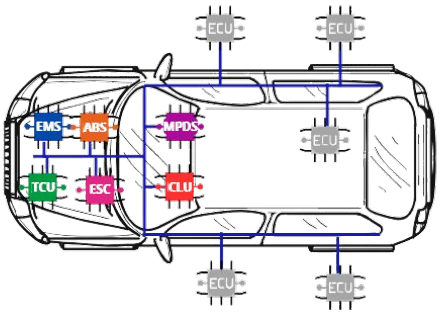
Modelos Interpretáveis com Inteligência Artificial Explicável (XAI) na Detecção de Intrusões em Redes Intra-Veiculares Controller Area Network (CAN)

Felipe N. Dresch, Felipe H. Scherer,
Silvio E. Quincozes, Diego Kreutz

Até que ponto a inteligência artificial explicável ajuda a entender o comportamento de um atacante?

Redes Intra-veiculares

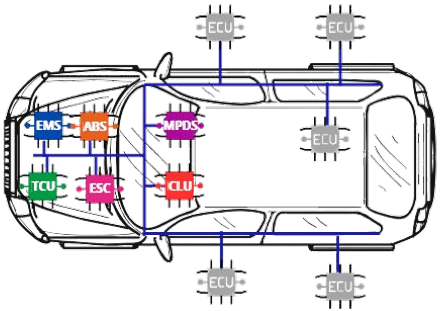
- **O protocolo Controller Area Network (CAN) foi proposto para a comunicação entre componentes dos veículos.**
  - **É largamente utilizado por veículos atuais para suprir a necessidade de interconexão entre as ECUs do veículo.**



Redes Intra-veiculares são vulneráveis!

- **O protocolo Controller Area Network (CAN) foi proposto para a comunicação entre componentes dos veículos.**
  - **É largamente utilizado por veículos atuais para suprir a necessidade de interconexão entre as ECUs do veículo.**

Vulnerabilidades!
Fuzzing, spoofing, DoS, etc.



Detecção de Intrusões em Redes CAN

- **Sistemas de Detecção de Intrusões (IDS) baseados em *machine learning* são promissores para identificar ataques às redes CAN!**
  - Mas eles tipicamente carecem de **explicabilidade**.

É fundamental que as decisões de IDS possam ser devidamente explicadas e que possam serem interpretadas!

Problema

- **Abordar a interpretabilidade das métricas XAI no domínio onde o IDS é implantado!**
  - X-IDSs limitam-se a **mostrar medidas** resultantes das ferramentas e **apontar características relevantes**!
  - As decisões práticas que podem ser tomadas a partir dessas métricas se limitam à seleção de características.

Problema de pesquisa: Faltam discussões sobre a origem do ataque e do modo de operação do atacante!

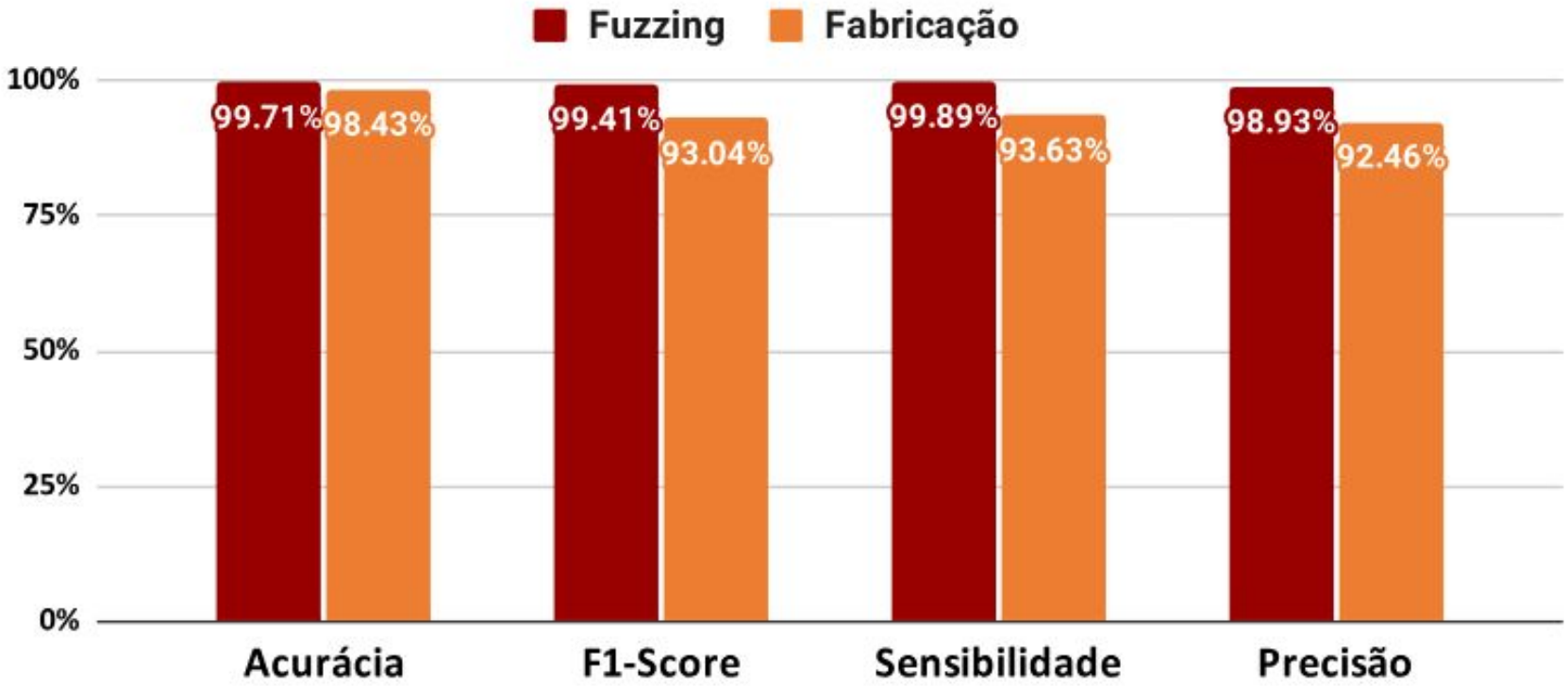
Potencializando técnicas de XAI em IDSs para redes intra-veiculares

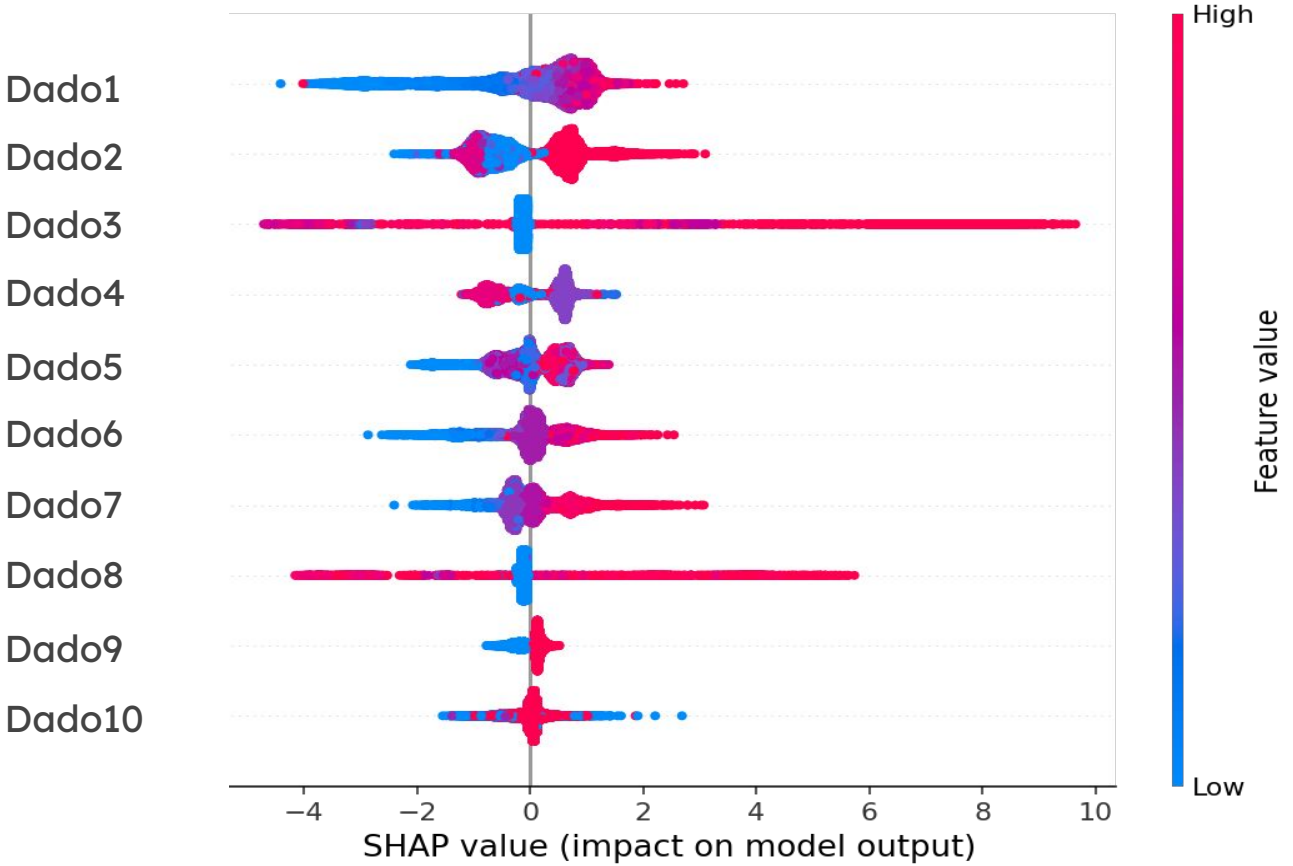
- **Avanços na interpretabilidade para melhorar a detecção de intrusões em redes CAN**
  - Aplicação do algoritmo **XGBoost**;
  - Integração da biblioteca **SHAP**;
  - **Mapeamento** das características CAN relevantes;.
  - Explicabilidade através da **análise técnica e engenharia reversa.**

Cenário de Experimentação

- **Materiais usados na implementação:**
  - Windows 11 Pro;
  - Intel Core i5 12° geração;
  - 16 Gb de memória RAM.
- **Métodos utilizados:**

Resultados

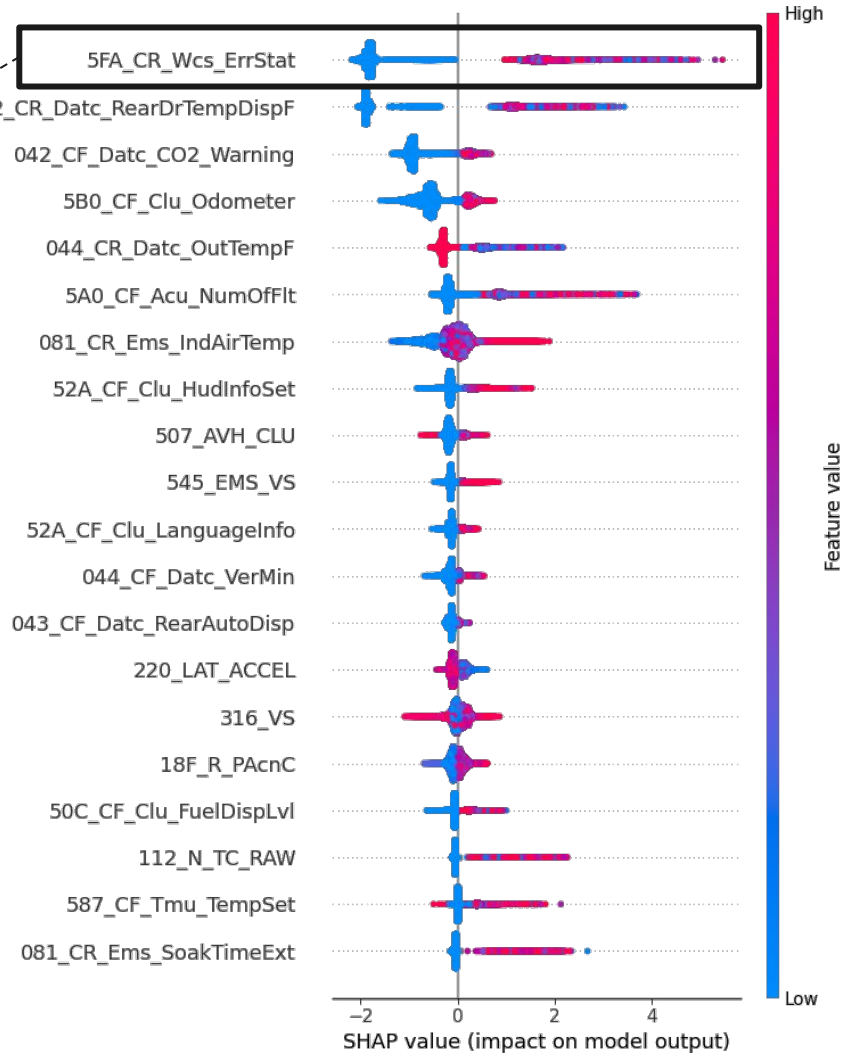

Fuzzing
Fabricação
100%
75%
50%
25%
0%
99.71%
98.43%
99.41%
93.04%
99.89%
93.63%
98.93%
92.46%
Acurácia
F1-Score
Sensibilidade
Precisão



Fonte: https://shap.readthedocs.io/en/latest/example_notebooks/api_examples/plots/beeswarm.html

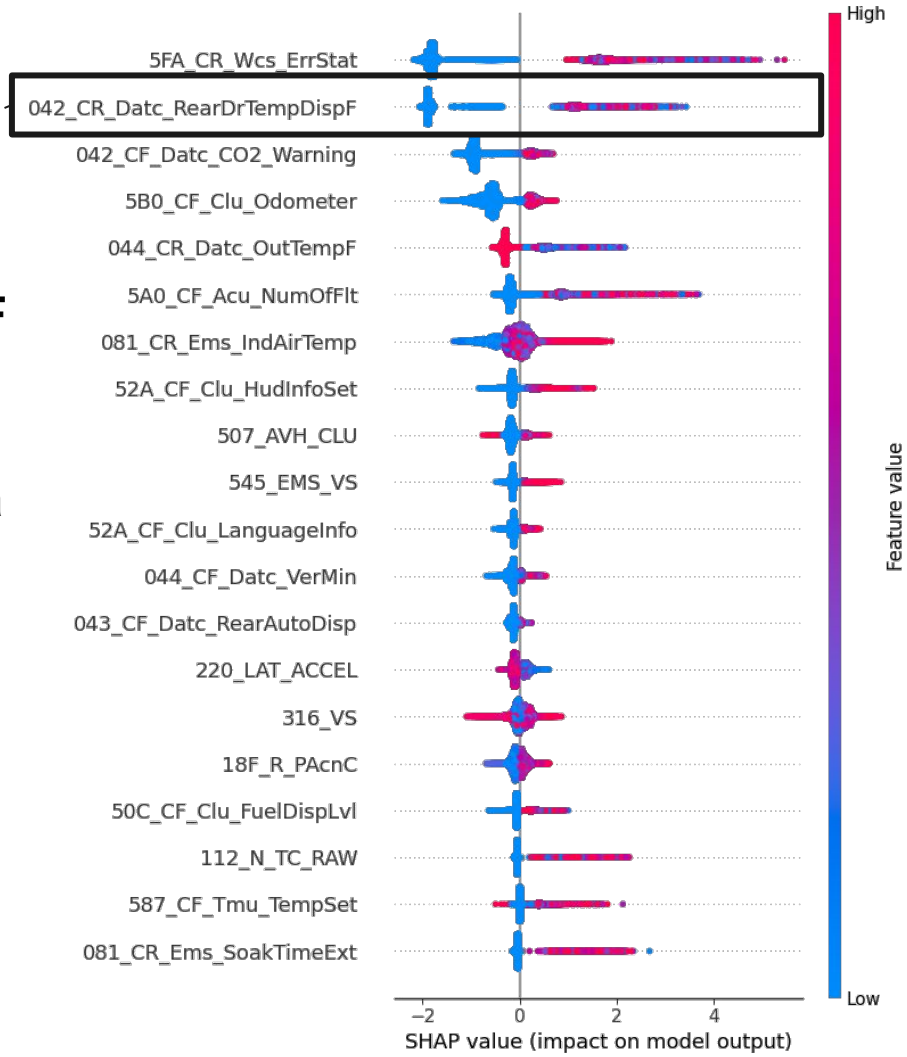
Ataque Fuzzing

- **5FA_CR_Wcs_ErrStat**
  - **Valores altos:** associados a valores SHAP positivos;
  - **Significado:** contribui mais para a identificação correta de ataques;
  - **Insights:** Indica erro do sistema de controle de carga do veículo.



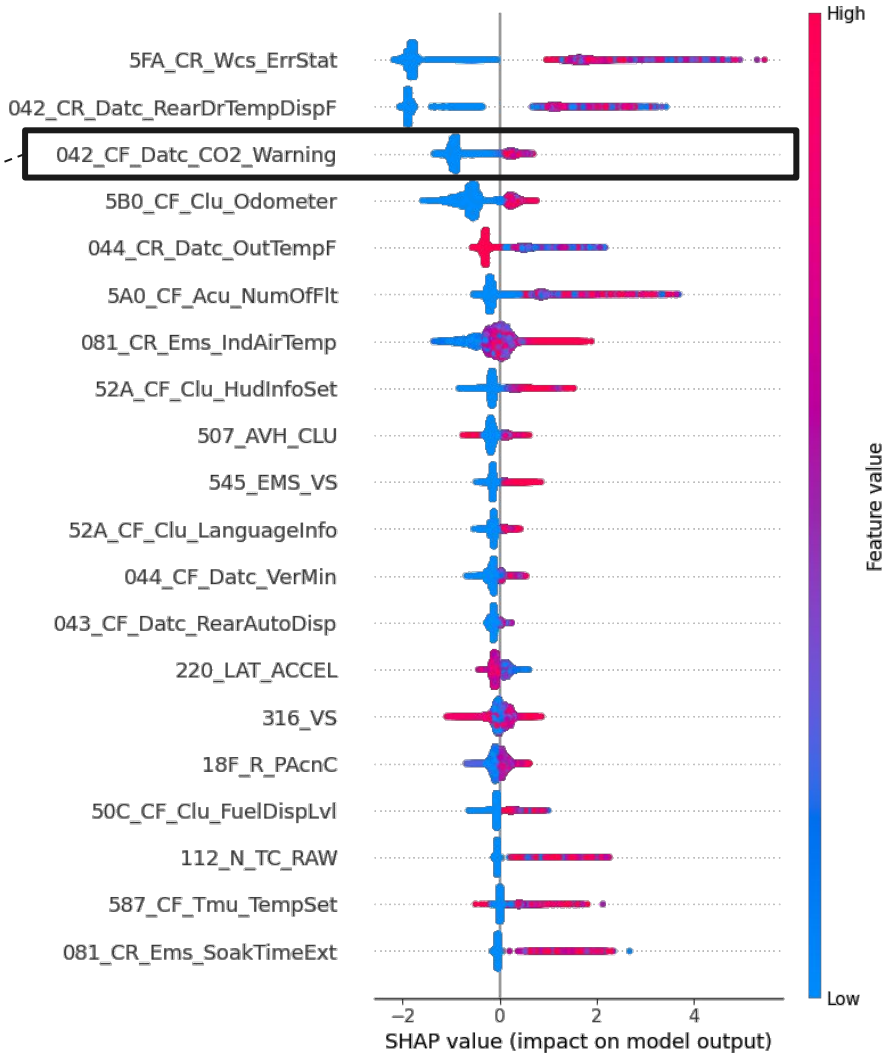
Ataque Fuzzing

- **042_CR_Datc_RearDrTempDispF**
  - **Valores altos:** associados a valores SHAP positivos;
  - **Significado:** contribui para a identificação de ambas as classes, com mais impacto em cenários de ataque;
  - **Insights:** relacionada ao sistema de controle climático (temperatura exibida na porta traseira).



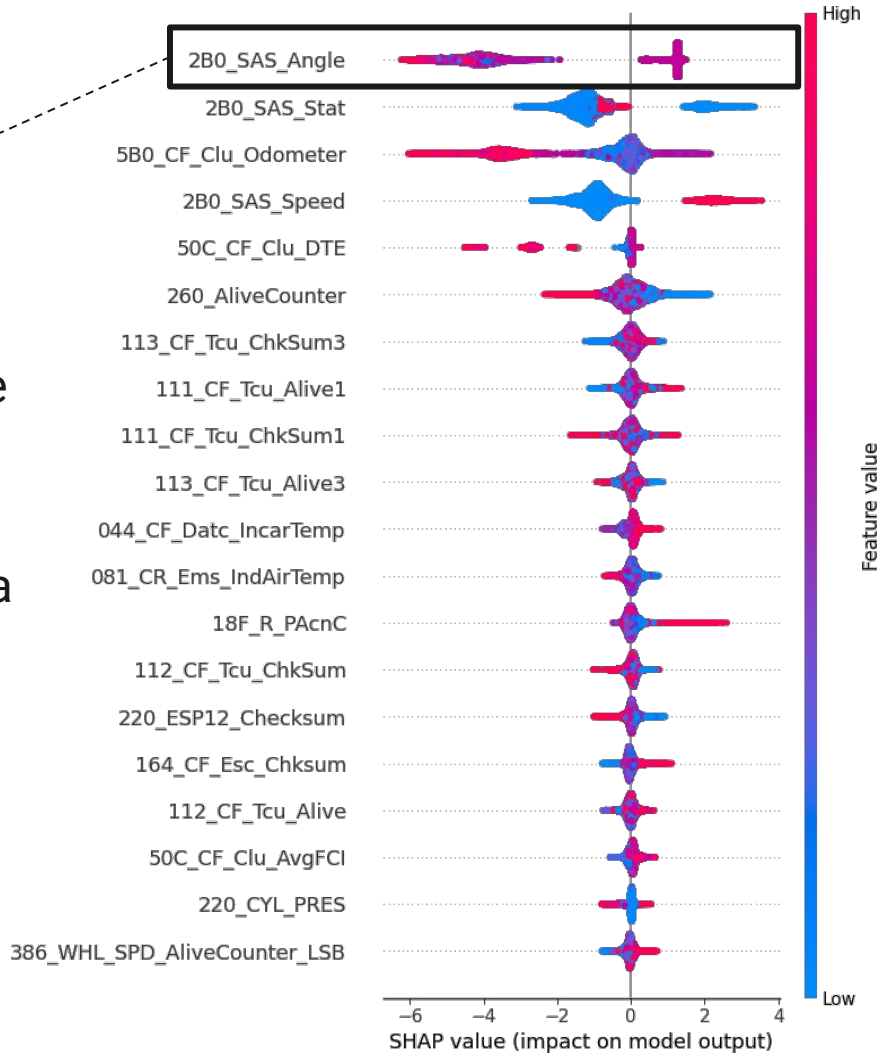
Ataque Fuzzing

- **042_CF_Datc_CO2_Warning**
  - **Valores altos:** associados a valores SHAP positivos, mas de menor impacto;
  - **Significado:** contribui mais para a identificação de cenários sem ataque.
  - **Insights:** associada ao alerta de emissão de CO2 emitido pelo sistema de controle climático.



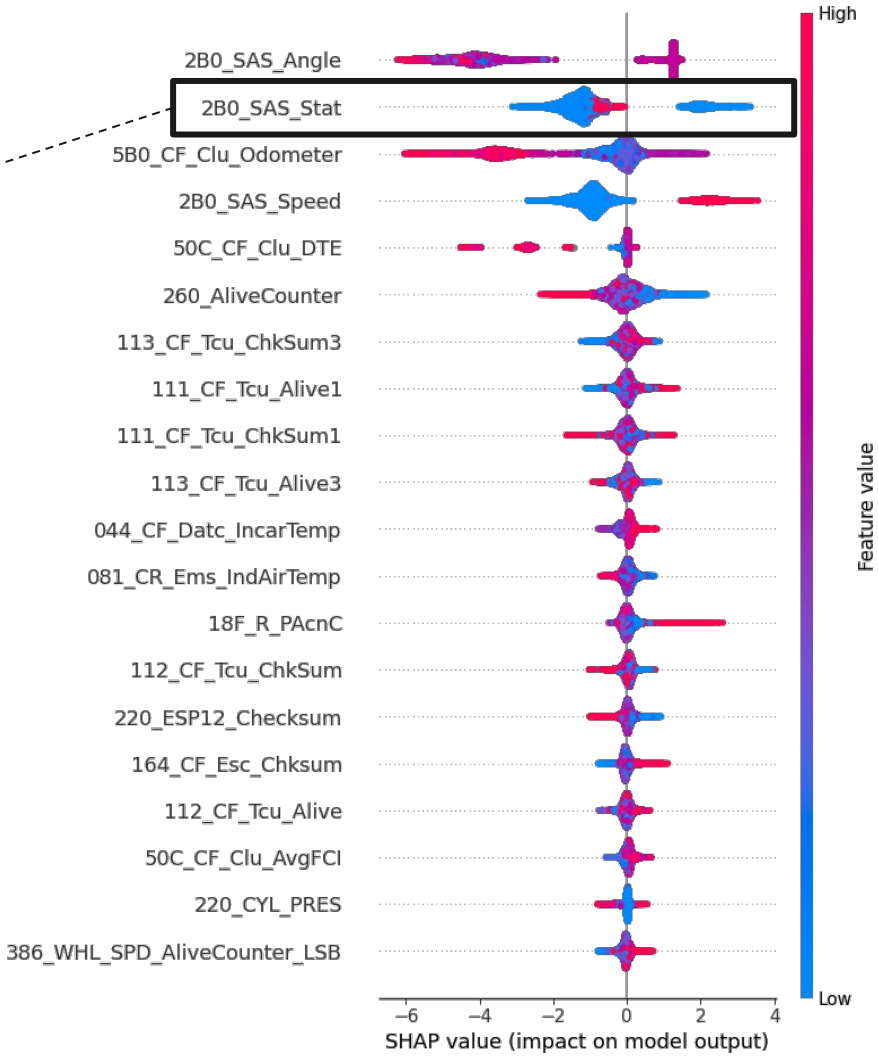
Ataque de Fabricação

- **2B0_SAS_Angle**
  - **Valores altos:** majoritariamente associados a valores SHAP negativos;
  - **Significado:** contribui mais para a identificação da classe normal (sem ataque);
  - **Insights:** o ataque está direcionado ao ângulo do sensor de direção.



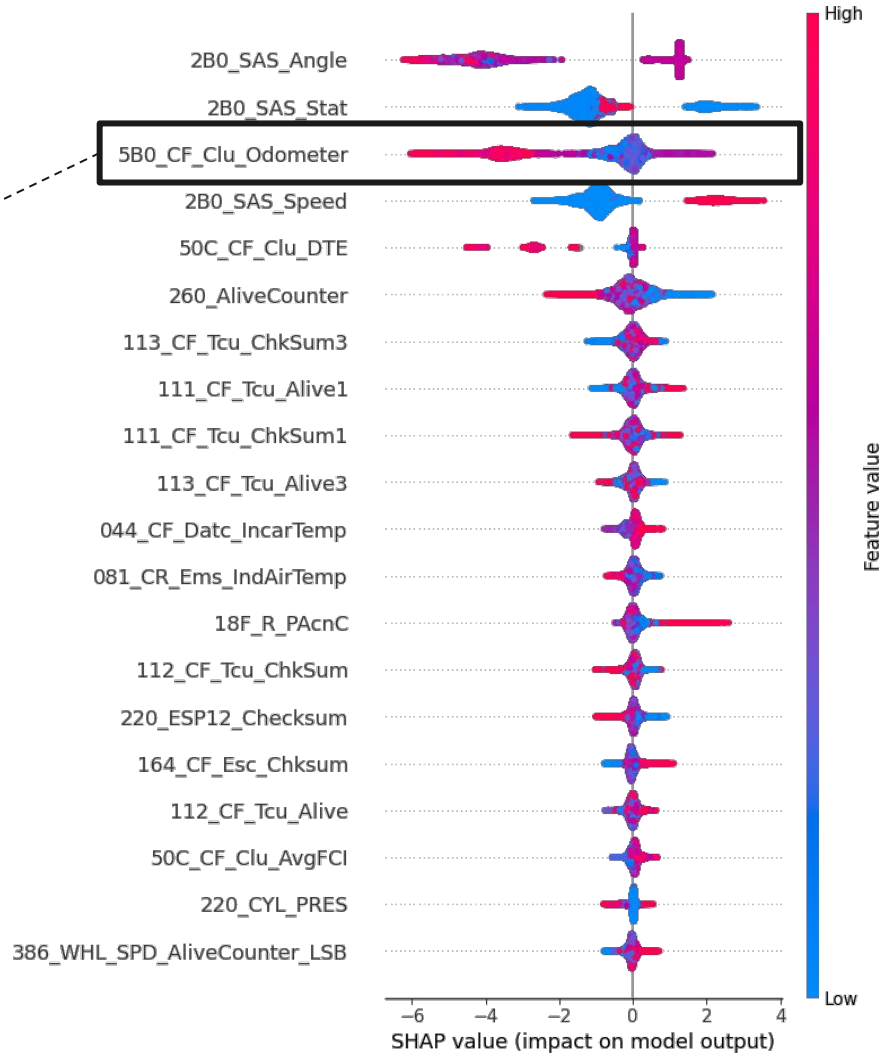
Ataque de Fabricação

- **2B0_SAS_Stat**
  - **Valores altos:** São poucos neste caso, mas todos ligeiramente negativos;
  - **Significado:** Contribui de maneira mais equilibrada nas predições das classes de ataque e normal;
  - **Insights:** possui relação com o status/desempenho do sistema de direção



Ataque de Fabricação

- **5B0_CF_Clu_Odometer**
  - **Valores altos:** mais associados a valores SHAP negativos;
  - **Significado:** contribui mais para a predição da classe normal, ainda que possua contribuições mistas;
  - **Insights:** relacionada ao painel e ao odômetro, que informa a distância percorrida pelo veículo.



Ataque	Característica	ECU	Função Principal
Fuzzing	5FA_CR_Wcs_ErrStat	ODS	Detecção de Ocupantes
Fuzzing	042_CR_Datc_RearDrTe_mp DispF, 044_CR_Datc _OutTempF, 042_CF_Datc_CO2_Warning	DATC	Aquecimento, Ventilação, Ar-condicionado

Relação ECU x Sinal

Ataque	Característica	ECU	Função Principal
Fuzzing e Fabricação	5B0_CF_Clu_Odometer	CLU	Painel de Instrumentos
Fabricação	2B0_SAS_Angle, 2B0_SAS_Stat, 2B0_SAS_Speed	MDPS	Módulo de Direção Assistida
Fabricação	50C_CF_Clu_DTE	CLU	Nível de Combustível

Relação ECU x Sinal

Considerações finais

- Buscou-se **aprimorar a explicabilidade** de um IDS;
- Empregaram-se as bibliotecas **SHAP e XGBoost**;
- Identificou-se as **características-chave** presentes em cada um dos ataques;
- Empregou-se técnicas de **engenharia reversa** para permitir o **mapeamento** das ECUs comprometidas.

Trabalhos futuros

- **Ainda existem desafios significativos devido à natureza confidencial das informações em redes CAN**
  - Integração de ferramentas de processamento de linguagem natural com XAI para aprimorar ainda mais a interpretabilidade.

Obrigado!

- Felipe N. Dresch - felipedresch.aluno@unipampa.edu.br
- Felipe H. Scherer - felipescherer.aluno@unipampa.edu.br
- Silvio E. Quincozes - silvioquincozes@unipampa.edu.br
- Diego Kreutz - diegokreutz@unipampa.edu.br

Perguntas?

Anexos

- **Redes externas podem representar ameaças à rede CAN?**
  - Com certeza!
  - Montadoras levam isso em consideração.

Emad Aliwa, Omer Rana, Charith Perera, and Peter Burnap. 2021. **Cyberattacks and Countermeasures for In-Vehicle Networks.** ACM Comput. Surv. 54, 1, Article 21 (January 2022), 37 pages. https://doi.org/10.1145/3431233.

Anexos

- **Posso usar um destes ataques para alterar a quilometragem do meu veículo?**
  - Odômetro mecânico ou eletrônico?
  - Crime em diversas localidades.

Fakhfakh, F., Tounsi, M. and Mosbah, M. (2022). **Cybersecurity attacks on CAN bus based vehicles: a review and open challenges**. Library Hi Tech, Vol. 40 No. 5, pp. 1179-1203. https://doi.org/10.1108/LHT-01-2021-0013.

Anexos

- **Como é feito o processo de desserialização dos dados?**
  - Repositório OpenDBC;
  - Informações sobre os sinais de veículos;
  - Algoritmo próprio que torna os sinais mais acessíveis.

Jeong, S., Lee, S., Lee, H., and Kim, H. K. (2024). **X-CANIDS: Signal-aware explainable intrusion detection system for controller area network-based in-vehicle network**. IEEE Transactions on Vehicular Technology, 73(3):3230–3246.

Anexos

- **Por que SHAP e não LIME?**
  - Considera todas as possíveis ordenações de características;
  - Consistência de resultados;
  - Fundamentação teórica mais robusta;

Lundberg, S. (2017). **A unified approach to interpreting model predictions**. arXiv preprint arXiv:1705.07874.

Anexos

- **Qual o critério de ordenamento das características nos gráficos?**
  - Magnitude média absoluta das contribuições.

Lundberg, S. (2017). **A unified approach to interpreting model predictions**. arXiv preprint arXiv:1705.07874.

Referências

Ding, W., Alrashdi, I., Hawash, H., and Abdel-Basset, M. (2024). **DeepSecDrive: An explainable deep learning framework for real-time detection of cyberattack in in-vehicle networks**. Information Sciences, 658:120057.

Hoang, T.-N., Islam, M. R., Yim, K., and Kim, D. (2023). **CANPerFL: Improve in-vehicle intrusion detection performance by sharing knowledge**. Applied Sciences, 13(11).

Jeong, S., Lee, S., Lee, H., and Kim, H. K. (2024). **X-CANIDS: Signal-aware explainable intrusion detection system for controller area network-based in-vehicle network**. IEEE Transactions on Vehicular Technology, 73(3):3230–3246.

Lundberg, H., Mowla, N. I., Abedin, S. F., Thar, K., Mahmood, A., Gidlund, M., and Raza, S. (2022). **Experimental analysis of trustworthy in-vehicle intrusion detection system using explainable artificial intelligence (XAI)**. IEEE Access, 10:102831102841.

Swetha, H., R., R. R. R., R., P. R., and Thomas Ciza, B. N. (2023). **XAI for intrusion detection system: comparing explanations based on global and local scope**. Journal of Computer Virology and Hacking Techniques.

Wickramasinghe, C. S., Marino, D. L., Mavikumbure, H. S., Cobilean, V., Pennington, T. D., Varghese, B. J., Rieger, C., and Manic, M. (2023). **RX-ADS: Interpretable anomaly detection using adversarial ml for electric vehicle CAN data.** IEEE Transactions on Intelligent Transportation Systems, 24(12):14051–14063.

Le, T.-T.-H., Suryanto, N., Kim, H., Ji, J., and Heo, S. (2023). **Enhancing intrusion detection and explanations for imbalanced vehicle can network data.** In Proceedings of the 12th International Symposium on Information and Communication Technology, pages 777–784.

Metwaly, A. A. and Elhenawy, I. (2023). **Sustainable intrusion detection in vehicular controller area networks using machine intelligence paradigm.** Sustainable Machine Intelligence Journal, 4:(4):1–12.