AI-Assisted Electronic Digital Twin for Liquid Level Sensors under Hardware Trojan Attack: An Experimental Validation

Alessandro Massaro, *Senior Member, IEEE*, Nicola Epicoco*, *Senior Member, IEEE,* Francesco Giannuzzi, Giuseppe Loseto, *Senior Member, IEEE,* and Filippo Gramegna

***Abstract*— This work proposes an Electronic Digital Twin (EDT) circuit model for Hardware Trojan (HT) attack characterization. Specifically, the goal is to compare circuit simulations, Artificial Intelligence (AI)-based attack classification and measurements proving a liquid level sensor attack by coupling an undesired resistance between the output nodes of the sensor. The paper describes all the steps to create the AI-assisted EDT model and validates the proposed modelling approach through experimental and numerical results. The paper adopts open-source tools, namely LTSPICE for circuit simulation, KNIME for AI classification, and Arduino IDE for firmware implementation. Concerning AI classification, the Random Forest (RF) supervised algorithm exhibited good performance for the classification of electrical resistance attacks. Theoretical and experimental behavior are investigated to validate the approach thus proposing a general protocol to realize an AI-assisted EDT framework.**

I. Introduction

Hardware Trojan (HT) attacks are an emergent topic about intentional circuit modification in Internet of Things Industrial (IIOT) networks [1]-[3], industrial applications [4], and Printed circuit Boards (PCBs) fabrication [5]. These attacks are mainly characterized by the change of the electrical/electronic behavior of electronic components due to the modification of the environment near the hardware region under attack or by coupling other undesired electrical or electronic components. This intentional modification relies on adding heat or humidity sources or coupling electrical elements such as resistance or capacitance to modify the signal read from Proportional–Integral–Derivative (PID) controllers [6] changing the gain coefficients, temperature/humidity sensors [7] altering the read values, or smart meters [8]. In these situations, the use of Artificial Intelligence (AI) algorithms, such as Support Vector Machines (SVM) and Random Forest (RF), can assist the digital modelling and control [9], interpreting and classifying circuit failures [10]. The use of Simulation Program with Integrated Circuit Emphasis (SPICE)-based simulators is also useful to detect possible attacks, as for side-channel leakage verification of masked cryptographic circuits [11]. Circuit simulations are also adopted to train RF models to predict disturbed signals in industrial Programmable Logic Controller (PLC) systems [12]. Hence, the combined use of circuit simulations with the RF optimizes the Electronic Digital Twin (EDT) model to interpret and classify attacks of electronic circuits. This work focuses on the application of liquid level sensing, which, compared to traditional sensors, present additional problems, due to the corrosive environment in which they are used, especially if immersed in direct contact with aggressive, chemical, or conductive fluids [13], [14].

In this context, the paper proposes an AI-assisted EDT model to reproduce the corresponding cyber physical model, which is then validated by comparing the experimental results of a testbed matching with the realized EDT. As shown in Fig. 1, the HT attack is simulated by coupling to the circuit a HT resistance that can modify the output voltage and currents of the liquid level sensor. Specifically, the proposed work addresses the following tasks: 1) construction of the EDT model that reproduces the physical behavior of a liquid level sensor detecting a liquid increase under a HT attack (see Fig. 1 (a) and Fig. 1 (b)); 2) interpretation of SPICE-based circuit simulations of the studied EDT by means of a parametric analysis by varying the resistance parameter of the undesired resistance R_{HT} (see Fig. 1 (a)) coupled to the output nodes of the sensor to simulate the HT attack; 3) definition of a RF model by means of the EDT training the supervised algorithm to classify the read signal thus supporting the interpretation of the output voltage signals for unknown R_{HT} values (i.e., the most common type of attack); 4) modeling and experimental validation of the EDT through a testbed replying the analyzed cyber-physical model; 5) formalization of the protocol for the construction of a suitable EDT integrating circuit simulations, AI data processing, and hardware testbeds.

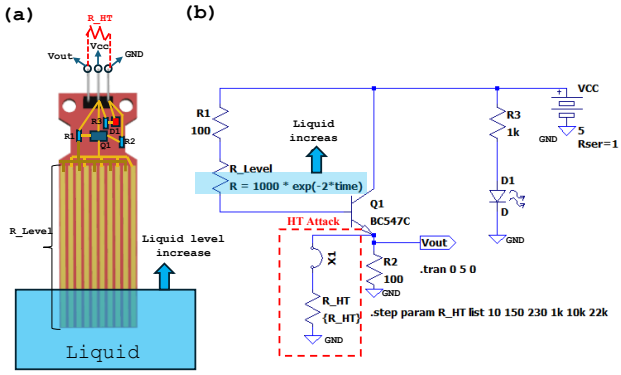

Figure 1. (a) Liquid level sensor and related electronic components integrated into the PCB having three output nodes (the R_{HT} resistance is coupled to V_{out} and GND pin). (b) Corresponding LTSPICE EDT model.

* Corresponding author (epicoco@lum.it).

All authors are with Department of Engineering of LUM "Giuseppe Degennaro" University, Strada Statale 100 km 18, Casamassima, 70010 Bari, Italy.

The remainder of this paper is as follows. Section II reviews the related literature on Digital Twins (DTs) and AI-based HT detection. Section III presents the proposed EDT model, integrating the physical principles of liquid level sensing and AI-driven data processing. Section IV details the experimental evaluation and discusses the obtained results, whereas Section V introduces the formalization adopted to define the EDT validation framework. Finally, Section VI concludes the paper and outlines directions for future research.

II. Related Work

As discusses in [15], research on AI-assisted HT detection and digital twin modelling has grown in recent years, combining circuit-level simulation, machine learning classification, and hybrid frameworks that integrate physical and virtual models. The goal of these works is often to improve detection performance, enhance modelling fidelity, or expand applicability to industrial cyber-physical systems [7]. In particular, a recent study on robust HT detection [16] integrates dual-domain feature extraction (time and frequency) and a stacked ensemble of AI models to detect trojan signatures in side-channel power traces. This approach outperforms individual classifiers by combining model strengths, but it typically requires detailed side-channel instrumentation and may not directly incorporate circuit simulation or sensor-specific perturbations. Similarly, Machine Learning (ML) techniques were applied in [17] to identify HT in open hardware designs analyzing real-world data from existing benchmarks and synthetic data generated with Large Language Models.

Although these works emphasize classification performance and feature extraction, demonstrating that AI-based algorithms can capture complex patterns introduced by HTs, they lack sensor modeling, circuit-level simulation, and consideration of sensor output perturbations, which are closer to practical embedded sensor attacks. In the context of cyber-physical security, recent works have extended the digital twin concept beyond passive replication toward active security assessment environments. Digital twin technologies are adopted in [18] for enhancing cybersecurity in production systems, demonstrating how DT frameworks can simulate diverse attack scenarios to evaluate vulnerabilities without disrupting physical operations. The works in [19]-[21] also validated the role of digital twins as tools for vulnerability analysis, simulation-based testing, and quantitative risk mitigation proposing reference threat model architectures for cyber-physical systems. Nevertheless, current DT implementations basically rely on either physics-based simulation or data-driven models in isolation. This separation hinders the development of comprehensive DT frameworks capable of simultaneously modelling electronic behavior and learning subtle attack signatures introduced by HTs.

The proposed approach addresses these gaps by introducing an AI-assisted EDT specifically tailored to liquid level sensors. Unlike conventional DTs, the presented EDT explicitly incorporates circuit-level modelling, where HT behavior is emulated by coupling undesired components between sensor output nodes. This enables direct representation of attack-induced electrical perturbations (adding the R_{HT} resistance as attack source), bridging the gap between abstract DT concepts and realistic electronic threats.

III. Electronic Digital Twin (EDT) Model

According to the sensor layout reported in Fig. 1 (a), the liquid level sensing is characterized by a series of parallel conductive traces. The Solu SL067 device (i.e., a liquid level sensor) has been used for both simulated and experimental measurements of the liquid sensor, characterized by three nodes of the whole PCB circuit (i.e., the voltage source V_{cc}, the output voltage V_{out}, which will be transduced into an intensity signal, and the ground reference GND). The parallel traces behave as a variable resistor whose resistance varies with the liquid level, behaving as a potentiometer. Since these traces conduct electricity, they short-circuit in the presence of a liquid, leading to a lower sensor resistance of the sensitive region. Hence the resistance of the copper traces decreases as the liquid level increases. Consequently, the current increases, which also leads to a voltage increase. The signal pin of the liquid level sensor provides the analogic output voltage. This helps detecting the liquid level, since a higher output voltage indicates a higher liquid level. The Spice-based circuit modelling the principle to detect the liquid level is illustrated in Fig. 1 (b). The variable resistance R_{Level} simulating the liquid level increases assumes the following deterministic trend (being t the time variable):

$$R_{Level} = 1000 \cdot exp(-2 \cdot t) \quad (1)$$

The base current of the Q1 NPN Bipolar Junction Transistor (BJT) changes with the variation of the R_{Level} by modifying the voltage output signal (V_{out}). The sensor is continuously powered by a battery having V_{cc} = 5 Volts according with the experimental power source (the power led D1 indicates the correct powering of the sensor). The HT attack is simulated by adding the undesired resistance R_{HT} at the output node (see Fig. 1 (b)): this resistance can abruptly or progressively change the current of the load R_2 (and consecutively the output voltage V_{out}) according with the current Kirchhoff law. The circuit simulations are performed by the LTSPICE open-source tool (version x64: 24.1.9), by setting a transient analysis between 0 and 5 seconds (time required to reproduce the liquid increase trend in Fig. 2 (a)).

IV. Results

Three main results are compared for the EDT model validation, namely, circuit simulations, RF data processing, and experimental results reproducing the circuit configuration of Fig. 1 (b) by means of the corresponding hardware components.

A. LTSPICE circuit and HT parametric results

The resistance decrease in eq. (1) due to the increase of the liquid level increases the output voltage V_{out} according with the trend in Fig. 2 (a). The parametric simulation of Fig. 2 (b) is useful to understand the voltage output sensitivity versus the attack resistance R_{HT}.

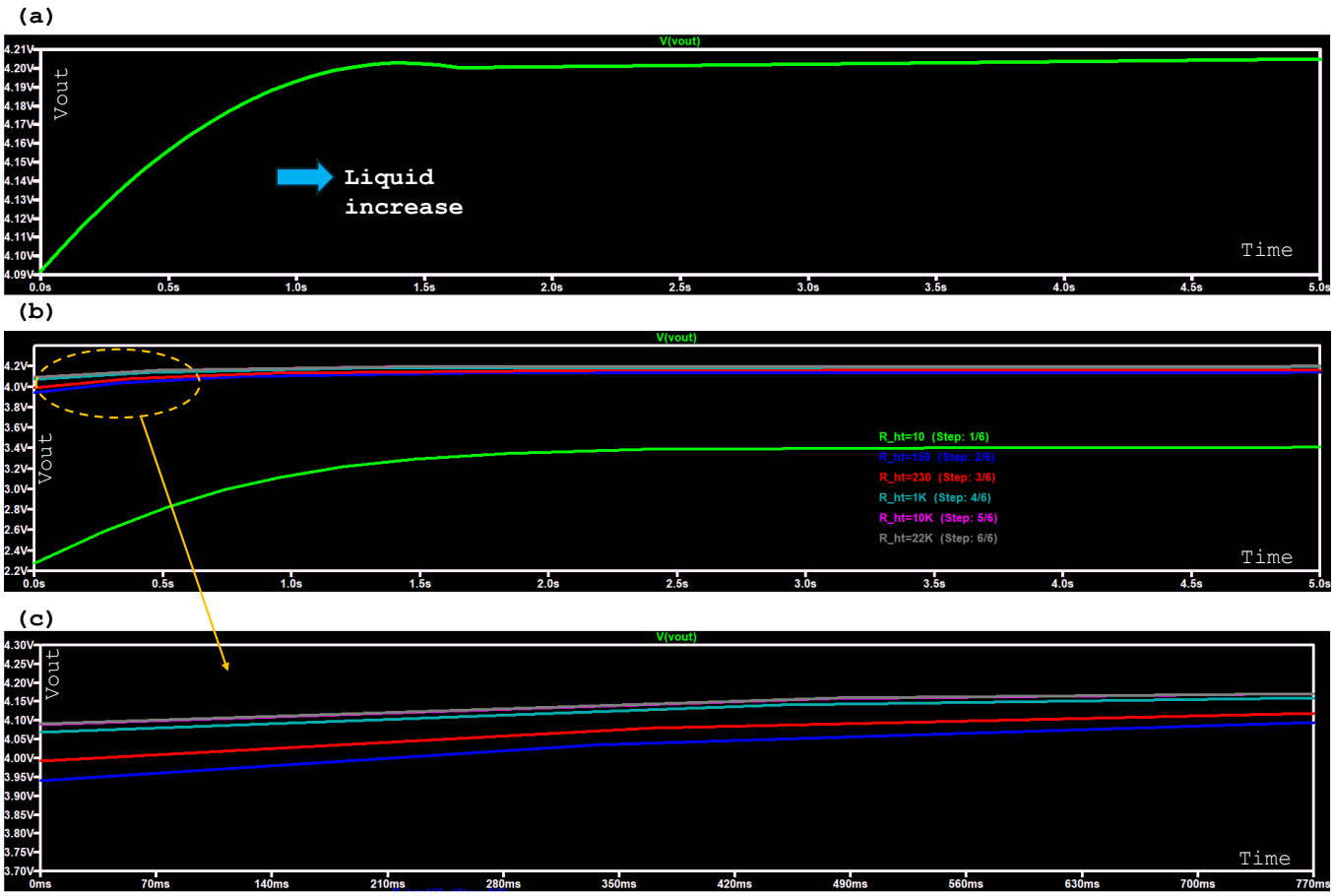

Figure 2. LTSpice EDT simulations: (a) V_{out} voltage signal following the simulation of the increase of the liquid level; (b) parametric analysis by varying the R_{HT} ranging from 10 Ω to 22 kΩ; (c) Zoomed part of Fig. 2 (b) highligthing the senstivity of 0.15 Volts by varying R_{HT} from 150 Ω to 22 kΩ.

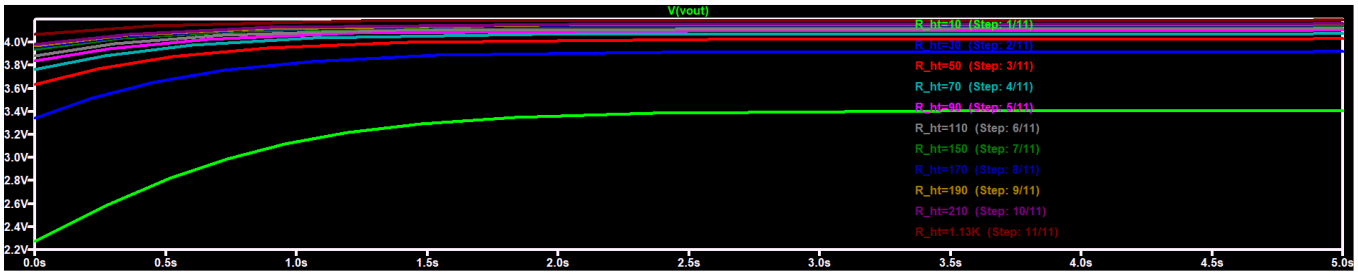

Figure 3. Parametric simulation by varyng the R_{HT} parameter from 10 Ω to 1.13 kΩ.

The parametric simulation is performed assuming R_{HT} = 10 Ω, 150 Ω, 230 Ω, 1 kΩ, 10 kΩ, and 22 kΩ, observing a high sensitivity ranging between 10 Ω and 150 Ω (output voltage variation of 0.6 Volts) and a low sensitivity ranging from 150 Ω to 22 kΩ (see the zoomed part reported in Fig. 2 (c) estimating an output voltage variation of 0.15 Volts).

B. AI data processing: Random Forest classification

Data classification is performed through a Random Forest (RF) algorithm based on multiple decision trees extracting features from the analyzed dataset. This choice is due to its suitability to provide good signal classification using the output results of the circuit simulations to construct a stable training model with a low statistical computational error [10]. The training model is constructed by considering the circuit simulations with R_{HT} = 10 Ω, 30 Ω, 50 Ω, 70 Ω, 90 Ω, 110 Ω, 150 Ω, 170 Ω, 190 Ω, and 210 Ω performed by the LTSPICE parametric analysis, providing the results shown in Fig. 3. A large variability of the R_{HT} values allows a more precise training model thus making the classification more accurate.

The RF algorithm is executed using the Konstanz Information Miner (KNIME) open-source tool (version 5.5.2) implementing the workflow of the inset in Fig. 4. The test is performed by considering as target the circuit with R_{HT} = 20 Ω (simulating an unknown HT behavior). As expected, the RF results of Fig. 4 highlights for the classified output a trend similar to the case of R_{HT} = 10 Ω rather than that with R_{HT} = 30 Ω, according with the gap observed in Fig. 3: the classified V_{out} signal deviates from the testing one in accordance with the created training model storing all the voltage gaps of the parametric simulation.

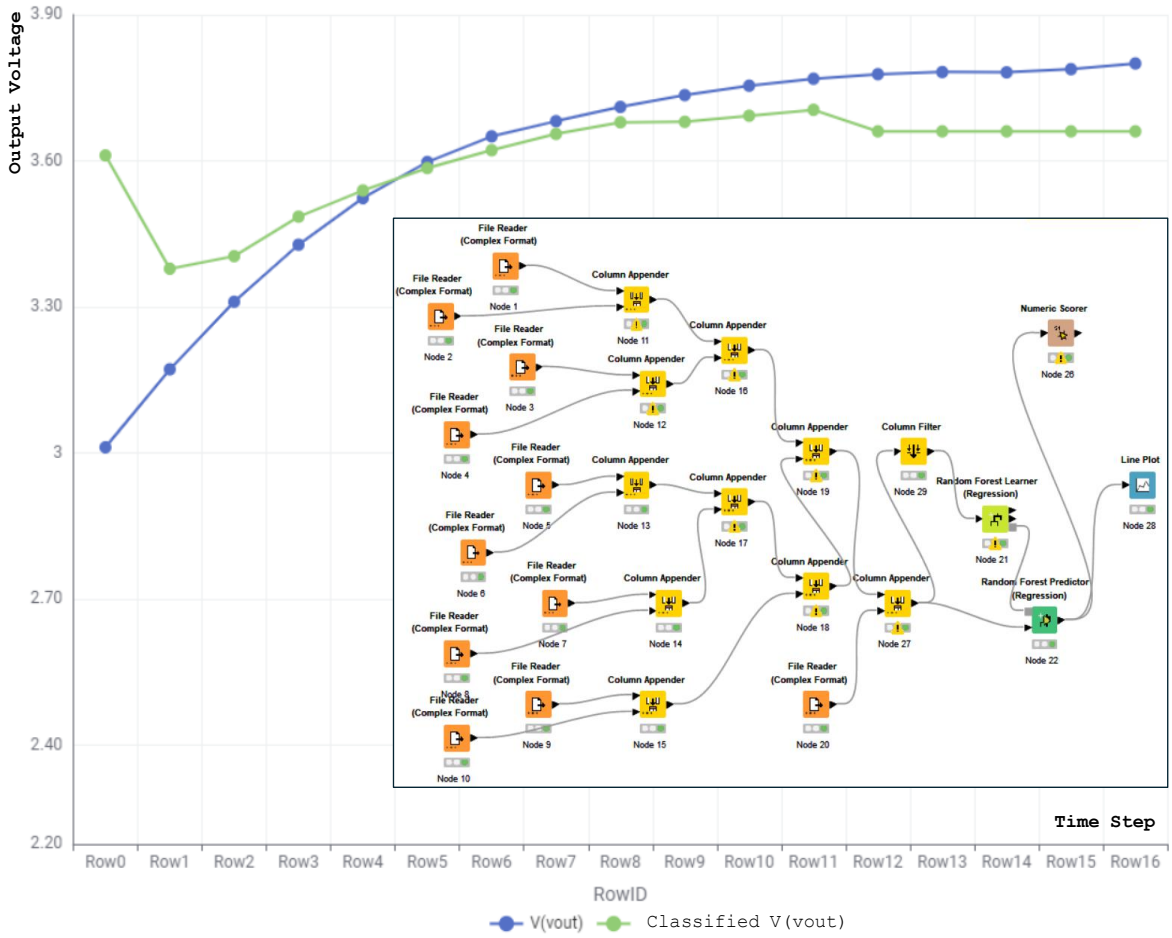

Figure 4. Comparison between the simulation and the classified signal (RF classification) considering the R_{HT} = 20 Ω. Inset: KNIME RF workflow implemented for the voltage classification (the row numbers indicate the time steps).

The optimized RF model (lower statistical error) is executed using a static random seed and considering 100 tree models.

The RF model is optimized to obtain low statistical errors. The KNIME workflow implementing the RF algorithm, is constructed by the following blocks named 'Nodes' (see the inset of Fig. 4):

1. *File Reader*, importing in the KNIME local repository the training and the testing dataset;
2. *Column Appender*, merging all input dataset into a unique table;
3. *Column Filter*, selecting the column (attributes) to be processed by the RF algorithm;
4. *Random Forest Learner*, training the RF supervised algorithm selecting the simulation with R_{HT}= 20 Ω (unknown attack hypothesis) with V_{out} as target class;
5. *Random Forest Predictor*, classifying the testing dataset corresponding to the output voltage with R_{HT} = 20 Ω;
6. *Numeric Scorer*, providing the statistical errors of the RF data processing;
7. *Line Plot*, plotting the RF processed results.

The RF statistical error estimations are listed in Table I, showing good results, with a MAE=0.111, corresponding to a maximum error of ±0.111 Volts for the classified V_{out} voltage. Clearly, the statistical errors may be different depending on the adopted training dataset.

TABLE I. RF STATISTICAL ERRORS.

Parameter	Value
Mean Absolute Error (MAE)	0.111
Mean Squared Error (MSE)	0.03
Root Mean Squared Error (RMSE)	0.173
Mean Signed Difference (MSD)	0.004
Mean Abdolute Percentage Error (MAPE)	0.033

C. Digital Twin experimental validation

The experimental testbed is realized using an Arduino Uno microcontroller connected to the SL067 liquid level sensor and transducing the output intensity of the liquid level correlated with the output voltage V_{out} of Fig. 1 (a). The open-source Arduino IDE platform is adopted for the implementation of the firmware allowing the coding of the measured voltage as an intensity signal in arbitrary units. The different HT resistances are coupled with the sensor nodes by using electrical jumpers and a breadboard able to not move the sensor system during the detection of the water level. The sensor is immersed into a container with water maintaining an immersion of the active length of the parallel conductive traces (constituting the variable resistance R_{Level}) equals to 80% (corresponding to 3 cm of immersed sensitive part) and measuring an intensity coded output of 616 (transduced signal of the sensor coding the resistance variation as provided by Arduino). Successively, the HT

resistances (R_{HT}) are gradually added by measuring the output intensity (see Fig. 5). By adding a low resistance value (R_{HT} = 10 Ω) the output intensity decreases due to a reduction in the transmitted power (that is, the power absorbed by the R_{HT} resistance). This is a consequence of the Ohm law (I_{HT}=V_{HT} / R_{HT}) and follows a principle different from that in eq. (1). For larger values of the resistances R_{HT}, this power absorption effect is weak but still a false reading of the measured intensity occurs (corresponding to a weak HT attack). Each measure of Fig. 5 corresponds to an average of 10 measurements (at each minute a measurement is performed, since a minute is sufficient to observe a stable intensity). The comparison between the gap decreases Δ (experimental gap of the intensity decrease) and δ (calculated RF and LTSPICE gaps of the voltage decrease) due to the coupling of the R_{HT} resistances is reported in Fig. 6, showing a good matching between all the trends. All the gaps are estimated with respect to the reference value of the case without attack (i.e., without adding the R_{HT} resistance)"

V. EDT Validation Framework

This section describes the procedure adopted to properly realize the AI-assisted framework. To this aim, the proposed approach integrates the Business Process Model and Notation (BPMN), a standard notation (ISO/IEC 19150:2013) useful to map procedures.

The BPMN workflow is here used to summarize the steps to construct an accurate EDT and describe the presented procedure. The BPMN workflow is reported in Fig. 7, consisting in the following phases:

- **Phase (1)**: the LTSPICE circuit layout is realized reproducing the physical behavior of the sensor understanding the circuital behaviors in terms of electrical currents and voltages;
- **Phase (2)**: the output of the LTSPICE circuit is used to train the AI supervised model and define the testing dataset for the output classification; if the algorithm error is high (low AI performances) the algorithm's hyper-parameters can be modified. Alternatively, more circuit outputs can be considered to train the model construction (by increasing the steps of the circuit parametric analysis, and, consecutively, increase the training dataset);
- **Phase (3)**: the hardware experimental test-bed is constructed by replying the cyber-physical behavior of the analyzed sensor and fixing experimental conditions for a stable measurements (in the proposed case the detected water level is fixed); the test-bed simulates the HT attacks as in the EDT model; the EDT model is validated in case of a good matching between results of phase (1), (2), and (3);
- **Phase (4)**: in the case of mismatching, the whole cyber-physical model (liquid level increase modelling, circuital elements, circuit layout, etc.) is revised.

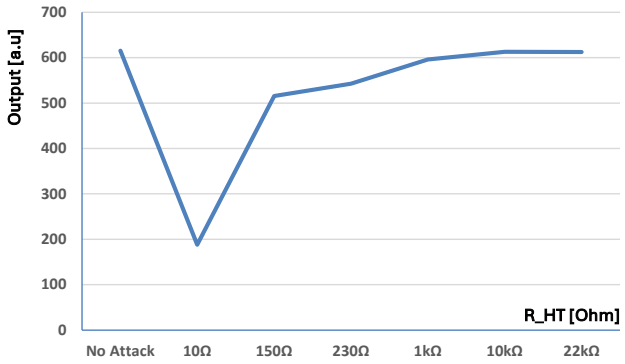

Figure 5. Experimental measurements of the intensity output versus the R_{HT} values representing the HT attacks. Each measure corresponds to an average of 10 measurements. The liquid in the experimental setup is kept fixed so as to maintain unchanged detection conditions when varying R_{HT}.

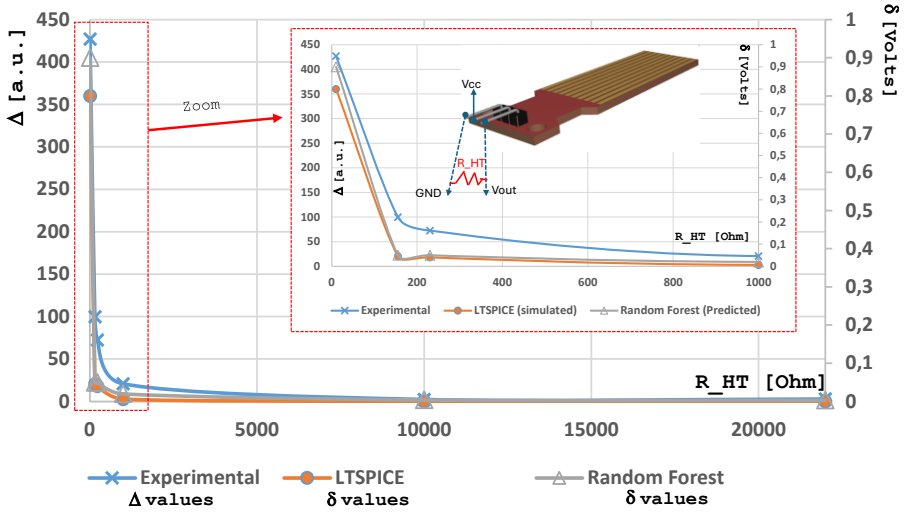

Figure 6. Comparison between AI-RF classification, measurements, and LTSPICE simulated data indicating the relative changes compared to the case without HT attack. Inset: zoomed plot and schematic configuration of the R_{HT} coupling.

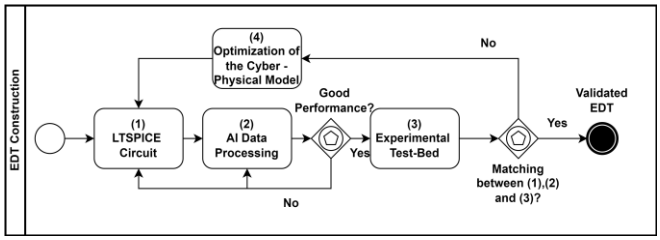

Figure 7. BPMN workflow summarizing the steps to realize the AI-assisted EDT.

The protocol illustrated in Fig. 7 can be adopted for the construction of different types of sensors integrating circuit simulations, AI data processing, and hardware testbed validating the whole cyber-physical model.

VI. Conclusion

This work proposes a baseline approach to validate EDT modelling, including SPICE-based circuit simulation and AI data processing, by characterizing the physical behavior of hardware attacks executed on a liquid level sensor. The HT attack is simulated by coupling an undesired electrical resistance at the output of the sensor PCB. The good matching between experimental results, AI-RF, and LTSPICE simulations proves that the circuit outputs are suitable to efficiently construct the AI-RF training model that can properly classify unknown hardware attacks. Finally, the work proposes a BPMN procedure to correctly realize the EDT model by integrating the hardware components and measuring the related experimental testbed. The proposed approach is adopted to validate the whole cyber-physic model and includes all the electrical and physical variables modelling of the sensing. The discussed procedure can be adopted to efficiently construct digital twin models using open-source tools.

The testbeds and the EDTs can be extended to more complex models including capacitive and inductive attacks and combined effects as well as different typologies of sensors. Future works will include a comparison with other benchmarking methods including the application of unsupervised algorithms and other recent DT frameworks.

References

[1] M. M. R. Monjur, J. Heacock, J. Calzadillas, M. S. Mahmud, J. Roth, K. Mankodiya, E. Sazonov, Q. Yu, "Hardware Security in Sensor and its Networks", *Frontiers Sensors*, vol. 3, art. 850056, 2022.

[2] M. R. Monjur, S. Sunkavilli, and Q. Yu, "ADobf: Obfuscated Detection Method Against Analog Trojans on I2C Master-Slave Interface", *IEEE 63rd International Midwest Symposium on Circuits and Systems* (MWSCAS), Springfield, MA, USA, 2020, pp. 1064-1067.

[3] M. Monjur, J. Calzadillas, and Q. Yu, "Hardware security risks and threat analyses in advanced manufacturing industry", *ACM Transactions on Design Automation of Electronic Systems*, vol. 28 (5), pp. 1-22, 2023.

[4] A. K. Verma, C. B. Mellado, C. G. Castaño, S. Sharma, "Analysis of Hardware Trojan for Multilevel Cascaded Converters", *2025 IEEE 5th International Conference on Sustainable Energy and Future Electric Transportation* (SEFET), Jaipur, India, 2025, pp. 1-7.

[5] J. Huan, S. D. Paul, S. Mandal, S. Bhunia, "PRISTINE: An Emulation Platform for PCB-Level Hardware Trojans", *IEEE Access*, vol. 12, pp. 49291-49305, 2024.

[6] A. Massaro, N. Epicoco, G. Loseto, C. Ardito, "IIOT Cyberattacks in Industrial Field Control Systems: PID Attack and Hardware Trojan Effects", *IEEE 20th International Conference on Automation Science and Engineering.* (CASE), Bari, Italy, 2024, pp. 1702-1707.

[7] A. Massaro, N. Epicoco, G. Loseto, C. A. Ardito, "AI-Based Detection and Simulation of Hardware Trojan Attacks in Industrial Applications", *IEEE Access*, vol. 13, pp. 50967-50979, 2025.

[8] A. Massaro, N. Epicoco, G. Loseto, G. Starace, C. A. Ardito, "Smart Metering and Hardware Trojan Electronic Digital Twin based on Artificial Intelligence", *IFAC-PapersOnLine*, vol. 59 (9), pp. 247-252, 2025.

[9] M. Ragnoli, *et al.*, "A condition and fault prevention monitoring system for industrial Computer Numerical Control machinery", *IEEE Access*, vol. 12, pp. 20919-20930, 2024.

[10] G. Akilarasu, V. Mariselvam and M. Parvathi, "AI-based Circuit Fault Detection and Classification using Python Simulation", *2025 4th International Conference on Automation, Computing and Renewable Systems* (ICACRS), Pudukkottai, India, 2025, pp. 174-179.

[11] K. Monta, M. Nagata, J. Balasch, I. Verbauwhede, "On the Unpredictability of SPICE Simulations for Side-Channel Leakage Verification of Masked Cryptographic Circuits", *60th ACM/IEEE Design Automation Conference* (DAC), San Francisco, CA, USA, 2023, pp. 1-6.

[12] A. Massaro, "Artificial Intelligence Signal Control in Electronic Optocoupler Circuits Addressed on Industry 5.0 Digital Twin", *Electronics*, vol. 13, art. 4543, 2024.

[13] M. Pavone, N. Epicoco, F. Magliocca, G. Pola, "A generalized approach for Feature Selection in Water Quality Monitoring", *31st Mediterranean Conference on Control and Automation* (MED), Limassol, Cyprus, 2023, pp. 599-604.

[14] M. Pavone, N. Epicoco, G. Pola, A. Manno, "Modelling sensors degradation for water quality monitoring: A data-driven approach", 20th IEEE International Conference on Automation Science and Engineering (CASE), Bari, Italy, 2024, pp. 1696-1701.

[15] K. I. Gubbi, B. S. Latibari, A. Srikanth, T. Sheaves, S. A. Beheshti-Shirazi, S. Manoj PD, S. Rafatirad, A. Sasan, H. Homayoun, S. Salehi. "Hardware Trojan Detection Using Machine Learning: A Tutorial", *ACM Transactions on Embedded Computing Systems*, vol. 22 (3), art. 46, 2023.

[16] S. N. Puspa, A. Enan, R. Majumder, M. S. Salek, G. Comert, M. Chowdhury, "Robust hardware Trojan detection leveraging dual-domain features and stacked ensemble learning", *Cybersecurity*, vol. 9 (1), art. 111, 2026.

[17] V. T. Hayashi and W. Vicente Ruggiero, "Hardware Trojan Detection in Open-Source Hardware Designs Using Machine Learning", *IEEE Access*, vol. 13, pp. 37771-37788, 2025.

[18] Y. Jiang, W. Wang, J. Ding, X. Lu, Y. Jing, "Leveraging Digital Twin Technology for Enhanced Cybersecurity in Cyber–Physical Production Systems", *Future Internet*, vol. 16 (4), art. 134, 2024.

[19] G. Erceylan, A. Akbarzadeh, V. Gkioulos, "Leveraging digital twins for advanced threat modeling in cyber-physical systems cybersecurity", *International Journal of Information Security*, vol. 24, art. 151, 2025.

[20] A. Massaro and N. Epicoco, "AI-Assisted Electronic Digital Twin for Capacitive Pressure Sensors Controlling Robotic Gripping", *IEEE Sensors Journal*, vol. 26 (9), pp. 13338-13346, 2026.

[21] G. Loseto, A. Massaro and N. Epicoco, "Federated Learning meets Electronic Digital Twins for Privacy-Preserving Hardware Trojan Detection in Smart Meter Infrastructures", *IEEE Sensors Journal*, Early Access, doi: 10.1109/JSEN.2026.3675872.