Hardware Proposal of Evolutionary Algorithm for Outlier Detection in Streaming Applications

1st Lucileide M. D da Silva
Academic Directorate of Sciences - DIAC/SC/IFRN
Federal Institute of Rio Grande do Norte
Santa Cruz, RN, Brazil
lucileide.dantas@ifrn.edu.br

2nd Marcelo A. C. Fernandes
Department of Computer Engineering and Automation
Federal University of Rio Grande do Norte
Natal, RN, Brazil
mfernandes@dca.ufrn.br

General Information

Thesis Title: Hardware Proposal of Evolutionary Algorithm for Outlier Detection in Streaming Applications.
Ph.D. Candidate: Lucileide Medeiros Dantas da Silva.
Advisor: Professor Dr. Marcelo Augusto Costa Fernandes.
Department: Post-Graduate Program in Electrical and Computer Engineering.
Institution: Federal University of Rio Grande do Norte (UFRN).

I. Motivation

Outlier or anomaly detection consists of detecting rare events in a data set. When data is captured and processed continuously in an online way they are considered as data streams [1]. Due to the increasing number of sensors in the most diverse areas and applications, streaming data is currently generated from many different sources and there is a huge rise in the availability of time-series data. It is a central problem in many application areas such as time-series forecasting, medical systems, industrial process monitoring, telecommunications, sensors networks, internet traffic, and others [2], [3]. These systems provide users with real-time information and continuously seek to extract knowledge from structures of unified analysis from massive data flows. Time-series anomaly detection helps to monitor the different metrics and parameters of industrial and corporate applications and services in real-time. It supervises the time-series continuously and sends alerts for probable risky events related to incidents instantly [4]. Consequently, outlier detection of data streams is a prominent research area in data mining [5], as well as an important task in various industrial applications.

Industry 4.0 is an area where anomaly detection has been increasingly applied to [3]. One of the challenges of Industry 4.0 is the detection of production failures, and defects [6]. New technologies aim to add value and increase process productivity, but face difficulties performing complex and massive-scale computing tasks due to the large amount of data generated [7]. Many solutions presented in the literature require knowledge of whole dataset processes and the systems for modelling and making a series of initial assumptions. These, in most cases, are not applicable [8]. Hence, traditional anomaly detection techniques may not be a practical approach to detecting anomalies in real-time streaming data series [9].

Beyond Industry 4.0, inspired by IoT, a new paradigm of medical devices and applications is emerging. The Internet of Medical Things, or IoMT, connects healthcare systems through networks. The IoMT is recreating a crucial role in the healthcare industry and can provide more flexible, accurate and reliable medical care services.

For example, ensuring all multimodal data streams in an ICU (Intensive Care Unit) are analysed as generated through automated data processing can improve patient care. It can also be helpful in medical research by systematically capturing and integrating massive amounts of data to evaluate treatments and interventions' effects [10]. Therefore, the demand for analysing, capturing, transmitting and storing health statistics from multiple resources will increase sharply. The quality and scale of traditional medical models can not meet the necessity of IoMT applications and services [11]. As a result, the medical domain faces challenges adapting to this new paradigm's constraints, like large amounts of physiological data, managing device diversity, scale, performance, available resources, and interoperability. In addition, IoMT needs real-time processing and low power consumption.

Typicality and Eccentricity Data Analytic (TEDA) algorithm has been presented as a possible solution for anomaly detection with streaming algorithms [12]–[14]. It is based on a new approach of outlier detection in a data stream context [15]. For example, it can be applied to detect autonomous behaviour in an industrial process operation. TEDA can be an alternative to the statistical framework for analyzing most data as it does not need prior information about the data distribution. It is very suitable for non-stationary data (changes in distribution over time) because it is an algorithm that can evolve, modelling the data distribution over time as new data arrives. Another essential TEDA aspect is that it is based on new metrics founded on the similarity/proximity of data in the data space, not in density or entropy, as per traditional methods.

Some essential aspects must be considered when choosing an anomaly detection method, such as the computational effort required to handle extensive streaming data. Since streaming is dynamic, unknown and unlimited, the received information needs to be stored and analyzed without compromising memory and run-time. Handling this type of data requires three fundamentals: high-throughput, ultra-low-latency, and low-power [16]–[21].

Present-day applications, whether in the industrial field, medical field, or other fields, require real-time processing of datasets that exceed the capacity of primary memory, which poses a challenge for many applications. Traditional data management systems using centralized structures could not be able to process big datasets in a timely manner. To achieve this, data transmission is done by streaming rather than by data transfer. However, some important aspects need to be considered when working with data streaming, such as the computational effort to handle large data. Since the passed information needs to be stored and analysed without compromising memory and runtime. Two key developments have emerged as critical considerations for these types of applications: the ever-increasing volume of data that requires processing and the growing significance of robust embedded systems.

Real-time analysis stream data is critically dependent on computational infrastructure. For example, the massive accumulation of real-time data can overload traditional computing systems due to the large amount of data generated by sensors, as well as the requirement for intensive processing and high performance. As a result, the ability to process massive data flows from different sources at high speed is a critical problem. Software-only approaches cannot keep up with the growing computational demands of real-time analysis, given the barriers to reducing latency in large volumes. In addition, as the number of nodes increases to handle the ever-growing amounts of data, performance is not scaled linearly [16], [20], [21].

Considering the challenges presented, this work proposes a specialized hardware architecture of TEDA in reconfigurable computing using FPGAs for Real-Time Anomaly Detection of Streaming Data. FPGA provides high performance with a low operating frequency by exploring parallelism. Programmable digital circuits on FPGAs offer the best compromise among flexibility, performance, cost and time-to-market. The development of the hardware technique allows systems to be made even faster than their software counterparts. This extends the possibilities of use, especially when time constraints are most severe.

Reconfigurable computing is an emerging area that allows the development of hardware architectures customised to the algorithm. The main objective is to fit hardware to the algorithm rather than the algorithm to hardware, as in the traditional models where the algorithms are adapted to fit the processor's instructions. Complex algorithms with a high degree of parallelisation may be inhibited regarding computational performance in a traditional model, as they often have to shape themselves sequentially to the target hardware [22], [23]. Developing specialised hardware architecture has emerged as an exciting alternative to overcome such bottlenecks as algorithm delays, data handling capacity, traffic volume, and system limits. This approach allows the creation of solutions that can handle massive data processing while satisfying essential requirements such as ultra-low-latency, low-power, high-throughput, security, and ultra-reliable operations [24], [25]. These features are essential for improving productivity and quality in Industry 4.0 processes.

In IoMT field, hardware implementation for streaming applications can be used for monitoring and interpreting ICU data, considering that in the IoMT paradigm, healthcare systems are recording a large amount of physiologic data continuously. Specialised hardware offers an option to stream multimodal time series data, combining EEG (electroencephalogram), intracranial pressure, ECG (electrocardiogram), and respiration rate. For example, a platform with reconfigurable computing can enable real-time data analysis with automated data processing. Even when monitoring large volumes of data, thus improving ICU care by ensuring that multimodal data streams are analysed as they are generated.

The motivation for this work is to accelerate the TEDA algorithm for time-constrained applications, massive data, and energy efficiency by using a hardware-based implementation [26], [27]. To validate and synthesise the results, an FPGA Virtex 6 xc6vlx240t-1ff1156 was used, which was chosen due to its high performance and flexible architecture.

II. Objectives

The primary objective of this study is to develop a dedicated hardware solution for outlier detection in streaming applications. These techniques have gained significant popularity across various domains and applications in today's society. The proposed hardware architectures in this thesis will facilitate the efficient utilization of the TEDA algorithm, with a specific focus on achieving low latency, low power consumption, and rapid response speed.

This thesis proposal aims to present a reference architecture for Real-Time Anomaly Detection of Streaming Data (TEDA) using reconfigurable computing with Field-Programmable Gate Arrays (FPGAs). By leveraging dedicated hardware resources, efficient management of time constraints can be achieved. Additionally, the research aims to investigate the impact of FPGA implementations, enabling the efficient deployment of algorithms based on criteria such as low latency, low power consumption, and high speed of response.

III. Contributions

This work presents a hardware architecture for FPGA implementation of the TEDA technique. The FPGA's low execution time and data flow parallelization allows the TEDA algorithm to be used in applications with significant data flow and processing time constraints where the use of generic processors such as CPUs does not meet the constraints imposed by some applications. Therefore, the main contributions of this work are:

1) Implementation of a specialised hardware architecture for Real-Time Anomaly Detection of Streaming Data (TEDA) using reconfigurable computing with FPGAs.

2) Introduction of flexible, reconfigurable, and FPGA-adaptable reference architecture.
3) Provision of lower and upper latency limits across various scenarios and resolutions.
4) Analysis of the area occupation based on the post-synthesis results concerning the hardware in the FPGA.
5) Analysis of dynamic power based on the post-synthesis results for different input sensor representations at their maximum throughput.
6) Qualitative and quantitative analysis of the architecture's impacts on reducing latency and power consumption.
7) Synthesis and analysis of the architecture for two different application scenarios: industry 4.0 and Internet of Medical Things (IoMT). The proposed implementation holds the potential to assist various fields with significant data flow and processing time restrictions, enabling high-speed real-time data processing for specific applications.

These contributions aim to enhance the TEDA algorithm's performance and applicability by leveraging FPGA-based hardware architecture, thereby addressing the unique requirements of applications with intensive data processing needs.

IV. Relevant Results

The proposed design was validated and evaluated on FPGA, with comprehensive implementation details provided and analysed. The design features a fully parallel input of N elements and a 3-stage pipelined architecture to reduce the critical path and optimize throughput.

Results regarding the hardware occupation, throughput, and power efficiency were presented to validate the proposal. In comparison to other software platforms, the design achieved a speed improvement of up to 693 times, with a throughput of up to 10.96 Mega Samples Per Second (MSPs) and a dynamic power consumption of 16 mW.

Furthermore, bit-accurate simulation results are presented for various application scenarios involving multiple sensors, spanning from Industry 4.0 environments to the Internet of Medical Things (IoMT).

This work is a pioneer in the hardware implementation of the TEDA technique in specialised hardware.

A. Sandwich Doctorate Program

This thesis work received mobility funding from the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), under the CAPES PrInt program, for a duration of 12 months at King's College London in London, United Kingdom. The exchange was further extended for an additional 6 months as a visiting researcher, totalling 18 months. CAPES PrInt is an internationalization program aiming to promote the formation of international research networks and promote the mobility of professors and students linked to graduate programs abroad. Consequently, a significant portion of the research was conducted at the Centre for Telecommunications Research of King's College London. This partnership between UFRN and King's College London, initiated through the collaboration of researchers, was further strengthened through the Sandwich program. Collaboration work was carried out during the exchange with Professor Dr. Toktam Mahmoodi.

B. Publications

The authors have published the proposed hardware approach presented in this work with the participation of international collaborators in the following international journal:

1) IEEE Access: L. M. D. Da Silva et al., "Hardware Architecture Proposal for TEDA Algorithm to Data Streaming Anomaly Detection," in IEEE Access, vol. 9, pp. 103141-103152, 2021, doi: 10.1109/ACCESS.2021.3098004.

Other results related to the development of solutions in reconfigurable hardware have also been published during this doctoral process. Here are the references for those works:

2) IEEE Access: L. M. D. Da Silva, M. F. Torquato and M. A. C. Fernandes, "Parallel Implementation of Reinforcement Learning Q-Learning Technique for FPGA," in IEEE Access, vol. 7, pp. 2782-2798, 2019, doi: 10.1109/ACCESS.2018.2885950.

3) Sensors: S. N. Silva, L. M. D. da Silva, L. A. Dias, and M. A. C. Fernandes, "Prediction Techniques on FPGA for Latency Reduction on Tactile Internet", in Sensors, vol. 22, no. 9, p. 3556, May 2022, doi: 10.3390/s22093556

These publications highlight the research outcomes and contributions made in the field of reconfigurable hardware during the doctoral process.

C. Thesis defense

The thesis document was written entirely in English, and the defence was held on January 12, 2023. The evaluation board was constituted by Professors Dr. Leonardo A. Dias (University of Warwick), Dr. Luiz Affonso Guedes (UFRN), Dr. Ignacio S. Gendriz (UFRN) and Dr. Sérgio N. Silva (External).

D. Post Doctorate Program

At the conclusion of the Ph.D. program, the successful completion of this thesis work led to the award of a Post Doctorate position at the Federal University of Rio Grande do Norte, Natal-RN, Brazil. I am currently engaged in research and development activities at the Research and Innovation Center in Information Technology (nPITI) within the Metrópole Digital Institute (IMD) - UFRN. My work primarily focuses on developing machine learning and artificial intelligence solutions, which also involve hardware design solutions.

References

[1] Subutai Ahmad, Alexander Lavin, Scott Purdy, and Zuha Agha. Unsupervised real-time anomaly detection for streaming data. *Neurocomputing*, 262:134 – 147, 2017. Online Real-Time Learning Strategies for Data Streams.

[2] Muhammad Fahim and Alberto Sillitti. Anomaly detection, analysis and prediction techniques in iot environment: A systematic literature review. *IEEE Access*, 7:81664–81681, 2019.

[3] Xiaokang Zhou, Yiyong Hu, Wei Liang, Jianhua Ma, and Qun Jin. Variational lstm enhanced anomaly detection for industrial big data. *IEEE Transactions on Industrial Informatics*, 17(5):3469–3477, 2021.

[4] Hansheng Ren, Bixiong Xu, Yujing Wang, Chao Yi, Congrui Huang, Xiaoyu Kou, Tony Xing, Mao Yang, Jie Tong, and Qi Zhang. Time-series anomaly detection service at microsoft. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, KDD '19, page 3009–3017, New York, NY, USA, 2019. Association for Computing Machinery.

[5] Nikos Giatrakos, Elias Alevizos, Alexander Artikis, Antonios Deligiannakis, and Minos Garofalakis. Complex event recognition in the Big Data era: a survey. *The VLDB Journal*, 29(1):313–352, July 2019.

[6] Paolo Napoletano, Flavio Piccoli, and Raimondo Schettini. Anomaly detection in nanofibrous materials by cnn-based self-similarity. *Sensors*, 18(1), 2018.

[7] Ibrahim Abaker Targio Hashem, Ibrar Yaqoob, Nor Badrul Anuar, Salimah Mokhtar, Abdullah Gani, and Samee Ullah Khan. The rise of "big data" on cloud computing: Review and open research issues. *Information Systems*, 47:98 – 115, 2015.

[8] Clauber Gomes Bezerra, Bruno Sielly Jales Costa, Luiz Affonso Guedes, and Plamen Parvanov Angelov. An evolving approach to unsupervised and real-time fault detection in industrial processes. *Expert Systems with Applications*, 63:134 – 144, 2016.

[9] M. C. Lee, J. C. Lin, and E. G. Gan. Rere: A lightweight real-time ready-to-go anomaly detection approach for time series. In *2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC)*, pages 322–327, 2020.

[10] Steven N. Baldassano, Shawniqua Williams Roberson, Ramani Balu, Brittany Scheid, John M. Bernabei, Jay Pathmanathan, Brian Oommen, Damien Leri, Javier Echauz, Michael Gelfand, Paulomi Kadakia Bhalla, Chloe E. Hill, Amanda Christini, Joost B. Wagenaar, and Brian Litt. Iris: A modular platform for continuous monitoring and caretaker notification in the intensive care unit. *IEEE Journal of Biomedical and Health Informatics*, 24(8):2389–2397, 2020.

[11] Gulraiz J. Joyia, Rao M. Liaqat, Aftab Farooq, and Saad Rehman. Internet of medical things (iomt): Applications, benefits and future challenges in healthcare domain. *Journal of Communication*, 12(4):240–247, 2017.

[12] B. S. J. Costa, C. G. Bezerra, L. A. Guedes, and P. P. Angelov. Online fault detection based on typicality and eccentricity data analytics. In *2015 International Joint Conference on Neural Networks (IJCNN)*, pages 1–6, July 2015.

[13] Dmitry Kangin, Plamen Angelov, Jose Antonio Iglesias, and Araceli Sanchis. Evolving classifier tedaclass for big data. *Procedia Computer Science*, 53:9 – 18, 2015. INNS Conference on Big Data 2015 Program San Francisco, CA, USA 8-10 August 2015.

[14] B. S. J. Costa, C. G. Bezerra, L. A. Guedes, and P. P. Angelov. Unsupervised classification of data streams based on typicality and eccentricity data analytics. In *2016 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pages 58–63, July 2016.

[15] Daniel Osherson and Edward E Smith. On typicality and vagueness. *Cognition*, 64(2):189 – 206, 1997.

[16] Paolo Cappellari, Mark Roantree, and Soon Ae Chun. A scalable platform for low-latency real-time analytics of streaming data. In Chiara Francalanci and Markus Helfert, editors, *Data Management Technologies and Applications*, pages 1–24, Cham, 2017. Springer International Publishing.

[17] P. Le Noac'h, A. Costan, and L. Bougé. A performance evaluation of apache kafka in support of big data streaming applications. In *2017 IEEE International Conference on Big Data (Big Data)*, pages 4803–4806, 2017.

[18] T. Aung, H. Y. Min, and A. H. Maw. Coordinate checkpoint mechanism on real-time messaging system in kafka pipeline architecture. In *2019 International Conference on Advanced Information Technologies (ICAIT)*, pages 37–42, 2019.

[19] O. Marcu, A. Costan, G. Antoniu, M. Pérez-Hernández, B. Nicolae, R. Tudoran, and S. Bortoli. Kera: Scalable data ingestion for stream processing. In *2018 IEEE 38th International Conference on Distributed Computing Systems (ICDCS)*, pages 1480–1485, 2018.

[20] K. Nakamura, A. Hayashi, and H. Matsutani. An fpga-based low-latency network processing for spark streaming. In *2016 IEEE International Conference on Big Data (Big Data)*, pages 2410–2415, 2016.

[21] Jian Fang, Yvo T.B. Mulder, Jan Hidders, Jinho Lee, and H. Peter Hofstee. In-memory database acceleration on fpgas: a survey. *The VLDB Journal*, 2019.

[22] Jihong Liu and Deqin Liang. A survey of fpga-based hardware implementation of anns. In *2005 International Conference on Neural Networks and Brain*, volume 2, pages 915–918, 2005.

[23] Ninnart Fuengfusin and Hakaru Tamukoh. Mixed-precision weights network for field-programmable gate array. *PLOS ONE*, 16(5):1–26, 05 2021.

[24] S. Ghose, A. Boroumand, J. S. Kim, J. Gómez-Luna, and O. Mutlu. Processing-in-memory: A workload-driven perspective. *IBM Journal of Research and Development*, 63(6):3:1–3:19, 2019.

[25] V. Sze. Designing hardware for machine learning: The important role played by circuit designers. *IEEE Solid-State Circuits Magazine*, 9(4):46–54, 2017.

[26] Muhammad Rashid, Malik Imran, and Atif Raza Jafri. Exploration of hardware architectures for string matching algorithms in network intrusion detection systems. In *Proceedings of the 11th International Conference on Advances in Information Technology*, IAIT2020, New York, NY, USA, 2020. Association for Computing Machinery.

[27] Lucileide M. D. Da Silva, Matheus F. Torquato, and Marcelo A. C. Fernandes. Parallel implementation of reinforcement learning q-learning technique for fpga. *IEEE Access*, 7:2782–2798, 2019.