Projeto de Pesquisa para Programa de Pré-Mestrado 2024

1 TEMA DA PESQUISA

Análise Formal de Segurança Crítica (Safety) de Sistemas de Medicina Assistida por Inteligência Artificial e Estudo de Caso de Sistema de Detecção de Arritmias Cardíacas em Eletrocardiogramas por Redes Neurais Residuais (ResNets) com Técnicas de Sobreamostragem e Subamostragem de Dados

2 QUANTIDADE DE VAGAS DISPONÍVEIS

Uma (1).

3 DESCRIÇÃO DO TEMA DE PESQUISA

A garantia de segurança crítica (*safety*) de sistemas com Inteligência Artificial (IA) é tida como um tema com lacunas a serem investigadas, em função do uso inédito e cada vez mais extensivo de IA para desempenhar funcionalidades que afetam diretamente a integridade de seres humanos e do meio ambiente (SILVA NETO et al., 2022).

Com o objetivo de preencher parte das lacunas identificadas na área, o método Safety ArtISt (*Safety Artificial Intellgence Structure*) foi desenvolvido no Grupo de Análise de Segurança (GAS) do Departamento de Engenharia de Computação e Sistemas Digitais da Escola Politécnica da Universidade de São Paulo (PCS/Poli-USP) e aplicado, até o momento, em quatro estudos de caso de diferentes áreas de aplicação. Estas, por sua vez, contemplam sistemas de controle para meios de transporte autônomos, sistemas eletrônicos e programáveis críticos para a segurança (*safety*) e sistemas de suporte ao diagnóstico médico.

Os sistemas desenvolvidos por Kozal e Ksieniewicz (2019a, 2019b) para identificar arritmias cardíacas em eletrocardiogramas por meio de Redes Neurais Residuais (ResNets) Profundas e técnicas de sobreamostragem e sbamostragem de dados representam parte dos objetos da pesquisa do método Safety ArtISt. Todavia, uma das subatividades obrigatórias do método, voltada à caracterização formal do conjunto imagem dos sistemas para caracterizar exaustivamento suas saídas, representa um dos temas em aberto para pesquisa futura (SILVA NETO; CUGNASCA, 2023) e é, portanto, proposto como temática para um projeto de pesquisa para o programa de Pré-Mestrado de 2024.

Tal trabalho de pesquisa envolve a construção de modelos matemáticos que permitam, ao menos,

sobreaproximar o conjunto imagem da IA por meio de regras lógico-aritméticas que descrevam a aplicação em análise (GILLULA; TOMLIN, 2012). No presente estudo de caso, portanto, tais equações devem contemplar modelos lógico-aritméticos que descrevessem as categorias de batimentos cardíacos (saudável e cada arritmia) em função de uma série temporal discreta de até 187 amostras no domínio [0; 1] e algum método para avalia-las (por exemplo, Teorias Módulo Satisfabilidade – SMT baseadas em programação linear). Esse traalho de pesquisa não é considerado trivial porque exige, no mínimo, conhecimento multidisciplinar apurado do domínio da aplicação para extrair as equações de interesse.

Avalia-se que o trabalho possa se beneficiar, por exemplo, do arcabouço ONNX2SMT, cujo objetivo é traduzir modelos de redes neurais para a linguagem SMT-LIB, de forma a viabilizar a transcrição da IA em um conjunto de regras e condições a serem avaliadas por meio de SMTs (GIRARD-SATABIN et al., 2020). Outras referências relacionadas à verificação formal de redes neurais e que também são consideradas promissoras para guiar o trabalho de pesquisa são as publicadas por Tran et al. (2019), Sha et al. (2021), Genin et al. (2022), Ivanov et al. (2019), Pulina e Tacchella (2011), Sidrane et al. (2022), Peruffo, Ahmed e Abate (2021) e Xiang et al. (2019).

4 REFERÊNCIAS BIBLIOGRÁFICAS

GENIN, D. et al. **Formal Verification of Neural Network Controllers for Collision-Free Flight**. 13th International Conference on Verified Software: Theories, Tools, and Experiments, VSTTE 2021 and 14th International Workshop on Numerical Software Verification, NSV 2021. **Anais…**The Johns Hopkins University Applied Physics Laboratory, Laurel, MD, United StatesSpringer Science and Business Media Deutschland GmbH, 2022. doi: 10.1007/978-3-030-95561-8_9

GILLULA, J. H.; TOMLIN, C. J. **Guaranteed safe online learning via reachability: Tracking a ground target using a quadrotor**. Proceedings - IEEE International Conference on Robotics and Automation. **Anais**...Saint Paul, MN, Saint Paul, MN, United States: 2012. doi: 10.1109/ICRA.2012.6225136

GIRARD-SATABIN, J. et al. **CAMUS: A framework to build formal specifications for deep perception systems using simulators**. **2**4th European Conference on Artificial Intelligence, ECAI 2020, including 10th Conference on Prestigious Applications of Artificial Intelligence, PAIS 2020. **Anais…**Cea List, FranceIOS Press BV, 2020. doi: 10.3233/FAIA200383

IVANOV, R. et al. **Verisig: Verifying safety properties of hybrid systems with neural network controllers**. 22nd ACM International Conference on Hybrid Systems: Computation and Control,

HSCC 2019. **Anais**...University of Pennsylvania, Philadelphia, PA, United States: Association for Computing Machinery, Inc, 2019. doi: 10.1145/3302504.3311806

KOZAL, J.; KSIENIEWICZ, P. Imbalance Reduction Techniques Applied to ECG Classification Problem. **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**, v. 11872 LNCS, p. 323–331, 2019a. doi: 10.1007/978-3-030-33617-2_33

KOZAL, J.; KSIENIEWICZ, P. **GitHub - jedrzejkozal/ecg_oversampling**. Disponível em: <https://github.com/jedrzejkozal/ecg_oversampling>. Acesso em: 20 dez. 2021b.

PERUFFO, A.; AHMED, D.; ABATE, A. **Automated and Formal Synthesis of Neural Barrier Certificates for Dynamical Models**. Tools and Algorithms for the Construction and Analysis of Systems: 27th International Conference, TACAS 2021, European Joint Conferences on Theory and Practice of Software, ETAPS 2021, Luxembourg City, Luxembourg, March 27 – April 1, 2. **Anais**...Berlin, Heidelberg: Springer-Verlag, 2021. doi: 10.1007/978-3-030-72016-2_20

PULINA, L.; TACCHELLA, A. NEVER: A tool for artificial neural networks verification. **Annals of Mathematics and Artificial Intelligence**, v. 62, n. 3–4, p. 403–425, 2011. doi: 10.1007/s10472-011-9243-0

SHA, M. et al. **Synthesizing Barrier Certificates of Neural Network Controlled Continuous Systems via Approximations**. Proceedings - Design Automation Conference. **Anais**...San Francisco, CA, United States: 2021. doi: 10.1109/DAC18074.2021.9586327

SIDRANE, C. et al. OVERT: An Algorithm for Safety Verification of Neural Network Control Policies for Nonlinear Systems. **Journal of Machine Learning Research**, v. 23, 2022.

SILVA NETO, A. V. et al. Safety Assurance of Artificial Intelligence-Based Systems: A Systematic Literature Review on the State of the Art and Guidelines for Future Work. **IEEE Access**, v. 10, p. 130733–130770, 2022. doi: 10.1109/ACCESS.2022.3229233

SILVA NETO, A. V.; CUGNASCA, P. S. **Relatório Técnico de Pesquisa - Desenvolvimento do Estudo de Caso de Sistema de Detecção de Arritmias Cardíacas por Aprendizado Supervisionado Profundo - Versão 3**. São Paulo (SP). doi: 10.5281/zenodo.7485108

TRAN, D. H. et al. Safety verification of cyber-physical systems with reinforcement learning control. **ACM Transactions on Embedded Computing Systems**, v. 18, n. 5s, p. 1–22, 2019. doi: 10.1145/3358230

XIANG, W. et al. **Reachable Set Estimation and Verification for Neural Network Models of Nonlinear Dynamic Systems**. **Unmanned System Technologies.** Department of Electrical Engineering and Computer Science, Vanderbilt University, Nashville, TN, United States, Springer, 2019. doi: 10.1007/978-3-319-97301-2_7