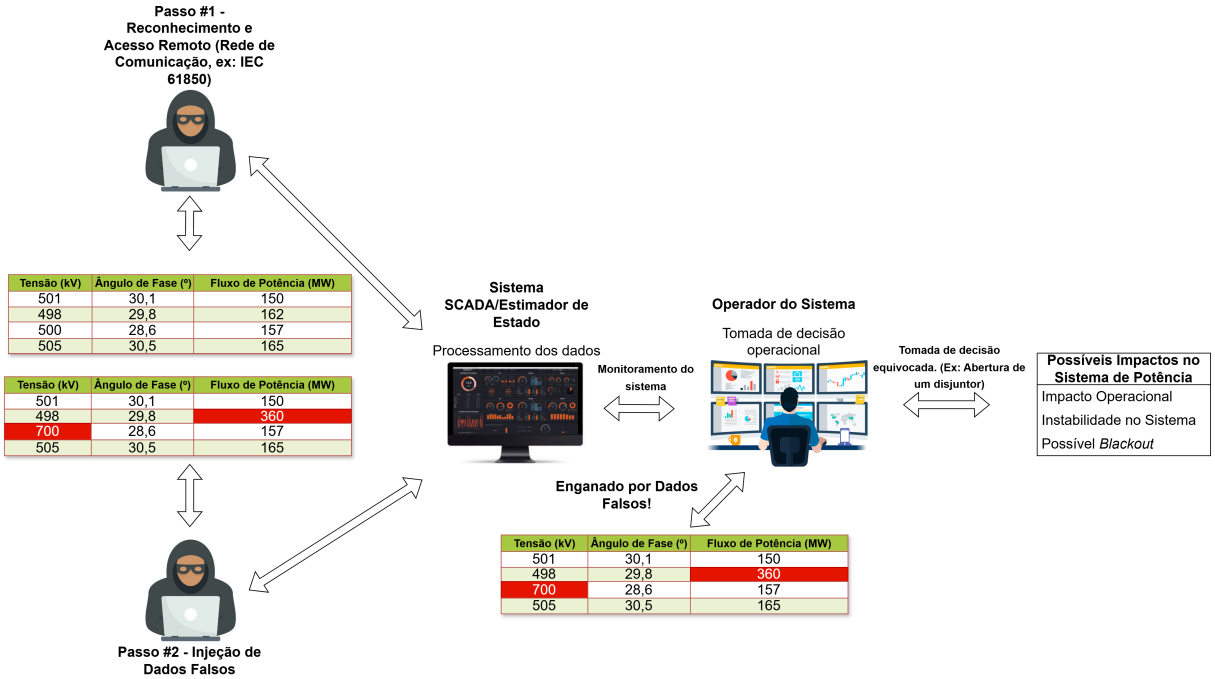
Ano	**Localização**	**Alvo do ataque**	**Tipo de ataque**	**Impacto**
2022	Finlândia	Ministro da Defesa e Relações exteriores	Ataque DoS Distribuído	Comunicações oficiais suspensas por 4 horas
2022	Israel	Provedores de Telecomunicações	Ataque DDoS	Vários sites oficiais israelenses *offline*
2021	Austrália	Sistemas de Telecomunicações	Código malicioso	Gravação dos dados de comunicação
2019	EUA	Operadores de rede	Ataque DDoS	Perturbação na operação sem causar interrupção
2015	Ucrânia	Acionamento de disjuntor por um *malware* denominado "*BlackEnergy*"	Ataque FDI	Mais de 225,000 consumidores afetados pleo apagão
2013	Nova York	Barragem Bowman	*Ataque FDI*	Obtenção do acesso remoto às informações sobre níveis de água e vazões
2010	Irã	Programa nuclear Iraniano	*Malware Stuxnet*	Destruição de cerca de 1000 centrífugas de enriquecimento de urânio

O incidente ocorrido na Ucrânia em 2015 (LIANG et al., 2017), foi um dos primeiros casos de ataque cibernético bem-sucedido com consequências graves aos sistemas de energia, resultando em um apagão elétrico de larga escala. Os hackers conseguiram comprometer as redes de três empresas de distribuição de energia, causando a interrupção do fornecimento de eletricidade para cerca de 225 mil pessoas. O ataque foi sofisticado, envolvendo o uso de *malware* conhecido como "*BlackEnergy*", que permitiu aos invasores obter acesso remoto aos sistemas de controle industrial (ICS, do inglês, *Industrial Control System*) das empresas. A ocorrência deste ataque destacou a vulnerabilidade das infraestruturas críticas a ameaças cibernéticas e serviu como um alerta global sobre a necessidade urgente de reforçar a segurança cibernética em sistemas de energia.

Entre as várias ameaças cibernéticas às quais os sistemas de energia são expostos, os ataques de injeção de dados falsos (FDI, do inglês, *False Data Injection*) têm se destacado

como um dos mais perniciosos e difíceis de detectar, pois visam a estimativa de estado dos sistemas de energia (WANG; BI; ZHANG, 2020). A estimativa de estado é uma aplicação em tempo real, utilizada para converter medições redundantes e outros dados disponíveis do sistema em uma estimativa de seu estado, onde entende-se como estado o valor das variáveis de interesse do sistema, como nível de tensão nos barramentos. Em outras palavras, a estimativa de estado atua eliminando erros de medição inevitáveis e correlacionando o conjunto de medições em uma estimativa confiável do estado do sistema. Tradicionalmente, medições incorretas têm sido identificadas e removidas utilizando técnicas analíticas de Detecção de Desvios de Dados (BDD, do inglês, *Bad Data Detection*). No entanto, pesquisas recentes demonstram que a manipulação inteligente e coordenada de múltiplas medições simultaneamente pode escapar das técnicas tradicionais de BDD, representando uma ameaça significativa à segurança do sistema (KARIMIPOUR; DINAVAHI, 2017). Esses ataques envolvem a manipulação maliciosa dos dados utilizados pelos sistemas de controle e monitoramento para estimar o estado do sistema elétrico. Consequentemente, os operadores do sistema podem ser induzidos a tomar decisões baseadas em informações incorretas, o que pode resultar em sobrecargas, apagões localizados, ou até mesmo em um colapso sistêmico de grandes proporções (WANG; LU, 2013). Para melhor compreensão das ameaças associadas ao cenário estudado é ilustrado na Figura 1 o modelo de ameaças considerando ataques de injeção de dados falsos em sistemas de potência.

Figura 1 – Modelo de ataque FDI em um Sistema de Potência.



Fonte: Autor.

A utilização da Inteligência Artificial (IA), especialmente através de técnicas de *deep learning*, tem sido fundamental no avanço da detecção de ataques FDI em

sistemas de potência (NIU et al., 2019). O *deep learning* permite o desenvolvimento de modelos altamente complexos e eficazes, capazes de analisar grandes volumes de dados operacionais e detectar padrões sutis de anomalias que poderiam passar despercebidos por métodos tradicionais. Entre essas técnicas, os *autoencoders* se destacam por sua habilidade em detectar anomalias ao aprender representações comprimidas dos dados e, em seguida, reconstruir essas informações (HINTON; SALAKHUTDINOV, 2006). Qualquer desvio significativo entre os dados reais e reconstruídos pode sinalizar a presença de um ataque. Além disso, os *autoencoders* destacam-se por sua abordagem não supervisionada, contornando um dos principais desafios no domínio de detecção de anomalias: a escassez de dados rotulados para treinamento. Ao aprender automaticamente uma representação compacta do comportamento normal do sistema a partir de dados não anotados, esses modelos dispensam a dependência de conjuntos de dados extensivamente classificados, que muitas vezes são difíceis de obter em aplicações do mundo real. Essa característica não apenas simplifica o processo de implementação, mas também aumenta a precisão na identificação de inconsistências provocadas por ataques ou falhas operacionais, reforçando assim a segurança e a resiliência dos sistemas de potência.

No presente trabalho, é investigada a aplicação de um *autoencoder*—uma arquitetura de rede neural profunda que comprime e reconstrói dados de forma hierárquica, aprendendo representações cada vez mais abstratas dos padrões de entrada—acoplado a um mecanismo de atenção *multi-head* para a detecção de ataques de injeção de dados falsos (FDI) em sistemas de potência. O *autoencoder* se diferencia dos modelos tradicionais por sua capacidade de extrair características complexas e não lineares dos dados operacionais, enquanto o mecanismo de atenção permite focalizar seletivamente em variáveis críticas do sistema, melhorando a sensibilidade na identificação de anomalias sutis induzidas por ataques. Os dados foram gerados por meio de simulações no MATPOWER[1], considerando sistemas de diferentes portes: IEEE-14, 30, 118 e 300 barras. A relevância desta pesquisa reside no potencial de avançar a detecção precoce de ameaças cibernéticas em infraestruturas críticas, como redes elétricas. Ao incorporar um mecanismo de atenção *multi-head* à arquitetura do *autoencoder*, busca-se aprimorar a identificação de anomalias causadas por ataques FDI, contribuindo para o desenvolvimento de técnicas mais eficientes. Dessa forma, este estudo visa fortalecer a segurança, a confiabilidade e a resiliência dos sistemas de potência contra ameaças emergentes.

[1] O **MATPOWER** é um pacote de ferramentas de código aberto, desenvolvido em arquivos M da linguagem MATLAB, empregado para a resolução de problemas de simulação e otimização de sistemas de potência em regime permanente. Para mais informações e acesso ao software, site eletrônico oficial: <https://matpower.org/>

1.1 Objetivos

O objetivo geral deste trabalho é avaliar a viabilidade da utilização de *autoencoders* com mecanismo de atenção para detecção de ataques de injeção de dados falsos em sistemas de potência. Destacam-se como objetivos específicos:

- Implementar um modelo de *autoencoder* com mecanismo de atenção adaptado à detecção de anomalias em sistemas de potência, explorando sua capacidade de capturar dependências temporais e espaciais nos dados.
- Gerar e caracterizar conjuntos de dados de ataques FDI para sistemas IEEE-14, 30, 118 e 300 barras utilizando o MATPOWER, considerando diferentes cenários de operação e padrões de ataque.
- Comparar o desempenho do *autoencoder* em diferentes cenários de ataques FDI, variando a intensidade dos ataques e os tipos de dados comprometidos, para avaliar a resiliência do modelo;
- Analisar a eficiência do *autoencoder* na detecção de anomalias em sistemas de potência com diferentes níveis de complexidade, utilizando métricas adequadas de avaliação de desempenho;
- Avaliar a capacidade de generalização do modelo de *autoencoder*, verificando sua performance ao ser treinado nos sistemas, analisando a portabilidade dos resultados entre diferentes topologias de sistema.

1.2 Relevância do Trabalho

A relevância do trabalho proposto é evidenciada por sua forte compatibilidade com os temas estratégicos definidos no Plano Estratégico Quinquenal de Inovação (PEQuI) 2024-2028 do Programa de Pesquisa, Desenvolvimento e Inovação (PDI) da Agência Nacional de Energia Elétrica (ANEEL). Esse plano, que orienta as direções estratégicas de inovação no setor energético brasileiro, destaca entre seus temas prioritários o TE3, focado em “Digitalização, Padrões, Interoperabilidade e Cibersegurança”, e o TE4, que aborda “Novas Tecnologias de Suporte”, incluindo Inteligência Artificial, Realidade Virtual e Aumentada e *Blockchain*.

A pesquisa se alinha com o TE3, já que explora a Cibersegurança no setor elétrico. Esse tema se refere ao desenvolvimento de padrões de rede, mantendo a segurança da rede e permitindo que a energia seja transportada entre as redes. Logo, essa pesquisa apresenta um avanço no desenvolvimento de padrões de rede, especialmente à medida que as redes de energia evoluem para sistemas mais interconectados e inteligentes, como as

smart grids. Com a crescente digitalização, o aumento da automação e a integração de diferentes redes elétricas, a proteção contra ciberataques tornou-se uma prioridade para garantir a segurança, estabilidade e confiabilidade do sistema de energia.

Esta pesquisa também se alinha com o TE4, uma vez que explora aplicação de IA para detecção de ataques FDI em sistemas de potência. Ao superar métodos tradicionais, os *autoencoders* identificam anomalias sutis nos dados, permitindo uma detecção mais precisa e em tempo real de manipulações maliciosas. Essa abordagem moderna aprimora a resiliência e confiabilidade das redes, trazendo avanços importantes na proteção contra ataques sofisticados, essenciais para a operação segura de sistemas de potência interconectados.

Assim, o alinhamento deste trabalho com as diretrizes do PEQuI da ANEEL reforça sua relevância e potencial impacto no cenário nacional de inovação em energia. Ao atender aos temas estratégicos propostos, a pesquisa não só impulsiona o avanço técnico e científico, como também contribui para os objetivos mais amplos de inovação e sustentabilidade no setor energético brasileiro. Esse alinhamento comprova que o estudo está na vanguarda do desenvolvimento tecnológico e em harmonia com as prioridades nacionais de inovação e desenvolvimento sustentável.

1.3 Contribuições

As principais contribuições deste trabalho são:

- **Avanço na Utilização de *Autoencoders* para Detecção de Anomalias em FDI**: Este trabalho avança no estado da arte ao propor um modelo baseado em *autoencoders* com mecanismo de atenção, capaz de detectar ataques FDI de forma não supervisionada, superando limitações de métodos tradicionais que dependem de dados rotulados ou conhecimento prévio da topologia do sistema. O mecanismo de atenção permite focar em medições críticas, melhorando a sensibilidade a anomalias sutis, como aquelas geradas por ataques FDI direcionados.
- **Aplicação em Sistemas de Diferentes Complexidades**: A metodologia foi validada em quatro sistemas distintos (IEEE-14, 30, 118 e 300 barras), demonstrando sua eficácia tanto em redes menores quanto em sistemas mais complexos com múltiplos estados operacionais. Isso reforça a escalabilidade da abordagem proposta.
- **Metodologia de geração de dados**: Essa dissertação avança no estado da arte apresentando uma metodologia para geração de dados para detecção de ataques FDI em sistemas de potência ao criar um conjunto de dados próprio, realista e controlável, que permite o treinamento e validação de modelos de detecção em cenários próximos ao mundo real, com variabilidade e inserção de ataques furtivos.

1.4 Organização do Trabalho

Este trabalho de dissertação está organizado em 6 capítulos. No primeiro capítulo, Introdução, foi apresentada uma contextualização na qual o trabalho está inserido, apresentando as justificativas, objetivos e a relevância do trabalho proposto. No segundo capítulo, Fundamentação Teórica, são apresentados conceitos importantes e necessários que servem de base para a elaboração da pesquisa. No terceiro capítulo, Revisão Bibliográfica, são apresentados os resultados do levantamento do estado da arte sobre os métodos de detecção de ataques de injeção de dados falsos, permitindo identificar quais técnicas já foram trabalhadas e identificar limitações e possíveis direcionamentos para a atual pesquisa. No quarto capítulo, Metodologia, são apresentados os passos desenvolvidos, bem como a ordem para que o trabalho seja realizado. No quinto capítulo, Resultados e Discussões, são apresentadas as análises e avaliações das técnicas e métodos utilizados, bem como as discussões para cada caso. No sexto capítulo, são apresentados as conclusões obtidas mediante a metodologia proposta e aplicada.

2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo de fundamentação teórica tem como objetivo apresentar a infraestrutura dos sistemas de energia ciberfísicos, a definição e o modelo matemático do ataque de injeção de dados falsos no contexto de estimativa de estado dos sistemas de energia apresentando também as condições e vulnerabilidades necessárias para ocorrência desses ataques. Além disso, as definições de *autoencoders*, bem como suas características utilizadas neste trabalho também são apresentadas.

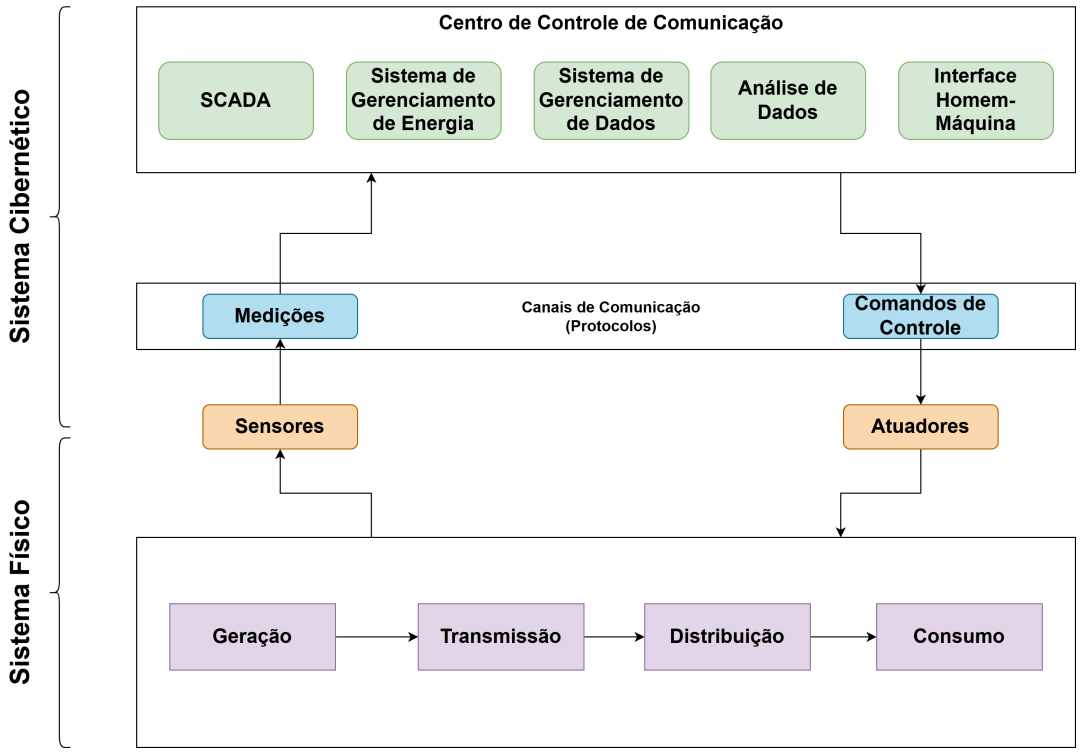
2.1 Infraestrutura dos Sistemas de Energia Ciberfísicos

Os sistemas de energia modernos, também conhecidos como *Smart Grids*, são constituídos de duas camadas independentes: Física e Cibernética (HABIB et al., 2023) como apresentado na Figura 2. A camada física do sistema é composta pelas fontes de geração de energia, sistema de transmissão, distribuição, consumidores, sensores, atuadores e dispositivos de Internet das Coisas (IoT, do inglês, *Internet of Things*). A camada cibernética é constituída por diferentes sistemas de comunicação e rede (Wi-Fi, Ethernet entre outros), além dos centros de controle. Em um sistema inteligente, sensores são usados para captar informações do ambiente físico, como temperatura, umidade ou movimento. Esses dados são enviados para uma central de controle, que processa e analisa as informações utilizando algoritmos específicos ou inteligência artificial. Após essa análise, a central decide quais ações devem ser tomadas e envia comandos para atuadores (dispositivos que realizam ações físicas, como ligar equipamentos, abrir válvulas ou mover motores).

Nos sistemas modernos, o Controle de supervisão e aquisição de dados (SCADA, do inglês, *Supervisory Control and Data Acquisition*) é utilizado para aquisição de dados provenientes de medições dos sistemas de potência, supervisão dos dispositivos eletrônicos inteligentes (DEI) por meio de sistemas de controle remoto, transmissão de comandos de controle e interconexão entre a aquisição de dados e o processo de supervisão (HASAN et al., 2023).

O centro de controle de comunicação ou unidade de controle central é responsável por regular a operação do sistema de energia. Para isso, o Sistema de Gerenciamento de Energia (SGE) atua como um sistema de automação capaz de controlar, operar, monitorar, otimizar e coordenar os dados de desempenho do sistema elétrico dentro da infraestrutura de redes inteligentes (SG) em tempo real. Utilizando o sistema SCADA, o SGE coleta, analisa e monitora continuamente os dados operacionais do sistema SG. As funções do SGE incluem a gestão do fluxo de energia ideal, planejamento das operações, estimativa de estado, gerenciamento de alarmes e otimização do desempenho do sistema de energia. A

Figura 2 – Rede inteligente física e Sistema Cibernético.



Fonte: Adaptado de (HASAN et al., 2024)

unidade de controle central, por sua vez, processa os dados de medição recebidos, realizando a estimativa de estado do sistema e identificando possíveis anomalias ou dados maliciosos que possam comprometer a segurança ou eficiência da operação (HABIB et al., 2023).

2.1.1 Segurança ciberfísica em Redes Inteligentes

Os principais objetivos do sistema de segurança são proteger contra ataques cibernéticos, garantir o monitoramento e a resposta em tempo real e aumentar a resiliência do sistema, buscando garantir assim a qualidade do serviço desde a geração até os consumidores finais (HABIB et al., 2023). Isso envolve a proteção da integridade, confidencialidade e disponibilidade dos dados e operações, para evitar interrupções que possam levar a apagões ou outras consequências graves. Ao permitir a detecção e resposta rápidas a incidentes de segurança, o sistema busca manter a estabilidade e a confiabilidade da rede, garantindo que serviços críticos permaneçam operacionais, mesmo diante de potenciais ameaças (HASAN et al., 2024).

O sistema de segurança para os principais componentes de uma rede inteligente opera por meio de uma combinação de tecnologias, protocolos e práticas projetadas para proteger cada componente contra ameaças cibernéticas, ao mesmo tempo, garantindo uma operação confiável:

- **Infraestrutura Avançada de Medição (AMI)**: O sistema de segurança para a AMI foca na proteção dos canais de comunicação entre os medidores inteligentes e os fornecedores de energia. Isso inclui o uso de criptografia para proteger a transmissão de dados, com o objetivo de que informações sensíveis, como padrões de consumo e dados de faturamento, não sejam interceptadas por partes não autorizadas. Mecanismos de autenticação são implementados para verificar as identidades tanto dos medidores quanto dos sistemas das concessionárias, prevenindo acessos não autorizados. Além disso, atualizações de software e avaliações de vulnerabilidades são realizadas regularmente para abordar possíveis fraquezas de segurança.

- **Sistemas de Supervisão e Aquisição de Dados (SCADA)**: O sistema de segurança para o SCADA inclui segmentação de rede para isolar as redes SCADA de outras redes de TI, reduzindo o risco de contaminação cruzada por ameaças cibernéticas. Sistemas de detecção de intrusão (IDS) são utilizados para monitorar atividades suspeitas e possíveis violações. Medidas de controle de acesso garantem que apenas o pessoal autorizado possa interagir com os sistemas SCADA, e verificações de integridade de dados são implementadas para assegurar que as informações processadas não foram adulteradas.

- **Sistemas de Gerenciamento de Energia (EMS)**: O sistema de segurança para o EMS foca na proteção dos dados e algoritmos usados para otimizar a geração e distribuição de energia. Isso inclui a implementação de controles de acesso e protocolos de autenticação buscando garantir que apenas usuários autorizados possam modificar as configurações do sistema ou acessar dados sensíveis. A criptografia de dados é utilizada para proteger as informações tanto em repouso quanto em trânsito. Além disso, o EMS pode utilizar algoritmos de aprendizado de máquina para detectar anomalias nos padrões de uso de energia, que podem indicar potenciais violações de segurança ou problemas operacionais.

- **Unidades de Medição Fasorial (PMUs)**: As PMUs fornecem monitoramento em tempo real de sinais elétricos. O sistema de segurança para as PMUs inclui medidas buscando garantir a autenticidade dos dados relatados. Isso envolve técnicas criptográficas para verificar que os sinais se originam de fontes legítimas. Protocolos de segurança de rede são implementados para proteger a comunicação entre as PMUs e os centros de controle, garantindo que os dados sejam transmitidos de forma segura. Auditorias e avaliações regulares são conduzidas para identificar e mitigar vulnerabilidades na infraestrutura das PMUs.

Além das medidas específicas para cada componente, a estrutura de segurança geral para redes inteligentes inclui processos de avaliação de riscos para identificar vulnerabilidades potenciais, planos de resposta a incidentes para lidar com violações de

segurança e monitoramento contínuo para detectar e responder a ameaças em tempo real. Nesse contexto, o *autoencoder* com mecanismo de *multi-head attention* proposto neste trabalho atua como uma camada avançada de detecção de anomalias, integrada ao sistema de monitoramento. Ao analisar fluxos de dados operacionais em tempo real (ex.: medições de PMUs, estados do sistema), o modelo identifica padrões sutis de ataques de falsificação de dados – que burlam métodos convencionais baseados em regras – graças à sua capacidade de: (i) aprender representações não lineares do comportamento normal do sistema (via *autoencoder*), e (ii) ponderar dinamicamente a relevância de variáveis específicas (via *multi-head attention*), aumentando a precisão na detecção. Programas de treinamento e conscientização para o pessoal complementam essa abordagem, garantindo que os operadores interpretem adequadamente os alertas gerados. Juntas, essas estratégias possibilitam a criação de um sistema de segurança abrangente que protege a integridade, confidencialidade e disponibilidade das operações das redes inteligentes.

2.1.2 Vulnerabilidades dos Sistemas Ciberfísicos

A interconexão dos componentes físicos com redes de comunicação em um CPS amplia a superfície de ataque, oferecendo múltiplos vetores para exploração de vulnerabilidades nas diversas camadas do sistema, incluindo sensores, atuadores, dispositivos de controle e redes de comunicação. A seguir, são apresentadas as principais vulnerabilidades dos sistemas ciberfísicos em redes de potência:

1. **Vulnerabilidades em Sensores e Atuadores:** Sensores são responsáveis por coletar medições físicas, como tensão, corrente e frequência em redes de energia elétrica, enquanto atuadores executam comandos de controle para ajustar esses parâmetros. Em ataques FDI, um adversário pode comprometer os sensores e enviar dados manipulados para o sistema de controle, alterando o estado percebido da rede elétrica. Essa manipulação pode resultar em decisões incorretas, como ajustes indevidos de tensão ou ativação de geradores de reserva (ZHU; JOSEPH; SASTRY, 2011). Além disso, um ataque coordenado a atuadores pode alterar a dinâmica do sistema de potência, induzindo oscilações ou instabilidades na rede.

2. **Vulnerabilidades em Protocolos de Comunicação:** A comunicação entre os componentes de um CPS ocorre através de protocolos, como o DNP3 (Distributed Network Protocol) e o IEC 61850, que são frequentemente utilizados em redes de potência. Esses protocolos, se não protegidos adequadamente, podem ser alvos de ataques de interceptação e modificação de pacotes. Um atacante pode realizar ações como man-in-the-middle (MitM) para injetar comandos falsos, desviar informações críticas ou atrasar mensagens, afetando a sincronização entre os dispositivos de controle e comprometendo a estabilidade do sistema (KUNDUR et al., 2010).

3. **Vulnerabilidades em Software e Hardware de Controle:** Sistemas SCADA e outros dispositivos de controle são fundamentais para a operação de redes de energia, mas frequentemente apresentam vulnerabilidades que podem ser exploradas por atacantes. A exploração de falhas em software pode permitir que adversários obtenham acesso remoto a dispositivos críticos, alterando os parâmetros de controle de forma maliciosa. Da mesma forma, falhas em hardware, como a execução de *firmwares* vulneráveis ou comprometidos, podem ser exploradas para implantar *malware*, oferecendo aos atacantes controle sobre o funcionamento interno dos dispositivos (KHAZRAEI et al., 2022).

4. **Dependência da Coordenação entre Múltiplas Unidades**: A operação de sistemas de potência envolve a coordenação entre várias unidades de controle distribuídas, como geradores, subestações e centros de controle. Essa coordenação é essencial para a operação segura e estável da rede. Entretanto, um ataque a uma dessas unidades pode ter efeitos cascata, comprometendo outras áreas do sistema interligado. Em um cenário de ataque FDI, a manipulação de medições em uma subestação pode resultar em ajustes inadequados de tensão e corrente em outros pontos da rede, potencialmente levando a apagões regionais (TEIXEIRA et al., 2015).

5. **Ameaças Internas**: Além de ataques externos, os sistemas ciberfísicos são vulneráveis a ameaças internas, que podem ser introduzidas por indivíduos com acesso legítimo ao sistema. Essas ameaças são especialmente perigosas devido ao conhecimento privilegiado sobre a infraestrutura e as operações do sistema, permitindo a manipulação intencional de parâmetros de controle ou a introdução de falhas. Ataques internos podem ser difíceis de detectar, uma vez que o comportamento malicioso pode estar mascarado dentro das operações esperadas (CáRDENAS; AMIN; SASTRY, 2008).

Essas vulnerabilidades tornam evidente a necessidade de estratégias de defesa cibernética que sejam capazes de detectar e mitigar ataques em tempo real. Em particular, a detecção de ataques FDI requer métodos que sejam capazes de identificar discrepâncias entre o estado real do sistema e as medições manipuladas, utilizando técnicas avançadas de aprendizado de máquina, como *autoencoders*, que têm demonstrado eficácia na identificação de anomalias em dados multivariados de sistemas de potência.

2.2 Estimação de Estados do Sistema Elétrico e o Ataque de Injeção de Dados Falsos

Para um controle e operação seguros e econômicos dos sistemas de energia é necessário que haja um sistema de monitoramento eficiente e confiável. Desta forma,

para atingir esse objetivo, os sistemas de energia modernos apresentam estimadores de estado. Schweppe e Wildes (1970) definem um estimador de estado como uma ferramenta matemática que, a partir de um conjunto de medições coletadas em um sistema de energia, calcula as melhores estimativas possíveis das variáveis de estado do sistema. Essas variáveis de estado, tipicamente, incluem as tensões (magnitudes e ângulos de fase) em todas as barras (ou nós) da rede elétrica. Na formulação do problema de estimação de estado, os autores utilizam um modelo matemático baseado nas equações de fluxo de potência que descrevem o comportamento elétrico da rede.

A escolha da abordagem de estimação está intrinsecamente ligada ao modelo de fluxo de potência adotado, resultando em diferentes compromissos entre precisão e eficiência computacional. A literatura consolida principalmente três abordagens:

- **Estimador CA (fluxo de potência CA)**: também conhecido como estimador de fluxo de potência completo ou não-linear, representa a abordagem mais precisa e abrangente para a estimação de estado. Ele se baseia no modelo completo do sistema elétrico, incorporando as complexas relações não-lineares entre as potências ativa e reativa e as tensões nos barramentos (módulos e ângulos). A formulação matemática envolve a minimização de um resíduo ponderado das medições, utilizando um conjunto de equações não-lineares que descrevem o fluxo de potência.

  A principal vantagem do Estimador CA reside na sua alta precisão, pois ele captura fielmente o comportamento do sistema, incluindo perdas de potência e o acoplamento entre as parcelas ativa e reativa. Isso o torna ideal para aplicações onde a exatidão é primordial, como na detecção de erros de medição (BDD) e na análise detalhada da segurança e estabilidade do sistema. No entanto, essa precisão vem com um custo computacional significativo. A resolução do problema de minimização requer métodos iterativos, como o algoritmo de Newton-Raphson, que demandam maior tempo de processamento e recursos computacionais. A convergência pode ser um desafio em sistemas com condições operacionais extremas ou em redes muito grandes, o que limita sua aplicação em cenários que exigem respostas em tempo real muito rápidas.

- **Estimador Desacoplado**: alternativa que busca um equilíbrio entre precisão e eficiência computacional. Esta abordagem baseia-se na premissa de que os fluxos de potência ativa e reativa são fracamente acoplados em sistemas de transmissão, permitindo uma simplificação das equações do fluxo de potência. Essencialmente, as equações de potência ativa são predominantemente relacionadas aos ângulos de fase das tensões, enquanto as equações de potência reativa estão mais ligadas aos módulos das tensões.

  Essa separação permite resolver o problema de estimação em duas etapas desacopladas, ou de forma iterativa com matrizes jacobianas simplificadas. A principal

vantagem é a redução substancial da complexidade computacional em comparação com o Estimador CA, tornando-o mais adequado para sistemas de grande porte e para aplicações que exigem um tempo de resposta mais rápido. Embora seja menos preciso que o Estimador CA completo, a perda de precisão é frequentemente aceitável para muitas aplicações operacionais, especialmente quando a velocidade de cálculo é um fator crítico. Sua aplicação é comum em centros de operação de sistemas de potência, onde a necessidade de monitoramento contínuo e atualizações frequentes do estado do sistema justifica o compromisso com a precisão em favor da eficiência.

- **Estimador CC (fluxo de potência CC)**: ou estimador de fluxo de potência linearizado, representa a abordagem mais simplificada e computacionalmente eficiente. Ele se baseia em uma série de linearizações e simplificações do modelo de fluxo de potência, assumindo que: (i) as perdas de potência nas linhas são desprezíveis; (ii) os módulos das tensões nos barramentos são próximos de 1,0 pu; (iii) as diferenças angulares entre barramentos adjacentes são pequenas; e (iv) a potência reativa é desconsiderada na formulação do problema.

  Essas simplificações transformam o problema não-linear em um sistema de equações lineares, o que permite uma resolução direta e extremamente rápida. A principal vantagem do Estimador CC é sua excepcional eficiência computacional, tornando-o ideal para aplicações que exigem velocidade máxima, como estudos preliminares de planejamento, análises de contingência em tempo quase real, otimização de despacho econômico e em sistemas de distribuição onde a complexidade do modelo CA completo pode ser excessiva para os recursos disponíveis. No entanto, essa eficiência vem ao custo de uma menor precisão, uma vez que as simplificações podem introduzir erros significativos, especialmente em sistemas com grandes variações de tensão ou perdas consideráveis. É importante reconhecer suas limitações e aplicá-lo em contextos onde a precisão reduzida é aceitável em troca da agilidade.

Estes métodos diferem essencialmente no compromisso entre precisão e desempenho computacional, sendo a escolha dependente dos requisitos específicos da aplicação considerada. Como será apresentado no Capítulo 3 a maioria dos trabalhos desenvolvidos no contexto de ataques de injeção de dados falsos concentra-se nos estimadores de estado CC. Essa classe de estimadores é amplamente utilizado devido à sua simplicidade e eficiência computacional. Entretanto, apesar da análise do fluxo de potência CC ser útil em estudos de planejamento e análises de contingência, onde a rapidez e a simplicidade são essenciais, são dependentes de determinadas condições e, portanto os estimadores de estado CC podem introduzir erros significativos em certos casos, especialmente em redes com alta carga ou em situações de alta e extra alta tensão (STOTT; JARDIM; ALSAC, 2009).

Desta forma, com o avanço contínuo no poder computacional, as vantagens de

utilizar o modelo de fluxo de potência CC em vez do estimador de estado CA podem se tornar quase irrelevantes em termos de carga computacional. O estimador de estado CA é fundamentado em um modelo de medição não linear (ALMUTAIRY, 2022):

$$z = h(x) + e \tag{2.1}$$

onde $z = (z_1, z_2, ..., z_m)^T$ representa o vetor de medições (magnitude de tensão, ângulo de tensão, fluxo de potência, corrente etc.), $x = (x_1, x_2, ..., x_n)^T$ o vetor de estados, h é uma função vetorial não linear que relaciona as medições com o estados do sistema e $e = (e_1, e_2, ..., e_m)^T$ o vetor de erros aleatórios de medições que seguem uma distribuição gaussiana com média 0 e uma variância correspondente matriz $R_z = diag(\sigma_1^2, \sigma_2^2, ..., \sigma_m^2)$.

Como apresentado no trabalho de Schweppe e Wildes (1970), o problema de estimativa de estado pode ser resolvido como um problema de mínimos quadrados ponderados (WLS - *Weighted Least Squares*) cuja função objetivo a ser minimizada é dada por:

$$J(x) = \sum_{i=1}^{m} (\frac{z_i - h_i(x)}{\sigma_i^2})^2 \tag{2.2}$$

Ao aplicar a condição de otimalidade de primeira ordem ao índice de desempenho $J(x)$, é gerado um esquema iterativo de Gauss-Newton, que fornece a estimativa ótima do estado do sistema $\hat{x}$.

O problema WLS assume que os erros de medição (e) seguem uma distribuição gaussiana, entretanto os sistemas de energia são formados por ações dinâmicas que apresentam fatores muitas vezes incontroláveis o que pode ocasionar um aumento no erro dos dados de entrada. Desta forma, para reduzir os erros de medição uma grande quantidade de técnicas de detecção de dados ruins (BDD - *Bad Data Detection*) são apresentadas na literatura, onde as mais comuns são voltadas para a análise do vetor residual (r). Após o processo de determinação da estimativa ótima do estado do sistema, o vetor residual r é definido como:

$$r = z - h(\hat{x}) \tag{2.3}$$

As medições de sensores utilizadas na estimação de estado podem ser imprecisas devido a falhas de configuração, defeitos nos dispositivos, ações maliciosas ou outros erros, o que pode impactar negativamente a estimativa das variáveis de estado. Portanto, é de grande importância que os operadores de sistemas de potência detectem e identifiquem medições incorretas. Diversos esquemas já foram propostos para a detecção, identificação e correção dessas medições. Medições com alto nível de ruído, polaridade invertida ou falhas no medidor apresentam valores elevados de erro residual, o que as torna fáceis

de serem identificadas pela BDD e assim removidas do processo de estimação de estado (MONTICELLI, 1999).

Um ataque FDI é aquele onde o atacante manipula sensores ou dispositivos de medição induzindo uma mudança arbitrária no valor da variável de estado sem ser detectado pela BDD do estimador de estado do sistema, podendo ser de forma direcionada ou aleatória (BOBBA et al., 2010). Nos ataques aleatórios, o atacante injeta erros de forma aleatória nas estimativas das variáveis de estado, enquanto em ataques direcionados, erros específicos são injetados nas estimativas de variáveis de estado específicas. Porém, uma categorização mais relevante se baseia no nível de conhecimento que o adversário possui sobre o sistema antes de executar o ataque, sendo então definidos como completos ou incompletos (YU; HOU; LI, 2018). Os ataques FDI completos ocorrem quando o atacante tem conhecimento total sobre o sistema elétrico no qual está atacando e se apresentam como um cenário crítico em relação aos incompletos pois resultam em estimativas incorretas dos estados do sistema, conseguindo driblar as técnicas convencionais de detecção baseadas em resíduos BDD. Isso pode levar os operadores do sistema a adotar ações de controle que coloquem em risco a operação segura dos sistemas de energia. Devido à sua natureza furtiva, esse tipo de FDI é comumente chamado de "invisível" (ALMUTAIRY, 2022). Este tipo de ataque é altamente furtivo e pode causar inúmeros danos aos sistemas de energia, onde como exemplo tem-se o ataque ocorrido na Ucrânia em 2015 (LIANG et al., 2017).

O modelo matemático de FDIs invisíveis contra a estimação de estado CA é o seguinte: sendo $\hat{x}_a$ a estimativa do estado do sistema obtida a partir do processamento do conjunto comprometido de medições $z_a = z + a$. O vetor de ataque $a = (a_1, a_2, ..., a_m)^T$ representa os desvios nas medições causados por um FDI. Sob a influência de FDIs, o vetor residual pode ser calculado da seguinte forma:

$$\begin{aligned} r_a &= z_a - h(\hat{x}_a) \\ &= z + a - h(\hat{x}_a) + h(\hat{x}) - h(\hat{x}) \\ &= z - h(\hat{x}) + a - h(\hat{x}_a) + h(\hat{x}) \\ &= r + a - h(\hat{x}_a) + h(\hat{x}) \end{aligned} \tag{2.4}$$

que, em geral, difere do vetor residual padrão. Contudo, se o vetor de ataque for formulado como:

$$a = h(\hat{x}_a) - h(\hat{x}) \tag{2.5}$$

o vetor residual não é alterado quando comparado a uma situação normal ($r_a = r$). Ou seja, quando a função vetorial $h(\cdot)$ é conhecida, bem como a estimativa de estado do sistema $\hat{x}$ o atacante pode criar um vetor de ataque que resulta em uma estimativa falsa do estado do sistema $\hat{x}_a$ que burla as técnicas convencionais de detecção baseadas em resíduos BDD.

Os ataques FDI têm a capacidade de desestabilizar ou até mesmo causar a interrupção total dos sistemas de controle distribuído que sustentam as redes elétricas modernas. A falsificação de linhas de comunicação e a alteração de níveis de sinal em uma rede de comunicação distribuída são utilizadas para executar ataques FDI, que visam modificar a saída de sistemas de corrente alternada (CA), resultando em danos aos equipamentos. O intervalo de dados maliciosos injetados é escolhido dentro da faixa do estado operacional nominal do sistema, a fim de manter o ataque FDI indetectável. Diversos tipos de ataques FDI podem ser realizados contra os sinais de *feedback* dos controladores distribuídos, por meio da injeção de uma variedade de dados falsos na rede de comunicação.

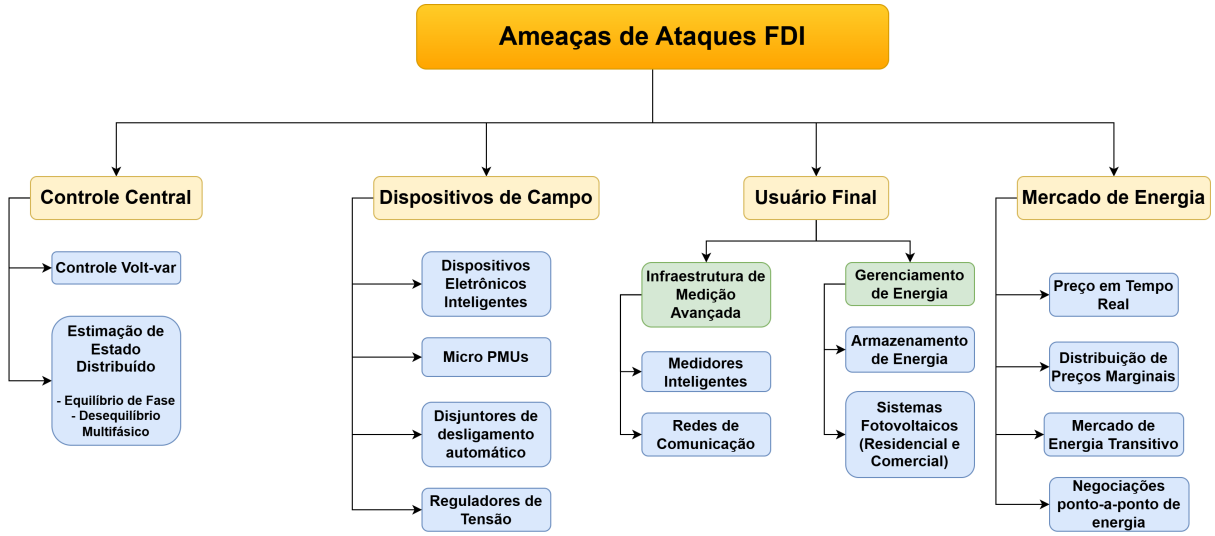
2.2.1 Ameaças dos ataques injeção de dados falsos

Os primeiros a estudar os ataques de dados falsos no contexto de sistemas de energia foram Liu, Ning e Reiter (2011), demonstrando que os ataques FDIs bem elaborados podem contornar os mecanismos tradicionais de detecção de dados incorretos nos sistemas de estimativa de estado de redes elétricas inteligentes. Esses ataques têm o potencial de causar consequências prejudiciais ao domínio físico da rede elétrica, comprometendo sua operação normal. Quando o atacante tenta realizar um ataque FDI, sensores ou dispositivos diversos baseados em IoT comprometem os conjuntos de dados de maneira discreta e introduzem procedimentos de agregação de dados. O atacante injeta um vetor de ataque e, ao mesmo tempo, apresenta a detecção baseada em resíduos BDD nas medições, evitando a detecção pelos operadores. Se o ataque FDI conseguir manipular os conjuntos de dados de medição originais, o conjunto de dados do vetor FDI pode ser apresentado, levando o sistema a interpretar erroneamente as informações.

Diversos ataques FDI podem ser introduzidos no sistema de Redes Inteligentes. Os principais pontos onde os ataques FDI podem acontecer nas Redes inteligentes estão ilustrados na Figura 3.

Os ataques FDI têm a capacidade de impactar severamente sistemas de energia, que possuem tanto poder físico quanto valores econômicos. Esses ataques podem ocorrer na distribuição de energia dentro de um sistema de redes inteligentes. Os invasores identificam os nós ideais ao longo da rota de fluxo de energia da rede, que estão interligados à geração, distribuição ou consumo de eletricidade. Neste sistema de distribuição, são utilizadas várias ferramentas de medição, como medidores inteligentes, relés inteligentes e reguladores de tensão, para diferenciar entre os diferentes nós. Todos esses nós comunicam e compartilham informações essenciais para o funcionamento do sistema. Durante a operação de distribuição de energia, os invasores podem aplicar técnicas de engano em diversos nós para falsificar relatórios. Eles tentam injetar dados maliciosos, como informações de energia adulteradas, mensagens de resposta ou solicitações. Quando os dados ou mensagens manipuladas são inseridos na rede, as ferramentas de medição podem produzir um sistema de energia

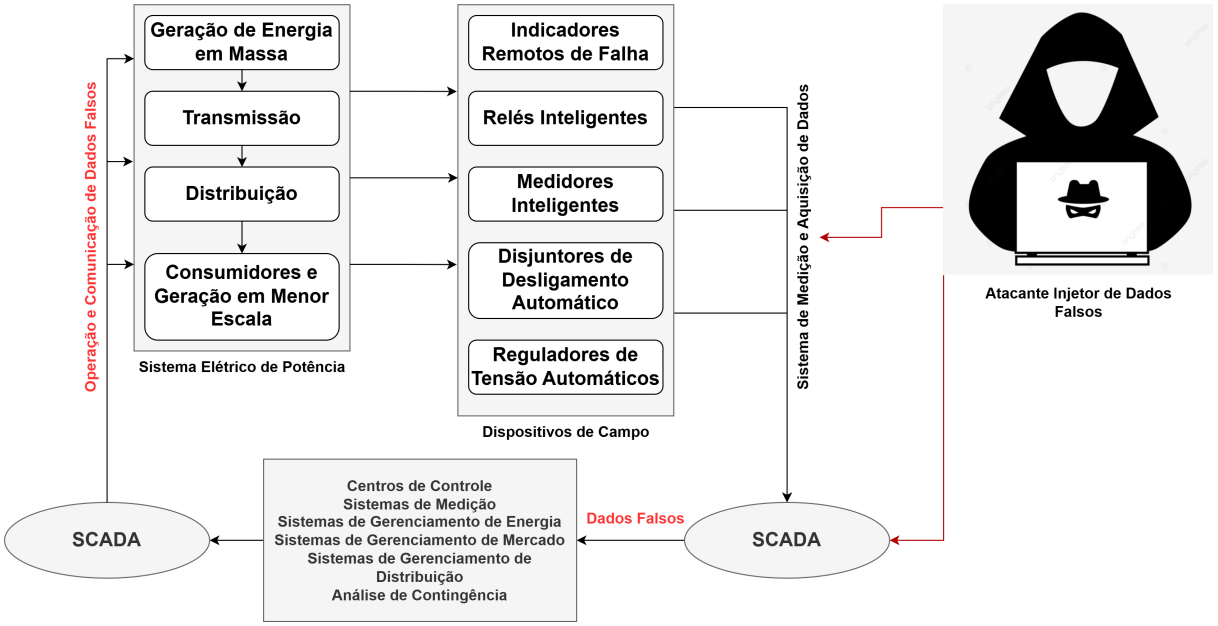
Figura 3 – Ameaças de ataques FDI e seus componentes.



Fonte: Adaptado de Husnoo et al. (2022).

distribuído desequilibrado, baseado em ofertas ou demandas falsas. Como consequência, o custo da energia distribuída pode aumentar significativamente. Quando um ataque de FDI é realizado no sistema de energia, o mercado elétrico sofre impactos, levando sistemas como SCADA, medidores inteligentes e Infraestrutura de Medição Avançada a recalcular os preços de liquidação da energia (HABIB et al., 2023). A Figura 4 apresenta um diagrama de ataques FDI em redes inteligentes.

Figura 4 – Diagrama de ataques FDI em redes inteligentes.



Fonte: Adaptado de Habib et al. (2023).

2.3 Detecção de ataques FDI em sistemas de energia

Diante da criticidade e impactos dos ataques FDI, várias abordagens foram tomadas para detectar esses ataques em redes inteligentes. Embora as abordagens sejam diferentes, dois cenários são observados: métodos baseados em modelos e métodos baseados em dados. A Tabela 2 apresenta um resumo desses métodos.

As primeiras técnicas de detecção de ataques FDI em sistemas de energia apresentadas utilizam métodos matemáticos baseados em modelos do sistema elétrico. Diante das incertezas, a falta de medições confiáveis e a necessidade do modelo exato do sistema são as principais limitações dos métodos de detecção de ataques cibernéticos desta categoria.

A outra categoria apresentada são os métodos baseados em dados. Esses métodos utilizam dados históricos e sinais de medições para aprender as características do sistema. São conhecidos como esquemas inteligentes de detecção e alguns dos modelos mais populares são: aprendizado por reforço, aprenzidado profundo, redes adversariais generativas. Essas técnicas constituem o que são conhecidos por métodos de aprendizado não-supervisionado, semi-supervisionado, supervisionado e por reforço.

Tabela 2 – Classificação dos Algoritmos de Detecção de Ataques FDI.

Categoria	**Tipo de Algoritmo**	**Métodos**
Baseados em Modelos	Algoritmos baseados em Estimação de estado	a) Estimação Estática
		b) Teste de Detecção
		c) Estimação Dinâmica
	Algoritmos sem Estimação de estado	–
Baseados em Dados	Baseados em Aprendizado de Máquina	a) Aprendizado Supervisionado
		b) Aprendizado não-supervisionado
		c) Aprendizado por reforço
	Baseados em mineração de dados e dados do sistema	–

2.3.1 Detecção baseada em modelos

As redes inteligentes são modeladas como sistemas ciberfísicos que integram medições em tempo real com dados estáticos do sistema. Esses dados estáticos incluem parâmetros do sistema, como a configuração das subestações, topologia de rede, características das linhas de transmissão e informações sobre os equipamentos envolvidos. Manipulações nestes dados são uma forma de ataque FDI (MUSLEH; CHEN; DONG, 2020). O modelo da rede pode variar de acordo com as condições operacionais, podendo ser representado de forma quase estática ou de natureza dinâmica (ZHAO et al., 2019).

O modelo quasi-estático representa um cenário em que os pontos operacionais do sistema mudam de forma gradual e suave, assumindo que os controladores no sistema respondem de maneira instantânea. Essa suposição minimiza ou elimina as respostas transitórias, que geralmente ocorrem em sistemas dinâmicos, permitindo uma análise simplificada do comportamento do sistema. Nesses casos, o sistema pode ser tratado como se

estivesse em um estado quase estacionário o tempo todo, com mudanças ocorrendo devagar o suficiente para que não introduzam grandes flutuações ou instabilidades. No contexto de redes inteligentes, onde vários componentes e subsistemas são fortemente integrados e exigem monitoramento constante, essa abordagem de modelagem é particularmente útil para capturar o desempenho geral do sistema sem a necessidade de considerar eventos dinâmicos de curta duração. Isso resulta em um modelo operacional mais estável e previsível (ZHAO et al., 2019).

Sob a suposição quasi-estática, os diferentes subsistemas da rede inteligente—desde a geração e distribuição de energia até o gerenciamento de cargas—podem ser modelados usando uma estrutura geral de medição. Essa estrutura normalmente incorpora tanto dados em tempo real (por exemplo, de sensores e medidores inteligentes) quanto parâmetros estáticos do sistema (como impedâncias de linha e configurações de transformadores), proporcionando uma visão abrangente do desempenho da rede em qualquer momento. O modelo geral de medição sob essa estrutura pode ser expresso como apresentado na Equação 2.1 (MUSLEH; CHEN; DONG, 2020; ALMUTAIRY, 2022).

O modelo de natureza dinâmica é adotado quando transientes ou mudanças rápidas no sistema precisam ser considerados. Nessa abordagem, os estados do sistema dependem não apenas das medições e dados atuais, mas também dos estados anteriores, permitindo uma análise mais precisa em cenários de variações rápidas, como alterações de carga ou inserções de geração distribuída. Esse modelo é especialmente útil em condições operacionais mais voláteis, acompanhando em tempo real as rápidas mudanças nas variáveis do sistema e interagindo com dispositivos inteligentes para otimização e controle eficiente. Esse modelo pode ser representado da seguinte forma (MUSLEH; CHEN; DONG, 2020):

$$\begin{cases} z_t = h(x_t) + e \\ x_t = f(x_{t-1}) + v \end{cases} \tag{2.6}$$

onde t representa o instante de tempo; $f(.) \in \mathbb{R}^m$ são funções não lineares dependentes do sistema que estabelecem a relação entre o vetor de estado x_t e o vetor de estado anterior x_{t-1}; e $v \in \mathbb{R}^m$ é o termo de erro que captura tanto a discretização do tempo quanto o erro de aproximação do modelo, possuindo média zero e variância $R \in \mathbb{R}^m$.

Dados errôneos nas medições da rede elétrica inteligente podem surgir de causas naturais, como falhas de dispositivos, ou de ataques intencionais, como ataques FDI . Esses erros podem ser modelados como um vetor de medição malicioso $z_t^\alpha = z_t + \alpha_t$, onde α_t representa os dados errôneos. O método tradicional para detectar esses erros envolve um teste de resíduo, onde o resíduo $r_t = |z_t - h(x_t)|$ é comparado a um limite pré-definido τ. Se $r_t \geq \tau$, um dado errôneo é identificado. Embora esse teste detecte efetivamente ataques observáveis (denominados "FDI Básico"), muitas vezes falha contra ataques mais sofisticados e não observáveis ("FDI *Stealth*"). Um FDI *stealth* consiste em

injeções coordenadas de vetores de medição maliciosos que imitam a covariância normal no sistema. Para que esses ataques tenham sucesso, o adversário deve acessar medições de múltiplos sensores e ter conhecimento da estrutura e parâmetros do sistema. Um exemplo inclui a manipulação das magnitudes de tensão em uma seção da rede elétrica inteligente. Para abordar um FDI, várias algoritmos de detecção foram propostos, categorizados em métodos de detecção baseados em estimativa—que consistem em estimativa de estados e comparação de medições—e métodos de cálculo direto que se baseiam apenas em medições do sistema, sem estimativas prévias.

2.3.2 Detecção baseada em dados

Apesar de os métodos baseados em modelos terem sido amplamente utilizados para a detecção de anomalias em sistemas de potência, sua eficácia é limitada em cenários que envolvem ataques sofisticados, como os ataques FDI. Esses métodos dependem de um conhecimento preciso da topologia da rede e dos parâmetros do sistema para construir modelos matemáticos que descrevem o comportamento esperado das medições. No entanto, em situações onde há incertezas nos parâmetros da rede, variações nas condições de operação ou a presença de dados manipulados que mantêm a consistência com as leis físicas do sistema, esses métodos enfrentam dificuldades em distinguir entre falhas genuínas e ataques (LIU; NING; REITER, 2011). Além disso, a necessidade de atualizações contínuas do modelo para acompanhar mudanças na rede pode ser um desafio operacional, tornando a abordagem menos flexível e mais suscetível a falhas na detecção de anomalias sutis (ZHANG et al., 2020). Devido a essas limitações, métodos baseados em dados têm sido cada vez mais explorados como alternativas para superar essas deficiências, utilizando técnicas de aprendizado de máquina e análise estatística para identificar padrões anômalos diretamente nos dados históricos e em tempo real, sem a mesma dependência de um modelo matemático.

Diferentemente dos algoritmos de detecção baseados em modelos, que utilizam parâmetros específicos do sistema, os algoritmos de detecção baseados em dados operam de forma independente de qualquer modelo do sistema elétrico. Isso significa que, no processo de detecção de FDIs, não são considerados nem os parâmetros do sistema nem os modelos subjacentes.

2.3.2.1 Algoritmos baseados em Aprendizado de Máquina

O aprendizado de máquina (ML) é um campo abrangente dentro da inteligência artificial (IA). Com o uso de ML, é possível empregar máquinas treinadas para executar tarefas complexas, como a detecção de FDIs em redes inteligentes. Diferentemente dos algoritmos de detecção de FDI que se baseiam em modelos, os algoritmos baseados em ML dependem dos dados coletados do sistema. Essa abordagem orientada por dados

implica uma forte dependência de dados históricos do sistema em análise para viabilizar o aprendizado da máquina. Os algoritmos de detecção que utilizam ML incluem métodos de aprendizado supervisionado, não supervisionado e por reforço.

O aprendizado supervisionado é um método em que modelos de máquina são treinados utilizando dados rotulados, onde cada entrada é associada a uma saída específica. Essa abordagem é representada como $\{(s_i, y_i)\}$, onde $s_i \in \mathbb{R}^n$ denota a i-ésima amostra de medições e $y_i \in \{-1, 1\}$ é o rótulo correspondente, identificando se a amostra é normal ou se contém um ataque FDI. Essa classificação é fundamental para a efetividade dos algoritmos de detecção, pois permite que a máquina aprenda a distinguir entre comportamentos normais e anômalos do sistema. No contexto da detecção de FDI em redes inteligentes, diversas abordagens de aprendizado supervisionado têm sido exploradas. Esses métodos utilizam um conjunto de dados rotulados, que geralmente consiste em medições históricas do sistema, para treinar modelos que podem reconhecer padrões associados a ataques FDI. As técnicas de aprendizado supervisionado podem incluir algoritmos como árvores de decisão (VIMALKUMAR; RADHIKA, 2017), máquinas de vetor de suporte (SVM) (ESMALIFALAK et al., 2017; YAN; TANG; HE, 2016) e redes neurais artificiais (KHANNA; PANIGRAHI; JOSHI, 2018; AYAD et al., 2018; VIMALKUMAR; RADHIKA, 2017), entre outros. Cada um desses métodos oferece diferentes vantagens em termos de precisão, eficiência computacional e capacidade de generalização.

O aprendizado não supervisionado é uma técnica de ML que utiliza dados não rotulados para identificar padrões e estruturas ocultas. Ao analisar as características intrínsecas dos dados, a máquina organiza os pontos em grupos com base em similaridades, permitindo a detecção de anomalias sem supervisão prévia. Em redes inteligentes, essa abordagem é particularmente útil para identificar ataques FDI, que podem se manifestar como anomalias sutis. Utilizando algoritmos de agrupamento ou detecção de anomalias, o aprendizado não supervisionado pode classificar os dados de medição em categorias normais e anômalas. Dessa forma, ele ajuda a identificar proativamente ameaças à integridade do sistema, contribuindo para a segurança e confiabilidade das operações em redes de energia. Exemplos de tais técnicas incluem K-Means *Clustering* (ZANETTI et al., 2019), que agrupa dados em clusters e identifica pontos fora dos grupos como anomalias; Redes Neurais *Autoencoder* (HONGSHAN et al., 2018), que identificam erros de reconstrução altos quando confrontadas com dados manipulados; e *Isolation Forest* (AHMED et al., 2019), que isola pontos de dados em árvores de decisão para identificar anomalias. Essas abordagens são particularmente valiosas em cenários com escassez de dados rotulados, permitindo a detecção de comportamentos anômalos e a identificação de potenciais ataques.

No aprendizado por reforço, a máquina busca aprender a ação ideal com base nas experiências adquiridas em ações anteriores. Diferentemente do aprendizado supervisionado, onde os dados das amostras são utilizados no treinamento, o aprendizado por reforço se

baseia em um processo de tentativa e erro. Assim, uma sequência de escolhas bem-sucedidas resultará no reforço do processo, à medida que resolve o problema de forma eficiente. Comparado aos tipos de aprendizado supervisionado e não supervisionado, o aprendizado por reforço ainda está em estágio inicial de investigação na área de detecção de ataques FDI em redes inteligentes e agregará vários benefícios nesse contexto (KURT et al., 2018).

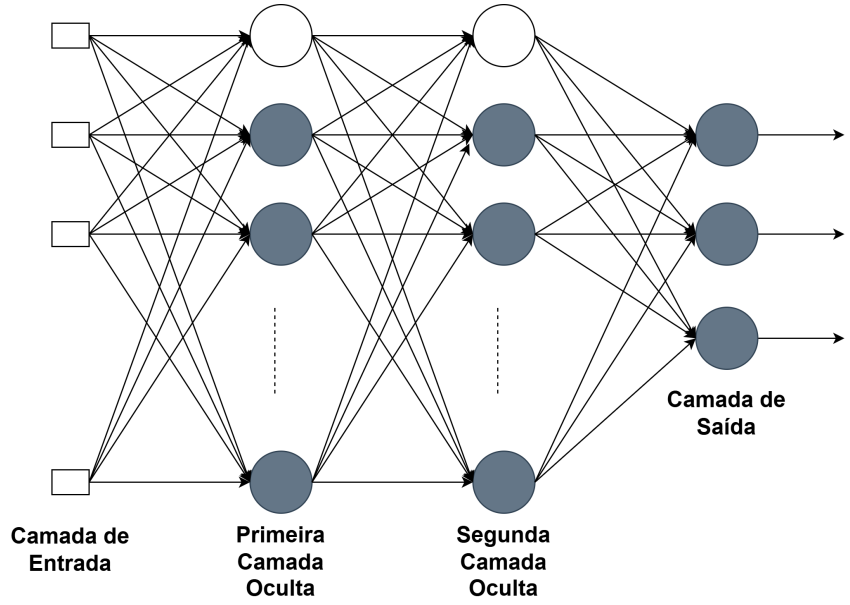
2.4 *Deep Learning*

Deep Learning, ou aprendizado profundo, é uma subárea do aprendizado de máquina (*Machine Learning*) que se baseia em redes neurais artificiais com múltiplas camadas para extrair representações de alto nível a partir de dados complexos. Essa abordagem ganhou notoriedade nas últimas décadas devido ao seu desempenho superior em tarefas de reconhecimento de padrões, como visão computacional, processamento de linguagem natural (PLN) e análise de séries temporais (CHOLLET, 2017). Diferentemente das abordagens de aprendizado de máquina convencionais, que geralmente dependem de extração manual de características (*feature engineering*), o *deep Learning* é capaz de aprender automaticamente representações diretamente a partir dos dados brutos, como imagens ou texto, o que reduz significativamente a necessidade de intervenção humana (GOODFELLOW; BENGIO; COURVILLE, 2016).

Uma das características fundamentais do *deep Learning* é a utilização de redes neurais artificiais (RNA), que possuem múltiplas camadas ocultas entre as camadas de entrada e saída, como apresentado na Figura 5. Essas camadas permitem que o modelo aprenda de forma hierárquica, extraindo características cada vez mais abstratas conforme os dados passam por diferentes níveis da rede. Por exemplo, em uma rede neural convolucional (CNN) usada para reconhecimento de imagens, as primeiras camadas podem detectar bordas e texturas simples, enquanto camadas mais profundas são capazes de identificar objetos inteiros e padrões complexos (LECUN; BENGIO; HINTON, 2015). A hierarquia de aprendizado é um dos principais fatores que tornam as redes neurais profundas capazes de lidar com a alta dimensionalidade dos dados e de capturar relações não lineares complexas.

O treinamento de redes neurais profundas é baseado no método de *backpropagation*, que ajusta os pesos dos neurônios ao longo das várias camadas, minimizando o erro do modelo em relação aos dados de treinamento. A combinação de *backpropagation* com algoritmos de otimização, como *gradient descent* e suas variantes, permite que a rede se adapte aos padrões dos dados de forma eficiente (RUMELHART; HINTON; WILLIAMS, 1986). Isso faz com que as redes profundas sejam capazes de alcançar uma performance elevada em problemas complexos, como visão computacional e processamento de linguagem natural (PLN). Além disso, funções de ativação não lineares, como a ReLU (*Rectified*

Figura 5 – Estrutura básica de uma RNA.



Fonte: Adaptado de Fleck et al. (2016).

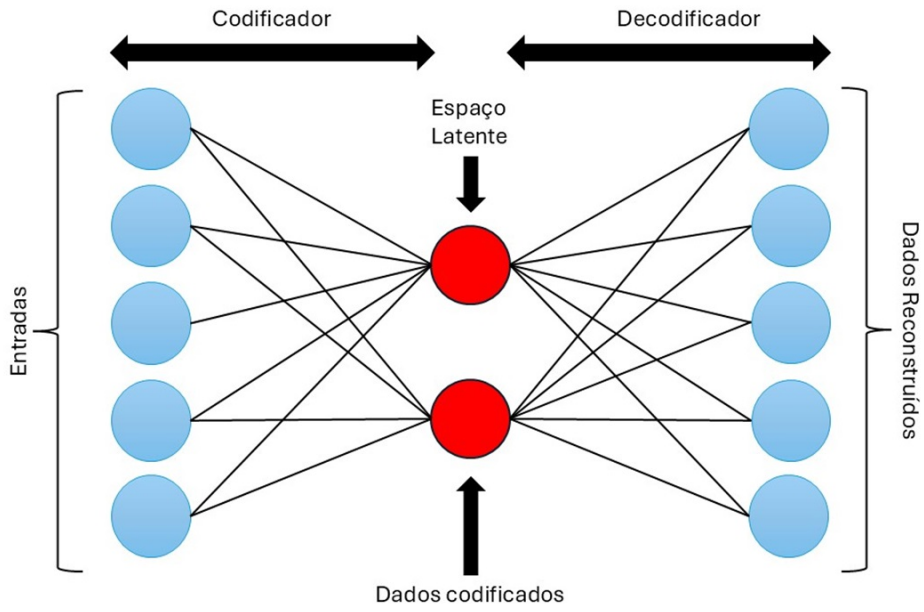
Linear Unit), são utilizadas para introduzir não linearidade ao modelo, o que amplia a capacidade da rede de aprender representações complexas (NAIR; HINTON, 2010).

O *deep Learning* tem sido amplamente aplicado na detecção de anomalias, oferecendo uma abordagem eficiente para identificar padrões incomuns em grandes volumes de dados. Redes neurais profundas, como *autoencoders* e redes convolucionais, são capazes de aprender representações complexas de dados normais e, assim, detectar desvios significativos que indicam anomalias. A detecção de anomalias por *deep Learning* é especialmente útil em contextos como segurança cibernética, monitoramento de sistemas industriais e análise de fraudes, pois permite identificar padrões sutis e não lineares que métodos tradicionais poderiam não detectar (CHALAPATHY; CHAWLA, 2019).

2.4.1 *Autoencoders*

Um *Autoencoder* (AE) apresentado na Figura 6 é um subconjunto do *deep learning*, caracterizado como uma rede neural não supervisionada e totalmente conectada proposta inicialmente por Hinton e Salakhutdinov (2006) para realizar aprendizado de máquina por meio de um processo de retropropagação. Originalmente, os AEs foram desenvolvidos como modelos para compressão de dados, onde transformam entradas em representações de dimensões reduzidas. Este processo de compressão é seguido pela descompressão, em que um decodificador é empregado para recuperar a entrada original a partir da versão codificada. O princípio fundamental que permite aos AEs realizar compressão e descompressão eficazmente reside na capacidade de explorar a correlação entre as diferentes entradas de dados.

Figura 6 – Arquitetura de um *Autoencoder*.



Fonte: Autor.

O processo de redução de dimensionalidade, que mapeia os dados de entrada de dimensão d_0 para o espaço latente B através das camadas ocultas H_1 a H_n, é chamado de *encoder* (codificador). O *encoder* transforma a entrada x em uma representação latente z. Subsequentemente, o *decoder* (decodificador) descomprime o código para produzir dados de saída com dimensão d_0. As matrizes de pesos W e o vetor de bias b são usados nos processos de codificação e decodificação, conforme (2.7) e (2.8):

$$Z = \sigma\left(W_e^{(n)}\left(\ldots\sigma\left(W_e^{(0)}X + b_e^{(0)}\right)\ldots\right) + b_e^{(n)}\right) \tag{2.7}$$

$$\hat{Y} = \sigma\left(W_d^{n}\left(\cdots\sigma\left(W_d^{0}Z + b_d^{0}\right)\cdots\right) + b_d^{n}\right) \tag{2.8}$$

onde W_e^n e W_d^n denotam as matrizes de pesos para os processos de codificação e decodificação, respectivamente. b_e^n e b_d^n são os vetores de bias, e σ representa uma função de ativação não linear aplicada elemento a elemento. X refere-se ao vetor de dados de entrada, Y é o dado no espaço latente, e $\hat{Z}$ representa os dados de saída.

Além de sua aplicação em compressão de dados, os *Autoencoders* têm se mostrado promissores na detecção de ataques FDI (KUNDU et al., 2020). A capacidade dos AEs de reconhecer padrões de correlação nas amostras de dados é valiosa, pois amostras atacadas mostram discrepâncias notáveis entre a entrada e a saída do modelo. Essa diferença se torna um indicador valioso de que os dados foram manipulados. Uma das principais vantagens

dos AEs em comparação com outros algoritmos de aprendizado de máquina é que eles não exigem dados rotulados para treinamento, o que os torna uma opção atraente em cenários onde os dados rotulados são escassos ou difíceis de obter. O uso de AEs permite a redução do tamanho do vetor de características de maneira não supervisionada, mantendo a essência dos dados. Além disso, a habilidade dos AEs em aprender com dados limpos não rotulados também os capacita a identificar outros tipos de ataques, aumentando ainda mais sua utilidade em sistemas de segurança cibernética. Essa combinação de características torna os *Autoencoders* uma ferramenta poderosa para a detecção de anomalias e proteção contra fraudes em ambientes complexos, como as redes inteligentes.

Quando essa arquitetura apresentada na Figura 6 é composta por várias camadas não-lineares, o *autoencoder* passa a ser chamado de *autoencoder* profundo (*deep autoencoder*) (HINTON, 2007), o que aumenta sua capacidade de representar dados complexos.

2.4.2 Mecanismo de Atenção *Multi-head*

Os *autoencoders* são redes neurais projetadas para aprender representações eficientes e compactas dos dados de entrada, através de uma arquitetura de codificação e decodificação. No entanto, a introdução de mecanismos de atenção nos *autoencoders* tem demonstrado melhorias significativas, especialmente em tarefas que exigem foco em características específicas dos dados, como processamento de linguagem natural, visão computacional e sistemas de recomendação. Os mecanismos de atenção foram inicialmente propostos em redes neurais para tradução automática, especialmente com o modelo *Transformer* (VASWANI et al., 2023), que destaca a capacidade de atenção de processar relações complexas nas entradas. Em um *autoencoder* com mecanismo de atenção, o objetivo é aprender quais partes da entrada devem receber mais "atenção" durante o processo de codificação e decodificação, permitindo que a rede se concentre em regiões ou características mais relevantes dos dados.

A estrutura do mecanismo de atenção em *autoencoders* pode ser resumida em três componentes principais:

- **Camada de Atenção**: Esta camada, inserida na parte do codificador, decodificador, ou ambos, calcula um peso de atenção para cada parte dos dados de entrada, indicando a importância relativa de cada parte no contexto da tarefa. Esses pesos são geralmente obtidos usando operações de similaridade, como produtos escalares ou redes *feedforward* simples (LUONG; PHAM; MANNING, 2015).
- ***Softmax***: Os pesos de atenção são normalmente processados com uma função *softmax*, que normaliza os pesos para que somem 1, criando uma distribuição de probabilidade que destaca quais partes da entrada são mais relevantes.

- **Agregação e Reajuste**: Após calcular e aplicar os pesos de atenção, o *autoencoder* ajusta a representação interna (geralmente no espaço latente) para refletir essas importâncias, permitindo uma reconstrução mais focada e detalhada dos dados.

A atenção multi-head (*Multi-Head Attention*), introduzida por (VASWANI et al., 2023) na arquitetura Transformer, representa um avanço significativo em mecanismos de atenção, permitindo que o modelo capture diversos tipos de dependências em sequências de dados de forma mais eficiente do que abordagens baseadas em atenção única.

O funcionamento da atenção multi-head tem como base o mecanismo de **atenção escalonada por produto interno** (*Scaled Dot-Product Attention*). Nessa abordagem, dadas as matrizes de consulta ($\mathbf{Q}$), chave ($\mathbf{K}$) e valor ($\mathbf{V}$), o cálculo da atenção é realizado através de:

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right)\mathbf{V}, \tag{2.9}$$

onde $\mathbf{Q} \in \mathbb{R}^{n \times d_k}$, $\mathbf{K} \in \mathbb{R}^{m \times d_k}$ e $\mathbf{V} \in \mathbb{R}^{m \times d_v}$, sendo n e m os comprimentos das sequências de entrada, e d_k e d_v as dimensões das chaves/consultas e valores. O termo $\sqrt{d_k}$ evita que os gradientes assumam valores extremamente pequenos durante o treinamento, contribuindo para a estabilidade numérica do modelo.

A atenção multi-head expande esse conceito através de:

$$\mathbf{Q}_i = \mathbf{Q}\mathbf{W}_i^Q, \quad \mathbf{K}_i = \mathbf{K}\mathbf{W}_i^K, \quad \mathbf{V}_i = \mathbf{V}\mathbf{W}_i^V \tag{2.10}$$

onde $\mathbf{W}_i^Q, \mathbf{W}_i^K \in \mathbb{R}^{d_{\text{model}} \times d_k}$ e $\mathbf{W}_i^V \in \mathbb{R}^{d_{\text{model}} \times d_v}$, com $d_k = d_v = \frac{d_{\text{model}}}{h}$.

Cada cabeça computa:

$$\text{head}_i = \text{Attention}(\mathbf{Q}_i, \mathbf{K}_i, \mathbf{V}_i) \tag{2.11}$$

resultando na saída final:

$$\text{MultiHead}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Concat}(\text{head}_1, \ldots, \text{head}_h)\mathbf{W}^O \tag{2.12}$$

com $\mathbf{W}^O \in \mathbb{R}^{hd_v \times d_{\text{model}}}$.

A utilização de atenção multi-head em *autoencoders* oferece vantagens significativas ao permitir a captura eficiente de dependências complexas em diferentes escalas e níveis hierárquicos dos dados. Cada cabeça de atenção pode focar simultaneamente em diferentes aspectos ou regiões da entrada, facilitando a modelagem tanto de padrões locais quanto de relações de longo alcance. Isso resulta em representações latentes mais ricas

e informativas, superando as limitações de *autoencoders* tradicionais na preservação de estruturas complexas durante a compressão e reconstrução dos dados.

A arquitetura ainda proporciona maior flexibilidade e interpretabilidade, adaptando-se facilmente a diversos tipos de dados (sequências, imagens ou grafos) e permitindo a análise dos padrões de atenção aprendidos. Essa abordagem não só melhora o desempenho em tarefas como redução de dimensionalidade e aprendizado de representações, mas também mantém a eficiência computacional característica dos Transformers, tornando-a particularmente útil para aplicações que exigem tanto precisão quanto escalabilidade.

2.4.3 Métricas de Avaliação de Desempenho

Na avaliação do desempenho de um *autoencoder* aplicado à detecção de ataques FDI em sistemas de potência, diversas métricas são fundamentais para medir a eficácia do modelo, pois elas fornecem uma visão detalhada sobre a capacidade do *autoencoder* de diferenciar eventos normais de anomalias nos dados. Esses ataques FDI podem comprometer a segurança e a operação de redes de energia, manipulando medições de sensores para enganar sistemas de controle. Portanto, uma avaliação precisa é essencial para assegurar que o *autoencoder* identifique corretamente essas irregularidades, minimizando alarmes falsos e falhas de detecção. As métricas como Acurácia, Verdadeiros Positivos (VP), Verdadeiros Negativos (VN), Falsos Positivos (FP) e Falsos Negativos (FN) permitem analisar tanto a precisão das detecções quanto a confiabilidade do modelo em um ambiente de alta criticidade, como o sistema de potência. Essa análise detalhada é essencial para ajustar o modelo e garantir sua efetividade na proteção contra ameaças potenciais, ajudando a manter a estabilidade e a segurança da infraestrutura elétrica:

- **Verdadeiro Positivo (VP)**: Refere-se aos casos em que um ataque FDI está presente e é corretamente identificado pelo *autoencoder* como anômalo. Isso demonstra a capacidade do modelo de detectar uma ameaça quando ela de fato ocorre.
- **Verdadeiro Negativo (VN)**: Representa as situações em que o sistema está em operação normal, sem a presença de ataques, e o *autoencoder* classifica corretamente esses casos como normais. Uma alta taxa de VN indica que o modelo não está gerando alarmes desnecessários.
- **Falso Positivo (FP)**: Ocorre quando o *autoencoder* identifica uma situação normal como se fosse um ataque FDI, gerando um alarme falso. Isso pode resultar em respostas desnecessárias e onerosas para investigar situações que não são realmente perigosas.
- **Falso Negativo (FN)**: Refere-se aos casos em que o ataque FDI ocorre, mas o modelo falha em detectá-lo, classificando a situação como normal. Este tipo de erro é

particularmente perigoso, pois permite que um ataque passe despercebido, podendo comprometer a segurança e a integridade do sistema de potência.

A acurácia é uma métrica de avaliação de modelos de classificação que mede a proporção de predições corretas em relação ao total de predições feitas pelo modelo. No contexto da detecção de anomalias, como ataques de FDI, a acurácia indica a capacidade do modelo de identificar corretamente tanto as situações anômalas (ataques detectados) quanto as normais (operações sem ataques). A fórmula para calcular a acurácia é:

$$\text{Acurácia} = \frac{\text{VP} + \text{VN}}{\text{VP} + \text{VN} + \text{FP} + \text{FN}} \tag{2.13}$$

onde:

- VP = Verdadeiros Positivos (ataques corretamente identificados).
- VN = Verdadeiros Negativos (situações normais corretamente classificadas).
- FP = Falsos Positivos (situações normais classificadas incorretamente como ataques).
- FN = Falsos Negativos (ataques não detectados).

A acurácia é útil para ter uma visão geral da *performance* do modelo, mas pode ser limitada em situações com classes desbalanceadas, onde a importância dos falsos negativos (como não detectar um ataque) é crítica, portanto utiliza-se ainda de forma complementar o *recall* e o *f1-score*.

O *recall* mede a capacidade do modelo em identificar corretamente as instâncias positivas entre todas as instâncias realmente positivas. É dado pela fórmula:

$$\textit{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

onde:

- TP é o número de verdadeiros positivos (*True Positives*);
- FN é o número de falsos negativos (*False Negatives*).

O *f1-score* é a média harmônica entre precisão e *recall*, equilibrando as duas métricas para avaliar o desempenho do modelo. É dado pela fórmula:

$$\textit{f1-score} = 2 \cdot \frac{\text{Precisão} \cdot \textit{Recall}}{\text{Precisão} + \textit{Recall}}$$

onde:

- Precisão $= \frac{\text{TP}}{\text{TP+FP}}$ (precisão);
- *Recall* $= \frac{\text{TP}}{\text{TP+FN}}$ (*recall*).

Essas fórmulas ajudam a avaliar a qualidade de um modelo de classificação, especialmente em cenários com classes desbalanceadas.

3 REVISÃO BIBLIOGRÁFICA

Neste capítulo são apresentadas a metodologia de revisão bibliográfica, os trabalhos mais relevantes encontrados e utilizados como referências, o levantamento do estado da arte e as principais lacunas encontradas na literatura.

3.1 Metodologia da Revisão Bibliográfica

A plataforma *Web of Science* foi utilizada como principal fonte de consulta bibliográfica, dado seu reconhecimento internacional como uma das mais abrangentes e respeitadas bases de dados multidisciplinares. Foram utilizados quatro padrões de buscas, apresentados na Tabela 3. Os padrões de busca utilizados visaram refinar os trabalhos acadêmicos que se alinham com o tema da pesquisa.

Tabela 3 – Padrões utilizados na busca na plataforma *Web of Science*.

Padrão	**Descrição da Busca**	**Código utilizado**
I	Detecção de ataques cibernéticos em Sistemas de Potência	Detection AND (of AND (cyber AND (attacks AND (on AND (power AND systems)))))
II	Detecção de ataques de injeção de dados falsos em Sistemas de Potência	Detection AND (of AND (false AND (data AND (injection AND (attacks AND (in AND (power AND systems)))))))
III	Detecção de ataques de injeção de dados falsos em Sistemas de Potência utilizando *Deep Learning*	Detection AND (of AND (false AND (data AND (injection AND (attacks AND (in AND (power AND (systems AND (using AND (deep AND learning))))))))))
IV	Detecção de ataques de injeção de dados falsos em Sistemas de Potência utilizando *Autoencoders*	Detection AND (of AND (false AND (data AND (injection AND (attacks AND (in AND (power AND (systems AND (using AND autoencoders)))))))))

Na Tabela 4 é apresentada a quantidade de documentos encontrados na *Web of Science* a nível geral e em nível Brasil, destacando que os primeiros trabalhos foram realizados e publicados a partir de 2004. Destaca-se também que dos padrões pesquisados, ao total foram encontrados 3.304 documentos e destes apenas 26 são publicações brasileiras. É possível observar que os trabalhos acadêmicos diminuem a cada padrão de busca realizado. O Padrão IV, que se refere ao foco principal deste trabalho apenas 9 trabalhos foram

encontrados, o que corresponde a apenas 0.41% em relação ao Padrão I, 0.88% em relação ao Padrão II e 9.38% em relação ao Padrão III. Além disso, percebe-se que não foram encontrados documentos de publicações brasileiras referentes aos padrões III e IV.

Tabela 4 – Resultados da busca na plataforma *Web of Science*.

Padrão	**Documentos encontrados**	**Publicações do Brasil**
I	2.195	24
II	1.021	7
III	97	0
IV	9	0

3.2 Panorama Geral dos Ataques de Injeção de Dados Falsos (FDIs)

Os ataques de injeção de dados falsos (FDIs) são atualmente um dos ataques cibernéticos mais danosos aos sistemas de potência pela sua capacidade de comprometimento da estimativa de estado do sistema, dificuldades de detecção, impacto direto em decisões operacionais, risco à estabilidade e segurança do sistema, além de alto custo e complexidade de métodos de mitigação. Diante desse cenário, pesquisadores, acadêmicos e membros da indústria vêm ao longo dos últimos anos desenvolvendo algoritmos de detecção de ataques FDI, com intuito de identificar os ataques por meio das vulnerabilidades do sistema a fim de diminuir o impacto e auxiliar nas tomadas de decisão operacionais.

Liu, Ning e Reiter (2011) no trabalho intitulado "*False data injection attacks against state estimation in electric power grids*", introduziram o estudo de ataque de injeção de dados falsos (FDI) em redes de energia inteligentes. O principal objetivo dos autores é propor um modelo teórico que examina como um invasor, ao conhecer a topologia e os parâmetros do sistema de potência, pode injetar dados falsos de maneira a passar despercebido pelos algoritmos de detecção baseados em estimativa de estado. O estudo se foca em modelos de fluxo de potência CC e analisa a eficácia de diferentes estratégias de ataque, considerando tanto o cenário em que o invasor tem conhecimento completo quanto parcial do sistema. O trabalho mostra que é possível construir vetores de ataque que violam a integridade dos dados sem serem detectados. Os autores demonstram matematicamente que, ao explorar a estrutura da matriz de sensibilidade do sistema, um invasor pode injetar dados falsos que escapam das técnicas de detecção de anomalias, comprometendo a acurácia da estimativa de estado. O estudo também mostra que o impacto desses ataques pode ser significativo, levando a decisões operacionais incorretas.

Métodos de detecção de ataques FDIs em sistemas de potência podem ser categorizados em baseados em modelos e baseados em dados. Os algoritmos baseados em modelos, como o método de mínimos quadrados ponderados (WLS) ou filtro de Kalman, requerem um modelo explícito do sistema e seus parâmetros (LI et al., 2020; LIU; HORNIKX, 2018). No entanto, esses métodos são computacionalmente caros e não escaláveis para sistemas de grande porte, além de serem sensíveis a incertezas nos parâmetros do modelo, o que pode levar a detecções falsas. Em contraste, os métodos de detecção baseados em dados se destacam por suprir essas lacunas. Eles aprendem as relações dos dados através de técnicas de mineração de dados e aprendizado de máquina, especialmente *deep learning*, oferecendo a vantagem de serem independentes dos parâmetros do sistema , o que reduz os erros de detecção causados por incertezas de modelagem e parâmetros.

A revisão bibliográfica apresentada neste capítulo tem como foco a análise de trabalhos publicados ao longo dos últimos anos que se voltaram para o estudo de métodos de detecção de ataques FDI baseado em dados, especialmente utilizando *machine learning*, *deep learning* e *autoencoders*.

3.3 Detecção baseada em Aprendizado de máquina (*Machine Learning*)

Esmalifalak et al. (2013) no trabalho intitulado "*Detecting stealthy false data injection using machine learning in smart grid*", abordam a detecção de ataques furtivos de injeção de dados falsos (FDIs) em redes elétricas inteligentes. Os dados foram gerados utilizando o programa MATPOWER para simular a operação de uma rede de potência. As cargas na rede foram consideradas estocásticas, seguindo uma distribuição uniforme. Medições de potência ativa foram coletadas de cada linha de transmissão. Para o sistema de teste IEEE 118-barras, foram geradas 304 características de medição. Utilizou-se simulação de Monte Carlo para registrar o vetor de medição em 1000 instâncias diferentes. A metodologia empregou principalmente Máquinas de Vetores de Suporte (SVM) supervisionadas, treinadas com dados rotulados, e uma abordagem de detecção de anomalias não supervisionada para identificar desvios nas medições. A Análise de Componentes Principais (PCA) foi aplicada para reduzir a dimensionalidade dos dados e otimizar a complexidade computacional. Os resultados em simulações com sistemas de teste padrão da IEEE demonstraram a eficácia das técnicas propostas na detecção de injeções de dados falsos furtivos. O *f1-score* foi a métrica de acurácia utilizada, com o SVM atingindo 0,956 (C = 5,648, sigma = 6,67). As principais limitações incluem a dependência de dados de treinamento rotulados para o SVM e a sensibilidade das abordagens à qualidade e complexidade dos dados. Os autores sugerem expansão para detectar outros tipos de falhas além dos ciberataques.

No estudo intitulado "*Detecting Stealthy False Data Injection Using Machine Learning in Smart Grid*" de Esmalifalak et al. (2017), propõem a detecção de ataques furtivos de injeção de dados falsos (FDIs) em redes elétricas inteligentes utilizando aprendizado de máquina. A metodologia emprega principalmente Máquinas de Vetores de Suporte (SVM) supervisionadas, treinadas com dados rotulados, e uma abordagem de detecção de anomalias não supervisionada para identificar desvios nas medições. A Análise de Componentes Principais (PCA) é aplicada para reduzir a dimensionalidade dos dados e otimizar a complexidade computacional. Os autores também utilizaram o MATPOWER para simular a operação da rede elétrica. Assim como no estudo anterior, as cargas foram estocásticas, seguindo uma distribuição uniforme. As medições de potência ativa foram coletadas de cada linha de transmissão no sistema de teste IEEE 118-barras, resultando em 304 características de medição. Foi empregada simulação de Monte Carlo para registrar o vetor de medição em 1000 instâncias diferentes.As simulações, realizadas em um sistema de teste IEEE 118-barras com cargas estocásticas, coletaram 304 características de medição de potência ativa, com a PCA retendo 99% da variância com apenas dois componentes principais. Os resultados demonstram a eficácia dos algoritmos na detecção de FDIs furtivos. O *f1-score* foi a métrica de acurácia utilizada, com o SVM atingindo 0,956 (C = 5,648, sigma = 6,67). O SVM distribuído mostrou convergência, embora mais lenta com o aumento do número de grupos devido ao *overhead* de comunicação. As principais limitações incluem a dependência de dados de treinamento rotulados para o SVM e a sensibilidade das abordagens à qualidade e complexidade dos dados. Os autores sugerem expansão para detectar outros tipos de falhas além dos ciberataques.

Liu et al. (2023) no trabalho intitulado "*Data-driven FDI Attacks: A Stealthy Approach to Subvert SVM Detectors in Power System Security*" propõem um ataque de injeção de dados falsos baseado em dados (DFDI) para subverter detectores SVM em sistemas de potência, buscando furtividade e impacto significativo na operação. Os dados para o estudo foram gerados a partir do sistema IEEE 14-barras, utilizando o MATPOWER e perfis de carga da Ercot Utility. O conjunto de treinamento incluiu 400 medições SCADA não comprometidas e 200 vetores comprometidos por ataques TFDI (Tradicional FDI) com μ entre 0.1 e 0.4, afetando ângulos de tensão em três barramentos aleatórios; 70% desses dados foram para treinamento e 30% para validação cruzada. O conjunto de teste utilizou 100 medições não comprometidas e 100 ataques DFDI. A metodologia DFDI emprega o método da Elipse de Área Mínima Envolvente (MEE) para restringir as medições comprometidas ao cluster de dados normais, garantindo a furtividade. Os resultados no IEEE 14-barras revelaram a limitação do SVM: para ataques TFDI, o *f1-score* variou de 0.00 (para μ=0.01) a 0.71 (para μ=0.30). Contra ataques DFDI, o SVM obteve *Recall* de apenas 0.05, Precisão de 0.50, Acurácia de 0.50 e *f1-score* de 0.09, indicando que o detector não distingue eficazmente dados normais de ataques DFDI. As limitações incluem a dependência do conhecimento prévio dos dados de treinamento do SVM, a validação

apenas em ambiente simulado e a falta de consideração de contramedidas avançadas.

3.4 Detecção baseada em Aprendizado profundo (*Deep Learning*)

Niu et al. (2019) no estudo "*Dynamic Detection of False Data Injection Attack in Smart Grid using Deep Learning*", propõem um framework de deep learning para detecção dinâmica de ataques de injeção de dados falsos (FDIs) em redes elétricas inteligentes. A abordagem emprega Redes Neurais Convolucionais (CNNs) e Redes Neurais Recorrentes (RNNs) com células Long Short-Term Memory (LSTM) para capturar padrões temporais e espaciais nos dados de medição da rede. O modelo é treinado com um grande conjunto de dados simulados, incluindo operações normais e cenários de ataque, para distinguir medições legítimas de manipuladas. Os dados foram gerados a partir do sistema IEEE 39-barras, utilizando cargas estocásticas para simular operações realistas. Medições de potência ativa foram coletadas de cada linha de transmissão. Ataques FDI foram gerados por invasores "*man-in-the-middle*" com uma estrutura de comunicação cliente-servidor , injetando k medições escolhidas aleatoriamente para gerar vetores de ataque com distribuição gaussiana. Cenários com vetores de ataque derivados de medições reais também foram testados. Os resultados demonstram a eficácia do modelo, superando métodos tradicionais baseados em estimativa de estado. A acurácia de detecção do mecanismo proposto atingiu mais de 90% para ataques FDI aleatórios quando a proporção de medições comprometidas (k/n) era alta. O uso de RNNs permitiu a captura de dependências temporais, e a combinação com CNNs auxiliou na identificação de padrões espaciais. As principais limitações incluem: a necessidade de grandes volumes de dados rotulados para treinamento; o alto poder computacional exigido pelo modelo, o que pode restringir sua aplicação em tempo real em sistemas de grande escala; e a complexidade do modelo, que pode dificultar sua interpretabilidade para os operadores de rede. Além disso, o sistema apresentou baixa acurácia quando a capacidade de ataque era baixa , embora isso possa ser resolvido pela incorporação de um detector de estimativa de estado (SE).

No trabalho intitulado "*A Deep Learning-Based Classification Scheme for False Data Injection Attack Detection in Power System*", Ding et al. (2021) desenvolvem propõem um esquema de classificação baseado em *deep learning* para detecção de ataques de injeção de dados falsos (FDIs) em sistemas de energia, visando maior precisão. Os dados foram simulados no sistema IEEE 14-barras, com cálculos de fluxo de potência em 30.000 momentos consecutivos e injeção de ruído gaussiano. Aleatoriamente, 15.000 FDIs foram lançados, assegurando que os resíduos estivessem abaixo de um limite para evitar a detecção de dados ruins. O modelo, uma Rede de Crença Profunda Condicional (CDBN), processou séries temporais para extrair características relevantes. Os resultados demonstraram que o esquema proposto alcançou uma precisão de detecção de até 97,3% em simulações, superando métodos tradicionais como Redes Neurais Artificiais (ANN) e Máquinas de

Vetor de Suporte (SVM). As principais limitações incluem a necessidade de grandes volumes de dados para treinamento, a sensibilidade ao ruído ambiental e a complexidade da otimização de parâmetros. Além disso, o estudo não considerou uma gama ampla de cenários de ataque, o que restringe a generalização dos resultados.

3.5 Detecção baseada em *Autoencoders*

Kundu et al. (2020) no artigo intitulado "*A3D: Attention-based auto-encoder anomaly detector for false data injection attacks*", desenvolveram o modelo de *autoencoder* com atenção (A3D) para detecção não supervisionada de ataques de injeção de dados falsos (FDIA) em sistemas de energia elétrica, comparando-o a métodos supervisionados e não supervisionados tradicionais. Os dados foram gerados a partir de simulações do sistema IEEE 14-barras usando dados de fluxo de potência do Texas, introduzindo ataques aleatórios de menor esforço para evitar detecção convencional. O estudo avaliou o A3D, *autoencoders* baseados em ANN e RNN, e métodos supervisionados e não supervisionados, utilizando o erro de reconstrução e *f1-score* como métricas. Os resultados mostraram que o A3D superou outros métodos, especialmente em cenários de intrusão aleatória, com *f1-scores* de detecção para A3D variando de 0.944 a 0.994 em diferentes níveis de comprometimento, e 0.984 para intrusões aleatórias. O desempenho de todos os modelos diminuiu com poucos dispositivos comprometidos, indicando dificuldade em distinguir pequenas alterações do ruído normal. As limitações incluem a dependência de dados simulados, que podem não capturar a complexidade do mundo real, a menor eficácia com poucos dispositivos comprometidos e a falta de análise da eficiência computacional e escalabilidade dos modelos. Além disso, a localização exata das intrusões apresentou baixa precisão devido à propagação do erro de reconstrução.

Yang et al. (2021) no trabalho intitulado "*Deep learning for online AC False Data Injection Attack detection in smart grids: An approach using LSTM-Autoencoder*", propõem um mecanismo online para detecção de ataques de injeção de dados falsos (FDIA) em sistemas de potência AC, empregando uma rede neural LSTM-*Autoencoder* para capturar padrões temporais e identificar anomalias nas medições de estado. Os dados foram gerados a partir dos sistemas IEEE 14 e 118-barras. Medições normais foram obtidas via MATPOWER e ataques simulados por um modelo de FDI que induz sobrecargas e alívio de carga. Após a estimativa de estados, os dados foram tratados como séries temporais e decompostos por Decomposição de Modo Variacional (VMD). Os resultados indicam alta eficácia, com acurácia de detecção de 93,56% para o sistema IEEE 14-barras e 92,75% para o IEEE 118-barras, e um tempo médio de teste de aproximadamente 2,86 ms por estado estimado. A arquitetura LSTM mostrou-se eficiente em capturar variações temporais. As limitações incluem a necessidade de grande volume de dados históricos para treinamento, a alta complexidade computacional para aplicação em tempo real em larga

escala e a exigência de uma infraestrutura de TI complexa.

No estudo "*FDI attack detection using extra trees algorithm and deep learning algorithm-autoencoder in smart grid* ", Majidi, Hadayeghparast e Karimipour (2022) propõem uma abordagem híbrida para detecção de ataques de injeção de dados falsos (FDIAs) em redes elétricas inteligentes, combinando o algoritmo *Extra Trees* e um *autoencoder* de *deep learning*. Os dados foram gerados via MATPOWER, simulando sistemas IEEE 14, 30, 57 e 118-barras, incluindo fluxos de potência, saídas de geradores e ataques com até 50% das medições comprometidas sob variadas condições de esparsidade. A metodologia utiliza o *Extra Trees* para seleção de características e classificação, e o *autoencoder* para identificar anomalias por meio de erros de reconstrução. Os resultados demonstram alta precisão na detecção de FDIs aleatórios. O modelo alcançou os seguintes *f1-score* e Acurácias: IEEE 14-Barras (96.97% e 97.8%), IEEE 30-Barras (96.32% e 95.18%), IEEE 57-Barras (99.39% e 99.39%), e IEEE 118-Barras (99.77% e 99.76%). Essa sinergia entre seleção de características e detecção baseada em reconstrução superou métodos isolados. As limitações incluem a necessidade de dados extensivos para treinamento em ambientes reais, a complexidade computacional para implementação em tempo real em larga escala e a exigência de ajustes finos na integração dos métodos.

No trabalho intitulado "*Autoencoder Based FDI Attack Detection Scheme For Smart Grid Stability*", G et al. (2022) propõem um esquema de detecção de ataques de injeção de dados falsos (FDI) em sistemas de energia inteligente utilizando *Autoencoders* (AEs) como técnica de aprendizado de máquina, visando identificar e classificar dados anômalos, além de reduzir a dimensionalidade dos dados. Os dados foram gerados para simular um ambiente de sistema de energia inteligente, coletando 60.000 instâncias de 13 características relacionadas à estabilidade do sistema, como tempo de reação, equilíbrio de potência e coeficiente de elasticidade de preço. O modelo AE foi treinado de forma não supervisionada com 80% desses dados limpos, enquanto 10% foram para validação e o restante para teste, e os ataques foram gerados aleatoriamente dentro de um intervalo predeterminado. Os resultados demonstraram a eficácia do *Autoencoder* na detecção de FDIs, com uma taxa de compressão de dados de até 76% e uma acurácia de 98%. A precisão para "*Normal Data*" foi de 0.98 e para "*Fault Data*" de 1.00, com *recall* de 1.00 e 0.87, respectivamente, e *f1-score* de 0.99 e 0.93. A curva ROC apresentou um AUC de 0.75, indicando boa capacidade de distinção. As limitações incluem a influência da qualidade e representatividade dos dados de treinamento, desafios de adaptação a diferentes topologias de rede, complexidade computacional para sistemas de grande escala e a possível ineficácia contra ataques mais sofisticados que exploram correlações complexas dos dados.

O trabalho intitulado "*A locational false data injection attack detection method in smart grid based on adversarial variational autoencoders*", Wang et al. (2024) propõem o AT-GVAE, um framework para detecção localizada de ataques de injeção de dados

falsos (FDI) em sistemas de energia elétrica, integrando VAEs e GANs para aprimorar a detecção de anomalias sutis. A metodologia emprega módulos VAE generativo (VAE-G) e discriminativo (VAE-D) com unidades GRU em múltiplas camadas para capturar correlações temporais, e um mecanismo de autoatenção para enfatizar características importantes. Os dados foram gerados via MATPOWER a partir dos sistemas IEEE 14 e 118-barras. Para o IEEE 14-barras, utilizaram-se cargas reais da NYISO (3456 amostras). Ataques foram modelados como alterações furtivas na estimativa de estado DC, com esparsidade de aproximadamente 10% e magnitudes Gaussianas (N(0,1)). Os resultados experimentais demonstram a superioridade do AT-GVAE, com melhorias significativas em AUC-ROC (ex: 5.02% no IEEE 118-barras vs. G-LVAE) e AUC-PR (ex: 10.43% no IEEE 118-barras vs. G-LVAE) em comparação com métodos baseados em GRU, VAE e treinamento adversarial. As limitações incluem a dificuldade de implementação prática em sistemas reais (validação simulada) e a falta de exploração extensiva da otimização dos parâmetros α e β.

3.6 Levantamento do Estado da Arte

Com o objetivo de fazer o levantamento o estado da arte, neste capítulo foram abordados diversos estudos que se dedicaram à detecção de ataques FDI em sistemas de potência, sendo empregadas técnicas de aprendizado de máquina e aprendizado profundo. Na Tabela 5, é apresentado um resumo dos trabalhos analisados.

Tabela 5 – Estado da Arte.

Referência	**Técnica utilizada**	**Sistema de Teste**	**Métricas**	**Limitações**
Esmalifalak et al. (2013)	SVM e PCA	Sistema IEEE-14 barras	*f1-score* e acurácia	Dependência de dados rotulados, Dependência da qualidade dos dados de entrada e Complexidade dos ataques.
Esmalifalak et al. (2017)	SVM e *Random Forests*	Sistema de teste padrão IEEE	*f1-score*, precisão e *recall*	Dependência de dados rotulados e escalabilidade dos modelos

Continua na próxima página

Tabela 5 – Continuação da página anterior

Referência	**Técnica utilizada**	**Sistema de Teste**	**Métricas**	**Limitações**
Niu et al. (2019)	RNN e CNN	Sistema IEEE-39 barras	Acurácia	Necessidade de grande volume de dados rotulados, Alto poder computacional e Complexidade do modelo
Kundu et al. (2020)	AEs com mecanismo de atenção	Sistema IEEE-14 barras	Precisão, *recall*, VP, FP e FN	Complexidade do modelo, alto poder computacional e restrição a um sistema de teste
Ding et al. (2021)	Rede CDBN	Sistema IEEE-14 barras	Acurácia, FP e curva ROC	Necessidade de um grande conjunto de dados, restrição a um único sistema de teste.
Yang et al. (2021)	LSTM-*Autoencoder*	Sistema IEEE-14 e 118 barras	Acurácia, VP, precisão e *recall*	Necessidade de um grande volume de dados e alta complexidade computacional
Majidi, Hadayeghparast e Karimipour (2022)	Extra Trees e AE	Sistema IEEE-14, 30, 57 e 118 barras.	Acurácia e FP	Necessidade de dados extensivos e variados, alta complexidade computacional, ajuste fino para integração de dois métodos.

Continua na próxima página

Tabela 5 – Continuação da página anterior

Referência	**Técnica utilizada**	**Sistema de Teste**	**Métricas**	**Limitações**
G et al. (2022)	AE	Sistema simulado	Acurácia e FP	Necessidade de dados extensivos e variados, alta complexidade computacional, ajuste fino para integração de dois métodos.
Liu et al. (2023)	SVM	Sistema IEEE-14 barras	Precisão, *recall*, Acurácia e *f1-score*	Necessidade de dados rotulados, restrição a um unico sistema de teste, ajuste fino dos parâmetros da SVM.
Wang et al. (2024)	AT-GVAE	Sistema IEEE-14 e 118 barras	FP, Acurácia e curva ROC	Falta de exploração extensiva da otimização dos parâmetros α e β. Dificuldade de aplicabilidade prática.
Este trabalho	*Autoencoder* com Mecanismo de Atenção *multi-head*	Sistema IEEE-14, 30, 118 e 300 barras	VP, VN, FP, FN, Acurácia, *recall*, *f1-score*, Precisão e curva ROC	

3.7 Lacunas encontradas na literatura

A revisão bibliográfica apresentada e o estado da arte apresentado na Tabela 5 neste capítulo permitiu identificar diversas lacunas e desafios persistentes na literatura referente

à detecção de ataques de injeção de dados falsos (FDI) em sistemas de potência. Embora os avanços nos métodos baseados em *Machine Learning* (ML) e *Deep Learning* (DL) sejam notáveis, especialmente o uso de *autoencoders*, ainda existem limitações significativas que motivam a presente pesquisa.

As principais lacunas observadas são:

- **Dependência de Dados Rotulados e Modelos Exatos:** Uma limitação recorrente nos trabalhos baseados em aprendizado de máquina supervisionado é a necessidade de um grande volume de dados rotulados para o treinamento eficaz dos modelos. No contexto de ataques cibernéticos em sistemas de potência, a obtenção de dados de ataque rotulados em cenários reais é extremamente desafiadora e, muitas vezes, inviável. Além disso, métodos baseados em modelos físicos frequentemente dependem de um conhecimento preciso da topologia da rede e de parâmetros do sistema, o que os torna mais frágeis diante de incertezas e variações operacionais, bem como computacionalmente caros e não escaláveis para sistemas de grande porte.

- **Complexidade Computacional e Escalabilidade para Grandes Sistemas:** Muitos dos algoritmos de detecção, especialmente aqueles que combinam técnicas extensivas de aprendizado profundo, demandam um alto poder computacional. Isso pode restringir sua aplicação em tempo real em sistemas de grande escala, onde a latência é um fator crítico. A complexidade do modelo também pode dificultar sua interpretabilidade para os operadores de rede. Embora alguns trabalhos abordem sistemas maiores, a eficiência computacional e a escalabilidade continuam sendo um desafio prático.

- **Sensibilidade a Ataques Sutis e Variações:** Observa-se que a eficácia de alguns modelos diminui diante de ataques mais sutis ou com baixa amplitude de alteração. Em particular, métodos que não conseguem capturar as complexas correlações dos dados podem falhar em distinguir pequenas alterações causadas por ataques do ruído normal ou de flutuações operacionais. A sensibilidade à qualidade e complexidade dos dados de entrada também é uma limitação em várias abordagens.

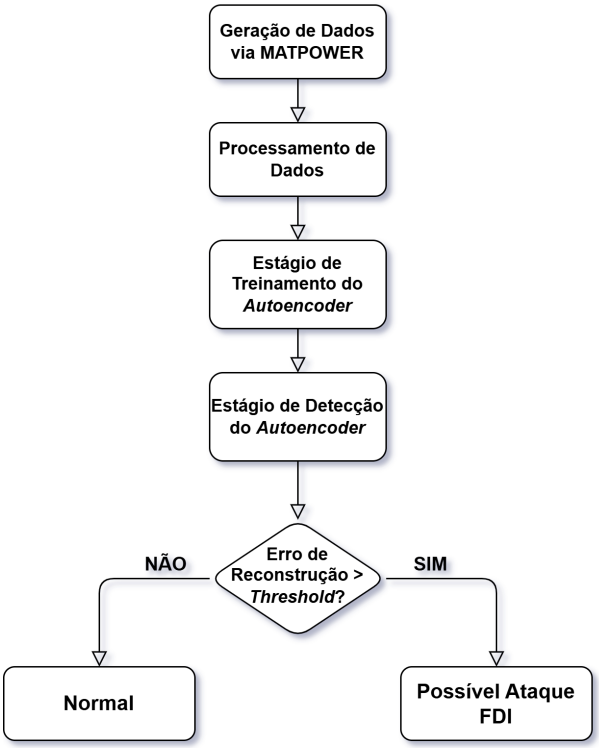
- **Generalização e Validação em Múltiplas Topologias de Sistema:** Uma parte considerável dos estudos encontrados na literatura foca na validação do modelo em apenas um sistema de teste (por exemplo, IEEE-14 barras ou IEEE-39 barras), o que limita e restringe a capacidade de generalização do modelo proposto. A falta de exploração extensiva da capacidade de portabilidade do modelo em diferentes topologias e níveis de complexidade de sistemas de potência representa uma lacuna para a aplicabilidade prática das soluções.

- **Falta de Dados Reais e Consideração de Outras Anomalias:** A maioria dos trabalhos analisados se baseia em dados simulados, o que pode não capturar a complexidade e as nuances do mundo real, incluindo ruídos, falhas de sensores genuínas, curtos-circuitos e outras anomalias físicas que podem interferir na detecção de ataques cibernéticos e gerar falsos positivos. A integração do modelo com métodos baseados em física ou a validação em cenários reais ainda são áreas que necessitam de mais investigação para aumentar a confiança e a aplicabilidade das soluções.

4 METODOLOGIA

No contexto das lacunas encontradas na literatura, este trabalho visa trazer uma evolução ao estado da arte na detecção de ataques FDI em sistemas de potência, propondo um algoritmo inovador baseado em *autoencoders* com mecanismo de atenção *multi-head*. A metodologia utiliza aprendizado não supervisionado, o que elimina a necessidade de dados rotulados de ataques, tipicamente exigidos por modelos supervisionados, e dispensa a construção de modelos de sistema extensos e altamente precisos, característicos das abordagens baseadas em modelos. O objetivo é que o algoritmo aprenda os padrões das medições de sistemas de energia em condições normais e, com base nesse conhecimento, possa identificar e rejeitar efetivamente dados falsos. Para isso, o estudo utiliza dados próprios gerados através do MATPOWER para simulação de sistemas de potência, garantindo um conjunto de dados consistente e representativo, e avalia a performance em sistemas de diferentes portes (IEEE-14, 30, 118 e 300 barras), com foco na sensibilidade à ataques sutis, preenchendo assim as lacunas identificadas. A Figura 7 apresenta o fluxograma geral da metodologia proposta.

Figura 7 – Metodologia Proposta.



Fonte: Autor.

O *autoencoder* é treinado exclusivamente com dados de medições de tensão e ângulo de fase, gerados artificialmente a partir de sistemas de energia em condições normais,

aprendendo seus padrões para identificar desvios que indicam anomalias. Entre os modelos comumente utilizados para reconhecimento de padrões, este trabalho investiga a viabilidade dos *autoencoders*, que tendem a apresentar um bom desempenho em tarefas de detecção de anomalias sem a necessidade de estruturas altamente complexas, oferecendo uma solução eficiente e adaptável para a detecção de anomalias.

4.1 Geração de dados

A geração de dados para a detecção de ataques de injeção de dados em sistemas de energia elétrica foi realizada utilizando o MATPOWER, uma ferramenta adotada para análise de sistemas de potência. No contexto desta dissertação, o processo de geração de dados envolve a simulação de fluxos de carga em sistemas de potência, com variações de demanda aleatórias, e a inserção de ataques FDI para avaliar a capacidade do modelo proposto de detectar esses ataques. A seguir, é descrito em detalhes o processo de geração de dados, com ênfase nas etapas relacionadas ao fluxo de carga e às equações utilizadas.

4.1.1 Definição do Sistema de Potência

O MATPOWER fornece diferentes modelos de sistemas de potência, foram escolhidos os sistemas IEEE 14-barras, IEEE 30-barras, IEEE 118-barras e IEEE 300-barras para esse estudo, tais sistemas foram escolhidos pois são tipicamente utilizados em diversos estudos de sistemas de potência, incluindo vários dos artigos reportados na revisão bibliográfica. Espera-se assim facilitar a comparação dos resultados produzidos com o de outros trabalhos. O sistema de potência consiste em geradores localizados em algumas barras e cargas distribuídas por outras barras, e a análise do fluxo de carga visa determinar as tensões e os ângulos de fase em todas as barras do sistema.

4.1.2 Análise de Fluxo de Carga

A análise de fluxo de carga é um processo fundamental na modelagem de sistemas de potência, visando determinar a distribuição de tensões e os ângulos de fase nas barras do sistema, sendo resolvida por um conjunto de equações não lineares que descrevem o comportamento das barras de potência. No MATPOWER, o fluxo de carga baseia-se no método de Newton-Raphson, uma técnica empregada para resolver esses sistemas de equações não lineares, cujo objetivo é equilibrar a potência gerada e consumida em cada barra, e as equações básicas para seu cálculo são expressas em termos de potências ativa e reativa:

$$P_i = V_i \sum_{j=1}^{N} V_j \left(G_{ij} \cos(\theta_{ij}) + B_{ij} \sin(\theta_{ij})\right) \tag{4.1}$$

$$Q_i = V_i \sum_{j=1}^{N} V_j \left(G_{ij} \sin(\theta_{ij}) - B_{ij} \cos(\theta_{ij})\right) \tag{4.2}$$

onde:

- P_i e Q_i são as potências ativa e reativa na barra i respectivamente,
- V_i é a tensão na barra i,
- G_{ij} e B_{ij} são as componentes de condutância e susceptância da matriz de admitância do sistema,
- θ_{ij} é a diferença de ângulo de fase entre as barras i e j,
- N é o número total de barras no sistema.

Essas equações são resolvidas iterativamente para obter os valores de V_i e θ_i para cada barra i no sistema, garantindo que as potências ativa e reativa sejam equilibradas em todas as barras. O processo de resolução é realizado utilizando o método de Newton-Raphson, que pode ser descrito pela atualização das tensões e ângulos de fase a cada iteração:

$$\begin{bmatrix} \Delta V \\ \Delta \theta \end{bmatrix} = - \left(\begin{bmatrix} J_{VV} & J_{V\theta} \\ J_{\theta V} & J_{\theta\theta} \end{bmatrix} \right)^{-1} \begin{bmatrix} f_V \\ f_\theta \end{bmatrix} \tag{4.3}$$

onde:

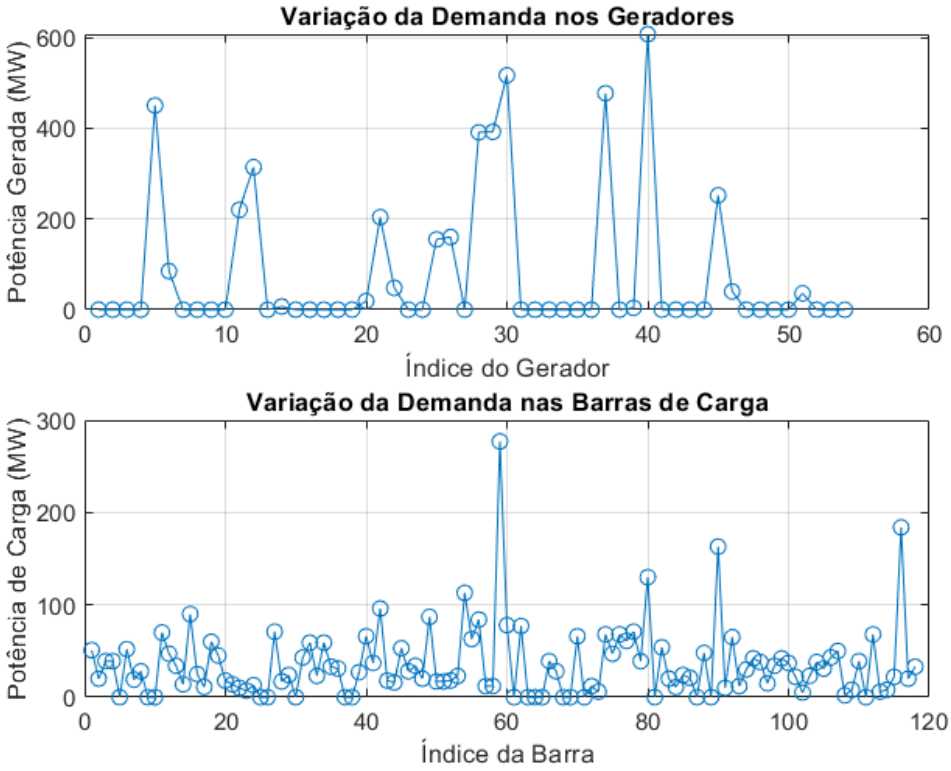
- ΔV e $\Delta \theta$ são as correções para as tensões e ângulos de fase nas barras,
- J_{VV}, $J_{V\theta}$, $J_{\theta V}$ e $J_{\theta\theta}$ são os elementos da matriz Jacobiana, que depende das características do sistema de potência,
- f_V e f_θ são as funções residuais associadas às potências ativa e reativa.

Esse processo é repetido até que as diferenças entre as correções (ΔV e $\Delta \theta$) sejam suficientemente pequenas, indicando que o sistema de fluxo de carga atingiu um ponto de equilíbrio. A inversão explícita da matriz Jacobiana na Equação (4.3) é computacionalmente onerosa para sistemas de grande porte. Assim, para otimizar o desempenho, o sistema linear equivalente ($J\Delta x = -f$) é resolvido por métodos eficientes como eliminação Gaussiana ou fatoração LU.

4.1.3 Geração das amostras

Após a resolução do fluxo de carga, são geradas amostras dos dados de tensão e ângulo de fase nas barras do sistema. Essas amostras representam o comportamento do sistema de potência sob diferentes condições de demanda. A demanda de potência nas barras é variada aleatoriamente em uma faixa de ±30%, valor esse apresentado como base nos trabalhos vistos na literatura (MAJIDI; HADAYEGHPARAST; KARIMIPOUR, 2022) para simular variações no consumo de energia, tanto nas barras de geração quanto nas barras de carga. Um exemplo é apresentado na Figura 8 para o sistema de 118 barras.

Figura 8 – Alteração na demanda nos geradores e nas barras de carga para o sistema IEEE-118 barras.



Fonte: Autor.

Cada amostra gerada contém um vetor de medições de tensão e ângulo de fase para todas as barras do sistema, que são extraídas dos resultados do fluxo de carga. Essas medições formam o conjunto de dados utilizado para treinar e testar os modelos de detecção de ataques. Foram geradas 16000 amostras para treinamento e 4000 para teste de dados normais.

4.1.4 Inserção de Ataques de Injeção de Dados Falsos (FDI)

A inserção de ataques FDI nos dados gerados segue uma abordagem sistemática para simular a manipulação das medições do sistema. O ataque consiste na modificação

de um subconjunto das amostras de dados de forma que as medições comprometidas sejam introduzidas com alterações nas tensões e ângulos de fase, simulando a ação de um atacante que altera as medições de sensores ou pontos de medição de forma furtiva. O objetivo é introduzir dados corrompidos no conjunto de dados sem que esses ataques sejam facilmente detectados, o que torna o modelo de detecção mais desafiador.

O Submódulo 2.6: Requisitos mínimos para subestações e seus equipamentos do ONS (ONS, 2022) no item 3.5.2.1, define os limites para equipamentos localizados nos terminais de uma linha de transmissão (LT). Tais equipamentos, que possam permanecer energizados após a manobra da LT, como reatores de linha, disjuntores, seccionadores e transformadores de potencial, devem suportar, no terminal em vazio, por um período de uma hora, as sobretensões à frequência industrial estabelecidas. Os limites são apresentados na Tabela 6.

Tabela 6 – Sobretensões Sustentadas Admissíveis a 60 Hz por 1 Hora em Terminais de LT em Vazio.

Tensão Nominal de Operação (kV)	**Máxima Tensão Sustentada Fase-Fase, Eficaz, por 1 Hora, em Terminais de LT a Vazio**	
	(kV Fase-Fase, Eficaz)	**(pu)[1]**
138	152	1,10
230	253	1,10
345	398	1,15
440	506	1,15
500	600	1,20
525	600	1,15
765	800	1,046

[1]Valores em pu tendo como base a tensão nominal de operação.

Tomando como base os valores acima determinados, para simular diferentes intensidades de ataques de falsificação de dados (FDI), as amostras selecionadas foram modificadas em três níveis distintos de amplitude: ±50%, ±25% e ±15%. Essas variações foram aplicadas aleatoriamente aos valores de tensão e ângulo de fase das medições originais através de um fator de modificação δ. O ataque foi simulado conforme a equação:

$$x_i^{\text{atacada}} = x_i \cdot (1 + \delta_i) \tag{4.4}$$

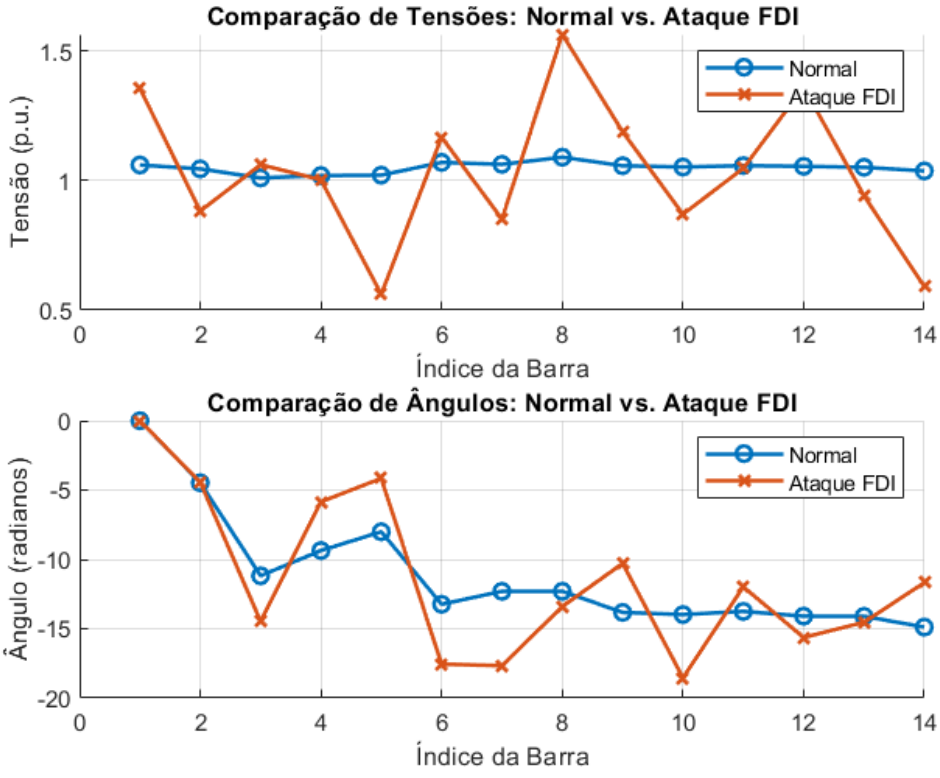
onde:

- x_i representa o vetor de medições originais (tensão e ângulo de fase) da amostra i
- δ_i é o fator de perturbação, gerado aleatoriamente conforme três intervalos distintos:
  - Variação alta: $\delta_i \in [-0.5, 0.5]$ (modificação de ±50%)
  - Variação média: $\delta_i \in [-0.25, 0.25]$ (modificação de ±25%)

– Variação baixa: $\delta_i \in [-0.15, 0.15]$ (modificação de $\pm 15\%$)

A seleção das amplitudes de ataque FDI em $\pm 50\%$, $\pm 25\%$ e $\pm 15\%$ visa a avaliar a eficiência do modelo de detecção perante diferentes níveis de severidade na manipulação de dados que podem comprometer a operação do sistema de potência. Ataques de $\pm 50\%$ simulam cenários extremos que, se não detectados, levariam as estimativas de estado a exceder drasticamente os limites de tensão em regime permanente e de sobretensão sustentada em terminais de LT definidos pelo Submódulo 2.6 do ONS. Já as variações de $\pm 25\%$ representam um risco significativo, podendo forçar o sistema a operar em condições de emergência que exigiriam dos equipamentos, como transformadores, a capacidade de suportar carregamentos de 120% a 140% da potência nominal por períodos específicos, conforme as diretrizes do mesmo submódulo. Por fim, ataques de $\pm 15\%$ são utilizados para testar a sensibilidade do modelo a manipulações sutis, projetadas para evadir detecções tradicionais; identificar esses desvios é necessário para manter o desempenho, confiabilidade e resiliência da Rede Básica dentro dos requisitos operacionais. Um exemplo da injeção de ataques é apresentado na Figura 9 para o sistema de 14 barras.

Figura 9 – Alteração nas medições de tensão e ângulo de fase em $\pm 50\%$ no sistema IEEE-14 barras.



Fonte: Autor.

4.2 Análise e Processamento dos Dados

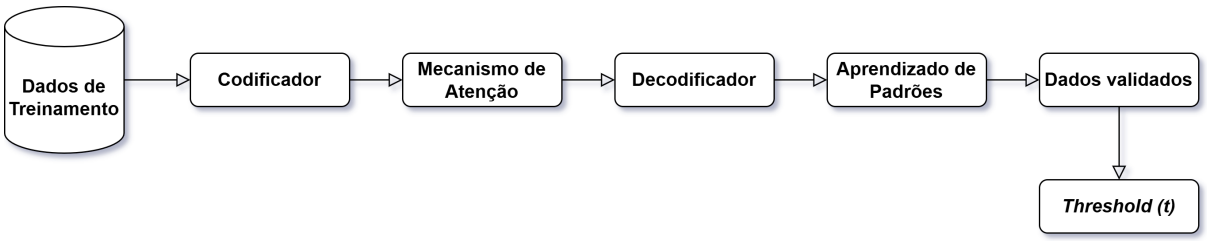
Após a geração dos dados, os mesmos foram analisados e padronizados. Para os conjuntos de dados tanto de medições normais como os vetores de ataque foram realizadas análises estatísticas descritivas, com objetivo de verificar dados faltantes ou NaN (*Not a Number*) o que pode interferir e ocasionar erros no processo de treinamento. Em nenhum conjunto de dados foi encontrado dado faltante ou NaN.

Para garantir a compatibilidade com a rede neural, os dados foram normalizados no intervalo [0, 1] utilizando o *MinMaxScaler* do *scikit-learn*. Essa etapa é importante para evitar viés no treinamento do modelo e melhorar a convergência durante o aprendizado.

4.3 Estágio de treinamento

Após a etapa de análise e processamento dos dados o *Autoencoder* foi implementado e treinado conforme o fluxograma apresentado na Figura 10.

Figura 10 – Estágio de Treinamento.



Fonte: Autor

O modelo implementado é um *deep autoencoder* com mecanismo de atenção multi-head, composto por três módulos principais:

- **Codificador**: Reduz a dimensionalidade dos dados de entrada através de duas camadas densas (128 e 64 neurônios, respectivamente), com funções de ativação *ReLU* e *dropout* (taxa de 0.2) para regularização. A saída do codificador é um vetor latente de dimensão 20, que captura as características essenciais dos dados.
- **Mecanismo de Atenção**: O vetor latente é remodelado para uma forma tridimensional (*batch_size*, 1, *latent_dim*) e processado por uma camada de atenção multi-head (4 *heads*). Esse mecanismo permite que o modelo identifique e pondere automaticamente as *features* mais relevantes para a detecção de anomalias. A saída da atenção é normalizada (*LayerNormalization*) e reduzida para um vetor unidimensional via *GlobalAveragePooling1D*.
- **Decodificador**: Reconstrói os dados a partir do vetor processado pela atenção, utilizando duas camadas densas (64 e 128 neurônios) e uma camada de saída com

ativação *sigmoid* para garantir valores no intervalo [0, 1]. O modelo é otimizado com o algoritmo *Adam* e função de perda *MSE* (Erro Quadrático Médio).

O treinamento foi realizado exclusivamente com dados normais (16000 amostras), visando ensinar o modelo a reconstruir adequadamente o comportamento esperado do sistema. Foram utilizados 200 épocas com *batch size* de 256 e 20% dos dados reservados para validação. Após o treinamento, o erro de reconstrução (diferença absoluta média entre entrada e saída) foi calculado para os conjuntos de teste normal e sob ataque. Um limiar de detecção foi estabelecido como o percentil 95 do erro de reconstrução dos dados normais para otimizar a relação entre a taxa de falsos positivos e a capacidade de detecção de anomalias. Essa escolha garante que a variabilidade inerente aos padrões de comportamento normais seja acomodada, minimizando classificações errôneas de desvios comuns. Consequentemente, qualquer amostra cujo erro de reconstrução supere esse valor é classificada como um ataque, permitindo identificar eventos que se desviam significativamente da normalidade e indicam potenciais intrusões, mantendo assim uma alta especificidade na detecção.

O residual para uma observação de treinamento Y_j é dado por $r_j = Y_j - \hat{Y}_j$. O erro de reconstrução R_j é definido como a razão entre o comprimento de r_j e a dimensão dos dados de entrada d_0. O objetivo do processo de treinamento é minimizar o valor médio da soma de todos os erros de reconstrução R_j, expresso como (4.5):

$$\min_{W,b} \mathcal{L} := \frac{1}{S} \sum_{j=1}^{S} \left(\frac{\|r_j\|^2}{d_0} \right) \tag{4.5}$$

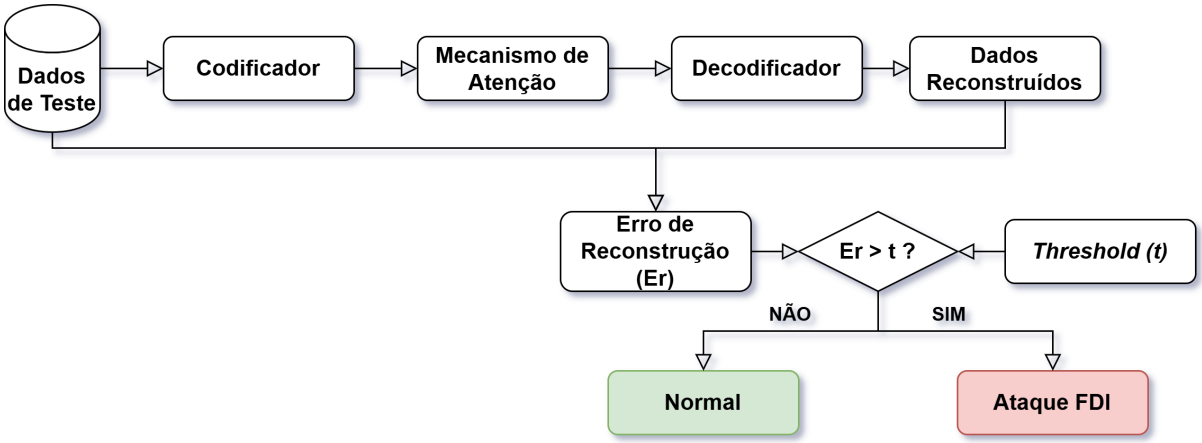
onde S representa o número total de observações usadas no treinamento.

4.4 Estágio de detecção

Após a realização do treinamento do modelo e definição do *threshold*, é realizada a etapa de detecção, conforme apresentado na Figura 11.

Nesta etapa, os dados de teste contendo 4000 amostras e sendo 50% das amostras comprometidas, são passados pelo algoritmo para reconstrução e cálculo do erro de reconstrução. Posteriormente, o erro de reconstrução de cada amostra é comparado com o *threshold* definido na etapa de treinamento. As amostras que apresentarem um erro de reconstrução maior que o limiar são apontadas como ataques FDI, caso contrário, são definidas como amostra normal (sem ataque FDI).

Figura 11 – Estágio de Detecção.



Fonte: Autor

4.5 Avaliação dos Modelos

A avaliação do modelo de detecção de ataques FDI foi realizada através de um conjunto abrangente de métricas e análises gráficas, cada uma com um propósito específico para garantir uma validação consistente do sistema proposto. Foram calculadas métricas de: acurácia, *recall* (Sensibilidade), Área sob a Curva ROC (AUC-ROC), matriz de confusão, precisão e *f1-score*.

A acurácia representa a porcentagem total de detecções corretas, incluindo tanto os casos normais quanto os sob ataque. Essa métrica é fundamental para avaliar o desempenho geral do modelo, pois indica sua capacidade de distinguir efetivamente entre operações legítimas e ataques. Uma alta acurácia sugere que o modelo está classificando corretamente a maioria das amostras, o que é essencial para aplicações em sistemas de potência, onde erros podem ter consequências críticas.

O *recall* mede a proporção de ataques FDI que foram corretamente identificados pelo modelo. Essa métrica é particularmente importante em cenários de segurança, onde deixar passar um ataque (falso negativo) pode ser mais prejudicial do que um alarme falso (falso positivo). Um *recall* elevado indica que o modelo está captando a maioria das amostras maliciosas.

A AUC-ROC avalia a capacidade do modelo de diferenciar entre operações normais e ataques, independentemente do limiar de classificação escolhido. Um valor próximo de 1 indica que o modelo possui uma excelente capacidade de discriminação, sendo capaz de separar claramente as duas classes. Essa métrica é especialmente útil para comparar diferentes abordagens ou ajustes do modelo, pois fornece uma medida consolidada de desempenho.

A matriz de confusão oferece uma visão detalhada dos acertos e erros do modelo,

destacando quantos ataques foram detectados corretamente (verdadeiros positivos) e quantos passaram despercebidos (falsos negativos), bem como quantas operações normais foram erroneamente classificadas como ataques (falsos positivos). Essa análise é essencial para identificar padrões de erro e ajustar o modelo para reduzir classificações incorretas que possam impactar a operação do sistema.

A curva ROC complementa a análise da AUC-ROC, mostrando como a taxa de verdadeiros positivos (*recall*) varia em relação à taxa de falsos positivos para diferentes limiares de decisão. Essa visualização ajuda a selecionar o ponto de operação ideal do modelo, equilibrando a detecção de ataques com a minimização de alarmes falsos.

A combinação dessas métricas e gráficos fornece uma avaliação abrangente do modelo, destacando seus pontos fortes e áreas para melhoria. Enquanto a acurácia e o *recall* garantem que o modelo está performando bem em termos de classificação, a AUC-ROC e a matriz de confusão oferecem *insights* sobre sua capacidade de generalização e equilíbrio entre detecção e falsos alarmes.

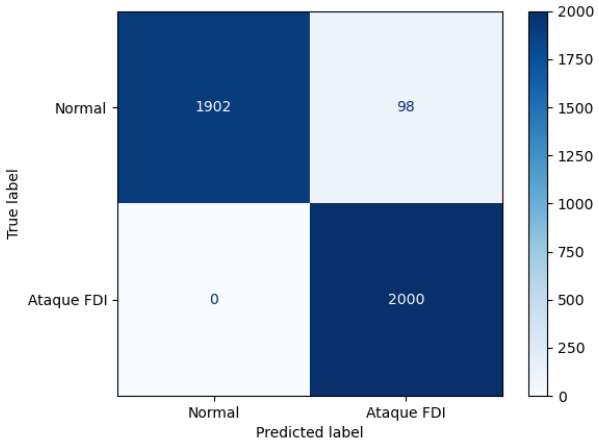
5 RESULTADOS E DISCUSSÕES

Neste capítulo, são apresentados os resultados obtidos após treinamento e detecção dos modelos para cada sistema testado.

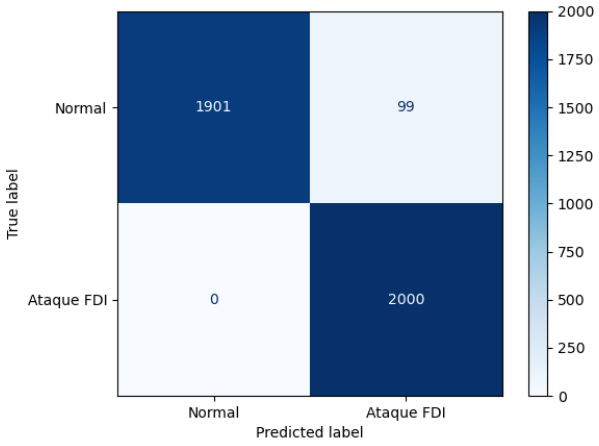
5.1 Sistema IEEE-14 barras

O modelo foi treinado com 16.000 amostras, geradas para cada amplitude de variação. Posteriormente, amostras com ataques foram usadas para calcular o erro de reconstrução e realizar a detecção.

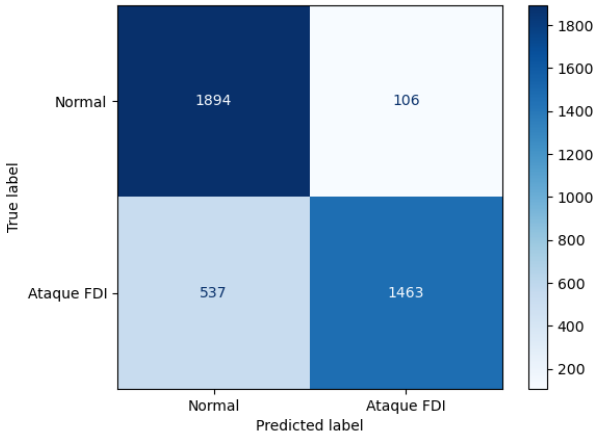
Figura 12 – Matriz de confusão - Sistema IEEE-14 Barras.



(a) Amplitude de ataque ±50%.



(b) Amplitude de ataque ±25%.



(c) Amplitude de ataque ±15%.

Fonte: Autor

As matrizes de confusão, apresentadas na Figura 12, para o sistema IEEE-14 revelam um desempenho diferenciado conforme a amplitude do ataque. Para variações de ±50% e ±25%, observa-se uma predominância de verdadeiros positivos, com taxas de detecção perfeitas (100% de *recall*), indicando que o modelo foi capaz de identificar todos os eventos de ataque sem registros de falsos negativos. No entanto, na amplitude de ±15%, verifica-se um aumento significativo de falsos negativos, com o *recall* caindo para 73,15%. Esse comportamento sugere que o modelo enfrenta dificuldades em distinguir ataques sutis das flutuações normais do sistema, apontando para uma limitação na sensibilidade do método para pequenas variações.

Na Tabela 7 são apresentadas as métricas de avaliação de desempenho para o sistema IEEE-14 barras.

Tabela 7 – Métricas de avaliação de desempenho para o sistema IEEE-14 Barras.

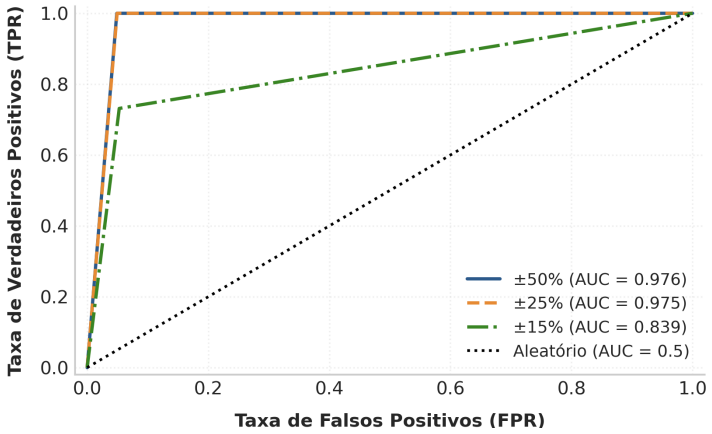
Amplitude de ataque	**Acurácia**	**Precisão**	***Recall***	***f1-score***	***threshold***
± 50%	97,55%	95,33%	100%	97,61%	0,057
± 25%	97,53%	95,28%	100%	97,58%	0,172
± 15%	83,93%	93,24%	73,15%	81,98%	0,635

As métricas apresentadas corroboram os padrões observados nas matrizes de confusão. Para amplitudes de ±50% e ±25%, o modelo exibe acurácia acima de 97,5% e *f1-score* próximo de 97,6%, demonstrando um equilíbrio adequado entre precisão e *recall*. Contudo, para ±15%, a acurácia cai para 83,93%, e o *f1-score* reduz-se para 81,98%, refletindo a queda no *recall*. O limiar de 0,635 nesse cenário indica que o modelo exigiu um ajuste mais conservador para minimizar falsos positivos, mas às custas de uma maior taxa de falsos negativos.

A análise da curva ROC, apresentada na Figura 13, complementa esses resultados, com AUCs de 0,976 (±50%) e 0,975 (±25%), demonstrando excelente capacidade discriminatória. Para ±15%, a AUC cai para 0,839, revelando uma performance significativamente inferior. Essa redução está alinhada com o aumento de falsos negativos observado na matriz de confusão, reforçando a necessidade de ajustes no modelo para melhorar sua sensibilidade a ataques menos intensos.

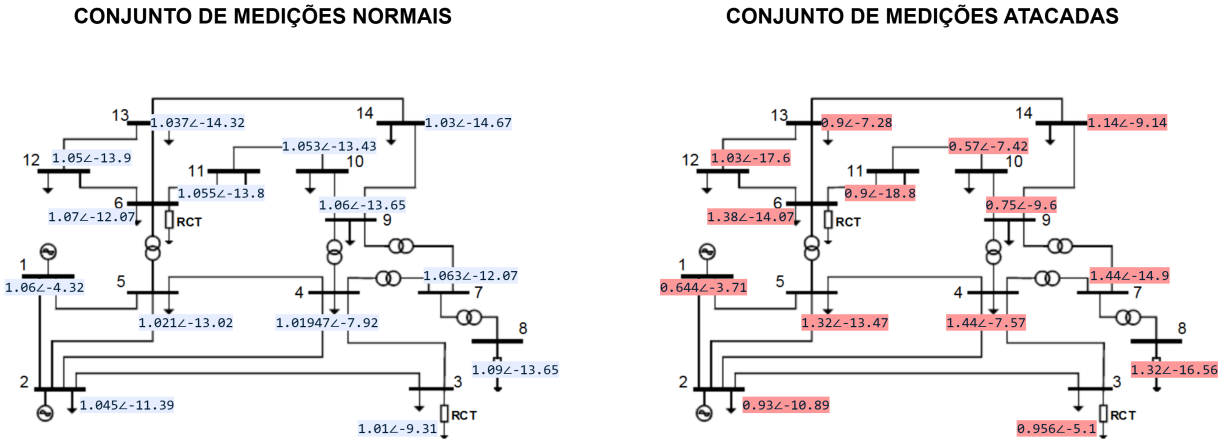
Para ilustrar o esquema de detecção, o sistema apresentado na Figura 14 apresenta as medições de tensão e ângulo de fase normais e atacadas com amplitude ±50% em cada barra de um sistema IEEE-14 barras. O erro de reconstrução retornado pelo *autoencoder* para as medições normais foi 0.01808 e para as medições atacadas o erro foi 2.45, como o limiar foi de 0.057 o *autoencoder* identifica como ataque FDI.

Figura 13 – Curva de ROC para o sistema IEEE-14 Barras



Fonte: Autor

Figura 14 – Exemplificação de medições normais e atacadas em cada barra do sistema IEEE-14 barras.



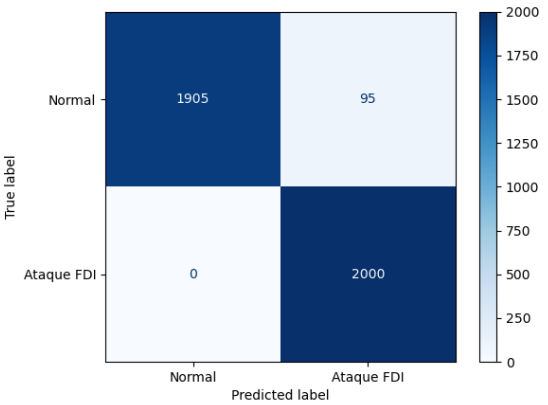
Fonte: Autor

5.2 Sistema IEEE-30 barras

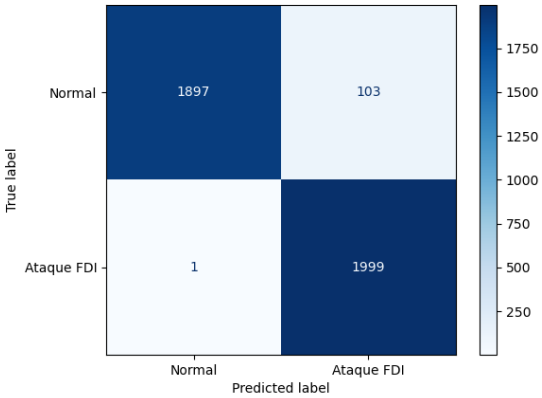
As matrizes de confusão, apresentadas na Figura 15, mostram um desempenho alto em todas as amplitudes de ataque. Destaque para o *recall* de 100% (±50% e ±25%) e 99,95% (±15%), indicando que o modelo manteve alta sensibilidade mesmo para variações sutis. A ausência de falsos negativos significativos sugere que o sistema possui uma capacidade superior de distinguir entre ataques e ruído operacional, comparado ao IEEE-14.

As métricas, apresentadas na Tabela 8, reforçam a consistência do modelo, com acurácia variando entre 97,40% e 97,70% e *f1-score* sempre acima de 97,4%. Os limiares apresentam variação expressiva (1,724 para ±50% e 0,009 para ±15%), indicando adaptabilidade do modelo a diferentes intensidades de ataque. Essa flexibilidade sugere que o método ajusta-se dinamicamente para manter alta precisão e *recall*, independentemente

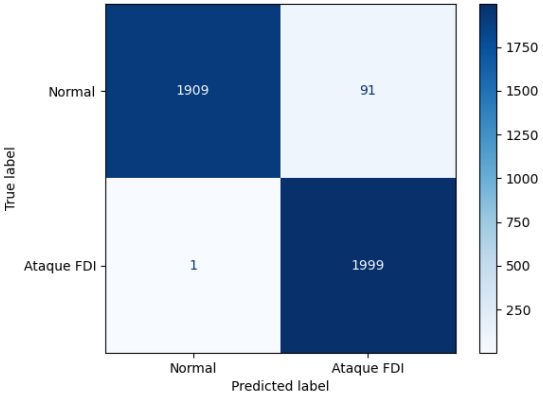
Figura 15 – Matriz de confusão - Sistema IEEE-30 Barras.



(a) Amplitude de ataque ±50%.



(b) Amplitude de ataque ±25%



(c) Amplitude de ataque ±15%

Fonte: Autor

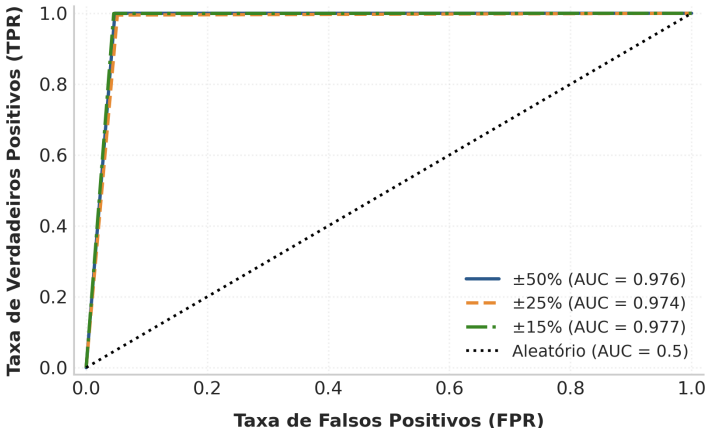
da amplitude.

Tabela 8 – Métricas de avaliação de desempenho para o sistema IEEE-30 Barras.

Amplitude de ataque	**Acurácia**	**Precisão**	***Recall***	***f1-score***	***thresholds***
± 50%	97,63%	95,46%	100%	97,68%	1,724
± 25%	97,40%	95,00%	99,95%	97,46%	0,013
± 15%	97,70%	95,64%	99,95%	97,75%	0,009

A curva ROC, apresentada na Figura 16, exibe AUCs superiores a 0,97 em todas as amplitudes, com curvas quase sobrepostas para ±50% e ±25%. Essa estabilidade confirma a solidez do modelo, que não apresenta a degradação observada no IEEE-14 para amplitudes menores. A performance consistente reforça a confiabilidade do método em sistemas de média complexidade.

Figura 16 – Curva de ROC para o sistema IEEE-30 Barras



Fonte: Autor

5.3 Sistema IEEE-118 barras

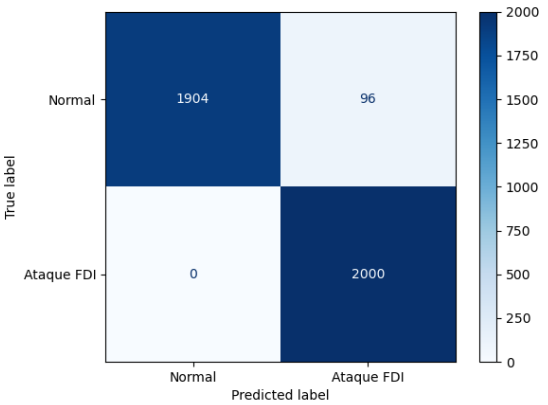
As matrizes de confusão do sistema IEEE-118, apresentadas na Figura 17, mostram padrões similares aos do IEEE-30, com altas taxas de verdadeiros positivos e *recall* próximo de 100% para amplitudes de ±50% e ±25%. Para a amplitude de ±15%, o recall mantém-se elevado (95,10%), com um ligeiro aumento de falsos negativos, o que sugere que, embora o sistema preserve boa sensibilidade, ataques sutis ainda representam um desafio, porém menos acentuado que no IEEE-14. Complementarmente, as métricas apresentadas na Tabela 9 revelam acurácia acima de 97,3% para ±50% e ±25%, com *f1-score* próximo de 97,5%. Para a amplitude de ±15%, a acurácia cai modestamente para 94,95%, enquanto o *f1-score* mantém-se em 94,96%, e o limiar de 0,120 para essa amplitude indica um equilíbrio adequado entre sensibilidade e especificidade, sem exigir ajustes extremos como no IEEE-14.

Tabela 9 – Métricas de avaliação de desempenho para o sistema IEEE-118 Barras.

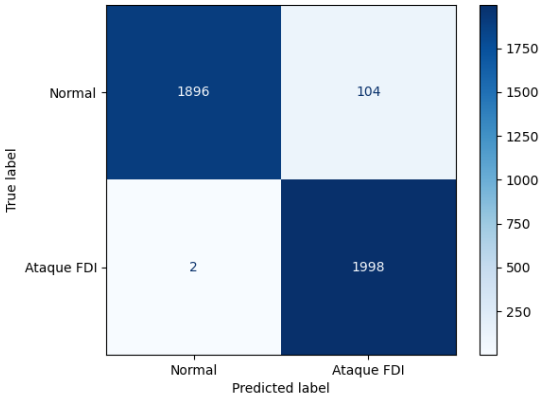
Amplitude de ataque	**Acurácia**	**Precisão**	***Recall***	***f1-score***	***threshold***
± 50%	97,60%	95,42%	100%	97,66%	0,100
± 25%	97,35%	95,05%	99,99%	97,42%	0,067
± 15%	94,95%	94,81%	95,10%	94,96%	0,120

Dentre as amplitudes avaliadas pela curva ROC, apresentada na Figura 18, o modelo ±50% se destaca com o maior AUC de 0.976, indicando um desempenho excelente e superior na distinção entre classes positivas e negativas; a amplitude ±25% apresenta um desempenho quase idêntico com um AUC de 0.974, mostrando uma diferença marginal, enquanto a amplitude de ±15%, apesar de um AUC ligeiramente menor de 0.950, ainda demonstra uma capacidade de classificação satisfatória.

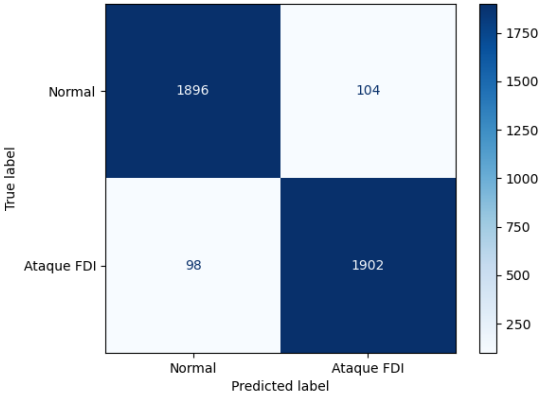
Figura 17 – Matriz de confusão - Sistema IEEE-118 Barras.



(a) Amplitude de ataque ±50%.



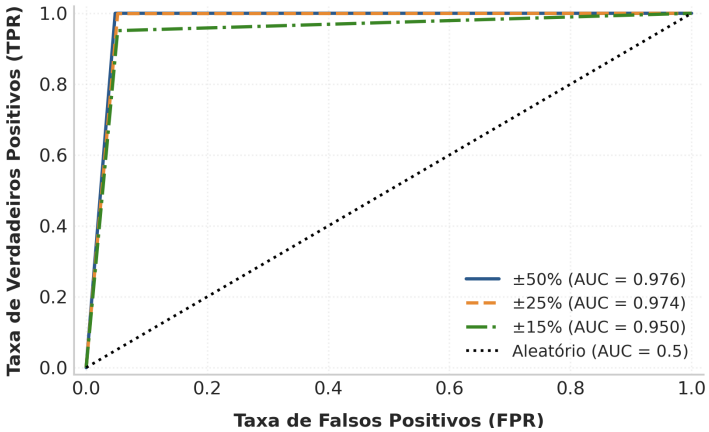
(b) Amplitude de ataque ±25%



(c) Amplitude de ataque ±15%

Fonte: Autor

Figura 18 – Curva de ROC para o sistema IEEE-118 Barras

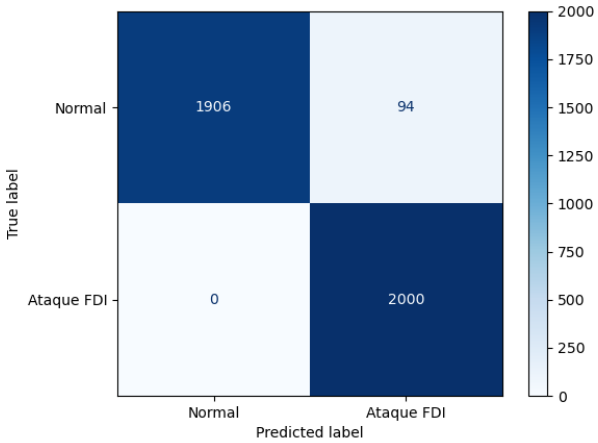


Fonte: Autor.

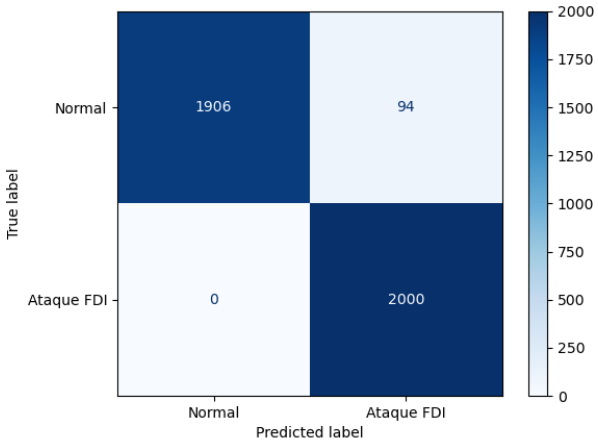
5.4 Sistema IEEE-300 barras

As matrizes de confusão do sistema IEEE-300, apresentadas na Figura 19, destacam-se pela performance quase perfeita, com *recall* de 100% (±50% e ±25%) e 99,95% (±15%). A ausência de falsos negativos significativos em qualquer amplitude demonstra que o modelo atinge máxima sensibilidade, mesmo para variações sutis, superando todos os demais sistemas analisados.

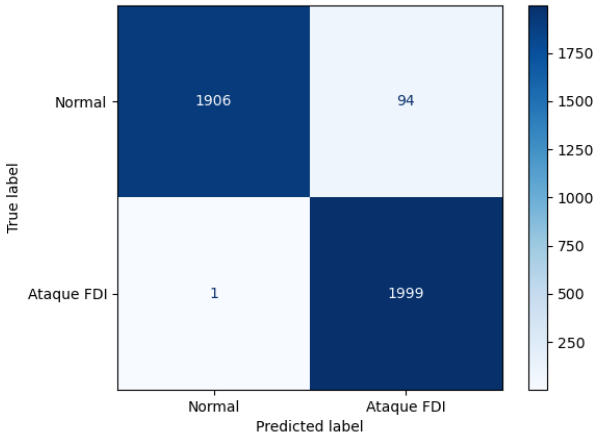
Figura 19 – Matriz de confusão - Sistema IEEE-300 Barras.



(a) Amplitude de ataque ±50%.



(b) Amplitude de ataque ±25%



(c) Amplitude de ataque ±15%

Fonte: Autor

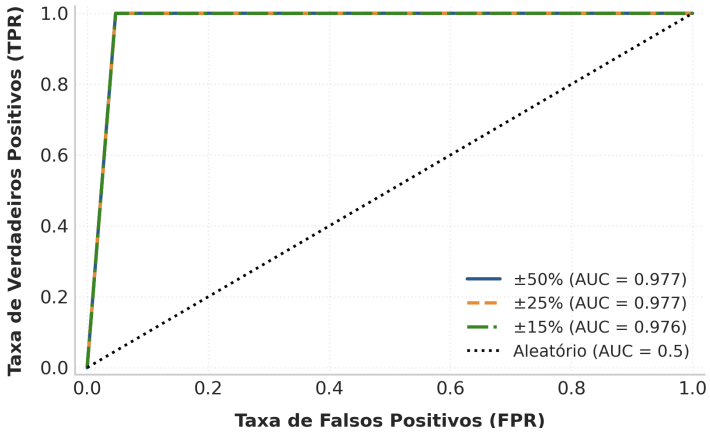
A Tabela 10 apresenta as métricas para o sistema IEEE-300 barras. As métricas exibem variações mínimas: acurácia entre 97,62% e 97,65%, e *f1-score* entre 97,67% e 97,70%. Os limiares mantêm-se estáveis (0,65), indicando parâmetros otimizados para todas as intensidades. Essa consistência reflete a maturidade do método em sistemas complexos.

Tabela 10 – Métricas de avaliação de desempenho para o sistema IEEE-300 Barras.

Amplitude de ataque	**Acurácia**	**Precisão**	***Recall***	***f1-score***	***threshold***
± 50%	97,65%	95,51%	100%	97,70%	0,647
± 25%	97,65%	95,51%	100%	97,70%	0,646
± 15%	97,62%	95,49%	99,95%	97,67%	0,655

A curva ROC, mostrada na Figura 20, apresenta AUC de 0,977 em todas as amplitudes, com curvas praticamente idênticas. Esse resultado consolida o IEEE-300 como o sistema com melhor desempenho global, sem trade-offs entre sensibilidade e especificidade, mesmo para ataques sutis.

Figura 20 – Curva de ROC para o sistema IEEE-300 Barras



Fonte: Autor

5.5 Considerações

A análise integrada dos resultados demonstra uma relação entre a complexidade do sistema e a eficácia do modelo de detecção: enquanto o sistema IEEE-14 apresentou limitações significativas na identificação de ataques sutis (±15%), refletidas na queda do *recall* (73,15%) e na AUC reduzida (0,839), os sistemas de maior porte (IEEE-30, IEEE-118 e IEEE-300) mantiveram desempenho consistente em todas as amplitudes, com *recall* acima de 99,95% e AUCs consistentemente superiores a 0,97. Essa disparidade sugere que a maior quantidade de dados e possivelmente a redundância inerente a sistemas complexos conferem maior resiliência ao modelo. Os limiares variáveis – desde valores extremamente baixos (0,009 no IEEE-30) até intermediários (0,635 no IEEE-14) – revelam a adaptabilidade do método, embora sistemas menores exijam ajustes mais conservadores que comprometem a sensibilidade. Esses resultados indicam que, para sistemas reduzidos, estratégias complementares (como amplificação de sinais ou *ensembles* de modelos) podem ser necessárias para melhorar a detecção de ataques sutis, enquanto o método atual

mostra-se plenamente adequado para redes de grande escala, mantendo precisão e *recall* elevados independentemente da intensidade do ataque.

6 CONCLUSÃO

A metodologia proposta nesta dissertação foi centrada no uso de *autoencoders* com mecanismo de atenção, particularmente com destaque para o uso da atenção *multi-head* para a detecção de ataques de injeção de dados falsos (FDI) em sistemas de potência. Este modelo se baseia no aprendizado profundo para detectar anomalias nos dados de medição, um problema presente em sistemas de energia modernos, que podem ser manipulados por ataques sofisticados como os FDI.

A abordagem metodológica consiste em treinar um *autoencoder*, uma arquitetura de rede neural que tenta aprender uma representação compactada e eficiente dos dados de entrada. O *autoencoder* tradicionalmente tem duas partes: um codificador que comprime os dados de entrada em uma representação de baixa dimensão (*latent space*) e um decodificador que tenta reconstruir os dados originais a partir dessa representação comprimida.

A principal inovação desta pesquisa reside na aplicação da camada de atenção *multi-head*, um avanço importante para aprimorar a análise de dados. Ao contrário de abordagens tradicionais que tratam todas as variáveis de forma homogênea, a atenção *multi-head* permite que o modelo identifique e pondere as informações mais relevantes em diferentes contextos, gerando múltiplas representações independentes dos dados. Essa capacidade de focar em aspectos diversos e críticos simultaneamente otimizou significativamente a detecção de anomalias, ao capturar com maior precisão as complexas dependências temporais ou espaciais presentes nos conjuntos de dados. Essa aplicação não só eleva a capacidade do modelo proposto, mas também abre novos caminhos para investigações futuras no campo.

No contexto de detecção de ataques FDI, cada *head* de atenção pode aprender diferentes padrões ou relações dentro dos dados de medição, como a correlação entre diferentes barras do sistema de energia ou as mudanças nas medições ao longo do tempo. Isso é essencial, pois, em um ataque FDI, os dados podem ser manipulados de forma a esconder a alteração por trás de informações aparentemente consistentes com o comportamento normal do sistema. A atenção *multi-head* melhora a detecção ao capturar essas múltiplas relações de forma eficiente.

Além disso, o mecanismo de atenção ajuda o modelo a ser interpretável, uma vez que podemos ver quais entradas ou variáveis o modelo está priorizando para identificar anomalias. Isso pode ser essencial em sistemas críticos, como redes de energia, onde os operadores precisam entender por que uma determinada medição foi considerada suspeita.

A geração de dados foi uma etapa fundamental para garantir que os modelos pudessem ser treinados e avaliados em condições realistas. Para isso, utilizou-se a ferramenta

MATPOWER, que permitiu a criação de cenários simulados de sistemas de potência com diferentes topologias e características. Foram gerados dados para quatro sistemas de diferentes complexidades: IEEE-14, IEEE-30, IEEE-118, e IEEE-300 barras.

Esses sistemas foram então sujeitos a ataques FDI, com injeções de dados falsos feitas de forma controlada para simular comportamentos de atacantes. A inserção de ataques foi feita de maneira a refletir tanto ataques direcionados (que visam manipular dados específicos) quanto aleatórios (que afetam um conjunto mais amplo de medições). Essa variação permitiu testar a resiliência do modelo em diferentes cenários de ataque e avaliar sua capacidade de generalizar para diferentes topologias e condições operacionais.

A validação da metodologia foi feita com base em métricas de avaliação de desempenho, como acurácia, precisão, *recall*, e AUC-ROC (curva de característica operacional do receptor). Esses testes foram realizados em todos os sistemas (IEEE-14, 30, 118 e 300 barras), e os resultados mostraram a eficácia do modelo proposto:

O modelo conseguiu detectar ataques FDI de forma precisa, mesmo em cenários onde as manipulações de dados eram sutis e difíceis de identificar por métodos tradicionais. A atenção *multi-head* desempenhou um papel relevante, permitindo que o modelo destacasse as medições mais relevantes e, portanto, aumentasse a taxa de detecção de ataques em comparação com modelos sem atenção.

O modelo foi testado em sistemas de diferentes complexidades e com diferentes padrões de ataque. Uma das vantagens do uso da atenção *multi-head* é que ela permitiu ao modelo aprender padrões de ataque que eram consistentes entre os sistemas, permitindo sua generalização para redes de diferentes escalas e características. Mesmo com topologias mais complexas, o modelo manteve um bom desempenho.

A utilização de atenção *multi-head* fez com que o modelo fosse mais eficiente em focar nas medições e eventos temporais mais críticos, melhorando significativamente sua capacidade de identificar anomalias associadas aos ataques FDI. Em cenários com dados ruidosos ou completamente manipulados, o modelo foi capaz de encontrar padrões ocultos nos dados, um desafio típico nos ataques de FDI, onde as distorções são deliberadamente sutis.

Embora os resultados tenham mostrado que a metodologia proposta, utilizando *autoencoders* com atenção *multi-head*, é altamente eficaz para a detecção de ataques FDI, existem algumas áreas de melhoria. O desempenho do modelo pode ser otimizado para grandes sistemas em tempo real, onde a latência e o uso de recursos computacionais são críticos. Além disso, a integração de outros tipos de aprendizado, como aprendizado por reforço, poderia melhorar a adaptabilidade do sistema em condições dinâmicas e variáveis.

Trabalhos futuros devem incluir a validação do modelo em cenários reais de operação de sistemas de potência, considerando a ocorrência de defeitos, curtos-circuitos e outras

anomalias físicas que podem interferir na detecção de ataques cibernéticos. Essa abordagem permitirá avaliar a eficiência do método sob condições operacionais mais complexas e ruidosas, além de verificar possíveis falsos positivos induzidos por falhas naturais do sistema. Adicionalmente, recomenda-se a investigação de técnicas de hibridização do modelo com métodos baseados em física para melhor discriminar entre eventos maliciosos e distúrbios convencionais, especialmente em sistemas de pequeno porte onde a detecção se mostrou mais desafiadora. Espera-se também otimizar o processo de geração de dados, utilizando outras abordagens de *softwares* especializados e simuladores em tempo real.

REFERÊNCIAS

ABDALLAH, A. Security and privacy in smart grid. In: *University of Waterloo*. [S.l.: s.n.], 2016. 14

AHMED, S.; LEE, Y.; HYUN, S.-H.; KOO, I. Unsupervised machine learning-based detection of covert data integrity assault in smart grid networks utilizing isolation forest. *Trans. Info. For. Sec.*, IEEE Press, v. 14, n. 10, p. 2765–2777, out. 2019. ISSN 1556-6013. Disponível em: <https://doi.org/10.1109/TIFS.2019.2902822>. 35

ALMUTAIRY, F. *Enhancing Cybersecurity of Power Systems using Machine Learning*. Dissertação (Mestrado) — University of Vermont, Vermont, United States of America, 2022. 28, 29, 33

AYAD, A.; FARAG, H. E. Z.; YOUSSEF, A.; EL-SAADANY, E. F. Detection of false data injection attacks in smart grids using recurrent neural networks. In: *2018 IEEE Power and Energy Society Innovative Smart Grid Technologies Conference (ISGT)*. [S.l.: s.n.], 2018. p. 1–5. 35

BOBBA, R.; DAVIS, K.; WANG, Q.; KHURANA, H.; NAHRSTEDT, K.; OVERBYE, T. Detecting false data injection attacks on dc state estimation. 01 2010. 29

CáRDENAS, A. A.; AMIN, S.; SASTRY, S. Research challenges for the security of control systems. In: *Proceedings of the 3rd Conference on Hot Topics in Security*. USA: USENIX Association, 2008. (HOTSEC'08). 25

CHALAPATHY, R.; CHAWLA, S. *Deep Learning for Anomaly Detection: A Survey*. 2019. 37

CHOLLET, F. *Deep Learning with Python*. First. [S.l.]: Manning Publications Co., 2017. 36

DING, Y.; MA, K.; PU, T.; WANG, X.; LI, R.; ZHANG, D. A deep learning-based classification scheme for false data injection attack detection in power system. *Electronics*, v. 10, n. 12, 2021. ISSN 2079-9292. Disponível em: <https://www.mdpi.com/2079-9292/10/12/1459>. 48, 52

ESMALIFALAK, M.; LIU, L.; NGUYEN, N.; ZHENG, R.; HAN, Z. Detecting stealthy false data injection using machine learning in smart grid. *IEEE Systems Journal*, v. 11, n. 3, p. 1644–1652, 2017. 35, 47, 51

ESMALIFALAK, M.; NGUYEN, N. T.; ZHENG, R.; HAN, Z. Detecting stealthy false data injection using machine learning in smart grid. In: *2013 IEEE Global Communications Conference (GLOBECOM)*. [S.l.: s.n.], 2013. p. 808–813. 46, 51

FLECK, L.; TAVARES, M. H. F.; EYNG, E.; HELMANN, A. C.; ANDRADE, M. A. de M. Redes neurais artificiais: princípios básicos. *Revista Eletrônica Científica Inovação e Tecnologia*, UTFPR, v. 7, n. 15, 2016. Disponível em: <https://doi.org/10.3895/recit.v7.n15.4330>. 37

G, A.; KH, V.; C, J. S. V.; NAIR, M. G. Autoencoder based fdi attack detection scheme for smart grid stability. In: *2022 IEEE 19th India Council International Conference (INDICON)*. [S.l.: s.n.], 2022. p. 1–5. 50, 53

GOODFELLOW, I. J.; BENGIO, Y.; COURVILLE, A. *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016. <http://www.deeplearningbook.org>. 36

HABIB, A. A.; HASAN, M. K.; ALKHAYYAT, A.; ISLAM, S.; SHARMA, R.; ALKWAI, L. M. False data injection attack in smart grid cyber physical system: Issues, challenges, and future direction. *Computers and Electrical Engineering*, v. 107, p. 108638, 2023. ISSN 0045-7906. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0045790623000630>. 10, 14, 15, 21, 22, 31

HASAN, M. K.; ABDULKADIR, R. A.; ISLAM, S.; GADEKALLU, T. R.; SAFIE, N. A review on machine learning techniques for secured cyber-physical systems in smart grid networks. *Energy Reports*, v. 11, p. 1268–1290, 2024. ISSN 2352-4847. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2352484723016323>. 22

HASAN, M. K.; HABIB, A. A.; SHUKUR, Z.; IBRAHIM, F.; ISLAM, S.; RAZZAQUE, M. A. Review on cyber-physical and cyber-security system in smart grid: Standards, protocols, constraints, and recommendations. *Journal of Network and Computer Applications*, v. 209, p. 103540, 2023. ISSN 1084-8045. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1084804522001813>. 21

HINTON, G. E. Learning multiple layers of representation. *Trends in Cognitive Sciences*, v. 11, n. 10, p. 428–434, 2007. ISSN 1364-6613. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1364661307002173>. 39

HINTON, G. E.; SALAKHUTDINOV, R. R. Reducing the dimensionality of data with neural networks. *Science*, v. 313, n. 5786, p. 504–507, 2006. Disponível em: <https://www.science.org/doi/abs/10.1126/science.1127647>. 17, 37

HONGSHAN, Z.; HUIHAI, L.; WENJING, H.; XIHUI, Y. Anomaly detection and fault analysis of wind turbine components based on deep learning network. *Renewable Energy*, v. 127, 05 2018. 35

HUSNOO, M. A.; ANWAR, A.; HOSSEINZADEH, N.; ISLAM, S. N.; MAHMOOD, A. N.; DOSS, R. *False Data Injection Threats in Active Distribution Systems: A Comprehensive Survey*. 2022. Disponível em: <https://arxiv.org/abs/2111.14251>. 31

KARIMIPOUR, H.; DINAVAHI, V. On false data injection attack against dynamic state estimation on smart power grids. In: *2017 IEEE International Conference on Smart Energy Grid Engineering (SEGE)*. [S.l.: s.n.], 2017. p. 388–393. 16

KHANNA, K.; PANIGRAHI, B. K.; JOSHI, A. Ai-based approach to identify compromised meters in data integrity attacks on smart grid. *IET Generation, Transmission & Distribution*, v. 12, n. 5, p. 1052–1066, 2018. Disponível em: <https://ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/iet-gtd.2017.0455>. 35

KHAZRAEI, A.; HALLYBURTON, S.; GAO, Q.; WANG, Y.; PAJIC, M. *Learning-Based Vulnerability Analysis of Cyber-Physical Systems*. 2022. Disponível em: <https://arxiv.org/abs/2103.06271>. 25

KUNDU, A.; SAHU, A.; SERPEDIN, E.; DAVIS, K. A3d: Attention-based auto-encoder anomaly detector for false data injection attacks. *Electric Power Systems Research*, v. 189, p. 106795, 2020. ISSN 0378-7796. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0378779620305988>. 38, 49, 52

KUNDUR, D.; FENG, X.; LIU, S.; ZOURNTOS, T.; BUTLER-PURRY, K. L. Towards a framework for cyber attack impact analysis of the electric smart grid. In: *2010 First IEEE International Conference on Smart Grid Communications*. [S.l.: s.n.], 2010. p. 244–249. 24

KURT, M. N.; OGUNDIJO, O. E.; LI, C.; WANG, X. Online cyber-attack detection in smart grid: A reinforcement learning approach. *CoRR*, abs/1809.05258, 2018. Disponível em: <http://arxiv.org/abs/1809.05258>. 36

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. *Nature*, v. 521, n. 7553, p. 436–444, 2015. ISSN 1476-4687. Disponível em: <https://doi.org/10.1038/nature14539>. 36

LI, B.; XIAO, G.; LU, R.; DENG, R.; BAO, H. On feasibility and limitations of detecting false data injection attacks on power grid state estimation using d-facts devices. *IEEE Transactions on Industrial Informatics*, v. 16, n. 2, p. 854–864, 2020. 46

LIANG, G.; WELLER, S. R.; ZHAO, J.; LUO, F.; DONG, Z. Y. The 2015 ukraine blackout: Implications for false data injection attacks. *IEEE Transactions on Power Systems*, v. 32, n. 4, p. 3317–3318, 2017. 15, 29

LIU, B.; WU, H.; YANG, Q.; LIU, X.; LIU, Y. Data-driven fdi attacks: A stealthy approach to subvert svm detectors in power system. In: *2023 IEEE Kansas Power and Energy Conference (KPEC)*. [S.l.: s.n.], 2023. p. 1–6. 47, 53

LIU, C.; HORNIKX, M. Effect of water content on noise attenuation over vegetated roofs: Results from two field studies. *Building and Environment*, v. 146, p. 1–11, 2018. ISSN 0360-1323. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0360132318305705>. 46

LIU, Y.; NING, P.; REITER, M. K. False data injection attacks against state estimation in electric power grids. *ACM Trans. Inf. Syst. Secur.*, Association for Computing Machinery, New York, NY, USA, v. 14, n. 1, jun 2011. ISSN 1094-9224. Disponível em: <https://doi.org/10.1145/1952982.1952995>. 30, 34, 45

LUONG, M.-T.; PHAM, H.; MANNING, C. D. *Effective Approaches to Attention-based Neural Machine Translation*. 2015. Disponível em: <https://arxiv.org/abs/1508.04025>. 39

MAJIDI, S. H.; HADAYEGHPARAST, S.; KARIMIPOUR, H. Fdi attack detection using extra trees algorithm and deep learning algorithm-autoencoder in smart grid. *International Journal of Critical Infrastructure Protection*, v. 37, p. 100508, 2022. ISSN 1874-5482. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1874548222000014>. 50, 52, 59

MONTICELLI, A. *State Estimation in Electric Power Systems*. 3Island Press, 1999. ISBN 9781461550006. Disponível em: <https://books.google.com.br/books?id=axAhswEACAAJ>. 29

MUSLEH, A. S.; CHEN, G.; DONG, Z. Y. A survey on the detection algorithms for false data injection attacks in smart grids. *IEEE Transactions on Smart Grid*, v. 11, n. 3, p. 2218–2234, 2020. 32, 33

NAIR, V.; HINTON, G. E. Rectified linear units improve restricted boltzmann machines. In: *Proceedings of the 27th International Conference on International Conference on Machine Learning*. Madison, WI, USA: Omnipress, 2010. (ICML'10), p. 807–814. ISBN 9781605589077. 37

NIU, X.; LI, J.; SUN, J.; TOMSOVIC, K. Dynamic detection of false data injection attack in smart grid using deep learning. In: *2019 IEEE Power and Energy Society Innovative Smart Grid Technologies Conference (ISGT)*. [S.l.: s.n.], 2019. p. 1–6. 17, 48, 52

ONS. *Submódulo 2.6: Requisitos mínimos para subestações e seus equipamentos*. Rio de Janeiro: ONS, 2022. Documento técnico. Revisão 2022.08. Disponível em: <http://www.ons.org.br>. 60

RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning representations by back-propagating errors. *Nature*, v. 323, p. 533–536, 1986. Disponível em: <https://doi.org/10.1038/323533a0>. 36

SAKHNINI, J.; KARIMIPOUR, H.; DEHGHANTANHA, A. Smart grid cyber attacks detection using supervised learning and heuristic feature selection. In: *2019 IEEE 7th International Conference on Smart Energy Grid Engineering (SEGE)*. [S.l.: s.n.], 2019. p. 108–112. 14

SCHWEPPE, F. C.; WILDES, J. Power system static-state estimation, part i: Exact model. *IEEE Transactions on Power Apparatus and Systems*, PAS-89, n. 1, p. 120–125, 1970. 26, 28

STOTT, B.; JARDIM, J.; ALSAC, O. Dc power flow revisited. *IEEE Transactions on Power Systems*, v. 24, n. 3, p. 1290–1300, 2009. 27

TEIXEIRA, A.; SHAMES, I.; SANDBERG, H.; JOHANSSON, K. H. A secure control framework for resource-limited adversaries. *Automatica*, v. 51, p. 135–148, 2015. ISSN 0005-1098. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0005109814004488>. 25

VASWANI, A.; SHAZEER, N.; PARMAR, N.; USZKOREIT, J.; JONES, L.; GOMEZ, A. N.; KAISER, L.; POLOSUKHIN, I. *Attention Is All You Need*. 2023. Disponível em: <https://arxiv.org/abs/1706.03762>. 39, 40

VIMALKUMAR, K.; RADHIKA, N. A big data framework for intrusion detection in smart grids using apache spark. In: *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. [S.l.: s.n.], 2017. p. 198–204. 35

WANG, S.; BI, S.; ZHANG, Y.-J. A. Locational detection of the false data injection attack in a smart grid: A multilabel classification approach. *IEEE Internet of Things Journal*, v. 7, n. 9, p. 8218–8227, 2020. 16

WANG, W.; LU, Z. Cyber security in the smart grid: Survey and challenges. *Computer Networks*, v. 57, n. 5, p. 1344–1371, 2013. ISSN 1389-1286. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1389128613000042>. 16

WANG, Y.; ZHOU, Y.; MA, J.; JIN, Q. A locational false data injection attack detection method in smart grid based on adversarial variational autoencoders image 1. *Applied Soft Computing*, v. 151, p. 111169, 2024. ISSN 1568-4946. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1568494623011870>. 50, 53

YAN, J.; TANG, B.; HE, H. Detection of false data attacks in smart grid with supervised learning. In: *2016 International Joint Conference on Neural Networks (IJCNN)*. [S.l.: s.n.], 2016. p. 1395–1402. 35

YANG, Y.; WANG, S.; WEN, M.; XU, W. Reliability modeling and evaluation of cyber-physical system (cps) considering communication failures. *Journal of the Franklin Institute*, v. 358, n. 1, p. 1–16, 2021. ISSN 0016-0032. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0016003218305994>. 49, 52

YU, J. J. Q.; HOU, Y.; LI, V. O. K. Online false data injection attack detection with wavelet transform and deep neural networks. *IEEE Transactions on Industrial Informatics*, v. 14, n. 7, p. 3271–3280, 2018. 29

ZANETTI, M.; JAMHOUR, E.; PELLENZ, M.; PENNA, M.; ZAMBENEDETTI, V.; CHUEIRI, I. A tunable fraud detection system for advanced metering infrastructure using short-lived patterns. *IEEE Transactions on Smart Grid*, v. 10, n. 1, p. 830–840, 2019. 35

ZHANG, Y.; KRISHNAN, V. V. G.; PI, J.; KAUR, K.; SRIVASTAVA, A.; HAHN, A.; SURESH, S. Cyber physical security analytics for transactive energy systems. *IEEE Transactions on Smart Grid*, v. 11, n. 2, p. 931–941, 2020. 34

ZHAO, J.; GóMEZ-EXPóSITO, A.; NETTO, M.; MILI, L.; ABUR, A.; TERZIJA, V.; KAMWA, I.; PAL, B.; SINGH, A. K.; QI, J.; HUANG, Z.; MELIOPOULOS, A. P. S. Power system dynamic state estimation: Motivations, definitions, methodologies, and future work. *IEEE Transactions on Power Systems*, v. 34, n. 4, p. 3188–3198, 2019. 32, 33

ZHU, B.; JOSEPH, A.; SASTRY, S. A taxonomy of cyber attacks on scada systems. In: *2011 International Conference on Internet of Things and 4th International Conference on Cyber, Physical and Social Computing*. [S.l.: s.n.], 2011. p. 380–388. 24